On Group-based Trajectory Modelling

Christopher Edward Davies

Thesis submitted for the degree of

Doctor of Philosophy

in

Statistics

at

The University of Adelaide

School of Mathematical Sciences

Faculty of Engineering, Computer and Mathematical Sciences

THE UNIVERSITY
of ADELAIDE

May 2018

Contents

List of abbreviations

1S	1-step
3S	3-step
AIC	Akaike information criterion
BCCR	Bayesian correct classification rate
BIC	Bayesian information criterion
CBCL	child behavior checklist
CPV	contamination predictive value
CCR	correct classification rate
GMM	growth mixture modelling
I3S	improved 3-step
ICC	intraclass correlation coefficient
LCGA	latent class growth analysis
LRT	likelihood ratio test
MGMM	multivariate Gaussian mixture modelling
PC3S	pseudo class 3-step
PR	probability regression
PW3S	probability weighted 3-step

Abstract

Group-based trajectory models are used for characteristics that, when followed longitudinally, may show subpopulations with distinct trajectories. This thesis describes three studies I undertook relating to these models.

Group-based trajectory models generally assume a certain structure in the covariances between measurements, for example conditional independence, homogeneous variance between groups, or stationary variance over time. Violations of these assumptions may result in poor model performance, but the extent and nature of this is not well understood. In the first study, I used simulation to investigate the effect of covariance misspecification on misclassification of trajectories in commonly used models under a range of scenarios. I found that the more complex models generally performed better over a range of scenarios. In particular, incorrectly specified covariance matrices could significantly bias the results, whereas using models with a correct but more complicated than necessary covariance matrix incurred little cost.

An underlying assumption of the group-based trajectory model is that it applies to all trajectories, and this does not allow for the possibility that outliers may be

present. Thus outlying trajectories may distort the estimated groups of these models and any subsequent analyses that use them. In the second study, I used simulations to assess the impact of outliers on group-based trajectory models. The presence of outliers tended to lead to an increased number of groups, and a reduction in the correct classification rate provided the group means were well separated. Following the simulations, I developed an algorithm for identifying outlying trajectories, and evaluated its performance on the simulated trajectory datasets. The application of my algorithm is recommended as part of sensitivity analyses to determine the effect that outliers may have.

One approach to modelling the influence of prior covariates in the group-based setting is to consider models wherein these covariates affect the group membership probabilities. In the third study, I compared six different methods from the literature for estimating the effect of covariates in this way. I found that when investigating the effects of covariates, the full likelihood approach minimised the bias in the estimates of the covariate effects. In this '1-step' approach, the estimation of the effect of covariates and the trajectory model are carried out simultaneously. Of the '3-step' approaches, where the the effect of the covariates are assessed subsequent to the estimation of the group-based trajectory model, only Vermunt's Improved 3-step resulted in bias estimates similar in size to the full likelihood approach. The remaining methods resulted in considerably higher bias in the covariate effect estimates, and should not be used.

This thesis provides guidance in the use of group-based trajectory models for

practising statisticians, focusing on the choice of covariance structures, the impact and identification of outlying trajectories, and the most appropriate methods for estimating the effects of covariates. Researchers should consider a wide range of models, and bearing in mind the assumptions they make, carefully choose that which fits best with the data.

Signed statement

I certify that this work contains no material which has been accepted for the award of any other degree or diploma in my name, in any university or other tertiary institution and, to the best of my knowledge and belief, contains no material previously published or written by another person, except where due reference has been made in the text. In addition, I certify that no part of this work will, in the future, be used in a submission in my name, for any other degree or diploma in any university or other tertiary institution without the prior approval of the University of Adelaide and where applicable, any partner institution responsible for the joint-award of this degree.

I give consent to this copy of my thesis, when deposited in the University Library, being made available for loan and photocopying, subject to the provisions of the Copyright Act 1968.

I acknowledge that copyright of published works contained within this thesis resides with the copyright holder(s) of those works.

I also give permission for the digital version of my thesis to be made available on

the web, via the University's digital research repository, the Library Search and also through web search engines, unless permission has been granted by the University to restrict access for a period of time.

I acknowledge the support I have received for my research through the provision of an Australian Government Research Training Program Scholarship.

SIGNED: .. DATE: 5/5/18

Acknowledgements

I would like to sincerely thank the many people who have helped me throughout my work towards this thesis.

To my supervisors, Gary Glonek and Lynne Giles, thank you for your advice, encouragement and patience throughout my candidature. I am incredibly grateful for all the time you have given to my research, and for your help in making this thesis the best it could be. Gary, your ability to think three steps ahead from a statistical perspective is a thing of wonder. Lynne, thank you for your initial conception of this research project and for your applied perspective.

I am thankful to have received scholarship support from the the University of Adelaide and the School of Mathematical Sciences. I would also like to thank everyone in the School of Mathematical Sciences. From the administrative staff who assist with anything and everything, including organising my attendance to conferences, to my fellow postgrads and postdocs for being there to bounce ideas off and for a welcome distraction.

I also wish to thank the families and researchers involved with the Generation

1 Study, particularly the chief investigators Vivienne Moore and Michael Davies for allowing my access to data from that study.

To my friends and family, thank you for your support and love throughout this journey, without which I would not have made it.

Mum and Dad, thank you for everything you've done for me, and for the opportunities you've given me. You've allowed me to feel that I could achieve anything.

Finally, to my wife Nikki and son Sam, I dedicate this thesis to you. Nikki, thank you for your support in the difficult times, and for celebrating my successes with me. For your patience as timelines stretched, and for giving me the perspective of what matters. Sam, I hope this thesis shows you what you can achieve if you set your mind to something. I'm looking forward to continuing the adventures of our family with you both.

Chapter 1

Introduction

1.1 Background

When a characteristic is measured longitudinally in a population, sometimes the paths of measurements, or trajectories, that individuals follow are similar to one another. An example of this is the growth of individuals through childhood and adolescence. In this situation, hierarchical modelling or latent curve analysis can be used. However, when some characteristics, such as childhood behaviour, are followed longitudinally they show subpopulations following distinct longitudinal trajectories.

One type of childhood behaviour is externalising behaviour, such as aggression, bullying or delinquency. Figure 1.1.1 shows a plot of externalising behaviour measurements of 10 children from the Generation 1 Study,[1,2] a prospective longitudinal study of South Australian children described in Section 1.1.2. In the plot there are distinct groups of trajectories visible, shown with different colours here. Some

individuals start with high externalising behaviour levels but their behaviour score improves as they age, shown in brown. Other individuals move from low levels to higher levels, before returning to lower levels of externalising behaviour, shown in black. There are also individuals with relatively stable low or high levels of externalising behaviour, shown magenta and cyan respectively. Although it is straightforward to assign the trajectories in this example into groups, doing so for the whole sample cannot be achieved by inspection alone.

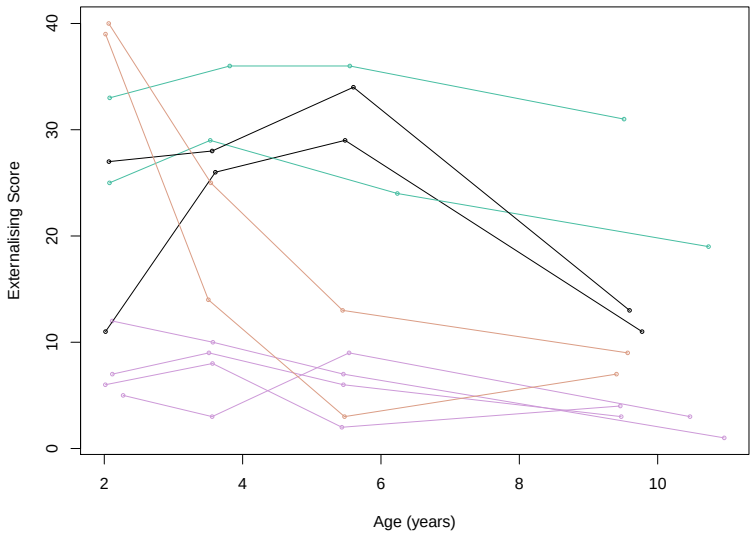


Figure 1.1.1: Externalising behaviour of 10 individuals following several distinct trajectories.

Early longitudinal studies with data believed to contain distinct subpopulations according to some characteristic were grouped using ad hoc assignment rules (for example in the child development literature).[3,4] For instance, Moffitt et al. assigned boys to an 'adolescence-limited' trajectory if they met certain criteria for extreme antisocial behaviour as adolescents, but not as children.[3] However, these methods

did not allow for the assumption of the existence of the groups to be tested, nor for the uncertainty about group membership to be estimated. Group-based trajectory modelling methods were developed to allow subpopulations to be estimated from longitudinal data. These methods include latent class growth analysis (LCGA)[5,6] and growth mixture modelling (GMM).[7–9] A detailed description of these methods is provided in Section 2.1.

Group-based trajectory modelling methods have been applied to a wide range of outcomes in criminology, clinical psychology and medicine.[5,10] Nagin and Land's paper first describing the LCGA model was motivated by questions about the concept of criminal careers.[5] Muthén's early papers describing GMM involved trajectories of heavy drinking, and the prediction of alcohol dependence.[7,8] More recent applications have included trajectories of childhood body mass index,[11] neurological improvement after stroke,[12] and marital conflict.[13] This thesis was motivated by analysis of data concerning childhood behaviour, which is another example of an area in which these models have been applied. The study drawn on for the data is outlined in Section 1.1.2.

1.1.1 Group-based trajectories of childhood behaviour

Childhood behaviour is an example of a characteristic for which individuals can plausibly be grouped into subpopulations following distinct mean trajectories. Using this setting for motivation, in this thesis I examine key aspects of the performance of such models. In this section, I describe previous research into group-based trajectories of

childhood behaviour, to illustrate some of the limitations in earlier applications.

Many childhood behaviours can be generally described as either externalising or internalising behaviours. Externalising behaviours are actions directed towards others, whereas internalising behaviours are actions directed towards the self, such as depression or anxiety. Typically, externalising behaviour tends to decrease throughout childhood, whereas internalising behaviour tends to increase with age, particularly for girls.[14] It has been widely observed that early internalising and externalising issues are a risk factor for academic and behavioural problems in later life.[15–17] Group-based trajectory modelling has been used to understand the different types of behavioural trajectories and to ascertain if it is possible to discriminate between those children who are on improving behavioural trajectories through childhood and those who are on chronically poor trajectories, thereby providing opportunities for targeted interventions.

In the previous research on childhood behaviour to date, the choice of models has been given little consideration. The only study in the applied literature to have modelled group-based trajectories of only internalising behaviours is Toumbourou et al.'s investigation of Victorian children.[18] A Gaussian, cubic, LCGA model was used and, although the covariance structure of the model was not described, it can be inferred to have been homogeneous and stationary through the software used for estimation. Consideration of other covariance structures was not described. The first study to model group-based trajectories of externalising behaviours was Nagin and Tremblay's 1999 investigation of physical aggression, opposition and hyperactivity

in boys.[19] A Gaussian, quadratic, LCGA model with homogeneous and stationary covariance structure was used, and again consideration of other covariance structures was not described. There are two studies that have investigated both internalising and externalising behaviours,[20,21] with Fanti and Henrich also examining the joint trajectories of the two types of behaviours.[21] Both studies used LCGA with a zero-inflated Poisson model to account for the skewed nature of the data, and polynomials up to quadratic order for the mean trajectories. The choice of model type in both studies was not justified by comparison with other possibilities. As seen in these examples, the assumptions made in fitting group-based trajectory models are not typically examined in detail, and the rationale for choosing a particular model is not generally provided.

1.1.2 The Generation 1 Study

Cohort formation

The Generation 1 Study is a prospective longitudinal cohort study of women and their children living in Adelaide, South Australia. The women were recruited in their first 16 weeks of pregnancy between 1998 and 2000 through the antenatal clinic at a public hospital or through three privately practising obstetricians. Details of the establishment of the cohort are described by Moore et al.[1] To be eligible to participate in the study the women had to be Caucasian, aged at least 18 years old, and free from certain conditions known to affect fetal growth. A total of 557 women

completed the pregnancy phase of the study and had a live singleton infant.

Longitudinal follow-up

Mothers and children have been followed-up using a structured protocol throughout the child's infancy and childhood. To date, data have been collected in early (<16 weeks gestation) and late (30-34 weeks gestation) pregnancy; when the children were born; at ages 3, 6, 9, and 12 months; and at ages 2, 3½, 5, 9½ and 12 years. All time points where measurements have been taken are shown below in Figure 1.1.2. Data pertaining to both the children and their wider family circumstances have been collected at each study wave. The study was reviewed and approved by the University of Adelaide Human Research Ethics Committee. The research study procedures conformed to the principles of the Declaration of Helsinki and all mothers in the study gave written informed consent.

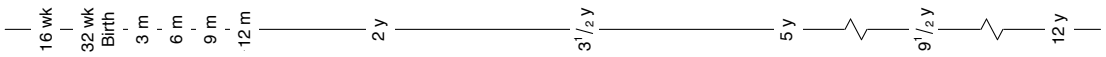


Figure 1.1.2: Timeline of measurements in the Generation 1 Study.

Childhood behaviour measurement

Childhood behaviour has been measured in the Generation 1 Study using the Achenbach Child Behavior Checklist (CBCL).[22] The CBCL records the parent's views on specific behavioural, emotional and social problems. Each item is reported to be not true (0), somewhat true (1), or very true (2). Total scores are derived as the

sum of the items. Internalising and externalising behaviour subscales have also been derived using factor analyses of the CBCL items.[22] The CBCL has good inter-rater (intraclass correlation coefficient (ICC) 0.96) and test-retest reliability (ICC 0.95) along with well supported content, criterion-related and construct validity.[22] In the Generation 1 Study, the CBCL has been completed by the child's main carer (usually mother) at ages 2, 3½, 5, 9½ and 12 years. In this thesis, only measurements up to 9½ years have been used as examples for the use of group-based trajectory models, as at the time of commencing analyses the 12 year data were not available.

1.2 Thesis aims

The overarching aim of this thesis is to understand the impact of violations of the assumptions underlying group-based trajectory models, and to provide a better understanding of their effectiveness for analysts and applied researchers who wish to use them. As this is a broad goal, I determined the following four specific aims that were motivated by consideration of the application of these models to the Generation 1 externalising data. I aimed to:

1. Investigate the effect of covariance misspecification on misclassification of trajectories in group-based trajectory models, including data with non-stationary covariance structure;

2. Explore the impact of outliers on group-based trajectory models, in terms of the number of groups estimated and the correct classification rate.

3. Develop an algorithm to identify outliers in the group-based trajectory modelling context, and to determine its effectiveness; and

4. Compare the performance of methods that estimate the effect of covariates on the group membership probabilities in group-based trajectory models.

1.3 Thesis outline

The remainder of this thesis is structured as follows. Chapter 2 details the formulation of group-based trajectory models and how the number of groups may be selected. The literature relevant to each of the four aims is then reviewed, providing motivation for each area of research. Chapter 3 addresses Aim 1 through a simulation study, and contains the first two publications arising from this thesis. Chapter 4 concerns outliers and group-based trajectory modelling, and addresses Aims 2 and 3. Chapter 5 contains the final publication arising from this thesis, and addresses Aim 4. Finally, Chapter 6 summarises the key results of this thesis, discusses limitations and potential future research, and presents some concluding remarks.

Chapter 2

Literature review and aims

In this literature review I will first provide an overview of three methods for modelling group-based trajectories. Second, I will describe methods that have been used to choose the number of groups in a group-based trajectory model. Finally, I will describe the relevant research and motivation for each of the aims of this thesis, as detailed in Section 1.2.

2.1 Group-based trajectory models

The two most commonly used methods for modelling group-based trajectories are latent class growth analysis (LCGA) and growth mixture modelling (GMM).[23] Additionally, multivariate Gaussian mixture modelling (MGMM) can be applied to model group-based trajectories as such an approach is a more general framework than either LCGA or GMM. All three methods are based on finite mixture mod-

elling, so that for K groups the marginal probability distribution of a randomly chosen trajectory is modelled by

$$P(\boldsymbol{y}) = \sum_{k=1}^{K} \pi_k P^k(\boldsymbol{y}),$$

where $P^k(\boldsymbol{y})$ is the conditional distribution of the trajectory, $\boldsymbol{y}$, given the individual is in group k (which I denote as $G = k$), and π_k is the group membership probability ($\pi_k = P(G = k)$) such that $\pi_k > 0$ for $k = 1, \ldots, K$ and $\sum_{k=1}^{K} \pi_k = 1$.

Finite mixture modelling requires K to be specified. In practice, the number of groups is rarely clear from the data. Methods to select the number of groups are described in Section 2.3.

Each of the probability distributions $P^k(\boldsymbol{y})$ are assumed to be multivariate Gaussian, as is common in these models for continuous outcomes. In what follows, I take the set of time points of observation to be the same for all subjects, so that for subject i in group k

$$\boldsymbol{y_i} \sim MVN(\boldsymbol{\mu}_i^k, \boldsymbol{\Sigma}^k) \tag{2.1.1}$$

where the $\boldsymbol{y_i}$ are the T responses for subject i, where $i = 1, \ldots, n$, and $\boldsymbol{\mu}_i^k$ and $\boldsymbol{\Sigma}^k$ are the mean vector of length T and $T \times T$ covariance matrix for group k. In this framework, all of the models can be seen to be special cases of MGMM, and constraints placed on $\boldsymbol{\mu}_i^k$ and $\boldsymbol{\Sigma}^k$ determine whether the method is LCGA or GMM, and whether the assumptions of conditional independence, homogeneity and stationarity apply.

Linear LCGA and GMM models are described below, but both can easily be extended to higher order polynomials if necessary. The inclusion of polynomial terms allows for non-linearity in the expected trajectories.

2.1.1 LCGA

The LCGA models[5] are all defined by taking:

$$\boldsymbol{\mu}_i^k = \alpha^k + \beta^k \boldsymbol{t}, \text{ and}$$

$$\boldsymbol{\Sigma}^k \text{ to be diagonal,}$$

where α^k and β^k are the group k intercept and slope respectively and $\boldsymbol{t}$ is the vector of time points of observation. The restriction that $\boldsymbol{\Sigma}^k$ is diagonal implies that conditional independence is assumed for all LCGA models.

LCGA models can be considered to have four different specifications of further constraints on $\boldsymbol{\Sigma}^k$, depending on whether the variances are assumed to be equal across times or between groups. I will describe these as follows, with L1 to L4 used to identify the four models in subsequent text:

L1: $\boldsymbol{\Sigma}^1 = \ldots = \boldsymbol{\Sigma}^K = \sigma^2 \boldsymbol{I}$, i.e. assuming equal residual variances between groups and across times;

L2: $\boldsymbol{\Sigma}^1 = \ldots = \boldsymbol{\Sigma}^K = \boldsymbol{\Sigma}$, with $\boldsymbol{\Sigma}$ diagonal, i.e. assuming equal residual variances between groups;

L3: $\boldsymbol{\Sigma}^1 = \sigma_1^2\boldsymbol{I}, \ldots, \boldsymbol{\Sigma}^K = \sigma_K^2\boldsymbol{I}$, i.e. assuming equal residual variances across times; and

L4: $\boldsymbol{\Sigma}^1, \ldots, \boldsymbol{\Sigma}^K$ with unconstrained diagonal elements, i.e. unconstrained residual variances.

Models L1 and L3 are stationary, while L1 and L2 have homogeneous variances between groups.

The LCGA model can be expanded in various ways to model more complicated relationships among longitudinal variables. Baseline covariates can be allowed to impact the group membership probabilities by relating the π_k to the predictors through polytomous regression, as described in Section 2.6. The effect of interventions occurring during the course of the trajectories can be estimated by adding time-varying covariates to the polynomial mean equation. The joint probabilities of membership in trajectories of two or more related outcomes can also be modelled in joint trajectory analysis.[24]

2.1.2 GMM

GMM has the same mean trajectories as LCGA but allows for correlation between the observations of individuals through the inclusion of random effects.[8,9] A GMM model can be specified by taking:

$$\boldsymbol{y}_i^k = \alpha^k + a_i^k + \beta^k\boldsymbol{t}_i + b_i^k\boldsymbol{t}_i + \boldsymbol{\epsilon}_i^k,$$

where $\boldsymbol{\epsilon}_i^k \sim MVN(0, \boldsymbol{R}^k)$ with $\boldsymbol{R}^k$ diagonal, and $(a_i^k, b_i^k)' \sim MVN(0, \boldsymbol{D}^k)$. Equivalently:

$$\boldsymbol{\mu}_i^k = \alpha^k + \beta^k \boldsymbol{t}, \text{ and}$$

$$\boldsymbol{\Sigma}^k = \boldsymbol{R}^k + \boldsymbol{Z}\boldsymbol{D}^k\boldsymbol{Z}',$$

where $\boldsymbol{Z} = [\mathbf{1}\ \boldsymbol{t}]$ and $\boldsymbol{Z}'$ is the transpose of $\boldsymbol{Z}$.

Unlike the LCGA model which is conditionally independent, non-zero covariances between measurements in the same group are implied by the GMM model. Although $\boldsymbol{R}^k$ is restricted to be diagonal, the random effects allow for dependence over time for individuals within the same group. Therefore this more complex model overcomes the conceptual difficulty with LCGA that measurements on individuals in the same group are exchangeable. As with more complex models in general, the GMM model is less biased, but has greater variance than the LCGA model. Through the added complexity of the group structures, a GMM model can possibly require a smaller number of groups than an LCGA model.

Estimation for these models involve iterative methods. The added complexity of the GMM model makes it more computationally intensive and more likely to suffer from problems with convergence than the LCGA model. Although the focus of this thesis is not on the computational aspects, but rather the statistical aspects of these models, analysts must be aware of potential issues with model fit for GMMs.

As for $\boldsymbol{\Sigma}^k$ in LCGA models, $\boldsymbol{R}^k$ can be considered to have four different specifications of further constraints, depending on assumptions concerning the equality of

variances across times or between groups. Additionally the covariance matrix of the random effects, $\boldsymbol{D}^k$, can be restricted to be equal between groups. The four models with the restriction $\boldsymbol{D}^1 = \ldots = \boldsymbol{D}^K = \boldsymbol{D}$ (denoted here by GA1-GA4) have the following constraints on $\boldsymbol{R}^k$:

GA1: $\boldsymbol{R}^1 = \ldots = \boldsymbol{R}^K = r^2\boldsymbol{I}$, i.e. assuming equal residual variances between groups and across times;

GA2: $\boldsymbol{R}^1 = \ldots = \boldsymbol{R}^K = \boldsymbol{R}$, with $\boldsymbol{R}$ diagonal, i.e. assuming equal residual variances between groups;

GA3: $\boldsymbol{R}^1 = r_1^2\boldsymbol{I}, \ldots, \boldsymbol{R}^K = r_K^2\boldsymbol{I}$, i.e. assuming equal residual variances across times; and

GA4: $\boldsymbol{R}^1, \ldots, \boldsymbol{R}^K$ with unconstrained diagonal elements, i.e. unconstrained residual variances.

The four GMM models with $\boldsymbol{D}^k$ free to vary between groups (GB1-GB4) have the following constraints on $\boldsymbol{R}^k$:

GB1: $\boldsymbol{R}^1 = \ldots = \boldsymbol{R}^K = r^2\boldsymbol{I}$, i.e. assuming equal residual variances between groups and across times;

GB2: $\boldsymbol{R}^1 = \ldots = \boldsymbol{R}^K = \boldsymbol{R}$, with $\boldsymbol{R}$ diagonal, i.e. assuming equal residual variances between groups;

GB3: $\boldsymbol{R}^1 = r_1^2\boldsymbol{I}, \ldots, \boldsymbol{R}^K = r_K^2\boldsymbol{I}$, i.e. assuming equal residual variances across times; and

GB4: $\boldsymbol{R}^1, \ldots, \boldsymbol{R}^K$ with unconstrained diagonal elements, i.e. unconstrained residual variances.

2.1.3 MGMM

MGMM is the most general approach I consider, and as mentioned LCGA and GMM are both special cases of MGMM. MGMM has been applied in wide variety of fields, including astronomy, biology, economics, engineering, genetics, marketing, medicine and psychiatry.[25] In this case I take $\boldsymbol{\mu}_i^k$ to be unrestricted and consider ten restrictions for $\boldsymbol{\Sigma}^k$ (denoted here by M1-M10). The first six covariance restrictions are based directly on the $\boldsymbol{\Sigma}^k$ matrices:

M1: $\boldsymbol{\Sigma}^1 = \ldots = \boldsymbol{\Sigma}^K = \sigma^2\boldsymbol{I}$, i.e. assuming equal variances between groups and across times, with no covariance;

M2: $\boldsymbol{\Sigma}^1 = \ldots = \boldsymbol{\Sigma}^K$ diagonal, i.e. assuming equal variances between groups, with no covariance;

M3: $\boldsymbol{\Sigma}^1 = \sigma_1^2\boldsymbol{I}, \ldots, \boldsymbol{\Sigma}^K = \sigma_K^2\boldsymbol{I}$, i.e. assuming equal variances across times, with no covariance;

M4: $\boldsymbol{\Sigma}^1, \ldots, \boldsymbol{\Sigma}^K$ all diagonal, i.e. unconstrained variances, with no covariance;

M5: $\boldsymbol{\Sigma}^1 = \ldots = \boldsymbol{\Sigma}^K$ unstructured, i.e. assuming equal variance and covariances between groups;

M6: $\boldsymbol{\Sigma}^1, \ldots, \boldsymbol{\Sigma}^K$ unstructured, i.e. unconstrained variances and covariances.

The final four covariance restrictions are based on the following eigenvalue decomposition of $\boldsymbol{\Sigma}^k$:

$$\boldsymbol{\Sigma}^k = \lambda_k \boldsymbol{Q}_k \boldsymbol{A}_k \boldsymbol{Q}_k',$$

with $\boldsymbol{Q}_k$ an orthogonal matrix of eigenvectors, $\boldsymbol{A}_k$ a diagonal matrix such that $|\boldsymbol{A}_k| = 1$, whose elements are the normalised eigenvalues of $\boldsymbol{\Sigma}^k$, and $\lambda_k = |\boldsymbol{\Sigma}^k|^{1/T}$. T is the number of time points. Where $\boldsymbol{A}_k$, $\boldsymbol{Q}_k$ and λ_k are used without a subscript (i.e. $\boldsymbol{A}$, $\boldsymbol{Q}$ and λ), this implies they are constrained to be equal between groups. This eigenvalue decomposition allows the volume, shape and orientation of the groups to be controlled independently of each other. Figure 2.1.1 shows an example set of 95% probability ellipses for each of models M7-M10 to illustrate the concepts of volume, shape and orientation. The final four covariance restrictions are:

M7: $\boldsymbol{\Sigma}^1 = \lambda_1 \boldsymbol{A}, \ldots, \boldsymbol{\Sigma}^K = \lambda_K \boldsymbol{A}$, i.e. assuming equal shape between groups, with variable volume and no covariance;

M8: $\boldsymbol{\Sigma}^1 = \lambda \boldsymbol{A}_1, \ldots, \boldsymbol{\Sigma}^K = \lambda \boldsymbol{A}_K$, i.e. assuming equal volume between groups, with variable shape and no covariance;

M9: $\boldsymbol{\Sigma}^1 = \lambda \boldsymbol{Q}_1 \boldsymbol{A} \boldsymbol{Q}_1', \ldots, \boldsymbol{\Sigma}^K = \lambda \boldsymbol{Q}_K \boldsymbol{A} \boldsymbol{Q}_K'$, i.e. assuming equal shape and volume between groups, but variable orientation; and

M10: $\boldsymbol{\Sigma}^1 = \lambda_1 \boldsymbol{Q}_1 \boldsymbol{A} \boldsymbol{Q}_1', \ldots, \boldsymbol{\Sigma}^K = \lambda_K \boldsymbol{Q}_K \boldsymbol{A} \boldsymbol{Q}_K'$, i.e. assuming equal shape between groups, but variable volume and orientation,

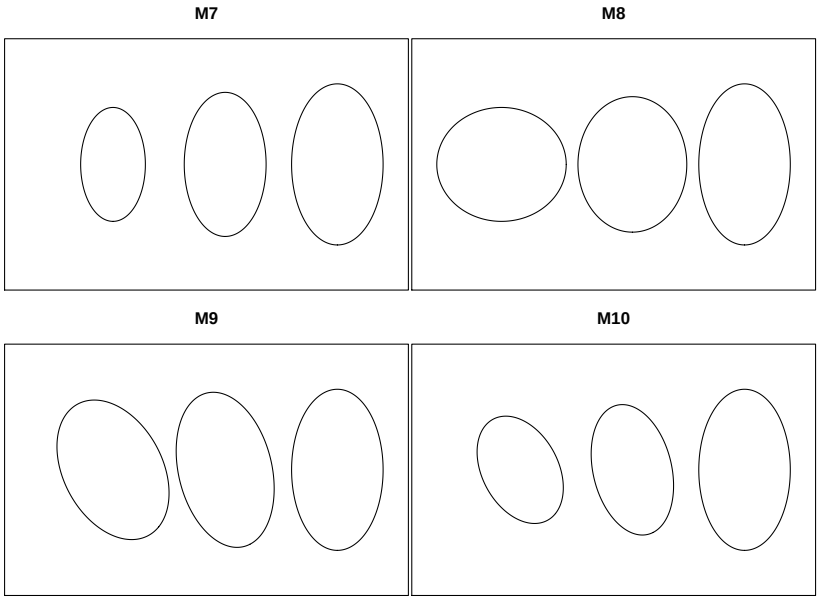


Figure 2.1.1: Examples of 95% probability ellipses for models M7-M10.

2.1.4 Summary of models

Table 2.1.1 summarises the characteristics of the different models, according to whether they exhibit conditional independence, homogeneous $\boldsymbol{\Sigma}^k$ between groups or stationary $\boldsymbol{\Sigma}^k$ over time. For the GMM models, Table 2.1.1 also shows whether the variance components $\boldsymbol{R}^k$ and $\boldsymbol{D}^k$ are homogeneous between groups.

2.2 Estimation of group-based trajectory models

Estimation of the parameters in a group-based trajectory model is typically performed using maximum-likelihood, as is the case for all finite mixture models.[5,8,25] As the maximum-likelihood estimates of the group-membership probabilities and group means and variances cannot be written in closed form, they need to be com-

Table 2.1.1: Characteristics of models, according to different properties of their covariance matrix or variance components.

Model	All models			GMM variance components	
	Conditional Independence	Homogeneous $\boldsymbol{\Sigma}^k$	Stationary $\boldsymbol{\Sigma}^k$	Homogeneous $\boldsymbol{R}^k$	Homogeneous $\boldsymbol{D}^k$
L1	✓	✓	✓	-	-
L2	✓	✓	×	-	-
L3	✓	×	✓	-	-
L4	✓	×	×	-	-
GA1	×	✓	✓	✓	✓
GA2	×	✓	×	✓	✓
GA3	×	×	✓	×	✓
GA4	×	×	×	×	✓
GB1	×	×	✓	✓	×
GB2	×	×	×	✓	×
GB3	×	×	✓	×	×
GB4	×	×	×	×	×
M1	✓	✓	✓	-	-
M2	✓	✓	×	-	-
M3	✓	×	✓	-	-
M4	✓	×	×	-	-
M5	×	✓	×	-	-
M6	×	×	×	-	-
M7	✓	×	×	-	-
M8	✓	×	×	-	-
M9	×	×	×	-	-
M10	×	×	×	-	-

puted iteratively.[25] Advances in the estimation of finite mixture models were made possible by the development of the Expectation-Maximisation (EM) algorithm by Dempster, Laird and Rubin in 1977.[26] The EM algorithm is used to compute the maximum-likelihood estimates in almost all applications of finite mixture modelling, although Bayesian estimation is also possible.[25]

In this thesis the `mclust` package in R[27,28] and Mplus[9] were both used to estimate the parameters of group-based trajectory models. Mclust uses the EM algorithm for maximum-likelihood estimation[27]. In Mplus, the estimator used for the group-based trajectory modelling was a maximum-likelihood estimator with robust standard errors, with an accelerated EM algorithm that utilises Quasi-Newton and Fisher Scoring optimisation steps when needed.[9]

Estimation of finite mixture models can be sensitive to starting points as the mixture likelihood functions often have local maxima. Therefore multiple random starts are needed to reduce the likelihood of finding a local optimum as the solution.

2.3 Selection of the number of groups

As described in the previous section, LCGA, GMM and MGMM rely on a given number of groups K. This is rarely obvious from the data, and the methods described in this section can be used to guide the selection of the value of K.

There has been considerable discussion in the literature around methods to estimate the appropriate number of groups K.[29] In other modelling contexts, more

complex models can be tested against simpler models through the likelihood ratio test (LRT). The LRT statistic is defined as twice the difference between the maximised log likelihood of the two models under consideration. For many applications, the LRT statistic has an asymptotic chi-squared distribution, however for mixed models the necessary regularity conditions do not hold.[25] This is because the null hypothesis involves the membership probability being on the boundary of the parameter space.

An alternative approach is to assess how a measure of model fit changes for different values of K. The chosen K is that which results in the optimum value of the measure of model fit. For mixture modelling, the Bayesian information criterion (BIC)[30] is commonly used to compare between models, as it trades complexity against model fit. The BIC is defined as

$$\text{BIC} = -2\log(L) + K\log(n), \tag{2.3.1}$$

where L is the value of the maximised likelihood. The derivation of equation (2.3.1) uses the Bayesian framework with equal prior probability on each model and vague priors on the parameters, given the model.[31] The difference between two BIC values is then an estimate of the ratio of the posterior probabilities of the two models. In this situation a BIC difference of 6 corresponds to a posterior probability of about 0.95 for the model with the higher BIC, relative to the model with the lower BIC. For this reason, Raftery suggests considering only the models with a BIC within 6

of the best fitting model, and also for parsimony to remove any models that have more likely sub-models nested within them.[31]

An alternative method to using the BIC is to use bootstrap samples to estimate the distribution of the LRT statistic, as described by McLachlan and Peel.[25] This avoids the assumption of the chi-squared distribution, as the distribution is estimated empirically. An advantage of this bootstrap method over the BIC is that it provides p-values for comparisons between models. Whereas a disadvantage is that the bootstrap method is more computationally intensive. A recent simulation study showed that the bootstrap LRT performed better than the BIC and other criteria for selecting the true number of groups.[32]

In addition to assessing the criteria described above, it has been suggested by Jung and Wickrama that when choosing the number of groups, it is important to consider the interpretability of the groups.[23] When there are two candidate numbers of groups with similar model fit, these authors advise analysts to consider whether the additional groups reveal any new important features in the data.

2.4 Covariance assumptions in group-based trajectory models

In group-based trajectory models, simplifying assumptions such as conditional independence, homogeneous variance between groups, or stationary variance over time are often made.

However, little research into the sensitivity to misspecification of these assumptions has been undertaken. Heggeseth and Jewell investigated the effect of covariance misspecification for models with an assumption of stationarity.[33] They showed that bias in estimates of model parameters occurs due to misspecification unless the component distributions are well separated. The bias in parameter estimates resulting from misspecifying heterogeneous variances as homogeneous has previously been found in univariate mixture models,[34,35] and in growth mixture models.[36] Gilthorpe et al. and Diallo et al. both recently investigated the effect of misspecifying the covariance structure in growth mixture models on the choice of the number of groups, and identified that models with more groups were favoured in misspecified situations.[37,38]

2.4.1 Aim 1

In previous literature, the effect of violations of assumptions of stationarity have not been investigated, and the extent of the impact of such violations has not been compared with that of violations of homogeneity. The first aim of this thesis is to investigate the effect of covariance misspecification on misclassification of trajectories in group-based trajectory models, including for data with a non-stationary covariance structure. This aim is addressed in Chapter 3.

2.5 Outlier impact and identification for group-based trajectory models

2.5.1 Outliers and their impact

Outliers are data points that are far from the rest of the distribution. In a univariate or bivariate setting, such observations are often easily identified as points separated from the majority of other observations. In a multivariate setting, such as with repeated observations for a group of individuals, outliers may also occur as more complex patterns of observations. A trajectory may be considered an outlier because a single observation is an outlier with respect to its univariate distribution. Alternatively, an outlier can also occur when the combination of its components are unusual rather than any single component.

Outliers may arise through normal variation, experimental error or contamination.[39] Often statistical distributions allow for observations to appear far from the rest of the observations with low probability. Therefore an outlier may arise through the natural variability in responses that are part of a particular distribution. For example, in a Normal distribution an observation of three or more standard deviations from the mean can be expected with probability 0.0027 (or once in every 371 observations). In a small sample, an outlier far from the rest of the distribution could arise by chance alone. If a large sample is observed, one can expect the sample to include more observations far from the mean, but these extreme observations

are unlikely to be separated from the majority of other observations. Outliers can also arise through experimental errors or other mistakes in data collection. Finally, outliers can be observed when there is contamination in the dataset from a different process. Errors and contamination do not necessarily lead to outliers, as these observations may happen to lie within the normal range of observations.

Outlying observations can distort the results of analyses undertaken when they are present. In 1953 Dixon described the impact that outlying observations can have on summary statistics such as the mean and standard deviation.[40] These results have been extended to multivariate settings where outliers can have a large impact on parameter estimates.[41] Therefore outliers can lead to different conclusions to those made if such observations are excluded.

2.5.2 Aim 2

The effects of outliers on group-based trajectory models have not been studied. The second aim of this thesis is to explore the impact of outliers on these models, in terms of the number of groups estimated and the correct classification rate. This aim is addressed in Section 4.2.

2.5.3 Outlier identification

As outliers can have an impact on the results of analyses, it is important that they are identified for further investigation. Once identified, an outlier can be removed if there is a substantive reason for doing so, such as additional information that

indicates a likely error. Alternatively, where outliers are plausibly true observations, their impact should be investigated through a sensitivity analysis. This involves comparing the results with and without the outliers included, or comparing the results with that of a robust method, when available. In the mixture modelling setting, an additional complexity is that besides outlying observations potentially being errors or arising from natural variation, they can also be thought of as arising from a sub-population that corresponds to a new group.

The topic of outlier identification has been extensively discussed in the extant literature for many decades and in a variety of statistical contexts and subject domains. There are now a large range of methods that exist for the detection of outliers in the statistical and computer science literature including, for example, statistical methods (both parametric and nonparametric), clustering based methods, and classification based methods. Recent comprehensive reviews of this subject are available.[39,42] A general review of these methods is beyond the scope of this thesis. As this thesis relates to group-based trajectory models, I will consider only outlier methods applicable to mixture models.

In the context of mixture models, there have been relatively few methods proposed for dealing with outliers. One approach has been to provide a single contamination component of the mixture specifically for outliers, with a different distribution to the rest of the components.[43,44] Alternatively, Zhuang et al. proposed separating the data into a normal distribution and 'outliers' repeatedly until all normal distributions have been found.[45] Another approach is to use mixture modelling

procedures that are robust to outliers, so as to limit the impact of the outliers on resulting estimates. One robust procedure is to use mixtures of distributions with heavy tails, such as multivariate t or skew t distributions.[46–50] Other procedures that have been proposed are to either trim outliers from the mixture components, or to 'grow' the mixture components in size until they reach a point where only outliers are excluded.[51,52]

One possible approach to identify outliers is to use a model selection criterion. The method proposed by Kitagawa, for example, uses the Akaike information criterion (AIC).[53] Kitagawa's method applies to univariate data, and uses the AIC to compare models where the highest and/or lowest values arise from distributions with a different mean or variance. This is achieved through a distribution function defined as three sub-functions, one for each of the lowest observations, the majority of the observations and the highest observations. The AIC is calculated for various combinations of numbers of high and low outliers, and the minimum AIC combination indicates which observations should be considered as outliers. Kadota extended Kitagawa's method to apply to two-dimensional data,[54] however these single distribution methods are not suited to the mixture modelling setting as there are too many possible alternate distributions from which outliers may have arisen to make them feasible. Outliers may be considered to come from any of the estimated mixture components, or from a new component with an increase in the number of groups. As methods to date do not recognise the possibility of the number of groups changing, this led to the development of the following aim.

2.5.4 Aim 3

Methods proposed for outlier identification in mixture models do not account for the number of groups changing with the identification of outliers. The third aim of this thesis is to develop an algorithm to identify outliers in the group-based trajectory modelling context, and to determine its effectiveness. This aim is addressed in Section 4.3.

2.6 Predictors of group-based trajectory models

2.6.1 Modelling the effect of covariates on group-based trajectory models

Longitudinal studies can provide an understanding of how characteristics at earlier points in time are related to subsequent outcomes. One approach to modelling the influence of covariates measured at earlier time points in the group-based setting is to consider models wherein these covariates affect the group membership probabilities.[55] Models in which prior covariates impact the shape of the expected trajectories directly are also possible, but are not considered in this thesis.

In a group-based model without covariates, π_k is modelled as:

$$P(G = k) = \pi_k = \frac{e^{\theta_k}}{\sum_{k=1}^{K} e^{\theta_k}}.$$

This ensures $0 < \pi_k < 1$ and $\sum_{k=1}^{K} \pi_k = 1$, while the θ_k are free to vary without restriction. As a result of the summation constraint, only $K - 1$ estimates of θ_k are required, and by convention $\theta_1 = 0$.

The effect of a covariate vector $\boldsymbol{x} = (x_1, \ldots, x_p)$ can be incorporated through the multinomial logistic regression model:

$$P(G = k|\boldsymbol{x}) = \frac{e^{\theta_{0k} + \boldsymbol{x}^T \boldsymbol{\theta}_k}}{\sum_{l=1}^{K} e^{\theta_{0l} + \boldsymbol{x}^T \boldsymbol{\theta}_l}}, \tag{2.6.1}$$

where $\boldsymbol{\theta}_k = (\theta_{1k}, \ldots, \theta_{pk})$. The effect of x_1 in group k relative to group 1 is estimated by θ_{1k}, with $e^{\theta_{1k}}$ providing an estimate of the odds ratio.

2.6.2 Methods for estimating the effect of covariates on group membership probabilities

A number of ways to estimate the multinomial logistic regression on the covariates have been proposed, with different approaches to account for the uncertainty in the group membership assignment. Here I outline six different methods that have been proposed in the wider literature:

1-step (1S)

In the 1S method,[56–58] estimation of the effects of covariates and the trajectory model are carried out simultaneously. That is to say, full maximum-likelihood is used for all parameters of the group-based model ($\boldsymbol{\mu}^k$, $\boldsymbol{\Sigma}^k$) and the θs of the multi-

nomial logistic regression. While this is a reasonable approach from an estimation perspective, there are various reasons why researchers have preferred other methods in which the group-based model is estimated first, and the effects of the covariates are assessed subsequently. As argued by Vermunt,[58] the 1S method may be impractical if there are many covariates to consider, in combination with a complex group-based trajectory model that would need to be re-estimated for each covariate. Vermunt also reasons that the 1S method complicates model building decisions, may not fit with the logic of applied researchers who view introducing covariates as a step that comes after the classification model has been built, and assumes the group-based model has not yet been constructed. According to Vermunt 'in many applications, it is more natural to use a stepwise approach and, moreover, ... sometimes it is the only reasonable way to proceed'.[58] The remaining methods considered in this thesis are such stepwise approaches, where the effect of the covariate is estimated subsequent to the estimation of the group-based model.

3-step (3S)

In the 3S method,[57,58] the trajectory model is first estimated without inclusion of covariates. In the second step, each individual is allocated to groups according to the maximum estimated posterior probability:

$$m_i = \arg\max_k \widehat{P}(G_i = k|\boldsymbol{y}_i).$$

In the third step, a multinomial logistic regression model, as in equation 2.6.1, is estimated using maximum-likelihood, with m_i as the response instead of G_i.

Pseudo class 3-step (PC3S)

The PC3S method[59,60] is similar to the 3S method, however at the second step, individuals are allocated randomly to groups J times according to the posterior probabilities:

$$P(m_{ij} = k) = \widehat{P}(G_i = k|\boldsymbol{y}_i),$$

for $j = 1, \ldots, J$. As in the 3S method, multinomial logistic regressions with m_{ij} as the response are estimated with maximum-likelihood to obtain J estimates of the covariate effect, with their average providing the PC3S estimate.

Improved 3-step (I3S)

The first two steps of the I3S method[57,58] are the same as for the 3S method. The third step differs, as it takes into account the misclassification error in the second step when individuals were allocated to m_i according to the maximum estimated posterior probability. This is achieved through a latent class model where the estimated classification errors are treated as known errors of classification. To estimate the classification error in M, $\widehat{P}(M = k|G = l)$, the classification uncertainty rate for M is calculated as:

$$\widehat{P}(G = l|M = k) = \frac{1}{n_k} \sum_{m_i = k} \widehat{P}(G_i = l|\boldsymbol{y}_i),$$

where n_k is the number of observations classified in group k by the most likely class variable M. The classification uncertainty rate is the average of the posterior probabilities for each of the classes among each of the allocated classes. Bayes' Theorem is then used to calculate the classification error in M:

$$\widehat{P}(M = k|G = l) = \frac{\widehat{P}(G = l|M = k)n_k}{\sum_{s=1}^{K} \widehat{P}(G = l|M = s)n_s}.$$

These classification errors allow M to be treated as an imperfect measurement of G in a latent class model, with a regression of G on x, and in this way the measurement error in M is taken into account. This involves maximising the log-likelihood function:

$$\sum_{i=1}^{n} \sum_{k=1}^{K} \log \sum_{l=1}^{K} P(G_i = l|\boldsymbol{x}_i)P(M = k|G = l).$$

This approach was first described by Vermunt,[58] and expanded on improvements made to the 3S method by Bolck, Croon and Hagenaars.[57] More details are provided in Vermunt's paper.[58]

Probability regression (PR)

In the PR method,[59] the posterior probabilities from step one of the 3S method are transformed using the logit function, and the transformed probabilities are subsequently used in a linear regression on the covariate x. That is, the following linear

regression is performed:

$$\log \frac{\widehat{P}(G_i = k|\boldsymbol{y}_i)}{1 - \widehat{P}(G_i = k|\boldsymbol{y}_i)} = \gamma_{0k} + \gamma_k x_i.$$

For two groups, the effect of x in group k relative to group 1 is estimated by γ_k, with e^{γ_k} providing an estimate of the odds ratio.

Probability weighted 3-step (PW3S)

The PW3S method[59] is based on the 3S method, however the posterior probabilities of membership of the chosen class, $\widehat{P}(m_i = k|\boldsymbol{y}_i)$, are used as weights in the multinomial logistic regression of step three, thereby accounting for the differing certainty in the assignments.

2.6.3 Previous studies comparing methods for estimating the effect of covariates

Bolck *et al.* compared the performance of the 1S, 3S and the I3S methods in their study introducing the I3S method,[57] and demonstrated both analytically and through simulation that the classification error in the second step of the 3S method leads to attenuation of parameter estimates. This was confirmed by Vermunt in a study extending the I3S method,[58] in which the extension was also compared with the other methods presented by Bolck *et al.*. In a simulation study, Clark and Muthén compared all methods considered here except for I3S, and found that the

1S method performed best.[59] The studies by Bolck *et al.* and Clark and Muthén considered only a single continuous covariate,[57,59] while Vermunt used three discrete numeric covariates but did not compare with results for a single covariate.[58] Thus it remains unclear which is the best method for estimating the effect of one or multiple covariates on the group membership probabilities.

2.6.4 Aim 4

No study has considered all six of these methods, nor compared how the inclusion of additional covariates affects resulting bias. The nature of the covariates, in terms of whether they are continuous or categorical, has also yet to be considered. The fourth aim of this thesis is to compare the performance of methods that estimate the effect of covariates on the group membership probabilities in group-based trajectory models. This aim is addressed in Chapter 5.

2.7 Conclusion

There are a number of areas where questions remain to be addressed in the group-based trajectory modelling literature, as described above. These identified gaps led to the development of four aims for this thesis. In order to better our understanding of these models, I will address these aims in the following three chapters.

Chapter 3

The impact of covariance misspecification in group-based trajectory models for longitudinal data with non-stationary covariance structure

3.1 Preface

This chapter contains the first of two articles contributing to this thesis which have been published in peer reviewed journals:

- Davies CE, Glonek GFV and Giles LC. The impact of covariance misspecification in group-based trajectory models for longitudinal data with non-stationary covariance structure. *Statistical Methods in Medical Research* 2017; 26: 1982-1991. First published date: August 17, 2015.

It addresses the first aim of this thesis by examining the impact of covariance misspecification on misclassification of trajectories in group-based trajectory models, including models for longitudinal data with non-stationary covariance structure.

In the range of group-based trajectory models described in Section 2.1, and summarised in Table 2.1.1, many models make the common assumptions of conditional independence, homogeneous variance between groups and stationary variance over time. However, there has been limited investigation into the sensitivity to misspecification of these assumptions. The importance of understanding the impact that misspecification of these assumptions can have on one's ability to correctly classify individuals into groups is shown in this chapter.

Of the covariance structures described in Section 2.1, models M7-M10 were not considered in this chapter as they did not correspond to the range of LCGA and GMM models also considered.

Supplementary Figure 1 referred to in the article is reproduced in Section 3.4.

Section 3.6 contains a letter to the editor published subsequently in the same journal relating to the above article:

- Davies CE, Glonek GFV and Giles LC. Letter to the editor. *Statistical Methods*

in Medical Research. Prepublished May 24, 2017.

I identified through my ongoing research some potential issues with the calculation of the estimates for the article. The impact of correcting these issues is elaborated on in the letter.

3.2 Statement of authorship

Statement of Authorship

Title of Paper	The impact of covariance misspecification in group-based trajectory models for longitudinal data with non-stationary covariance structure.
Publication Status	☑ Published ☐ Accepted for Publication ☐ Submitted for Publication ☐ Unpublished and Unsubmitted work written in manuscript style
Publication Details	Davies CE, Glonek GFV and Giles LC. The impact of covariance misspecification in group-based trajectory models for longitudinal data with non-stationary covariance structure. Statistical Methods in Medical Research 2017; 26: 1982-1991. First published date August 17, 2015. DOI: 10.1177/0962280215598806. <http://journals.sagepub.com/doi/10.1177/0962280215598806> The final, definitive version of this paper has been published in Statistical Methods in Medical Research, 2017 by SAGE Publications Ltd, All rights reserved. Copyright © 2015 Davies CE, Glonek GFV and Giles LC. Reprinted by permission of SAGE Publications.

Principal Author

Name of Principal Author (Candidate)	Christopher E Davies		
Contribution to the Paper	Designed the study, simulated the data, performed all analyses, interpreted the results, drafted the manuscript and acted as corresponding author.		
Overall percentage (%)	80		
Certification:	This paper reports on original research I conducted during the period of my Higher Degree by Research candidature and is not subject to any obligations or contractual agreements with a third party that would constrain its inclusion in this thesis. I am the primary author of this paper.		
Signature		Date	06/12/17

Co-Author Contributions

By signing the Statement of Authorship, each author certifies that:

i. the candidate's stated contribution to the publication is accurate (as detailed above);
ii. permission is granted for the candidate in include the publication in the thesis; and
iii. the sum of all co-author contributions is equal to 100% less the candidate's stated contribution.

Name of Co-Author	Gary FV Glonek		
Contribution to the Paper	Contributed to the design of the study and interpretation of the results, and reviewed the manuscript.		
Signature		Date	06/12/17

Name of Co-Author	Lynne C Giles		
Contribution to the Paper	Contributed to the design of the study and interpretation of the results, and reviewed the manuscript.		
Signature		Date	06/12/17

3.3 Article

Article

SMMR
STATISTICAL METHODS IN MEDICAL RESEARCH

Statistical Methods in Medical Research
2017, Vol. 26(4) 1982–1991

DOI: 10.1177/0962280215598806
journals.sagepub.com/home/smm
SAGE

The impact of covariance misspecification in group-based trajectory models for longitudinal data with non-stationary covariance structure

Christopher E Davies,[1,2] Gary FV Glonek[1] and Lynne C Giles[2]

Abstract

One purpose of a longitudinal study is to gain a better understanding of how an outcome of interest changes among a given population over time. In what follows, a trajectory will be taken to mean the series of measurements of the outcome variable for an individual. Group-based trajectory modelling methods seek to identify subgroups of trajectories within a population, such that trajectories that are grouped together are more similar to each other than to trajectories in distinct groups. Group-based trajectory models generally assume a certain structure in the covariances between measurements, for example conditional independence, homogeneous variance between groups or stationary variance over time. Violations of these assumptions could be expected to result in poor model performance. We used simulation to investigate the effect of covariance misspecification on misclassification of trajectories in commonly used models under a range of scenarios. To do this we defined a measure of performance relative to the ideal Bayesian correct classification rate. We found that the more complex models generally performed better over a range of scenarios. In particular, incorrectly specified covariance matrices could significantly bias the results but using models with a correct but more complicated than necessary covariance matrix incurred little cost.

Keywords

Covariance, model misspecification, mixture models, longitudinal data, group-based trajectory modelling

1 Introduction

Longitudinal studies can give insights into how outcomes of interest change over time. Data arising from such studies can be used to define discrete paths of measurements, or trajectories, for each individual within a given population. Group-based trajectory modelling methods seek to identify subgroups of individuals within a population with trajectories that are more similar to each other than to trajectories in distinct groups.

In group-based trajectory models, assumptions of conditional independence, homogeneous variance between groups and stationary variance over time are commonly made. However, there has been limited investigation into the sensitivity to misspecification of these assumptions. Heggeseth and Jewell assessed the effect of covariance misspecification for models with an assumption of stationarity.[1] They found that bias in estimates of model parameters occurs due to misspecification unless the component distributions are well separated. The bias in parameter estimates resulting from misspecifying heterogeneous variances as homogeneous has previously been identified in univariate mixture models[2,3] and in growth mixture models.[4] Gilthorpe et al. recently investigated the effect of misspecifying the random structure in growth mixture models on the choice of the number of groups, and showed that models with more groups were favoured in misspecified situations.[5] However, the effect of violations of assumptions of stationarity has not been investigated, and the extent of the impact has not been compared with that of violations of homogeneity.

[1]School of Mathematical Sciences, The University of Adelaide, Adelaide, Australia
[2]School of Public Health, The University of Adelaide, Adelaide, Australia

Corresponding author:
Christopher E Davies, School of Mathematical Sciences, The University of Adelaide, Adelaide, SA 5005, Australia.
Email: chris.davies@adelaide.edu.au

Here we investigate, through a simulation study, the impact of covariance misspecification on misclassification of trajectories in commonly used models under a range of scenarios. We define a measure of performance relative to the ideal Bayesian correct classification rate (BCCR). This article is structured as follows. Section 2 presents the model specifications including the different covariance structures considered. Section 3 describes the suite of simulations that were undertaken, broadly based on serial measurements of behaviour from a cohort of South Australian children. Section 4 presents the results of the simulations, and the implications of our findings are discussed in Section 5.

2 Model specification

The two most commonly used methods for modelling group-based trajectories are latent class growth analysis (LCGA) and growth mixture modelling (GMM).[6] Additionally, multivariate Gaussian mixture modelling (MGMM) can be applied to model group-based trajectories as it is a more general framework than either LCGA or GMM. All three methods are based on finite mixture modelling, so that for K groups the marginal probability distribution of a randomly chosen trajectory is modelled by

$$P(\boldsymbol{y}) = \sum_{k=1}^{K} \pi_k P^k(\boldsymbol{y})$$

where $P^k(\boldsymbol{y})$ is the conditional distribution of the trajectory, $\boldsymbol{y}$, given the individual is in group k, and π_k is the group membership probability such that $\pi_k > 0$ for $k = 1, \ldots, K$ and $\sum_{k=1}^{K} \pi_k = 1$.

Finite mixture modelling requires K to be specified. In practice, the number of groups is rarely clear from the data. There has been considerable discussion in the literature around methods to estimate the appropriate number of groups.[7] We assume, as in similar studies,[1–3] that the correct number of groups is known.

Each of the probability distributions $P^k(\boldsymbol{y})$ is assumed to be multivariate Gaussian, as is common in these models for continuous outcomes. In what follows, we take the set of time points of observation to be the same for all subjects, so that for subject i in group k

$$\boldsymbol{y_i} \sim MVN(\boldsymbol{\mu}^k, \boldsymbol{\Sigma}^k) \tag{1}$$

where $\boldsymbol{y_i}$ are the responses for subject i, $i = 1, \ldots, n$, and $\boldsymbol{\mu}^k$ and $\boldsymbol{\Sigma}^k$ are the mean vector and covariance matrix for group k. Constraints placed on $\boldsymbol{\mu}^k$ and $\boldsymbol{\Sigma}^k$ determine whether the method is LCGA, GMM or MGMM, and whether the assumptions of conditional independence, homogeneity and stationarity apply.

Linear LCGA and GMM models are described below, but both can easily be extended to higher order polynomials if necessary.

2.1 Latent class growth analysis

The LCGA model[8] can be defined by taking $\boldsymbol{\mu}^k = \alpha^k + \beta^k \boldsymbol{t}$ and $\boldsymbol{\Sigma}^k$ as diagonal, where $\boldsymbol{t}$ is the vector of time points of observation. The assumption that $\boldsymbol{\Sigma}^k$ is diagonal implies conditional independence for all LCGA models.

LCGA models can be considered to have four different specifications of further constraints on $\boldsymbol{\Sigma}^k$, depending on whether the variances are assumed to be equal across times or between groups. We will describe these as: L1 ($\boldsymbol{\Sigma}^1 = \ldots = \boldsymbol{\Sigma}^K = \sigma^2 \boldsymbol{I}$, i.e. assuming equal residual variances between groups and across times), L2 ($\boldsymbol{\Sigma}^1 = \ldots = \boldsymbol{\Sigma}^K = \boldsymbol{\Sigma}$, i.e. assuming equal residual variances between groups), L3 ($\boldsymbol{\Sigma}^1 = \sigma_1^2 \boldsymbol{I}, \ldots, \boldsymbol{\Sigma}^K = \sigma_K^2 \boldsymbol{I}$, i.e. assuming equal residual variances across times) and L4 ($\boldsymbol{\Sigma}^1, \ldots, \boldsymbol{\Sigma}^K$ with unconstrained diagonal elements, i.e. unconstrained residual variances). Models L1 and L3 are stationary, while L1 and L2 have homogeneous variances between groups.

2.2 Growth mixture modelling

GMM extends the mean trajectories of LCGA to allow for the variation of trajectories of individuals through the inclusion of random effects.[9,10] A GMM model is often specified by taking

$$\boldsymbol{y}_i^k = \alpha^k + a_i^k + \beta^k \boldsymbol{t}_i + b_i^k \boldsymbol{t}_i + \boldsymbol{\epsilon}_i^k$$

where $\boldsymbol{\epsilon}_i^k \sim MVN(0, \boldsymbol{R}^k)$ with $\boldsymbol{R}^k$ diagonal, and $(a_i^k, b_i^k)' \sim MVN(0, \boldsymbol{D}^k)$. In our notation, the GMM model is expressed as

$$\boldsymbol{\mu}^k = \alpha^k + \beta^k \boldsymbol{t}, \text{and}$$
$$\boldsymbol{\Sigma}^k = \boldsymbol{R}^k + \boldsymbol{Z}\boldsymbol{D}^k\boldsymbol{Z}'$$

where $\boldsymbol{Z} = [\boldsymbol{1}\ \boldsymbol{x}]$.

Unlike the LCGA model which is conditionally independent, non-zero covariances between measurements in the same group are implied by the GMM model. Although $\boldsymbol{R}^k$ is assumed to be diagonal, the random effects allow for dependence over time for individuals within the same group.

As for $\boldsymbol{\Sigma}^k$ in LCGA models, $\boldsymbol{R}^k$ can be considered to have four different specifications of further constraints, depending on assumptions concerning the equality of variances across times or between groups. Additionally the covariance matrix of the random effects, $\boldsymbol{D}^k$, can be assumed to be equal between groups. The four models with the assumption $\boldsymbol{D}^1 = \ldots = \boldsymbol{D}^K = \boldsymbol{D}$ have the following constraints on $\boldsymbol{R}^k$: GA1 ($\boldsymbol{R}^1 = \ldots = \boldsymbol{R}^K = r^2\boldsymbol{I}$, i.e. assuming equal residual variances between groups and across times), GA2 ($\boldsymbol{R}^1 = \ldots = \boldsymbol{R}^K = \boldsymbol{R}$, i.e. assuming equal residual variances between groups), GA3 ($\boldsymbol{R}^1 = r_1^2\boldsymbol{I}, \ldots, \boldsymbol{R}^K = r_K^2\boldsymbol{I}$, i.e. assuming equal residual variances across times) and GA4 ($\boldsymbol{R}^1, \ldots, \boldsymbol{R}^K$ with unconstrained diagonal elements, i.e. unconstrained residual variances).

The four GMM models with $\boldsymbol{D}^k$ free to vary between groups have the following constraints on $\boldsymbol{R}^k$: GB1 ($\boldsymbol{R}^1 = \ldots = \boldsymbol{R}^K = r^2\boldsymbol{I}$, i.e. assuming equal residual variances between groups and across times), GB2 ($\boldsymbol{R}^1 = \ldots = \boldsymbol{R}^K = \boldsymbol{R}$, i.e. assuming equal residual variances between groups), GB3 ($\boldsymbol{R}^1 = r_1^2\boldsymbol{I}, \ldots, \boldsymbol{R}^K = r_K^2\boldsymbol{I}$, i.e. assuming equal residual variances across times) and GB4 ($\boldsymbol{R}^1, \ldots, \boldsymbol{R}^K$ with unconstrained diagonal elements, i.e. unconstrained residual variances).

2.3 Multivariate Gaussian mixture modelling

MGMM is the most general approach that we consider. In this case, we take $\boldsymbol{\mu}^k$ to be unrestricted and consider the following six assumptions for $\boldsymbol{\Sigma}^k$: M1 ($\boldsymbol{\Sigma}^1 = \ldots = \boldsymbol{\Sigma}^K = \sigma^2\boldsymbol{I}$), M2 ($\boldsymbol{\Sigma}^1 = \ldots = \boldsymbol{\Sigma}^K$ with unconstrained diagonal elements), M3 ($\boldsymbol{\Sigma}^1 = \sigma_1^2\boldsymbol{I}, \ldots, \boldsymbol{\Sigma}^K = \sigma_K^2\boldsymbol{I}$), M4 ($\boldsymbol{\Sigma}^1, \ldots, \boldsymbol{\Sigma}^K$ with unconstrained diagonal elements), M5 ($\boldsymbol{\Sigma}^1 = \ldots = \boldsymbol{\Sigma}^K$ unstructured) and M6 ($\boldsymbol{\Sigma}^1, \ldots, \boldsymbol{\Sigma}^K$ unstructured).

2.4 Summary of models

Table 1 summarises the characteristics of the different models, according to whether they exhibit conditional independence, homogeneous $\boldsymbol{\Sigma}^k$ between groups or stationary $\boldsymbol{\Sigma}^k$ over time. For the GMM models, Table 1 also shows whether the variance components $\boldsymbol{R}^k$ and $\boldsymbol{D}^k$ are homogeneous between groups.

2.5 BIC model selection strategy

The Bayesian information criterion (BIC)[11] is a measure of model fit that can be used to guide the selection of mixture models. The BIC provides a trade-off between model fit and model complexity. A commonly used model selection strategy is to select the model, from amongst those under consideration, which has the maximum BIC. To investigate how well this strategy performed in comparison to any single model, the model with the maximum BIC of the 18 considered was tested in each scenario, referred to here as model B.

3 Simulations

3.1 Specification of simulations

We used simulation to evaluate the performance of the 18 models described in Section 2 and the maximum BIC model, with respect to correct classification rate (CCR) as defined in Section 3.2, under a range of scenarios. The impact on bias of the model parameters, in terms of the difference of the estimates from the true values, can also be considered. However, we focus on a classification-based performance measure, as misclassification of trajectories leads to bias in the estimation of the model parameters. Although misclassification is not the only possible source of bias, in the scenarios we considered the level of bias was consistent with the level of misclassification. This is elaborated on in Section 4.

Table 1. Characteristics of models.

	All models			GMM variance components	
Model	Conditional independence	Homogeneous Σ^k	Stationary Σ^k	Homogeneous R^k	Homogeneous D^k
L1	✓	✓	✓	-	-
L2	✓	✓	×	-	-
L3	✓	×	✓	-	-
L4	✓	×	×	-	-
M1	✓	✓	✓	-	-
M2	✓	✓	×	-	-
M3	✓	×	✓	-	-
M4	✓	×	×	-	-
GA1	×	✓	✓	✓	✓
GA2	×	✓	×	✓	✓
GA3	×	×	✓	×	✓
GA4	×	×	×	×	✓
GB1	×	×	✓	✓	×
GB2	×	×	×	✓	×
GB3	×	×	✓	×	×
GB4	×	×	×	×	×
M5	×	✓	×	-	-
M6	×	×	×	-	-

The simulated scenarios are motivated by the childhood behaviour data from the Generation 1 Study, a prospective longitudinal cohort study of women and their children living in Adelaide, South Australia. The women were recruited in pregnancy between 1998 and 2000.[12] Data were collected from a total of 536 children at ages 2, 3.5, 5 and 9.5 years; 354 contributed data at all four time points. Childhood behaviour was measured in the Generation 1 Study using the Achenbach Child Behavior Checklist,[13] which records the parent's views on specific behavioural, emotional and social problems. Total and sub-scale scores can be derived as the sum of the ordinal items, with higher scores suggesting worse behaviour. We focus here on the externalising sub-scale derived from the subset of questions relating to behaviours such as aggression and bullying. The externalising sub-scale calculated from the Generation 1 Study has been observed to exhibit non-stationary covariance, and therefore motivated our investigation into the effect this had on the appropriate choice of group-based trajectory model. Figure 1(a) presents the externalising data from these four time points of the Generation 1 Study, with Figure 1(b) and (c) showing the groups identified from a two-group GA4 model fit to these data. Each of the plots has a random 25 trajectories from each group joined. As can be seen in part (a) of Figure 1, the variance decreases over time, with the variance at 9.5 years much lower than at the earlier time points. This non-stationarity is only accounted for in some of the models described in Section 2, and our simulations allow the effect of accommodating this to be determined.

In the present study, datasets of 500 observations were simulated to have properties approximating those for the externalising behaviour data from the Generation 1 Study, with four time points aligned with the standardised time points derived from child ages 2, 3.5, 5 and 9.5. Standardised time points were used to aid the convergence of the group-based models, and the same four time points were used for each subject to allow the MGMM models to be compared with the LCGA and GMM models, i.e. for $i = 1, \ldots, n_k$ $\boldsymbol{t}_i = (-1, -0.5, 0.1, 1.3)'$. To investigate the effect of more time points on conclusions about misclassification we also repeated the simulations with eight time points. In a general review of the literature, studies using group-based trajectory models were found to have a number of time points between 3 and 16, and sample sizes ranged from hundreds to several thousands, for example.[14–16] Our simulations have been designed to be of relevance to applied researchers, in keeping with typical studies.

For comparison with previous research,[1] the simulation design had two groups. The mean and covariance structure varied between scenarios in the 'high' group (group 1) while the mean and covariance structure for the 'low' group (group 2) remained fixed. The three dimensions of variation for group 1 were: the separation between the mean trajectories of the groups (i.e. the vertical position of the group 1 mean trajectory), the dependence across the time points and the stationarity of the covariance matrix. We altered these three aspects in scenarios assessing classification performance for various combinations of separation, dependence and stationarity.

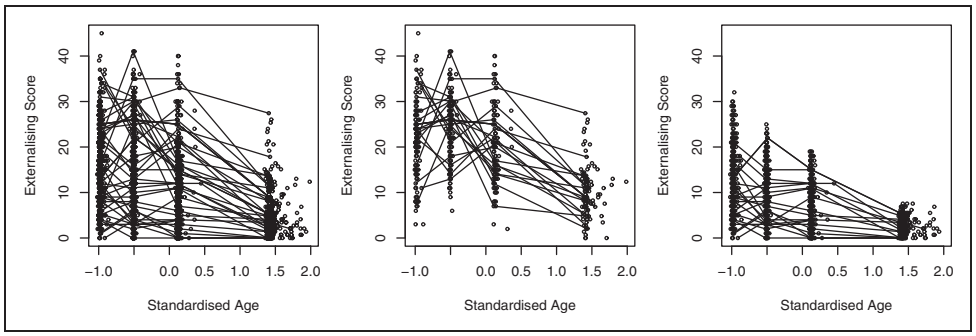


Figure 1. (a) Externalising scores versus standardised age in the Generation 1 Study. (b) The 'high' group (group 1) from a GA4 model fitted to these data. (c) The 'low' group (group 2) from a GA4 model fitted to these data.

For each scenario, with $k = 1, 2$ and $i = 1, \ldots, n_k$, a linear model was used to simulate 200 datasets

$$\boldsymbol{y}_i = \alpha_k \mathbf{1} + \beta_k \boldsymbol{t}_i + \boldsymbol{\epsilon}_i^k$$

where $\mathbf{1} = (1, 1, 1, 1)'$ and $\boldsymbol{\epsilon}_i^k \sim N(0, \boldsymbol{\Sigma}^k)$.

In the first scenario, with a stationary covariance structure without dependence, the following parameters were used: $\alpha_1 = 14,\ \alpha_2 = 7,\ \beta_1 = -5,\ \beta_2 = -5$; $\boldsymbol{\Sigma}^k = \sigma_k^2 \boldsymbol{I}$ with $\sigma_1^2 = 35$ and $\sigma_2^2 = 25$; and $n_1 = n_2 = 250$. As well as the number of time points and their relative positions being based on the Generation 1 Study, the shape of the mean trajectories and the coefficient of variation were based on the parameters estimated in the GA4 model fit to the Generation 1 Study data (shown in Figure 1).

Altering α_1 varies the vertical separation between the mean trajectories of the groups. The dependence and stationarity of group 1 are each varied using $\boldsymbol{\Sigma}^1$. To implement this, $\boldsymbol{\Sigma}^1$ was parameterised by

$$\boldsymbol{\Sigma}^1 = \sigma_1^2 \boldsymbol{V}(\boldsymbol{I} + \rho(\boldsymbol{J} - \boldsymbol{I}))\boldsymbol{V}$$

where $\boldsymbol{J}$ is the 4×4 matrix of ones and $\boldsymbol{V}$ is the diagonal matrix with the diagonal elements equal to 1, $\sqrt{1 + \frac{v-1}{3}}$, $\sqrt{1 + \frac{2v-2}{3}}$ and $\sqrt{v}$, respectively. The parameter ρ (referred to hereafter as dependence) defines an exchangeable correlation structure for group 1. The parameter v (referred to hereafter as taper) controls the ratio of $\boldsymbol{\Sigma}^1_{11}$ to $\boldsymbol{\Sigma}^1_{44}$. The other diagonal elements of $\boldsymbol{\Sigma}^1$ are scaled proportionally as 1/3 and 2/3 of the difference between 1 and v. When $v \neq 1$ and $\rho \neq 0$, the off diagonal elements of $\boldsymbol{\Sigma}^1$ are jointly scaled by ρ and v.

A total of 180 scenarios were considered by varying the parameters α_1, ρ and v across 6, 6 and 5 levels, respectively. α_1 was varied in increments of 2 from a baseline of $\alpha_{1,\rho,v}$ to $\alpha_{1,\rho,v} + 10$. $\alpha_{1,\rho,1}$ took the value 14 for all values of ρ. For $v \neq 1$, $\alpha_{1,\rho,v}$ depended on ρ and v in order to avoid certain unrealistic scenarios as discussed in Section 3.3. ρ took values 0, 0.25, 0.5, 0.75, 0.9 and 0.99, and v took values 10, 2, 1, 0.5 and 0.1.

3.2 Measuring classification performance

The CCR is defined as the proportion of observations correctly assigned into their group. As mixed models are 'unsupervised', the result is a grouping of trajectories rather than assignment to predefined classes. This results in the known issue of 'label switching' in mixture models.[17] To obtain a CCR, we considered both labellings of the groups, and chose the labelling that resulted in the higher CCR value. For each scenario and model combination, we calculated the mean CCR over the 200 simulations.

Even with a correctly specified model with known parameters, changes in the separation, dependence and stationarity affect the classification performance. This is because the difficulty of the classification task is altered. In order to remove this variation from our comparisons between models, we evaluated the CCR relative to the BCCR

$$\text{relative CCR} = \frac{\text{CCR}}{\text{BCCR}}$$

where the BCCR is the optimal classification proportion that can be expected with the knowledge of the underlying distribution function, as calculated using the Bayes minimum error classifier.[18] A relative CCR of 1 is achieved when the CCR meets the optimum level of the BCCR, whereas the lowest possible relative CCR is 0.5, with a CCR of 0.5 and BCCR of 1. The Bayes minimum error classifier is based on the log of the posterior odds

$$g(x) = \log\frac{p(x|c_1)}{p(x|c_2)} + \log\frac{P(c_1)}{P(c_2)}$$

where $p(x|c_k)$ is the probability density function for group k, and $P(c_k)$ is the prior probability of group k. The observation is classified as group 1 (c_1) if $g(x) > 0$, and group 2 (c_2) otherwise. As we have groups of equal size in all scenarios, $P(c_1) = P(c_2) = 1/2$ here. We estimated the BCCR values for each of the scenarios by calculating the average of the classification rates according to the decision rule from 100 simulations. In the scenarios considered, the BCCR values ranged from 0.83294 to 0.99993, with the lowest values observed for the minimum separation, moderate dependence and little taper. Although the theoretical maximum of the relative CCR is 1, it is possible for the estimates obtained here to be greater than 1 by chance as both the numerator and denominator were estimated from simulations. We observed this only occurred in a small number of cases (3.0%), and the maximum relative CCR value obtained was 1.00114.

3.3 Adjusting for BCCR differences across scenarios

In addition to the introduction of the relative CCR, a further adjustment was applied to reduce the differences in BCCR values between the scenarios with different levels of taper. This was because the range of BCCR values was reduced in scenarios with taper values of 10, 0.5 and 0.1. For example, the minimum BCCR value for scenarios with $v = 0.1$ was 0.93304. This meant all scenarios with this taper level were dramatically easier to classify, reducing our ability to compare differences between models.

To control the BCCR values across scenarios with different taper levels ($v \neq 1$), we adjusted the separation through α_1. In particular, for each scenario with $v \neq 1$, we chose the separation baseline for group 1 ($\alpha_{1,\rho,v}$) so that the average BCCR for each level of ρ was as close as possible to the average BCCR for that level of ρ in scenarios with $v = 1$.

3.4 Analysis software

R[19] was used for the data simulations and Mplus[10] was used via R with the MplusAutomation[20] package for estimation of the LCGA (L1–L4) and GMM models (GA1–GA4 and GB1–GB4). The Mclust package in R was used to estimate the MGMM models (M1–M6).[21,22] The model with the maximum BIC for each scenario (B) was identified using R. Estimation of mixture models can be sensitive to starting points.[10] Therefore, 200 random starts were used for each model estimation in Mplus, with the maximum log-likelihood solution being chosen to reduce the likelihood of finding local optima.

4 Results

Figure 2 summarises the distribution of the relative CCR values from all 180 scenarios, for each of the 18 different models considered and the maximum BIC model. The variation in the relative CCR values was dependent on the type of scenario used for the simulation, with low dependence and large separation typically resulting in higher relative CCR values due to lower classification difficulty.

As can be seen from Figure 2, the four LCGA models and those MGMM models that assumed conditional independence (M1–M4) had relative CCR values much lower than many of those for models that allowed for dependence between observations within each group.

The GMM models that assumed equal random effects covariance matrices between groups (GA) had higher relative CCR values than the LCGA models, when the residual variances were free to vary between groups (i.e. models GA3 and GA4 compared with L3 and L4). However, these models were more negatively affected by the assumption of equal residual variances between groups (i.e. models GA1 and GA2 compared with L1 and L2).

The GMM models that allowed for different random effects covariance matrices between groups (GB) resulted in the highest relative CCR values. The median relative CCR values of the GB1–GB4 models were 0.9845, 0.9888, 0.9967 and 0.9981, respectively, due to the generality of the model reducing the bias associated with the more

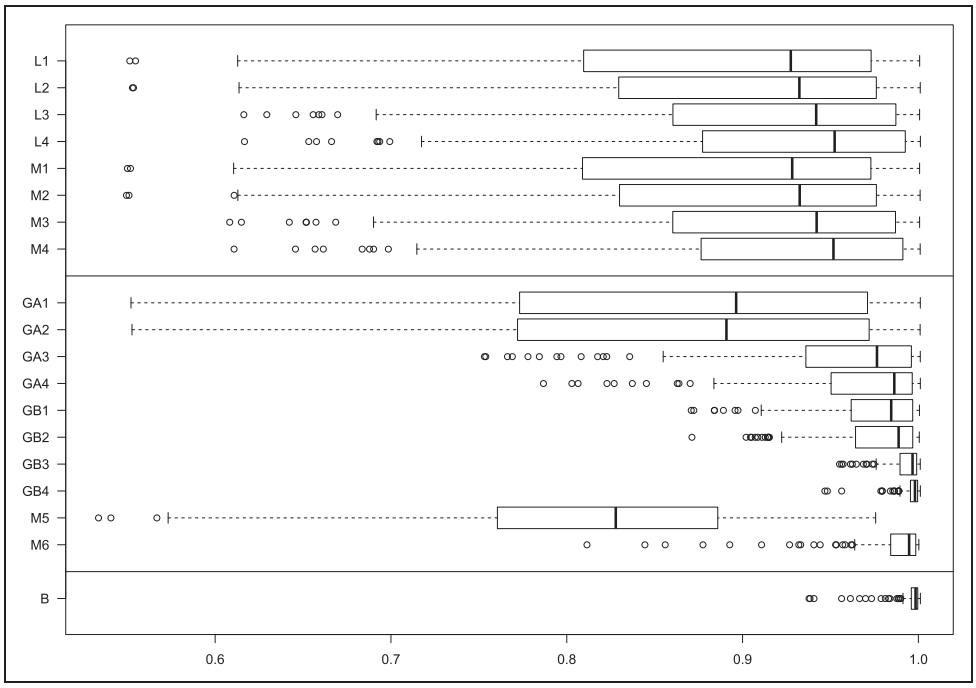


Figure 2. Boxplots of relative CCR values from simulations with four time points.

restrictive models. The highest relative CCR values across all scenarios were observed for the GB4 model with residual variances free to vary between groups and across times.

Considering the residual variances of the GMM and LCGA models, it can be seen from Figure 2 that assuming they are equal between groups (models L2, GA2 and GB2) resulted in lower relative CCR values than the models without this assumption (models L4, GA4 and GB4), and this effect was greater than that observed when they were assumed to be equal across times (models L3, GA3 and GB3).

Of the MGMM models, the fully unrestricted model (M6) performed best (median relative CCR 0.9947). The poorest performing model in terms of relative CCR was M5, as although it allowed for dependence between the observations, the homogeneity assumption resulted in poor classification.

Choosing the model according to the maximum BIC among all 18 different models for each scenario (model B) resulted in a median relative CCR of 0.9984. Using the maximum BIC typically chose an LCGA model when the dependence was low (0 or 0.25), an M1 model when there was any non-stationarity and very high dependence (0.99) and also when taper was 10 and dependence was 0.75 or greater. In all other cases, one of the GB models was chosen most frequently, with the GB3 model being chosen most often in 49 of 180 scenarios.

As a result of being able to select less complex models for scenarios without complicated dependence, model B performed slightly better than the best performing model otherwise, GB4. This difference was most apparent in the scenarios with the lowest separation, as the classification was the most difficult. For these scenarios (i.e. with dependence of 0, lowest separation), model B classified six more trajectories correctly than GB4, on average. Furthermore, the standard deviation of CCR values was lower for model B in these scenarios by an average of 0.021, compared with GB4. This increased precision of CCR estimation is a result of the models with maximum BIC being appropriately complex for these scenarios without dependence. This bias-variance trade-off is typical in such situations of increasing model complexity.[23]

In the simulations with eight time points rather than four, the GB models were chosen even more frequently, with these models typically chosen unless the dependence was 0 or if the dependence was 0.99 and the taper was either 0.1 or 10. This meant the GB models were chosen in 138 or 180 scenarios. Supplementary Figure 1 shows the distribution of the relative CCR values in the scenarios with eight time points. The results for these simulations are broadly similar to those from four time points, with details given in the supplementary material (available at: http://smm.sagepub.com/). As the number of time points increases, it could be expected that performance would

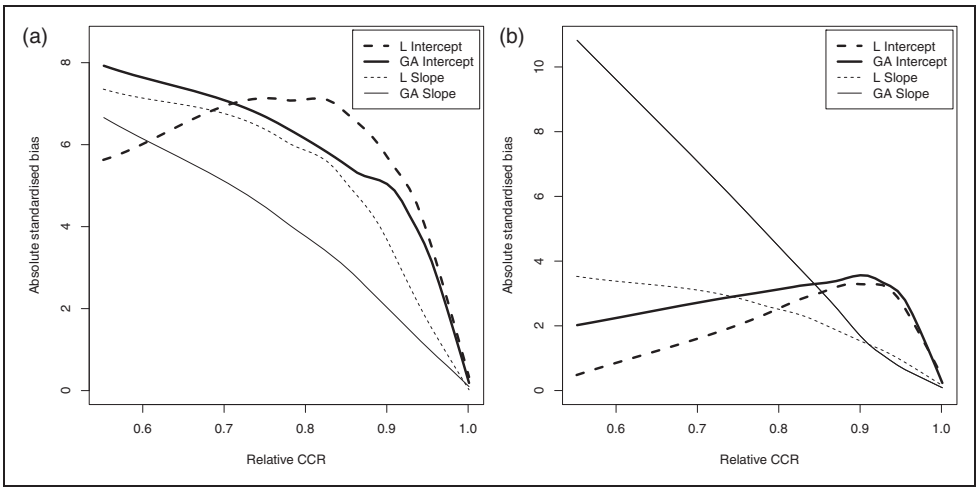


Figure 3. Lowess fits to the absolute standardised bias of the intercept and slope parameters against the relative CCR aggregated over models L1–4 and GA1–4. (a) Group 1, (b) group 2.

decline for models with more complex variance structures. However, only a modest effect was observed for the models considered in going from four to eight time points.

As mentioned in Section 3.1, the level of misclassification according to the relative CCR was consistent with the level of bias in the estimation of the model parameters. Figure 3 shows lowess fits to the absolute standardised bias of the intercept and slope parameters against the relative CCR in models L1–4 and GA1–4. The four different L models have been combined as the distributions of bias against relative CCR were very similar. The same aggregation was done for the four GA models. As can be seen in Figure 3, the level of bias generally increases as the relative CCR increases. When the bias is seen to reduce for very small CCR values, this is due to a small number of scenarios at the tail of the relative CCR distribution. The curves for the GB models are not shown, due to the fact that they showed effectively zero bias, regardless of the relative CCR. This is not surprising, as can be seen in Figure 2, the GB models have the vast majority of their relative CCR value over 0.95. The MGMM models are not included as they do not estimate an intercept and slope parameter.

As an example of how the model parameter estimates can change under the different methods, Figure 4 shows the mean trajectories of two group models estimated according to the 18 methods in Section 2, when fitted to the externalising scores in the Generation 1 Study. The trajectories for the low and high groups are observed to be generally similar, but in four cases the high group shows a markedly different path. These four cases are for the GMM models with residual variance matrix the same between groups (i.e. GA1, GA2, GB1 and GB2). This difference of the GA1 and GA2 models is consistent with their poor performance in our simulations. Furthermore, the most inferior model in our simulations, M5, assigned all observations into one group. As such, it only has a mean trajectory appear in the low plot, because the overall mean is closest to the low group means.

5 Discussion

The results of this study demonstrate that the use of group-based trajectory models with only limited examination of how well underlying assumptions are met should be avoided. In particular, the assumptions regarding the random effects covariance matrix should be considered, contrary to the fact that they are sometimes neglected as these parameters are not of substantive interest.[4]

The findings show that the best single method to handle data with non-stationary covariance or non-homogenous variances was GB4. Although this is the most general GMM, it does not theoretically provide a perfect fit to the covariance structure of the simulated data. The GB4 model is best able to accommodate for the scenarios with different variances across the time points, the different variances between the groups and also high levels of dependence across the time points. This complex model was however susceptible to slight over-fitting in scenarios without complex covariance structure. In the situations without dependence, the performance of the GB4 model's CCR estimates was worse, and their variability was greater, compared to models chosen according to

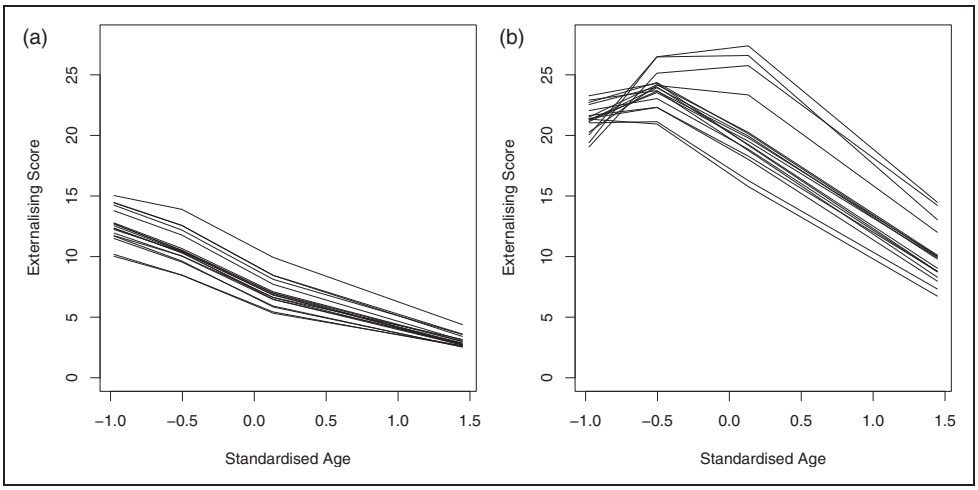


Figure 4. Mean trajectories from two group models using all 18 methods fitted to the externalising scores in the Generation 1 Study. (a) Low group, (b) high group.

the BIC to have appropriate complexity. Therefore, in situations with few time points, where the covariance structure of a dataset is to be determined, use of the BIC to determine the best model to fit provides a good way to assess which covariance assumptions may be appropriate. With eight time points, the GB models became more favourable, and in these situations GB4 should be considered first, and model diagnostics assessed as simpler models are subsequently considered.

More assumptions than necessary may be applied by analysts if the default group-based trajectory models are fit in statistical packages. For example, the default in Mplus for the covariance assumptions is equal covariance between groups, but different covariance over time.[10] The results of this study demonstrate that poor classification can result from using an assumption of equal covariance between groups, when it does not apply, and care should be taken by researchers to ensure they are not making this assumption erroneously. The bias in parameter estimates resulting from misspecifying heterogeneous variances as homogeneous has previously been identified in univariate mixture models[2,3] and in GMMs.[1,4] However, no previous study has investigated the effect of misspecifying the variance as stationary. We have shown that although the misclassification resulting from inappropriate homogeneity assumptions is much greater than from inappropriate stationarity assumptions, neither type can be ignored. Our simulation showed that the use of models that do not account for the dependence between observations, such as the LCGA and conditionally independent MGMM models, also give rise to poor classification in a wide variety of scenarios.

In this study we have only considered data simulated from two groups, with only one group subject to dependence and non-stationary covariance. Although we expect the results of our study could generalise more broadly, future research is needed to investigate results with three or more groups, where all groups involve some dependence and non-stationarity. In addition, we have only considered models which do not include covariates impacting on either the mixing proportions or the trajectories. Further research is needed to extend the results of this paper to such cases.

When using group-based mixture models, one must ensure that the covariance matrix assumptions are appropriate for the situation. If a more restrictive model is used, the CCR will be much below the optimal value under the correct covariance structure. We therefore recommend carefully selecting a model for use in group-based trajectory applications that is appropriate for the research questions and dataset of interest. The models considered should have the flexibility to account for different covariance structures across time points and especially between groups. However, after investigating between these models and simpler candidates, the model with the best fit to the particular dataset should be chosen. Further research is also warranted to determine which method is best for assessment of model fit.

Acknowledgements

The support of the Generation 1 Study Chief Investigators, Prof Vivienne Moore and Prof Michael Davies, in allowing access to data from that study is gratefully acknowledged.

Declaration of conflicting interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: Christopher Davies received financial support for this research through a postgraduate scholarship from the University of Adelaide.

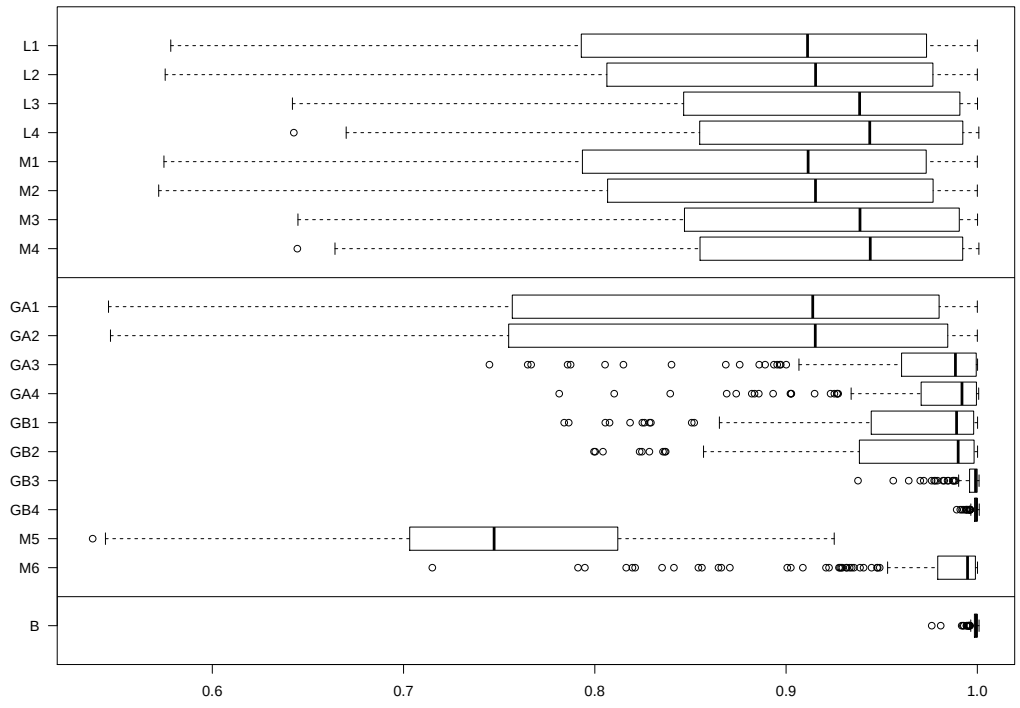
Data collection for the Generation 1 study was supported by grants from the Faculty of Health Sciences at the University of Adelaide, Dairy Research and Development Corporation, Channel 7 Children's Research Foundation Grant 13745, and National Health and Medical Research Council (Grants 465455 and 465437 and Australian Based Public Health Training Fellowship 627033 to LCG).

References

1. Heggeseth BC and Jewell NP. The impact of covariance misspecification in multivariate Gaussian mixtures on estimation and inference: an application to longitudinal modeling. *Stat Med* 2013; **32**: 2790–2803.
2. Gray G. Bias in misspecified mixtures. *Biometrics* 1994; **50**: 457–470.
3. Lo Y. Bias from misspecification of the component variances in a normal mixture. *Comput Stat Data Anal* 2011; **55**: 2739–2747.
4. Enders CK and Tofighi D. The impact of misspecifying class-specific residual variances in growth mixture models. *Struct Equ Modeling* 2008; **15**: 75–95.
5. Gilthorpe MS, Dahly DL, Tu YK, et al. Challenges in modelling the random structure correctly in growth mixture models and the impact this has on model mixtures. *J Dev Orig Health Dis* 2014; **5**: 197–205.
6. Jung T and Wickrama KAS. An introduction to latent class growth analysis and growth mixture modeling. *Soc Personal Psychol Compass* 2008; **2**: 302–317.
7. Oliveira-Brochado A and Martins FV. Assessing the number of components in mixture models: a review, https://ideas.repec.org/p/por/fepwps/194.html (2005, accessed 4 May 2015).
8. Nagin DS and Land KC. Age, criminal careers, and population heterogeneity: specification and estimation of a nonparametric, mixed Poisson model. *Criminology* 1993; **31**: 327–362.
9. Muthén B and Muthén LK. Integrating person-centered and variable-centered analyses: growth mixture modeling with latent trajectory classes. *Alcohol Clin Exp Res* 2000; **24**: 882–891.
10. Muthén LK and Muthén BO. *Mplus user's guide*, 6th ed. Los Angeles, CA: Muthén & Muthén, 1998–2010.
11. Schwarz G. Estimating the dimension of a model. *Ann Stat* 1978; **6**: 461–464.
12. Moore VM, Davies MJ, Willson KJ, et al. Dietary composition of pregnant women is related to size of the baby at birth. *J Nutr* 2004; **134**: 1820–1826.
13. Achenbach TM and Rescorla LA. *ASEBA school-age forms & profiles*. Burlington, VT: Aseba, 2001.
14. Bowers AJ and Sprott R. Examining the multiple trajectories associated with dropping out of high school: a growth mixture model analysis. *J Educ Res* 2012; **105**: 176–195.
15. Lanza ST and Collins LM. A mixture model of discontinuous development in heavy drinking from ages 18 to 30: the role of college enrollment. *J Stud Alcohol Drugs* 2006; **67**: 552.
16. Hser YI, Huang D, Chou CP, et al. Trajectories of heroin addiction: growth mixture modeling results based on a 33-year follow-up study. *Eval Rev* 2007; **31**: 548–563.
17. Redner RA and Walker HF. Mixture densities, maximum likelihood and the EM algorithm. *SIAM Rev* 1984; **26**: 195–239.
18. Berger JO. *Statistical decision theory and Bayesian analysis*, 2nd ed. New York: Springer, 1985, pp.357–358.
19. R Core Team. *R: a language and environment for statistical computing*. Vienna, Austria, 2014.
20. Hallquist M and Wiley J. MplusAutomation: automating Mplus model estimation and interpretation (Version 0.6-3), 2014.
21. Fraley C and Raftery AE. Model-based clustering, discriminant analysis and density estimation. *JASA* 2002; **97**: 611–631.
22. Fraley C, Raftery AE, Murphy TB, et al. mclust version 4 for R: normal mixture modeling for model-based clustering, classification, and density estimation. Technical Report No. 597, Department of Statistics, Seattle, WA: University of Washington, 2012.
23. Hastie T, Tibshirani R and Friedman J. *The elements of statistical learning*, 2nd ed. New York: Springer, 2009.

3.4 Supplementary Figure 1

As shown in Supplementary Figure 1, the relative CCR distributions for the simulations with eight time points are broadly similar to those with four time points. The main difference between these results is the improvement in relative CCR of the GMM models with residual variances free to vary between groups, relative to the others. This is particularly the case for the GB3 and GB4 models. The M5 model is inferior to the other models, to a greater extent than observed for four time points.



Supplementary Figure 1: Boxplots of relative CCR values from simulations with eight time points.

3.5 Statement of authorship

Statement of Authorship

Title of Paper	Letter to the editor.
Publication Status	☑ Published ☐ Accepted for Publication ☐ Submitted for Publication ☐ Unpublished and Unsubmitted work written in manuscript style
Publication Details	Davies CE, Glonek GFV and Giles LC. Letter to the editor. Statistical Methods in Medical Research. Prepublished May 24, 2017; DOI: 10.1177/0962280216648050. <http://journals.sagepub.com/doi/10.1177/0962280216648050> The final, definitive version of this paper has been published in Statistical Methods in Medical Research, 2017 by SAGE Publications Ltd, All rights reserved. Copyright © 2017 Davies CE, Glonek GFV and Giles LC. Reprinted by permission of SAGE Publications.

Principal Author

Name of Principal Author (Candidate)	Christopher E Davies		
Contribution to the Paper	Designed the supplementary series of simulations, performed all analyses, interpreted the results, drafted the letter and acted as corresponding author.		
Overall percentage (%)	80		
Certification:	This paper reports on original research I conducted during the period of my Higher Degree by Research candidature and is not subject to any obligations or contractual agreements with a third party that would constrain its inclusion in this thesis. I am the primary author of this paper.		
Signature		Date	06/12/17

Co-Author Contributions

By signing the Statement of Authorship, each author certifies that:

i. the candidate's stated contribution to the publication is accurate (as detailed above);
ii. permission is granted for the candidate in include the publication in the thesis; and
iii. the sum of all co-author contributions is equal to 100% less the candidate's stated contribution.

Name of Co-Author	Gary FV Glonek		
Contribution to the Paper	Contributed to the design of the supplementary simulations and interpretation of results, and reviewed the letter.		
Signature		Date	06/12/17

Name of Co-Author	Lynne C Giles		
Contribution to the Paper	Contributed to the design of the supplementary simulations and interpretation of results, and reviewed the letter.		
Signature		Date	06/12/17

3.6 Letter to the editor

SMMR
STATISTICAL METHODS IN MEDICAL RESEARCH

Letter to the Editor

Statistical Methods in Medical Research
0(0) 1–3

DOI: 10.1177/0962280216648050
smm.sagepub.com
SAGE

Letter to the Editor

Christopher E Davies,[1,2] Gary FV Glonek[1] and Lynne C Giles[2]

Dear Professor Everitt,

Davies et al.[1] presented the impact of covariance misspecification in group-based trajectory models for longitudinal data with non-stationary covariance structure. Our ongoing research has brought to light some potential difficulties in the calculation of the estimates used in that article.

The estimates concerned arose from growth mixture models estimated with non positive-definite random effect covariance matrices or residual covariance matrices. The estimation software Mplus[2] produced at least one warning of these types in 66% of the growth mixture models estimated with the random effect covariance matrices equal between groups. In growth mixture models with the random effect covariance matrices free to vary between groups, 97% of the models estimated had at least one of these warnings produced by Mplus. Strictly speaking, the theory used to justify inference requires those covariance matrices to be positive definite. Notwithstanding this, we believe our conclusions to be sound for two reasons.

First, the simulation studies showed, independently of the statistical theory, that the estimates for the recommended models GB3 and GB4 (with random effect covariance matrices and residual covariance matrices free to vary between groups) had desirable statistical properties in terms of the relative correct classification rate (CCR).

Second, we have conducted a supplementary series of simulation experiments by slightly modifying the models to avoid the above difficulties observed in the original study. The conclusions of those simulations are, in practical terms, the same as for the original study.

Specifically, we constrained the covariance of the random intercepts and slopes for the growth mixture models to be zero and constrained their variances to be positive. The residual variances were also constrained to be positive. Following the use of these constraints, the frequency of warnings reduced to 0.18% for the homogeneous random effect covariance matrix models, and 0.45% for the heterogeneous random effect covariance models. These warnings related to the maximum likelihood solution having a unique likelihood value among those resulting from the random starts and therefore being at risk of a local maximum.

Table 1 shows the median relative CCR values in the growth mixture models with these constraints applied and the simulations that generated warnings excluded. The differences in the median relative CCRs are all ≤ 0.0183,

Table 1. Median relative CCR values of the growth mixture models with and without constraints.

Model	Supplementary simulations with constraints	Original study without constraints
GA1	0.9021	0.8964
GA2	0.9025	0.8908
GA3	0.9621	0.9765
GA4	0.9687	0.9863
GB1	0.9662	0.9845
GB2	0.9771	0.9888
GB3	0.9902	0.9967
GB4	0.9940	0.9981

[1]School of Mathematical Sciences, The University of Adelaide, Adelaide, Australia
[2]School of Public Health, The University of Adelaide, Adelaide, Australia

Corresponding author:
Christopher E Davies, School of Mathematical Sciences, The University of Adelaide, Adelaide, SA 5005, Australia.
Email: chris.davies@adelaide.edu.au

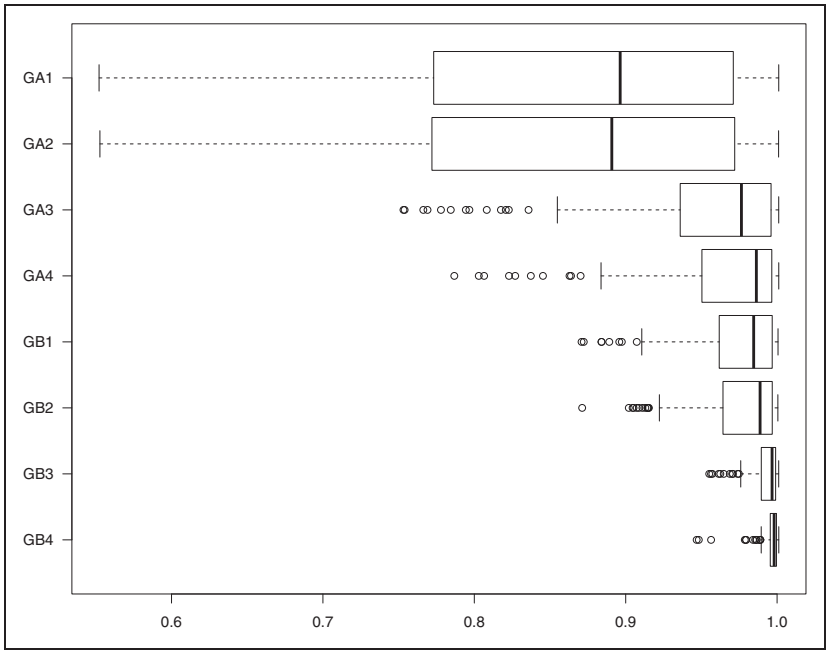


Figure 1. Relative CCR values for the growth mixture models without constraints.

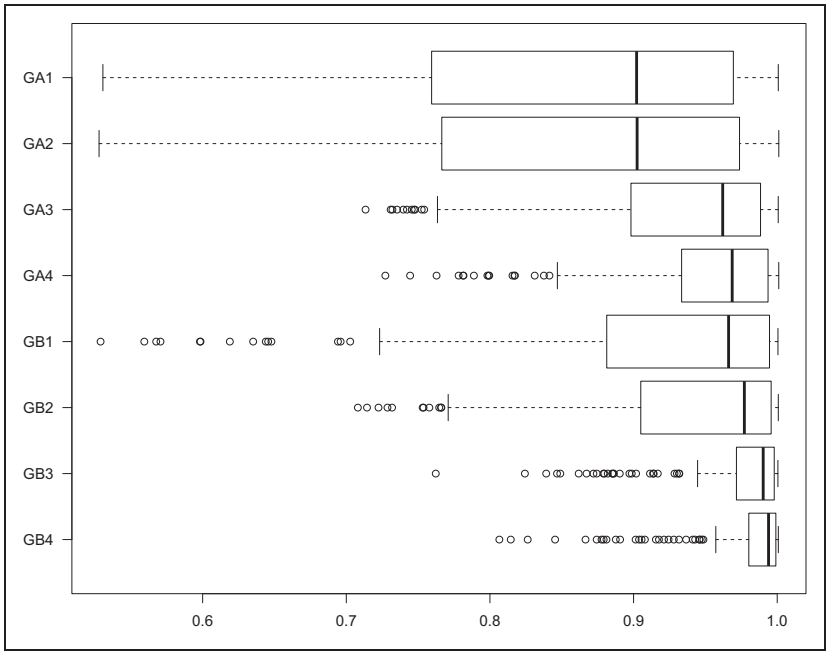


Figure 2. Relative CCR values for the growth mixture models with the constraints.

and the rank order of the models is maintained, with the exception of the GA1 and GA2 models, which have very similar median relative CCR values in either case.

Figure 1 shows the distributions of the relative CCR values for the growth mixture models under their original parameterisation, whereas Figure 2 shows the relative CCR values for the re-estimated models with constraints.

The greatest effect of application of the constraints is for the GB1 and GB2 models. This is to be expected, as in the revised models the random effect covariance matrices were constrained to diagonal, but unconstrained in our original presentation.

Comparison of Figures 1 and 2 demonstrates that the relative CCR values are largely similar in the two sets of simulations. Most importantly, our original conclusion still holds that GB3 and GB4 perform best.

References

1. Davies CE, Glonek GFV and Giles LC. The impact of covariance misspecification in group-based trajectory models for longitudinal data with non-stationary covariance structure. *Stat Methods Med Res*. Epub ahead of print 17 August 2015. DOI: 10.1177/0962280215598806.
2. Muthén L and Muthén B. *Mplus user's guide*, 6th ed. Los Angeles, CA: Muthén & Muthén, 1998–2010.

Chapter 4

Outliers in group-based trajectory models

4.1 Introduction

As described in Section 2.5, outliers are data points that are far from rest of the distribution. Such observations may be easily identified in a univariate or bivariate setting as points that appear away from the majority of other observations. However in a multivariate setting, such as with repeated observations for a group of individuals, outliers may also occur as more complex patterns of observations. A trajectory may be considered an outlier because a single observation is an outlier with respect to its univariate distribution. Alternatively, the observations in a trajectory could be consistent with the distribution of values at each time point but, when considered in the context of the multivariate distribution of all time points, be a multivariate

outlier.

Outliers may arise through random variation, experimental error or contamination.[39] There are a variety of methods that exist for the detection of outliers in the statistical and computer science literature, covered in recent comprehensive reviews.[39,42] Those used in the context of mixture models are described in the literature review of Section 2.5. These methods fall into three categories: those robust to outliers, those with trimming or exclusion of outliers, and those with a separate mixture component for the outliers.

An underlying assumption of the group-based trajectory model is that it applies to all trajectories, and this does not allow for the possibility that outliers may be present. Thus outlying trajectories may distort the estimated groups of these models and any subsequent analyses that use them.

In the mixture modelling setting, an additional complexity is that besides outlying observations potentially being errors or arising from natural variation as part of an existing group, they can also be thought of as arising from a sub-population that corresponds to a new group. Choosing between these options is often subjective and, at present, researchers have limited tools to assist them.

This chapter describes simulation studies undertaken to explore both the impact of outliers in group-based trajectory models, and the development and effectiveness of an algorithm I designed to identify outliers.

4.2 Outlier impact

This section explores, through simulation studies, the impact of outliers on the conclusions reached when using group-based trajectory models. I demonstrate that the impacts of outliers in group-based trajectory models can vary in severity and in some situations can have unexpected consequences.

4.2.1 Examples of outlier impact

As the impact of outliers can be more complicated in the group-based trajectory setting than in other contexts, this section presents three examples from my simulation results to illustrate the range of effects that outliers can have. As a consequence of the introduction of outliers, the first shows an increase in the number of groups, the second shows an unexpected improvement in classification performance, and the third shows a change in the parameter estimates with the number of groups remaining unchanged. The first and second examples use the same simulated dataset, with the different effect of outliers the result of a different covariance assumption in the fitted mixture model.

For context, the design of the simulations is briefly described here, with the full description following in Sections 4.2.3 and 4.2.4. All simulations had two substantive groups of 140 trajectories, across four evenly spaced time points. I considered two sets of simulations: one where the mean trajectories of the two groups diverged and another where the mean trajectories crossed. After simulating the substantive group

trajectories, 20 contaminating trajectories were added to the datasets to provide potential outliers. At each time point, observations of the contaminating trajectories were sampled from a uniform distribution between 4 less than the lowest group mean at that time point to 4 more than the highest mean at that time point.

Example 1

The first example demonstrates that the introduction of outliers can lead to a large increase in the number of estimated groups. This is a common outcome when outliers are introduced in a model with covariances constrained to be equal between groups. To accomodate the outliers in the mixture model, additional groups are included in the best fitting model. The BIC, as described in Section 2.3, was used to select the best fitting number of groups. Figure 4.2.1 shows the simulated data comprising two substantive groups and 20 contaminating trajectories. The two groups are close together at the first time point (difference of 1 in the true group means), but relatively distant at the fourth time point, as the 'high' group has a mean trajectory with a positive slope of 0.5. In Figure 4.2.1, twenty-five randomly selected trajectories from each group are connected as joining all of the points makes it difficult to see individual trajectories. The mean trajectory from each group is shown as a thicker line of the same colour in each plot. This approach will be applied to all plots of group-based trajectories in this chapter.

Figure 4.2.2 presents the estimated model fit to the simulated data with the contaminating trajectories omitted. The colours of the points and individual tra-

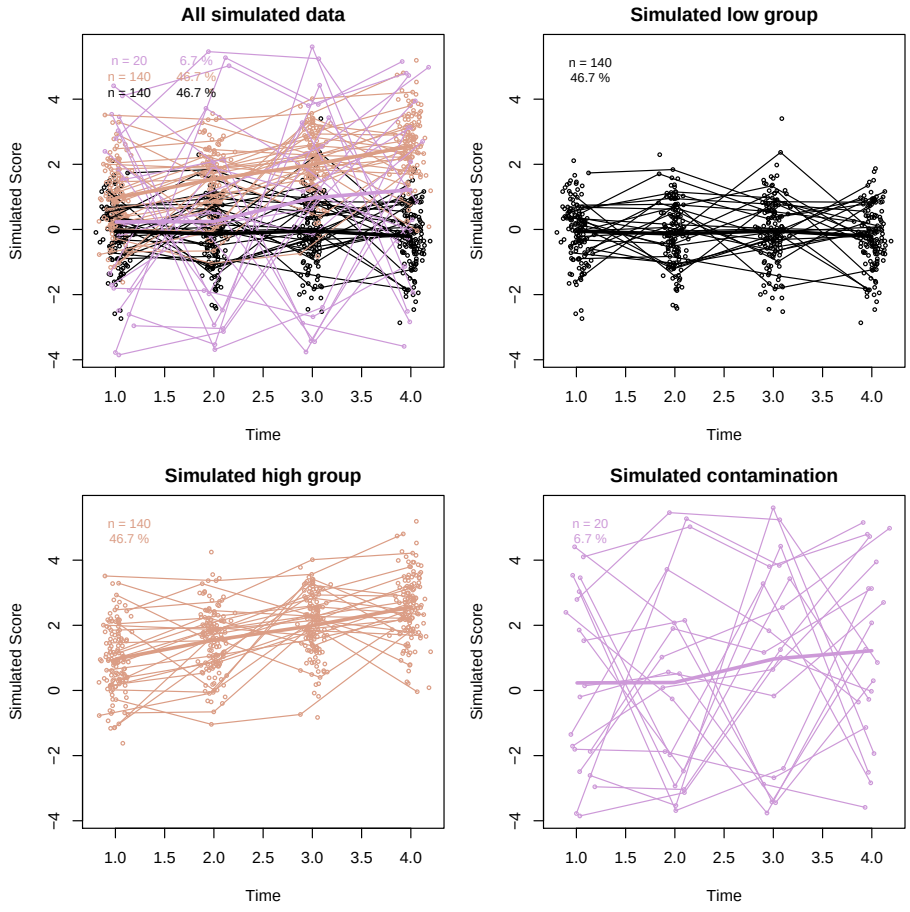


Figure 4.2.1: Simulated data for example 1.

jectories reflect their substantive group membership, while the colour of the mean trajectory and group sizes for each fitted group are chosen based on the largest simulated group represented. Trajectories assigned to an incorrect group are highlighted by being automatically selected to have their points connected in the plot. The model used was an MGMM with an unstructured covariance constrained to be equal between groups (model M5 as described in Section 2.1.3). The choice of two groups had the best fit according to the BIC. Under this model the simulated groups were identified with high accuracy, with 95.7% of trajectories assigned to their correct group.

Figure 4.2.3 shows the chosen M5 model fit to the simulated data with the contamination included. The number of groups in the best fitting model increased by three to account for the outliers. This is not a problematic outcome as the observations in the original groups were still as accurately classified, with 95.7% of the substantive group trajectories correctly allocated. Moreover, the contaminating trajectories were largely allocated to the three additional groups and thus were readily identified.

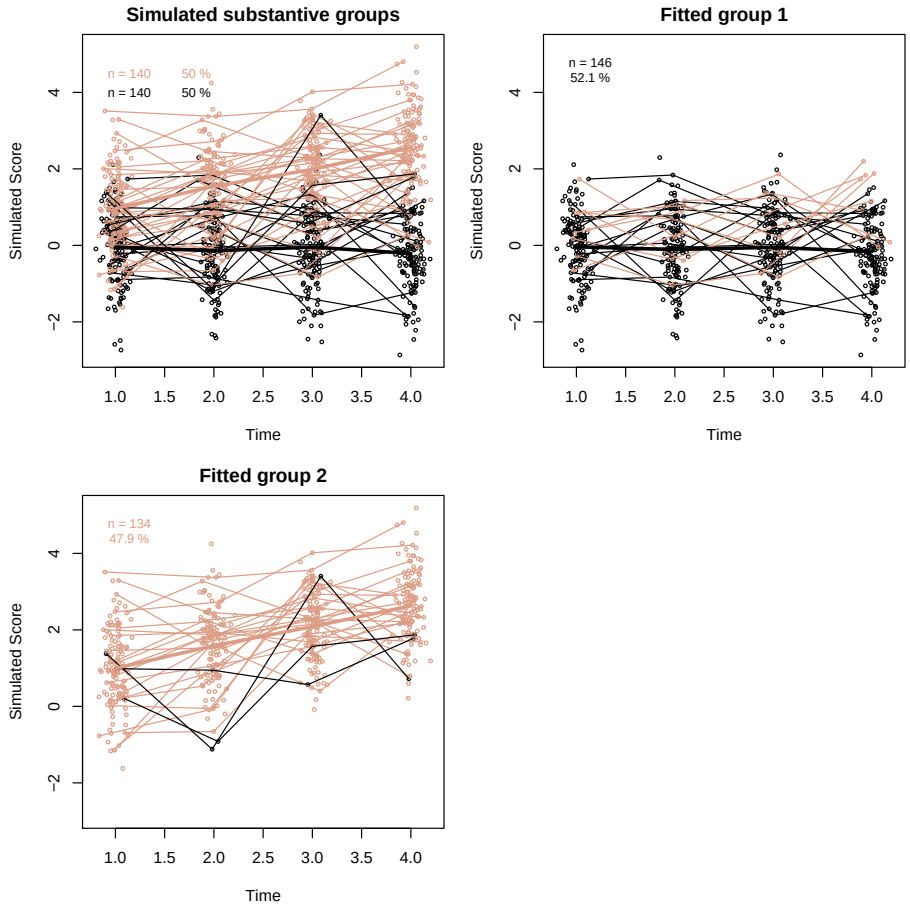


Figure 4.2.2: Estimated groups for example 1, with contamination omitted.

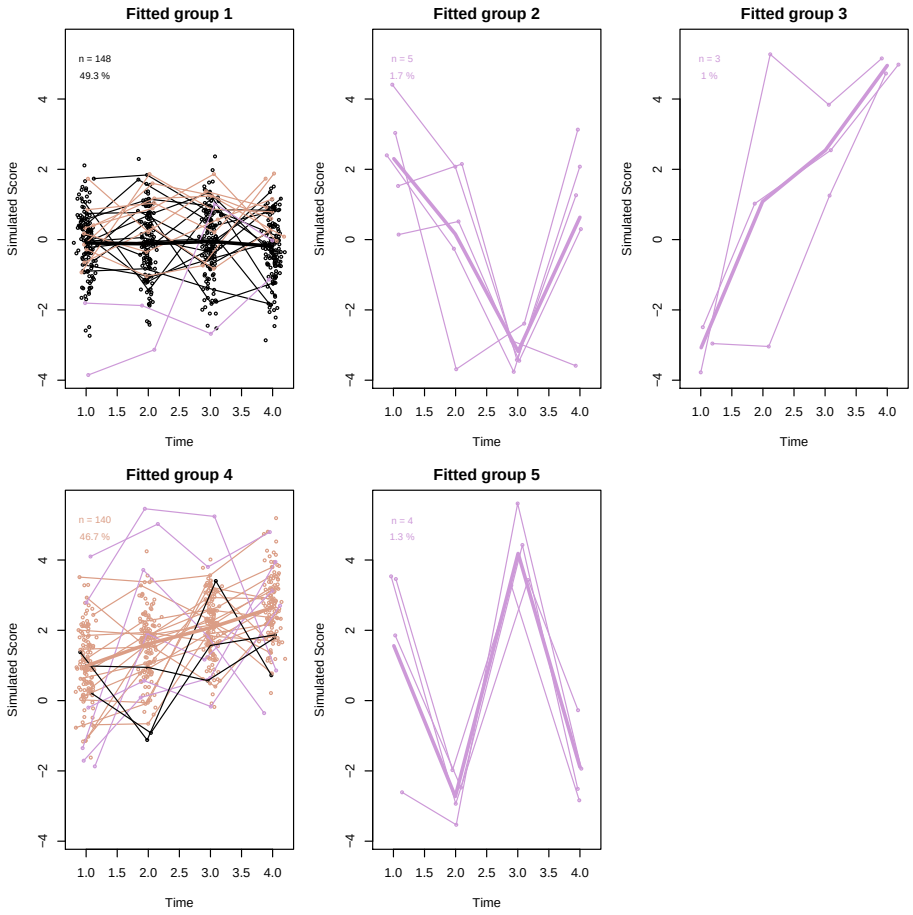


Figure 4.2.3: Estimated groups for example 1, with contamination included.

Example 2

The second example uses the same simulated dataset as the first example, but with the covariance matrices of the MGMM free to vary between groups (model M6 as described in Section 2.1.3). This example demonstrates that the introduction of outliers can surprisingly lead to an improved classification rate. When this model was fit to the data without contamination, the BIC identified that a single group had the most appropriate fit. Thus the classification of the true simulation groups was poor (50.0% accurately classified). However, when the contamination was included in the dataset, the best fitting model separated the two substantive groups and identified an extra group for the outliers. This model is shown in Figure 4.2.4. After the introduction of contamination, the substantive groups are now classified accurately, with 95.7% correctly classified. This is unexpected as the introduction of contamination would be expected to make classification more difficult, rather than easier. However, this is not a phenomenon that can be relied upon to always improve classification, as it could similarly lead to two groups being identified, when only a single group is required.

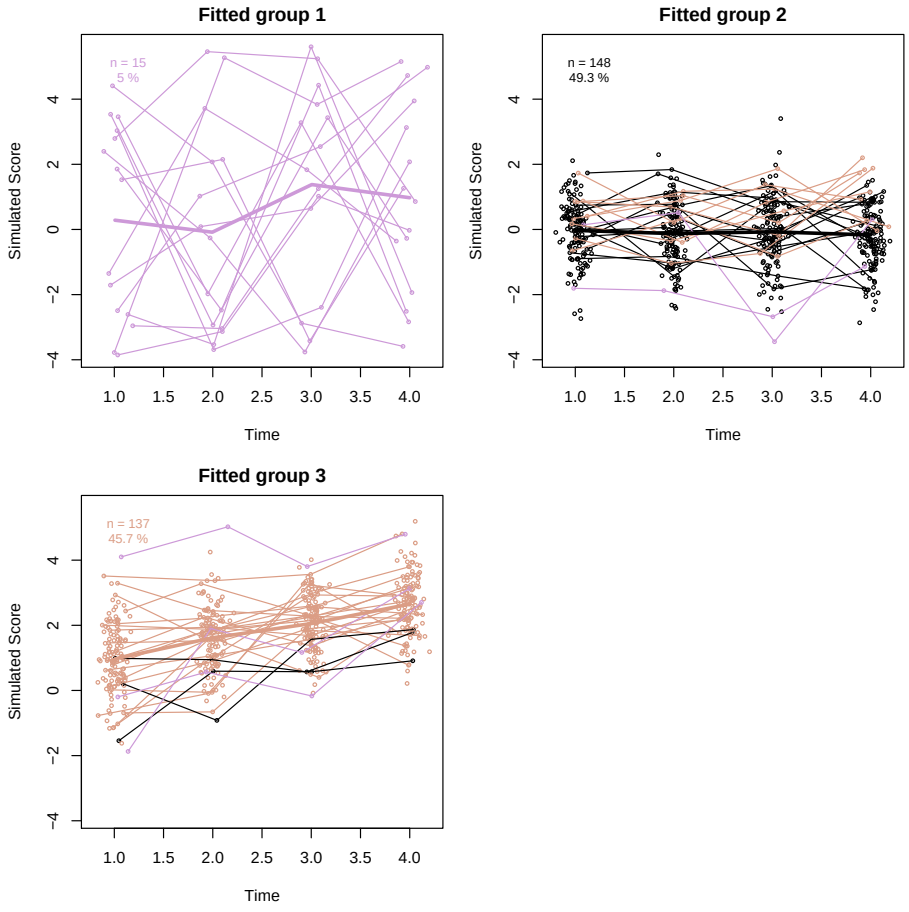


Figure 4.2.4: Estimated groups for example 2, with contamination included.

Example 3

The third example shows how the introduction of outliers can change the parameter estimates but not change the estimated number of groups. The dataset is shown in Figure 4.2.5 and is different from the previous two examples. In particular the two substantive groups are more separated than in examples 1 and 2, with the vertical separation at the first time point being 2 and the slope remaining at 0.5.

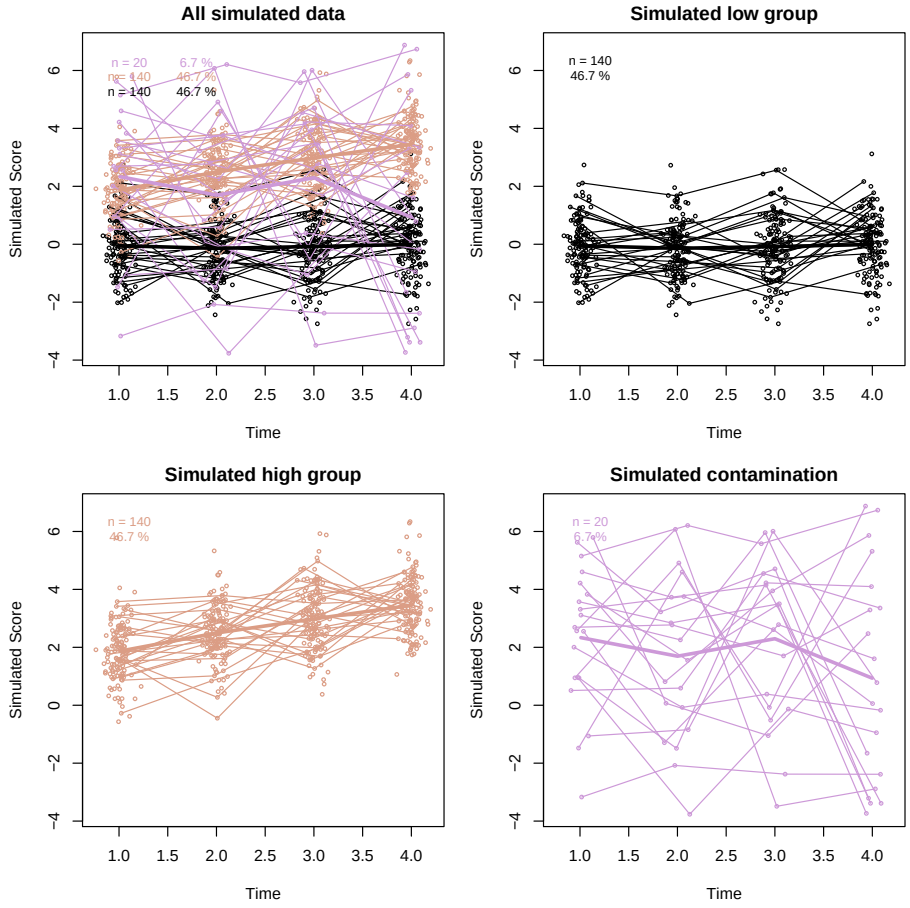


Figure 4.2.5: Simulated data for example 3.

When the M6 model was fit to this data, two groups provided the best fit both

when contamination was omitted and included. These fitted models are shown in Figures 4.2.6 and 4.2.7, respectively. The classification accuracy remained similar after the outliers were introduced, changing from 99.3% to 98.9% of substantive group trajectories allocated correctly, but the estimated parameters for the groups have changed considerably. In particular the estimated variance of each group increased at all time points. For the 'low' group the variance increased slightly from 0.97 to 1.02 on average, whereas for the 'high' group, the variance increased from 0.99 to 1.86 on average, because most of the contaminating trajectories were assigned to the 'high' group.

The above examples have demonstrated that outliers in the context of group-based trajectory models can have a range of effects. These include an increase in the number of groups, a surprising improvement in classification performance, and a change in the parameter estimates with the number of groups remaining unchanged. I will now describe how I assessed through simulation studies the impact of outliers when using group-based trajectory models.

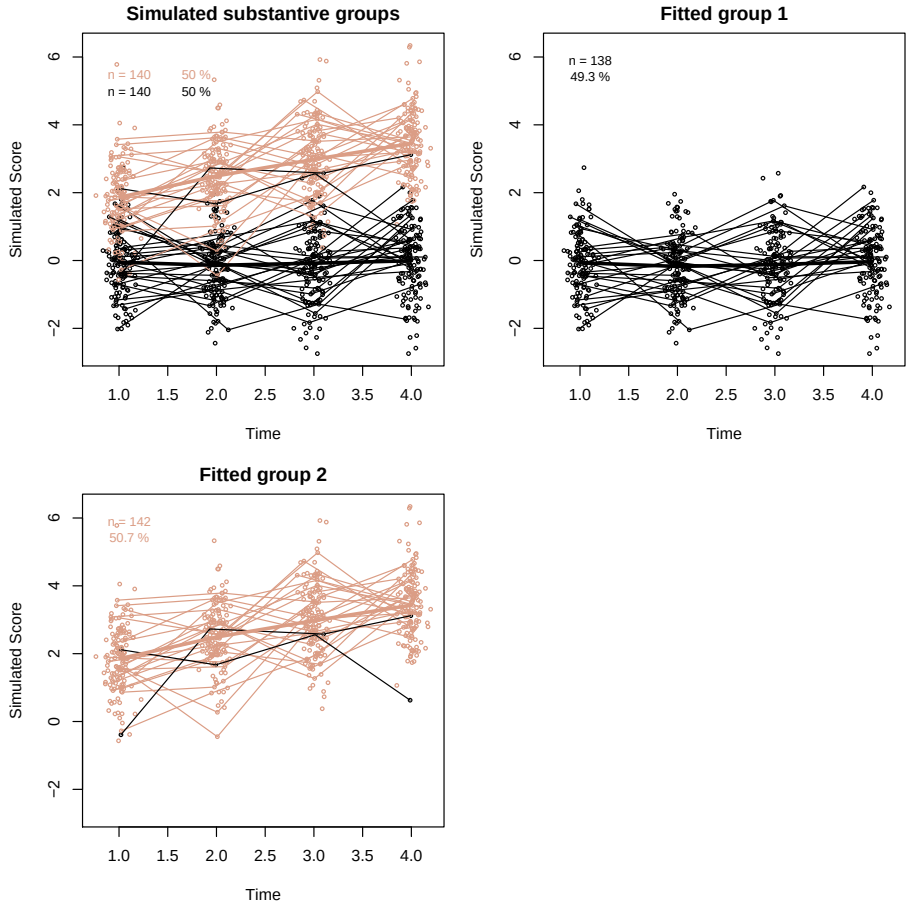


Figure 4.2.6: Estimated groups for example 3, with contamination omitted.

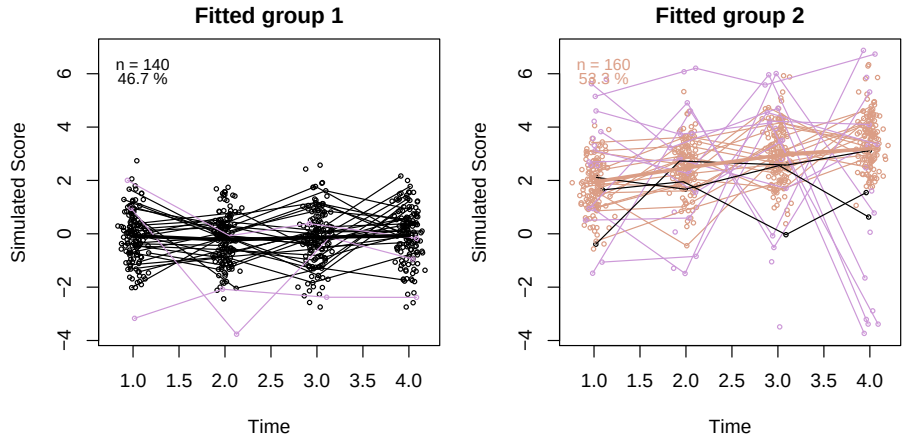


Figure 4.2.7: Estimated groups for example 3, with contamination included.

4.2.2 Quantifying outlier impact

In what follows the relative correct classification rate (CCR) of the substantive groups only will be considered as the primary measure of performance. The impact on the relative CCR will be used to quantify the effects of outliers on group-based trajectory models. The classification rate of the contaminating trajectories has not been included in the relative CCR as these trajectories may plausibly be assigned to the substantive groups, or to additional groups depending on their path. To define the relative CCR there are two technical issues that must be addressed.

First, as mixed models are 'unsupervised', the result is a grouping of trajectories rather than assignment to predefined classes. This results in the phenomenon of 'label switching' in mixture models.[66] To obtain a CCR for my simulations with two substantive groups, the resultant groups were relabelled with group 1 as the group with the most trajectories from substantive group 1, group 2 as the group with the most trajectories from substantive group 2, and the remaining trajectories (if any) labelled as belonging to outlier groups. In the case of the most trajectories from both substantive groups being allocated to the same group, this was relabelled as group 1 and the remaining trajectories (if any) were labelled as belonging to outlier groups.

Second, even with a correctly specified model with known parameters and no outliers, changes in the distance between mean trajectories affect the classification difficulty. Therefore, in order to compare the observed CCR with what would be

ideally possible, the CCR was evaluated relative to the Bayesian correct classification rate (BCCR):

$$\text{relative CCR} = \frac{\text{CCR}}{\text{BCCR}},$$

where the BCCR is the optimal classification proportion that can be expected with the knowledge of the underlying distributions, including the fact that there are two groups, as calculated using the Bayes minimum error classifier.[67]

The Bayes minimum error classifier is based on the log of the posterior odds:

$$g(\boldsymbol{x}) = \log \frac{f^1(\boldsymbol{x})}{f^2(\boldsymbol{x})} + \log \frac{P(c_1)}{P(c_2)},$$

where $f^k(\boldsymbol{x})$ is the probability density function for group k, with $\boldsymbol{x}$ being a trajectory, and $P(c_k)$ is the prior probability of group k. The trajectory is classified as group 1 (c_1) if $g(\boldsymbol{x}) > 0$, and group 2 (c_2) otherwise. As I have groups of equal size in all scenarios, $P(c_1) = P(c_2) = 1/2$ here. The BCCR can be calculated exactly in this case using Linear Discriminant Analysis:

$$\text{BCCR} = 1 - \Phi(-d/2),$$

where Φ is the distribution function of the standard normal distribution and d is the Mahalanobis distance between the group means.[71]

In this case, a relative CCR of 1 is achieved when the CCR meets the optimum level of the BCCR, whereas when one or two groups are estimated the lowest possible

relative CCR is 0.5, with a CCR of 0.5 and BCCR of 1. When there are three or more groups estimated a relative CCR below 0.5 is possible.

Alternative performance measures that could be considered to assess the impact of outliers are the estimated number of groups or bias in the parameter estimates. As seen in the examples, the number of groups estimated is not directly informative as additional groups can be appropriately formed for the contaminating trajectories. However, I will consider results on the estimated number of groups to assist in the understanding of the impact of outliers, particularly when the two substantive groups are not always able identified as distinct. The bias in parameter estimates is unlikely to add meaningful information to the conclusions one would make from the correct classification rate in a large number of circumstances. The bias would only provide additional information in situations such as in example 3 where the substantive classification is accurate, but the contamination has been assigned into the substantive groups. Furthermore, bias is difficult to define in circumstances where one substantive group has been divided into two groups. For this reason I do not present results on the bias of the estimates.

4.2.3 Outlier impact on divergent trajectories

Design of the divergent trajectories simulations

Here I describe the design of the first series of simulations to determine the effects of outliers on group-based trajectory models. In this series of simulations we consider

two groups. One group has a mean trajectory always equal to zero, with the other group having a variable linear mean trajectory. The mean trajectory of the variable group is altered through two parameters, the separation and the slope. Separation refers to the vertical separation of the group mean trajectories at the first time point. Slope refers to the slope of the variable group's mean trajectory. To allow a broader range of contexts for assessing the impact of outliers, both the separation and the slope were varied to alter the classification difficulty. When either the separation or the slope are large, the classification into two groups is easier. In what follows we will refer to these as the divergent trajectories simulations.

The simulation study consisted of 1000 simulated datasets of the two groups, each with 140 subjects across four evenly spaced time points, for 12 different scenarios. The 12 scenarios were all combinations in which the separation of the groups took possible values 1, 2 and 3 and the slope of the 'high' group took possible values 0, 0.25, 0.5 and 0.75. The slope of the 'low' group was constant at zero. The variance at each time point was 1 in both groups, and the off-diagonal elements of the covariance matrices were 0.25 in each group. The use of an exchangeable covariance structure was chosen so the impact of outliers could be investigated in a situation more complicated than conditional independence. Figure 4.2.8 shows the mean trajectories of the two groups across the various scenarios. Table 4.2.1 shows the Mahalanobis distances between the group mean trajectories, with the distance between the mean trajectories increasing with the separation and the slope.

To provide potential outliers, 20 contaminating trajectories (6.67%) were simu-

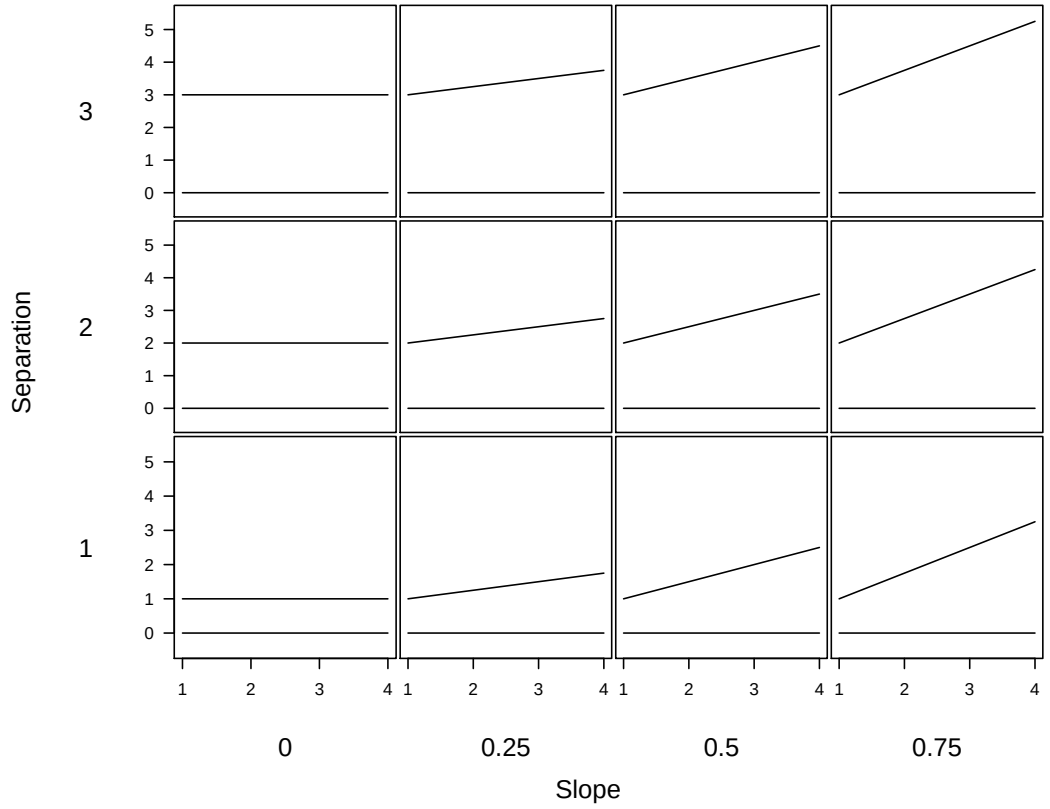


Figure 4.2.8: The mean trajectories of the 12 scenarios used in the divergent trajectories simulations.

Table 4.2.1: Mahalanobis distance between group means in the divergent trajectories simulations.

	Slope			
Separation	0	0.25	0.5	0.75
3	4.54	5.14	5.81	6.53
2	3.02	3.65	4.35	5.11
1	1.51	2.18	2.94	3.75

lated separately for each of the 1000 datasets. At each time point, observations of the contaminating trajectories were sampled from a uniform distribution between 4 less than the lowest group mean at that time point to 4 more than the highest mean at that time point. Models were fit to the datasets with the contamination omitted, and also with the contamination included, to determine the impact on the group-based trajectory model estimation.

Four different modelling approaches were used for each dataset. MGMMs were estimated with version 4 of the `mclust` package in R.[27,28]

1. Unstructured covariance matrix, fixed between the groups (model M5 from Section 2.1.3).

2. Unstructured covariance matrix, free to vary between the groups (model M6 from Section 2.1.3).

3. Using the BIC to choose between models 1 and 2.

4. Using the BIC to choose between all models available in `mclust`, including those with simpler covariance structures such as diagonal (models M1-M10 from Section 2.1.3).

These modelling approaches were chosen to provide a range of model assumptions, from restricted to flexible, for the assessment of outlier impact. For each model the number of groups was free to vary from 1 to 9 according to the minimum BIC. A larger number of groups was only chosen if the BIC improved by more than 6 over the BIC for the smaller number of groups, as described in Section 2.3.

For each scenario and model combination I calculated the mean CCR over the set of simulations, and divided it by the BCCR to derive the relative CCR.

The impact of the presence of outliers was compared using the number of groups estimated and the relative CCR of the substantive group trajectories, from the models fit to the datasets with the contamination omitted and included.

Computational difficulties

For 9 of the simulations there were computational issues, for at least one of the models. These computational issues led to 11 failures of `mclust` where it was unable to compute estimates. This either occurred with the initial dataset, or during outlier identification as will be described in Section 4.3. The 9 simulations were among 12,000, from 12 scenarios over 1000 simulations, and the 11 failures were among 48,000 combinations of the twelve scenarios, over 1000 simulations with 4 model fits. In these cases where computational difficulties occurred a fresh simulation was performed. In all instances it was not possible to determine the cause of these estimation failures. However, due to their low frequency, re-simulating those particular simulations could be expected to be of negligible impact.

Results for divergent trajectories simulations

In this section the impact of outliers on the number of groups estimated and the relative CCR are described.

The average number of groups estimated for the scenarios with the contaminating

trajectories omitted and included are shown in Table 4.2.2. Considering first the number of groups estimated when contamination was omitted, the correct number of groups was identified for models 1, 2 and 3 when the groups were well separated. This occurred when the separation was 3, and also when the separation was 2 and the slope was sufficiently large. It will be seen through the CCR and BCCR that high classification performance is also present in these well separated cases where two groups are identified. However, when the groups were not well separated, the 'best' model was often identified as having a single group. This suggests that for these scenarios with low separation and slope, the performance measure based on correct classification of the two substantive groups will provide limited information.

Table 4.2.2: Average number of groups estimated for various scenarios with and without contaminating trajectories.

		Slope							
		Without Contamination				With Contamination			
Model	Separation	0	0.25	0.5	0.75	0	0.25	0.5	0.75
1	3	2.000	2.000	2.000	2.000	5.183	5.478	5.527	5.633
	2	1.744	2.000	2.000	2.000	4.765	4.921	5.081	5.281
	1	1.000	1.007	1.608	1.999	2.666	3.186	4.188	4.421
2	3	2.000	2.000	2.000	2.000	3.195	3.213	3.242	3.218
	2	1.003	1.427	1.999	2.000	2.301	2.946	3.185	3.208
	1	1.000	1.000	1.003	1.539	1.981	1.982	2.250	2.979
3	3	2.000	2.000	2.000	2.000	3.497	3.546	3.503	3.443
	2	1.744	2.000	2.000	2.000	2.504	3.278	3.505	3.555
	1	1.000	1.007	1.608	1.999	2.021	2.014	2.505	3.447
4	3	2.042	2.036	2.045	2.027	3.900	3.815	3.781	3.733
	2	2.109	2.147	2.078	2.037	3.793	3.880	3.913	3.949
	1	1.164	1.329	2.004	2.100	3.037	3.300	3.593	3.748

With the inclusion of the contaminating trajectories, the average number of

groups estimated was higher for all scenarios, ranging from an increase of 0.76 to 3.6 groups. This increase could be due to:

- The substantive groups being divided into smaller groups by the presence of the contamination, which would lead to reduced classification performance. As will be seen in the relative CCR results, when classification performance decreased it did not do so to a large extent. This indicates that while some of the substantive group trajectories may often be incorrectly classified in the presence of contamination, it is unlikely that a substantive group will be close to evenly divided into multiple groups.

- A new group, or multiple groups, being estimated to contain the contaminating trajectories, as was seen in example 1 of Section 4.2.1. This would not impact on the classification performance as the CCR definition excludes contaminating trajectories. This outcome resulted in the most new groups for model 1, where the average number of groups estimated was higher than for the other models in almost all scenarios when contamination was present. This is because the constraint of equal covariance matrices between groups prevented a single group from being formed for the contaminating trajectories, as often occurred for the other models. This single group requires a large variance at each time point, which is inconsistent with the variances of the substantive groups. Thus, for model 1, many groups were estimated with a few contaminating trajectories in each.

- The substantive groups being identified as distinct, when the best fitting model had a single group in the absence of the contamination. This would lead to the classification performance improving, as was the case in example 2 of Section 4.2.1. Despite this phenomenon improving classification performance, it is likely to be an artifact of these scenarios and not to generalise to wider applications. My explanation for the cause of this phenomenon is returned to in Section 4.2.5.

The average relative CCR for the scenarios with the contaminating trajectories omitted and included are shown in Table 4.2.3. The underlying BCCR values are shown in Table 4.2.4.

Table 4.2.3: Average relative CCR for the 12 scenarios with and without contaminating trajectories.

		Slope							
		Without Contamination				With Contamination			
Model	Separation	0	0.25	0.5	0.75	0	0.25	0.5	0.75
1	3	0.999	1.000	1.000	1.000	0.934	0.932	0.935	0.934
	2	0.879	0.997	0.999	1.000	0.907	0.929	0.940	0.934
	1	0.645	0.583	0.818	0.997	0.705	0.748	0.879	0.944
2	3	0.998	0.999	1.000	1.000	0.987	0.989	0.991	0.994
	2	0.536	0.723	0.997	0.999	0.655	0.908	0.985	0.986
	1	0.645	0.580	0.539	0.776	0.641	0.584	0.654	0.944
3	3	0.999	1.000	1.000	1.000	0.986	0.986	0.989	0.992
	2	0.879	0.997	0.999	1.000	0.673	0.913	0.981	0.984
	1	0.645	0.583	0.818	0.997	0.643	0.588	0.685	0.954
4	3	0.992	0.994	0.992	0.995	0.841	0.852	0.856	0.870
	2	0.843	0.969	0.986	0.992	0.808	0.829	0.831	0.824
	1	0.692	0.645	0.810	0.978	0.869	0.808	0.833	0.845

When contamination was not present and the separation between the groups was

Table 4.2.4: BCCR without contaminating trajectories.

	Slope			
Separation	0	0.25	0.5	0.75
3	0.988	0.995	0.998	0.999
2	0.935	0.966	0.985	0.995
1	0.775	0.862	0.929	0.970

large, the relative CCR was high. This occurred in the same scenarios as which the correct number of groups was identified. As would be expected, the relative CCR decreased in the low separation scenarios described above where the two groups tended to be combined into a single group. When this occurred the CCR was defined to be 0.5, as half of the substantive trajectories were incorrectly grouped with the rest. For example, with a separation of 1, and a slope of 0, model 1 identified the best fitting model as having one group for all simulations. Therefore the average CCR was calculated to be 0.5, despite the BCCR being 0.775, for a relative CCR of 0.645. However with a slope of 0.5, and still a separation of 1, model 1 identified the best fitting model as having two groups for 60.8% of the simulations and one group for the remaining 39.2%. In the simulations where two groups provided the best fit, the average CCR was 0.928, close to the BCCR of 0.929. However across all simulations the average CCR was 0.760 for a relative CCR of 0.818.

In the presence of contamination, and when the groups were well separated, the relative CCR was lower than in the same scenarios without contamination. In the cases where the group means were close together and contamination allowed the two substantive groups to be identified as distinct, the relative CCR improved with

the introduction of the contamination. In the most extreme case, the relative CCR improved by 0.185 for model 2 when the separation was 2 and the slope was 0.25. However, as mentioned earlier, this occurred in the scenarios where the best fitting model in the absence of contamination had one group, so use of a classification based performance measure provides limited information.

4.2.4 Outlier impact on crossed trajectories

To investigate whether the unexpected results for close divergent trajectories generalised to other situations, I assessed whether similar results occurred for trajectories with crossed paths.

Design of the crossed trajectories simulations

The crossed trajectories simulation study consisted of 1000 simulated datasets of two groups of 140 subjects across four time points and with four scenarios. In these scenarios the mean trajectory of each group was zero half way between the second and third time points. The slope of one group (group 1) varied between 0.25, 0.5, 0.75 and 1, while the slope of the other group (group 2) was the negative of the slope in group 1. This formed a range of scenarios with crossed trajectories that become easier to classify as the slope increases. The covariance matrices were the same as for the divergent trajectories simulations. Figure 4.2.9 shows the mean trajectories of the two groups across the four scenarios. The Mahalanobis distances between the group means in the crossed trajectories simulations were 1.29, 2.58, 3.87 and 5.16 for the respective increasing slope values.

The same four model types and contamination simulation method were used for the crossed trajectories simulations as for the divergent trajectories simulations.

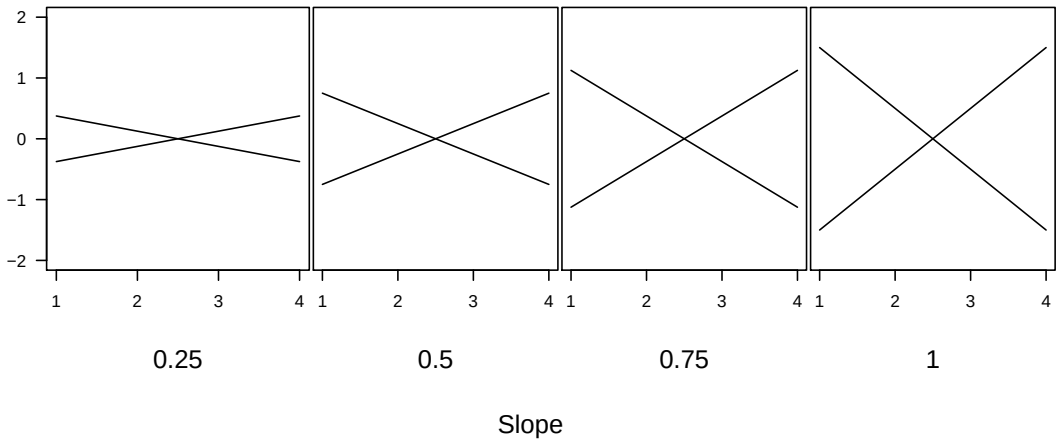


Figure 4.2.9: The mean trajectories of the four scenarios used in the crossed trajectories simulations.

Computational difficulties

As for the divergent trajectories simulations, estimation problems occurred in a small number of cases. For 17 of the simulations there were computational issues, for at least one of the models. These computational issues led to 24 failures of `mclust` where it was unable to compute estimates. This either occurred with the initial dataset, or during outlier identification as will be described in Section 4.3. The 17 simulations were among 4,000, from 4 scenarios over 1000 simulations, and the 24 failures were among 16,000 combinations of the 4 scenarios, over 1000 simulations with 4 model fits. In these cases where computational difficulties occurred a fresh simulation was performed. In all instances it was not possible to determine the cause of these estimation failures. However, due to their low frequency, re-simulating those particular simulations could be expected to be of negligible impact.

Results for crossed trajectories simulations

The average number of groups estimated for the crossed trajectories simulations, with the contaminating trajectories omitted and included are shown in Table 4.2.5. As with the divergent trajectories simulations, when the groups were well separated and contamination was not present, the correct number of groups was identified in all cases for models 1, 2 and 3. This occurred for models 1 and 3 when the slope was 0.75 or greater, and for model 2 when the slope was 1. However, when the slopes were low, and thus the classification problem was more difficult, the best fitting model was often identified as having a single group. In the presence of outliers, the average number of groups estimated was again higher than without outliers, as was the case for the divergent trajectories simulations.

Table 4.2.5: Average number of groups estimated for crossed trajectories simulations with and without contaminating trajectories.

	Slope							
	Without Contamination				With Contamination			
Model	0.25	0.5	0.75	1	0.25	0.5	0.75	1
1	1.000	1.111	2.000	2.000	1.678	1.805	3.403	3.818
2	1.000	1.000	1.786	2.000	1.805	1.665	2.502	2.902
3	1.000	1.111	2.000	2.000	1.845	1.774	3.176	3.475
4	1.326	1.182	2.030	2.032	2.772	2.879	3.798	3.883

The average relative CCR for the crossed trajectories simulations with the contaminating trajectories omitted and included are shown in Table 4.2.6. The underlying BCCR values are also shown in that table. For models 1, 2 and 3, unlike the previous scenarios, the relative CCR was roughly similar regardless of whether con-

tamination was present. In the scenarios where the number of groups was correctly identified with the contamination omitted, the relative CCR reduced slightly with the inclusion of outliers. The largest relative CCR reduction of 0.023 was for model 1 with a slope of 1. The relative CCR did again show unexpected improvements with the contamination included, although in this case for a slope of 0.5 only. The largest improvement in relative CCR after the contamination was included, among models 1 to 3, was 0.025 for model 1 with a slope of 0.5. Interpretation of this phenomenon is further discussed in Section 4.2.5. For model 4, the relative CCR was higher by 0.081 for a slope of 0.5 when contamination was present, whereas it was reduced by at least 0.166 when the slope was 0.75 or 1.

Table 4.2.6: BCCR without contamination and average relative CCR with and without contamination for crossed trajectories simulations.

		Slope							
		Without Contamination				With Contamination			
		0.25	0.5	0.75	1	0.25	0.5	0.75	1
Model	BCCR	0.741	0.902	0.974	0.995				
1	relative CCR	0.675	0.604	0.998	1.000	0.666	0.629	0.977	0.977
2	relative CCR	0.675	0.555	0.893	0.999	0.670	0.557	0.889	0.987
3	relative CCR	0.675	0.604	0.998	1.000	0.670	0.571	0.976	0.984
4	relative CCR	0.619	0.615	0.992	0.994	0.536	0.696	0.826	0.821

4.2.5 Discussion of outlier impact

From my simulation studies I have found that introducing outliers into group-based trajectory datasets tends to increase the number of groups estimated. This is often

due to extra groups being created for the outliers, that do not necessarily impact on the true group trajectories. For model 1, the constraint of equal covariance matrices between groups resulted in the outliers being assigned to several small groups. In contrast, for models 2, 3 and 4 the possibility of different covariance matrices between groups often allowed one group to be estimated to contain the outlying trajectories. These results were common between the divergent and crossed trajectories simulations.

If it happened that the contaminating trajectories were always confined to a new group or groups, it could be argued that they are less problematic here than in other statistical contexts. In particular the assignments of the substantive trajectories would be robustly preserved and the analyst would be readily able to identify the outlier group(s). However, this was not observed to be the case, as the introduction of outliers reduced the relative CCR when the group means were well separated and the number of groups was estimated correctly in the absence of outliers. This is because the new groups created for the outlying observations now also contained some of the substantive group observations that coincidentally had paths similar to those of the outliers.

In certain divergent trajectories simulations, where the group means were close together, the introduction of outliers could lead to the CCR actually improving, as described in Section 4.2.3. The same phenomenon was found to occur for the crossed trajectories when the slope was 0.5. This was due to the fact that in these circumstances the two substantive groups could have an appropriate model fit to

them with just one group. This group would have a covariance structure with high correlation over time, to account for the between group variation. For instance, in example 2 of Section 4.2.1, a single group was identified as having the best fit to the data without contamination. The covariance matrix and BIC for that single group model are shown in Table 4.2.7, as well as the covariance matrices and BIC for a two group model. Note the relatively high covariances in the off diagonal elements for the single group model. In contrast, the two group model had lower correlation over time within each group, as can be seen by the smaller covariances in the off diagonal elements. In these cases, the model selection using the BIC sees the penalty for the new parameters involved in an extra group being greater than the gain in the likelihood improvement. For the example 2 data, the BIC of the one group model was 21.9 below that for the two group model. The penalty included in the BIC for the 15 additional parameters required was greater than the improvement in the log-likelihood.

When contamination is included, with its uniform distribution that has an uncorrelated covariance structure, a model with a lower correlation between points over time is more likely. For the example 2 data with contamination included, the best fitting model identified the two substantive groups as distinct and also an extra group for the outliers, as shown in Figure 4.2.4. The covariance matrices and BIC for that three group model are shown in Table 4.2.8. In the three group model, the covariance matrix for the group fitted to the outliers had large variances and strong negative covariances in the off-diagonal elements.

Table 4.2.7: Covariance matrices and BIC of models with 1 and 2 groups fit to the data for example 2 without contamination.

Number of groups	Covariance Matrices	BIC
1	$\begin{pmatrix} 1.16 & 0.67 & 0.76 & 1.00 \\ 0.67 & 1.61 & 0.97 & 1.38 \\ 0.76 & 0.97 & 1.97 & 1.65 \\ 1.00 & 1.38 & 1.65 & 2.81 \end{pmatrix}$	3436.9
2	$\begin{pmatrix} 0.82 & 0.22 & 0.34 & 0.30 \\ 0.22 & 0.95 & 0.14 & 0.28 \\ 0.34 & 0.14 & 1.12 & 0.34 \\ 0.30 & 0.28 & 0.34 & 1.06 \end{pmatrix}$ $\begin{pmatrix} 1.06 & 0.33 & 0.17 & 0.38 \\ 0.33 & 0.85 & 0.08 & 0.20 \\ 0.17 & 0.08 & 0.70 & 0.18 \\ 0.38 & 0.20 & 0.18 & 0.84 \end{pmatrix}$	3458.8

Although it is not the preferred model it is interesting to consider what happens when we fit the two group model. One might expect the substantive groups would be identified together as they were in the absence of contamination, and the second group would consist of the contamination. Instead, the substantive trajectories were largely assigned to distinct groups reflecting how they were simulated, as shown in Figure 4.2.10. The covariance matrices and BIC for the two group model are shown in Table 4.2.8. The BIC of the two group model was 102.2 higher than that for the three group model. Thus, the improvement in the log-likelihood was now greater for the three group model than the penalty due to the 15 additional parameters.

In general the identification of the substantive groups as distinct occurs in situations where, in terms of the BIC, the likelihood improvement from fitting more groups is now greater than the penalty involved with the additional parameters. These unexpected results should act as a warning to researchers unfamiliar with

Table 4.2.8: Covariance matrices and BIC of models with 2 and 3 groups fit to the data for example 2 with contamination included.

Number of groups	Covariance Matrices	BIC
2	$\begin{pmatrix} 1.45 & 0.16 & 0.06 & -0.09 \\ 0.16 & 1.81 & 0.35 & 0.62 \\ 0.06 & 0.35 & 2.29 & 0.63 \\ -0.09 & 0.62 & 0.63 & 2.09 \end{pmatrix}$ $\begin{pmatrix} 0.90 & 0.36 & 0.26 & 0.31 \\ 0.36 & 0.92 & 0.21 & 0.31 \\ 0.26 & 0.21 & 0.73 & 0.29 \\ 0.31 & 0.31 & 0.29 & 0.84 \end{pmatrix}$	4152.9
3	$\begin{pmatrix} 6.73 & -0.23 & -2.00 & -3.23 \\ -0.23 & 7.95 & 0.24 & 1.56 \\ -2.00 & 0.24 & 7.96 & -0.58 \\ -3.23 & 1.56 & -0.58 & 6.67 \end{pmatrix}$ $\begin{pmatrix} 0.80 & 0.24 & 0.33 & 0.28 \\ 0.24 & 0.92 & 0.13 & 0.26 \\ 0.33 & 0.13 & 1.15 & 0.27 \\ 0.28 & 0.26 & 0.27 & 0.96 \end{pmatrix}$ $\begin{pmatrix} 1.18 & 0.43 & 0.29 & 0.45 \\ 0.43 & 0.97 & 0.20 & 0.29 \\ 0.29 & 0.20 & 0.80 & 0.27 \\ 0.45 & 0.29 & 0.27 & 0.93 \end{pmatrix}$	4050.7

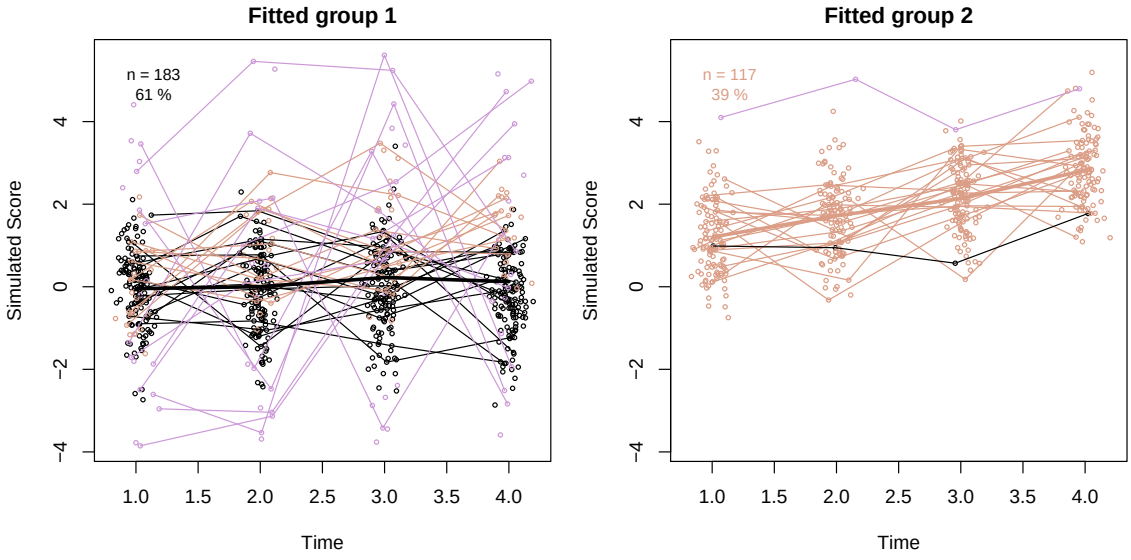


Figure 4.2.10: Two group model fit to the example 2 data with contamination included.

these models, and for researchers to be wary of situations when outlying trajectories may be present in their dataset.

4.3 Outlier identification

The previous section demonstrated that outliers in the data can lead to substantive groups being split, and thus can decrease the rate at which trajectories are correctly classified. To assess the impact of outliers, one can attempt to identify these trajectories and then investigate their impact. I have developed a method for the identification of outlying trajectories, based on the maximum change in the BIC due to the removal of a single trajectory.

This section describes the motivation for the algorithm as well as the algorithm itself, applies the algorithm to a real data set as an example, and demonstrates its performance in application to the simulated divergent and crossed trajectories datasets from the previous section.

4.3.1 Motivation for the outlier identification method

Group based trajectory modelling is a difficult setting for outlier identification. This is because the presence of a single outlying trajectory may not only influence the estimates for a particular group, but may also change the number of groups chosen or the allocation of trajectories to groups. Therefore an outlier cannot necessarily be determined just on the basis of fit of the current model. The BIC is the model

selection criterion commonly used for mixture modelling and it is natural to consider it as the basis for an outlier detection method in this setting. As described in Section 2.3, this is because the BIC accounts for the model fit but also penalises the number of parameters used in the model.

In this setting, a data set could be considered to not contain outliers if there are no outlying observations identified, as defined by low log-likelihood contributions of trajectories within each group. We say there are no outliers when the best fitting model is chosen according to the BIC. As an individual's log-likelihood contribution is the log of the probability of observing their particular combination of data characteristics, a low log-likelihood contribution implies a trajectory of low probability. Under the assumption of normality, the log-likelihood contribution of each trajectory is directly related to the Mahalanobis distance of that trajectory from the estimated group mean. The goal of outlier detection is then to find the maximal subset of the data for which this condition of no outlying observations holds.

It is computationally prohibitive to consider all combinations of subjects as potential outliers. For this reason a sequential approach is considered wherein one outlier is identified in each iteration of the algorithm. The algorithm identifies outlying trajectories based on the maximum change in BIC due to the removal of a single trajectory, with the number of groups in the model free to change from iteration to iteration.

Given that our definition of outlying observations relative to a given model is based on log-likelihood contributions of individual trajectories, a low log-likelihood

contribution within each group was used to select potential outlier candidates for comparison of the BIC after their removal. The algorithm considers o candidate outliers from each group for removal at each iteration. Since the smallest log-likelihood does not always correspond to the largest change in BIC, values of $o > 1$ are considered. For example if removing a certain trajectory results in a model with a lower number of groups having the best fit, this may provide additional gain in the BIC improvement due to the reduction in the number of parameters required. The choice of the number of outlier candidates is a trade-off between the additional time the algorithm will require and the desire to remove only the points which will improve the BIC by the largest extent. Typically we set o to be 2 or 3. Despite the use of multiple outlier candidates, it is still possible for a small number of outlying trajectories to mask each other from being identified as outliers as the algorithm only removes one trajectory per iteration.

When choosing between models with different numbers of groups, covariance structures or removed outlier candidates, the BIC is used as described in Section 2.3. As per the findings of the article in Chapter 3, the best fitting covariance structure for the full dataset is chosen at the start of the algorithm.[61] This covariance structure is then maintained throughout the algorithm.

Although the algorithm is explained in terms of the sequential removal of individual outlying trajectories it is not our recommendation that these trajectories by discarded. Rather, it is intended as a tool to help the analyst identify trajectories for further scrutiny.

4.3.2 Method of outlier identification using the BIC

Overview of the outlier identification method

The following is an overview of the process used in the outlier identification algorithm. The steps of the algorithm are described subsequently in more detail.

Testing for the presence of outliers Before identifying particular trajectories as outliers, the algorithm establishes whether outlying trajectories are present. In particular, the algorithm determines whether all trajectories could plausibly have come from the fitted group-based trajectory model. This is based on the log-likelihood contributions of the individual trajectories.

A small number of trajectories with the lowest log-likelihood contributions from each group are identified and hereafter referred to as the outlier candidates. To assess the significance of these individuals a simulation approach is used, with samples simulated from the fitted model. The log-likelihood contributions of the candidates are compared to the distribution of the corresponding order statistics of the simulated trajectories from each group. For example, the lowest log-likelihood contribution in the first group is compared to the lowest log-likelihood contributions from the samples simulated for that group. A P-value is calculated for each candidate, as the proportion of relevant likelihood contributions greater than the observed likelihood contribution. A Bonferroni adjusted P-value cut-off is used, and if all P-values exceed this threshold, and all groups have more than one trajectory, it is concluded no outliers are present and the algorithm terminates. Otherwise

the algorithm continues to identify a single 'outlying' trajectory. The Bonferroni adjustment is appropriate in this situation, as there will be a fairly small number of multiple comparisons, and only a single P-value over the threshold is required for the algorithm to continue.

Process of identifying an outlying trajectory The outlying trajectory is identified as the single trajectory, from among the outlier candidates, that causes the BIC to reduce by the largest amount when that trajectory is removed. To calculate the BIC, the model is re-fitted, and in this process the number of groups can also change.

Once this single trajectory is identified, it is removed from the dataset and the algorithm is repeated iteratively until a point is reached where no outliers are identified in the first step. The output of the algorithm is the subset of trajectories removed from the dataset, which forms the list of identified outliers. These outliers can then be considered by the analyst for evidence of errors or used in a sensitivity analysis to determine their impact on the results.

The outlier identification algorithm

Inputs The algorithm requires the following inputs:

- The T different covariance structures of the group-based trajectory models to be used.
- X, the data set of interest.

- $K_{\max}$, the maximum number of groups to be considered.
- α, the significance level.
- o, the number of trajectories to be considered as outlier candidates from each group.
- M, the number of simulated samples generated at each step.

The steps of the algorithm are now described in detail.

Initialisation

1. For $K \in \{1, \ldots, K_{\max}\}$ and $t \in \{1, \ldots, T\}$, fit group-based trajectory models to X with t as the covariance structure and K as the number of groups. Define a_{tK} to be the BIC obtained when the model has covariance structure t and K groups. Let $a = \min_{t,K} a_{tK}$.
2. Set $j \leftarrow 0$. $K_j = \arg\min_K \{a_{tK} : a_{tK} < a + 6, t \in \{1, \ldots, T\}\}$ is the initial number of groups. $t_{\text{init}} = \arg\min_t (a_{tK_j})$ is the initial covariance structure. This covariance structure is used until the termination criterion is met.

Iterative steps

3. Consider the o trajectories in each of the K_j groups that have the smallest log-likelihood contribution. These $o \times K_j$ trajectories are the outlier candidates. There may be fewer than $o \times K_j$ outlier candidates if there are fewer than o trajectories in some groups.

4. Simulate M samples from each group of the model and estimate the distributions of the first o order statistics of the log-likelihood contributions in each group.

5. For each of the outlier candidates, calculate the proportion $\{P_{ik} : i \in \{1, \ldots, o\}$ and $k \in \{1, \ldots, K_j\}\}$ of the M simulated samples for which the corresponding order statistic for that group is less than or equal to the observed log-likelihood contribution.

6. If every $P_{ik} \geq \frac{\alpha}{oK_j}$ and all groups have more than one trajectory the algorithm terminates. Otherwise continue.

7. Denote the outlier candidates $\{X_{ik} : i \in \{1, \ldots, o\}$ and $k \in \{1, \ldots, K_j\}\}$. For $K \in \{1, \ldots, K_{\max}\}$, $i \in \{1, \ldots, o\}$ and $k \in \{1, \ldots, K_j\}$, fit a model with covariance structure t_{init} to $X \setminus X_{ik}$ with K as the number of groups. Define b_{Kik} to be the BIC when the model with K groups is fit to $X \setminus X_{ik}$ and let $b = \min_{K,i,k} b_{Kik}$.

8. Set $j \leftarrow j+1$. $K_j = \arg\min_{K}\{b_{Kik} : b_{Kik} < b+6, i \in \{1, \ldots, o\}, k \in \{1, \ldots, K_{j-1}\}\}$ is the updated number of groups. Therefore, at this step the number of groups is free to change, whereas the covariance structure was fixed in the previous step according to that chosen in step 2.

9. Remove X_{ik} from X and return to step 3, where $(i, k) = \arg\min_{i,k}\{b_{K_j ik}\}$.

When the algorithm terminates the trajectories identified as outliers are those trajectories that have been removed.

Post iteration model fit for sensitivity analysis

To assess the impact of the identified outliers, a sensitivity analysis can be used in which the model fit with all data is compared to the model fit with the outliers removed. When conducting this sensitivity analysis, the following check of all model covariance structures and numbers of groups can be conducted to ensure the best model is fit to the remaining data after outlier removal.

After the algorithm terminates, for $K \in \{1, \ldots, K_{\max}\}$ and $t \in \{1, \ldots, T\}$, fit models to the resulting dataset with t as the covariance structure and K as the number of groups. Define c_{tK} to be the BIC obtained when the model has covariance structure t and K groups. Let $c = \min_{t,K} c_{tK}$. $K_{\text{end}} = \arg\min_{K}\{c_{tK} : c_{tK} < c + 6, t \in \{1, \ldots, T\}\}$ is the final number of groups. $t_{\text{end}} = \arg\min_{t}(c_{tK_j})$ is the final covariance structure.

4.3.3 Application of the outlier identification method to Generation 1 Study data

This section provides an example of the application of the outlier identification algorithm to the externalising data from the Generation 1 Study, as described in Section 1.1.2. Figure 4.3.1 shows the CBCL scores of externalising behaviour for all 557 children in the study, with a random 10% of the trajectories joined. This

was done as connecting all of the points would make it difficult to identify any individual trajectories. The general trend is for behaviour to improve from age 2 to age 9 (i.e. scores decrease). In this application of the outlier identification algorithm, I restricted the externalising data to complete cases only (n=354) for simplicity when using the `mclust` package.[27] Standardised time points were used to aid the convergence of the group-based models.

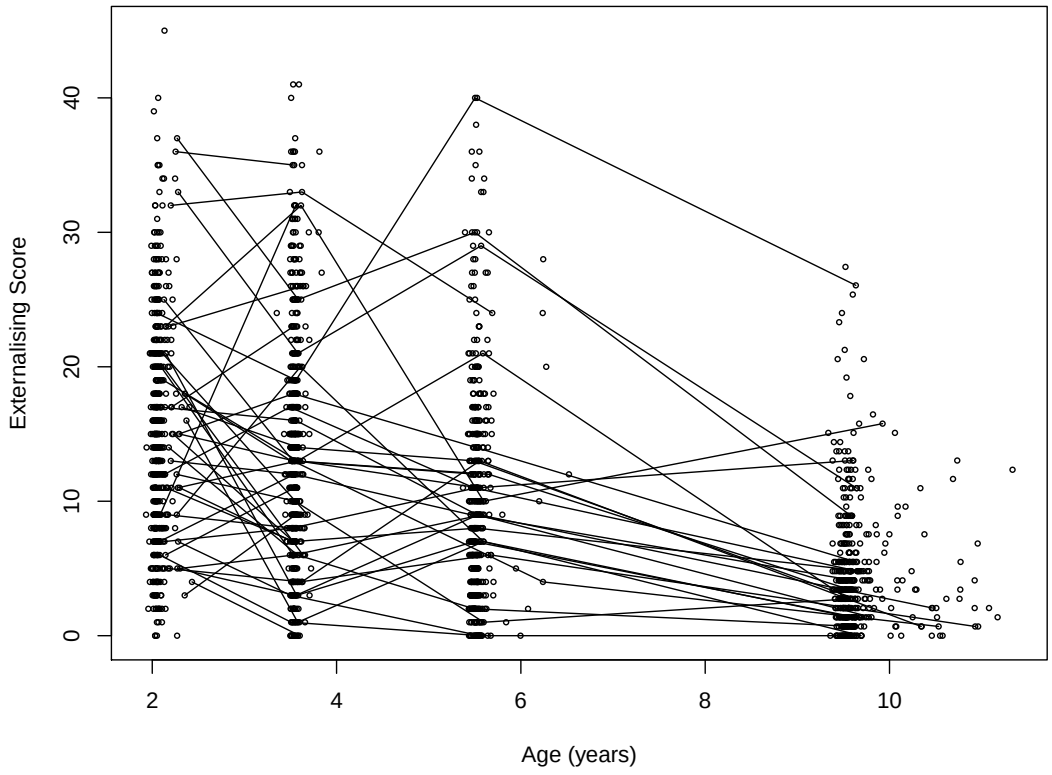


Figure 4.3.1: Externalising behaviour measurements from the Generation 1 Study, with a random 10% of the trajectories joined.

The following inputs were provided to the outlier identification algorithm for this example. All available covariance structures in version 4 of `mclust` were considered (models M1-M10 of Section 2.1.3). This was appropriate as the use of a variety

of possible covariance structures is recommended when investigating a dataset for group-based trajectories.[61] The numerical inputs had the following values: $K_{\max} = 9$, $\alpha = 0.05$, $o = 2$ and $M = 200$.

In the initialisation steps, the minimum BIC (8711.06) was achieved for a model with four groups ($K_c = 4$) and a covariance structure with $\boldsymbol{\Sigma}^k = \lambda_k \boldsymbol{A}$, meaning the covariance matrices are constrained to be diagonal, and proportional between groups (corresponding to model M7 of Section 2.1.3). No other combination of number of groups and covariance structure had a BIC within six of this minimum value. The fit of this model to the externalising data is shown in Figure 4.3.2. There is a 'very low' group with 57 observations, a 'low' group with 171 observations, a 'medium' group with 95 observations, and a 'high' group with 31 observations.

In the first of the iterative steps, the two candidate outlier trajectories with the smallest log-likelihood contributions in each group were identified. These trajectories, highlighted in red in Figure 4.3.3, are the trajectories with lowest probability in each group.

Steps 4, 5 and 6 use simulated samples from the model distribution to estimate the probability P_{ik} of obtaining the observed log-likelihood contributions, and compare them to the threshold $\frac{\alpha}{oK_c}$. The estimated P_{ik} are shown in Table 4.3.1. In this instance, $\frac{\alpha}{oK_c} = 0.00625$ and as at least one of the probabilities is less than that threshold the algorithm continues.

In steps 7, 8 and 9, the algorithm finds the single trajectory among the candidates whose removal results in a model with the smallest K such that the new model's BIC

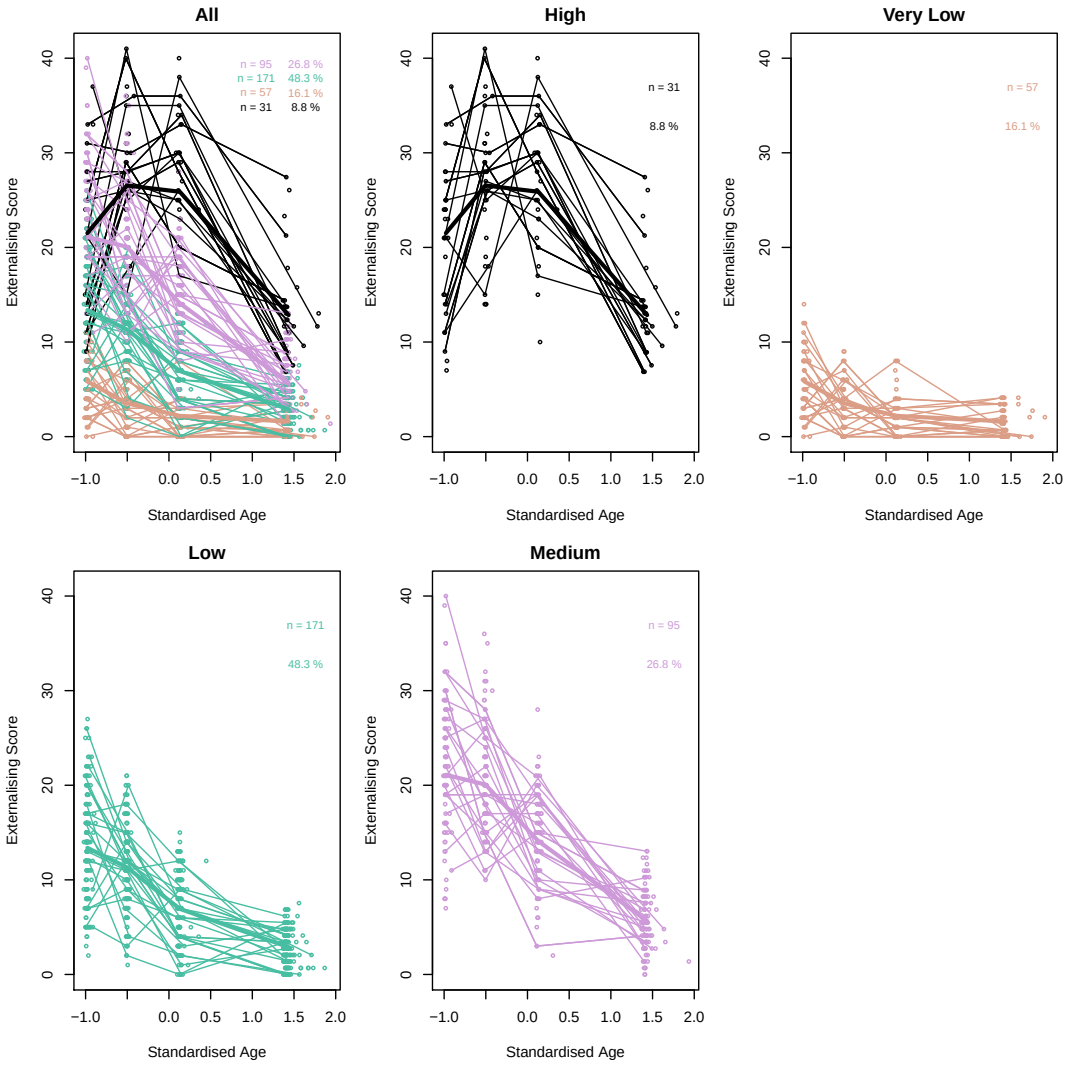


Figure 4.3.2: Four group M7 model fit to the externalising data.

Table 4.3.1: Estimated P_{ik} based on 200 simulations.

Group	Smallest	Second Smallest
Very Low	0.425	0.120
Low	0.795	0.530
Medium	0.500	0.550
High	0.005	<0.005

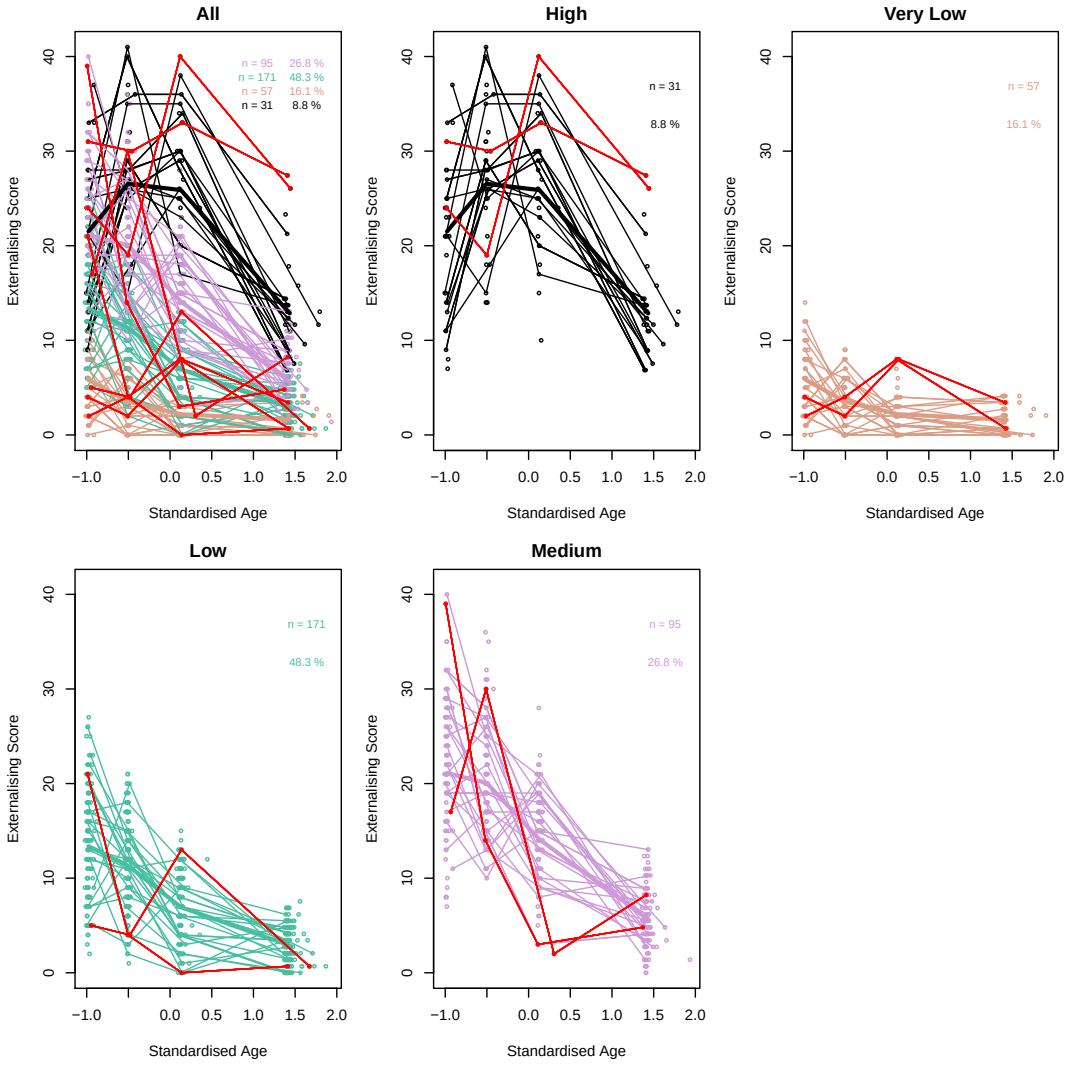


Figure 4.3.3: Four group M7 model fit to the externalising data with the two candidate outliers from each group highlighted (first step of the algorithm).

is within six of the lowest BIC, among possible new models with $K \in \{1, \ldots, K_{\max}\}$. This requires estimating the 72 models with between one and nine groups with each of the eight trajectories individually removed, to determine which model has the lowest resulting BIC. In this example, only the M7 covariance structure is used in the iteration steps as it was selected in the initialisation steps. In this case the BIC was reduced most by removal of the candidate trajectory from the 'high' group with the lower externalising score at the first time point. With that trajectory removed, the best fitting model according to the BIC also changed to have five groups instead of four. The algorithm then returns to step 3 with the data set that now has the single outlying trajectory removed.

The algorithm iterates until it finds no trajectory with a probability below the threshold at step six. For this example, it takes the removal of three trajectories and they are those identified as outliers, as shown in Figure 4.3.4.

To assess the impact of the identified outliers, a sensitivity analysis was conducted to compare the model fit with all data to that with the identified outliers removed. The final model chosen in the post iteration fit was a five group M7 model, as shown in Figure 4.3.5. No other covariance structure was found to achieve a BIC within six of the minimum achieved for the five group M7 model (8570.50). The four group M7 model was the second lowest with a BIC of 8578.12. In this case the best fitting model has changed from four groups to five groups by removing the outlying trajectories. The group memberships of the two models are cross-tabulated in Table 4.3.2. Although all three trajectories identified as outliers were from the

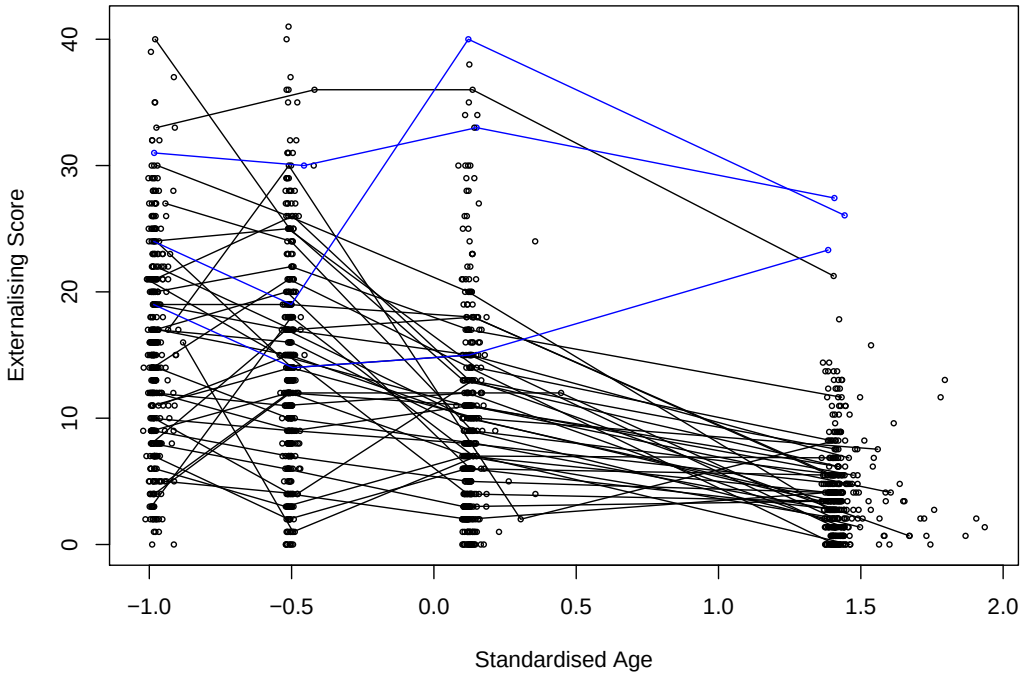


Figure 4.3.4: Externalising data with the three trajectories identified as outliers highlighted.

'High' group, the remaining 'High' group was maintained as a distinct group in the model fit with the outliers removed. However, the other three groups were split into four groups despite none of their group members being identified as outliers and removed. The use of my outlier identification algorithm has shown that the initial estimation of a four group model was quite unstable and dependent on a small number of influential outlying trajectories.

Table 4.3.2: Cross-tabulation of group memberships from models with all data and with outliers removed.

	Group with outliers removed					
Group with all data	Very Low	Low	Low-Medium	Medium	High	Outliers
Very Low	23	34	0	0	0	0
Low	0	42	129	0	0	0
Medium	0	0	14	81	0	0
High	0	0	0	0	28	3

4.3.4 Application of the outlier identification method to simulated data

To assess the effectiveness of the outlier identification algorithm, I applied it to the same simulated datasets as used for the simulations concerning the impact of outliers. The descriptions of the simulated datasets are provided in Sections 4.2.3 and 4.2.4.

The divergent trajectories simulation study consisted of 1000 simulated datasets of two groups of 140 subjects across four evenly spaced time points for each of 12

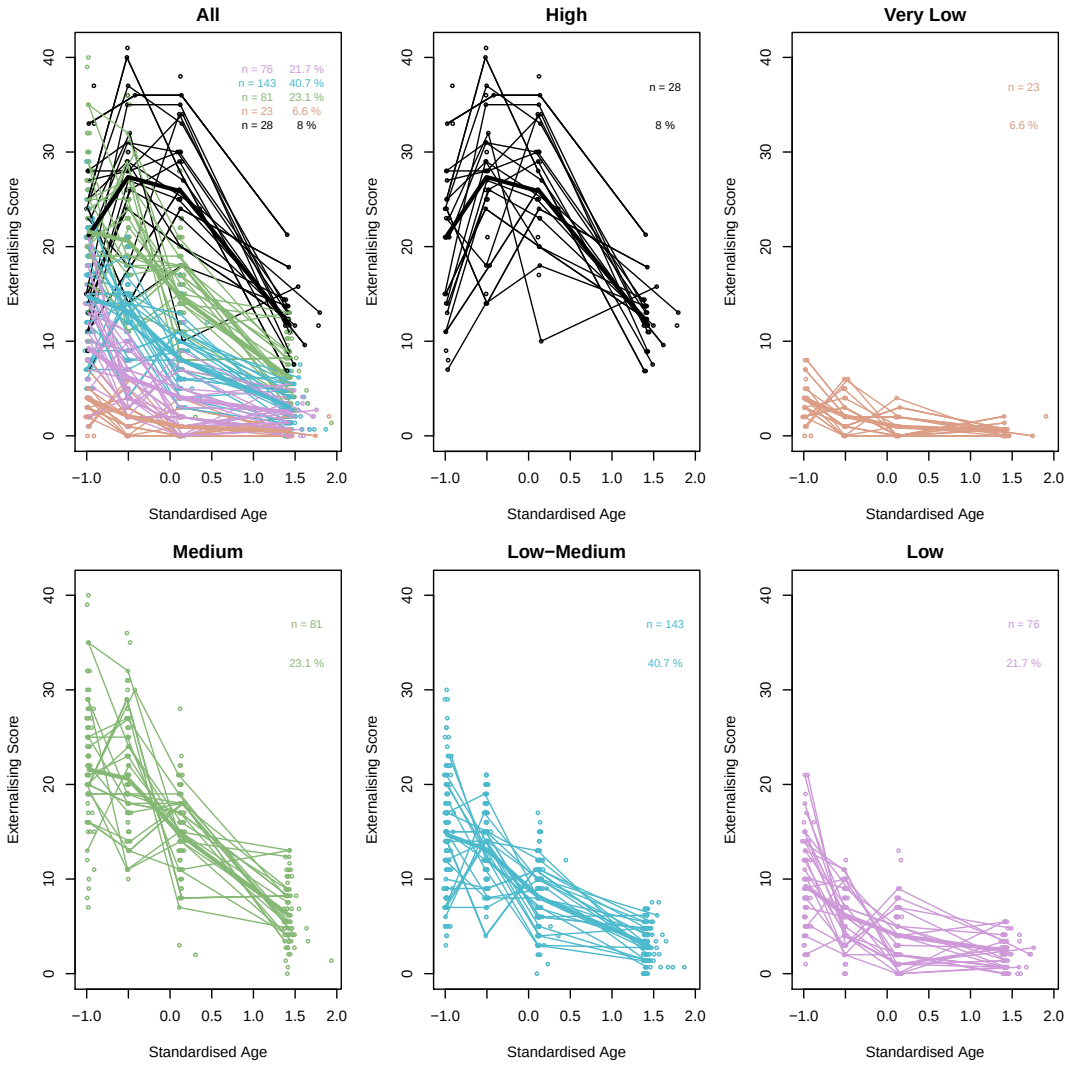


Figure 4.3.5: Five group M7 model identified as the best fit to the externalising data in the post iteration sensitivity analysis.

scenarios. The 'low' group has a mean trajectory always equal to zero, and thus a slope of zero, with the 'high' group having a variable linear mean trajectory. The mean trajectory of the 'high' group is altered through two parameters, the separation and the slope. Separation refers to the vertical separation of the group mean trajectories at the first time point. Slope refers to the slope of the 'high' group's mean trajectory. The 12 scenarios were all combinations in which the separation of the groups varied between 1, 2 and 3 and the slope of the 'high' group varied between 0, 0.25, 0.5 and 0.75. The variance at each time point was 1 in both groups, and the off-diagonal elements of the covariance matrices were 0.25 in each group.

To introduce outliers, 20 contaminating trajectories were added to the simulated data (6.67%). The contamination was sampled from a uniform distribution at each time point, between 4 less than the lowest group mean at that time point to 4 more than the highest mean at that time point.

The crossed trajectories simulation study also had 1000 simulated datasets of two groups of 140 subjects across four time points, but for only four scenarios. In these scenarios the mean trajectory of each group was zero half way between the second and third time points. The slope of one group (group 1) varied between 0.25, 0.5, 0.75 and 1, while the slope of the other group (group 2) was the negative of the slope in group 1. This formed a range of scenarios with crossed trajectories that become easier to classify as the slope increases. The covariance matrices were the same as for the divergent trajectories simulations. Twenty contaminating trajectories were added to the crossed trajectories data in the same way as for the divergent

trajectories simulations.

The MGMMs were estimated with version 4 of the `mclust` package in R.[27,28] The same four modelling approaches were used for this application of the outlier identification method as were used in the impact of outlier simulations. These were:

1. Unstructured covariance matrix, fixed between the groups (model M5 from Section 2.1.3).

2. Unstructured covariance matrix, free to vary between the groups (model M6 from Section 2.1.3).

3. Using the BIC to choose between models 1 and 2.

4. Using the BIC to choose between all models available in `mclust`, including those with simpler covariance structures such as diagonal (models M1-M10 from Section 2.1.3).

Outlier identification was applied with $\alpha = 0.05$, $M = 200$, $o = 2$ and $K_{\max} = 9$ for these simulations.

Performance measures To quantify the performance of the outlier identification method, various measures were used.

The *average number of trajectories identified*, according to whether they were substantive or contamination trajectories, was used to initially assess the extent to which the algorithm identified outliers among each type of trajectories.

An additional measure was used to determine the quality of the trajectory identification, which was the proportion of trajectories identified as outliers that had been introduced as contamination, referred to as the *contamination predictive value* (CPV). The CPV was calculated as the average of the numerator divided by the average of the denominator, to avoid the issue of many small denominators, including zero values as denominators. Ideally all identified trajectories would be contamination trajectories, and the CPV would be 1.

The impact of the outliers identified using the algorithm was assessed in the following performance measures by comparing the results for the model fit to the full dataset to the post iteration model fit to the dataset with trajectories identified as outliers removed. The impact of those identified as outliers reflects on the performance of the algorithm itself.

The *average number of groups* before and after outlier removal was used as a measure of performance, to show the extent to which the outlier identification method removed groups formed as a result of outliers being present.

The *relative CCR* provided the key measure of performance for how well the substantive group trajectories were classified, before and after outlier removal. The denominator of the CCR was the initial number of substantive group trajectories, even if some of these trajectories were removed.

The *number of contamination trajectories assigned to a substantive group*, before and after outlier removal, was used to assess how many contamination trajectories were incorrectly allocated. These contamination trajectories may impact the pa-

rameter estimates of the substantive groups, whereas contamination trajectories assigned to their own groups are not problematic.

Due to the method used to simulate them, contamination trajectories are not necessarily outliers. Therefore when contamination trajectories are assigned to a substantive group, this can be due to their trajectory path being close to the trajectories within that group by chance. To determine how frequently contamination trajectories were simulated close to a substantive group, the *number of contamination trajectories less than the 99th percentile of Mahalanobis distance from the closest substantive group mean* was calculated before model fitting and outlier removal. The population mean and covariance matrix were used for each substantive group when calculating the Mahalanobis distance. The Mahalanobis distances were converted to percentiles based on the χ^2 distribution with four degrees of freedom. The threshold level of 0.99 for the Mahalanobis distance percentile was chosen so that the vast majority of the substantive trajectories simulated within that group would be considered close to the other trajectories of that group by this standard.

To visualise the performance of the outlier identification algorithm, plots of *the Mahalanobis distance percentiles of all trajectories from their assigned group mean*, before and after outlier removal, were used. The estimated means and covariance matrices were used for each group when calculating the Mahalanobis distance. Trajectories in the following categories were represented distinctly: substantive trajectories, contamination trajectories assigned to an outlier group and contamination trajectories assigned to a substantive group.

Results for outlier identification from divergent trajectories simulations

In what follows the results of the outlier identification method for the complete set of divergent trajectories simulations are presented. To aid understanding of these results, an example of one particular simulation instance is described first. The example simulation uses model 1, with a separation of 3 and slope of 0. Figure 4.3.6 shows the simulated data before any model fitting.

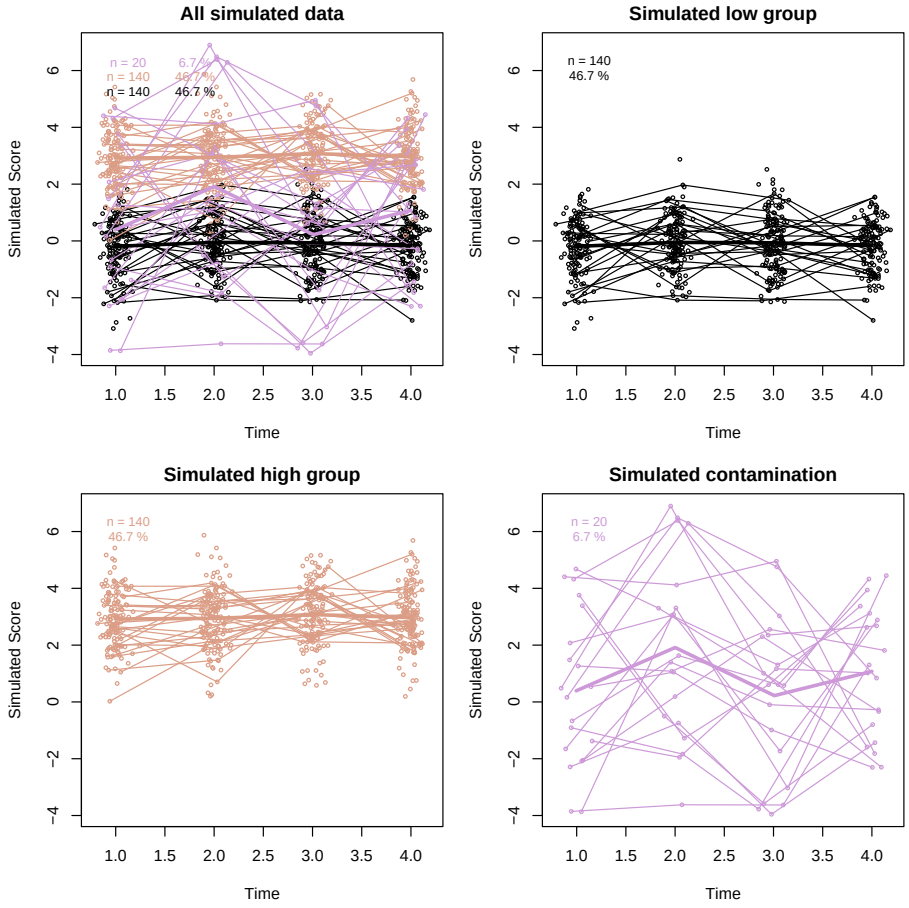


Figure 4.3.6: Simulated data for the results example.

Before outlier removal, the best fitting model had four groups and the CCR

was 0.986 as four substantive group trajectories were incorrectly classified. Also, 11 contamination trajectories were classified into the substantive groups, with six being assigned to the low group and five to the high group. Six of the contamination trajectories had a Mahalanobis distance percentile of less than 0.99 from the closest group mean trajectory. Figure 4.3.7 shows the model fit before outlier removal.

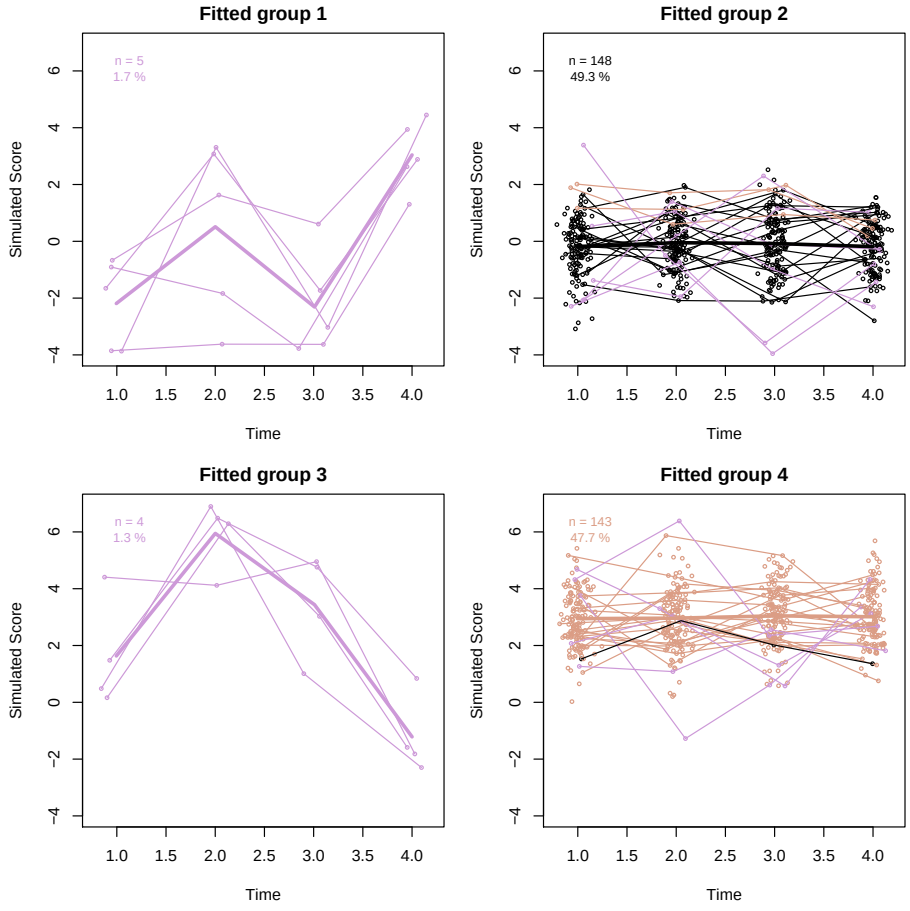


Figure 4.3.7: Model 1 fit to the results example data before outlier removal.

Application of the outlier identification algorithm resulted in the removal of 13 trajectories. This included 12 contamination trajectories and one substantive group trajectory, for a CPV of 0.923. After outlier removal, the best fitting model had two

groups, and of the remaining 279 substantive group trajectories, four were incorrectly assigned, giving a CCR of 0.982 as the denominator of the CCR is the initial number of substantive group trajectories. As there were no outlier groups estimated, all eight remaining contamination trajectories were assigned to a substantive group. Figure 4.3.8 shows the model fit after outlier removal.

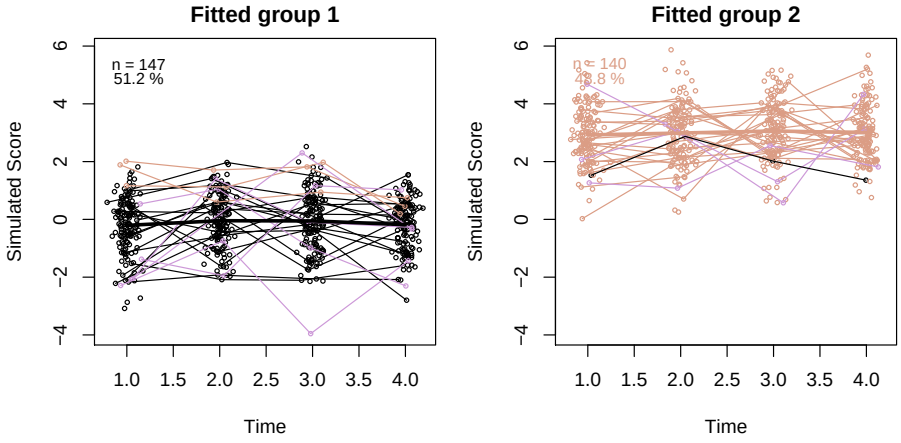


Figure 4.3.8: Model 1 fit to the results example data after outlier removal.

Table 4.3.3 shows the average number of trajectories identified by the outlier algorithm for the different divergent trajectories simulations, according to whether they were substantive or contamination trajectories. The number of trajectories identified varied considerably according to which model was used, and in some cases by separation and slope. For model 1, when the separation and slope were large enough the average number of contamination trajectories identified was greater than 15, though the number of substantive trajectories identified as outliers also increased with separation and slope. However, when the separation and slope were at their lowest less than half of the contamination trajectories were identified as outliers.

For models 2 and 4, very few trajectories were identified by the algorithm, with little impact of the separation or slope on the number of trajectories identified. The average number of trajectories identified in total was 1.15 for model 2 and 0.55 for model 4. As so few trajectories were identified by the algorithm for these models, the subsequent performance measures will not show any large changes for models 2 and 4 as a result of outlier removal. For model 3, the highest average number of contamination trajectories identified was 5.77, for a separation of 1 and a slope of 0.75. For this slope, the average number of contamination trajectories identified actually decreased as the separation increased. These results show the average number of trajectories identified by model 3 to be between those of models 1 and 2, which arises since model 3 is by definition the choice of models 1 and 2 according to the BIC.

Table 4.3.4 shows the CPV for the divergent trajectories simulations. The CPV varied greatly according to the model used, and also by separation and slope. It was lowest for model 1 when the separation and slope were at their highest, which was also when the number of trajectories identified was highest. For both models 1 and 3, the CPV decreased as the separation and slope increased. For both models 2 and 4, the CPV is high regardless of the separation and slope, however as mentioned, very few trajectories were identified for these models so a high CPV is not important.

Table 4.3.5 shows the average number of groups estimated before and after outlier removal for the different divergent trajectories simulations. As seen in Section 4.2.3, in the presence of contamination before outlier removal, the fitted group-based

Table 4.3.3: Average number of trajectories identified by the algorithm in divergent trajectories simulations, according to whether they were substantive or contamination trajectories.

		Slope							
		Substantive				Contamination			
Model	Separation	0	0.25	0.5	0.75	0	0.25	0.5	0.75
1	3	5.69	6.81	7.85	8.84	16.6	17.1	16.8	16.4
	2	2.56	3.24	4.19	5.82	15.0	15.3	16.2	16.0
	1	0.12	0.48	1.70	2.94	8.50	9.42	13.5	15.0
2	3	0.12	0.10	0.07	0.10	0.93	0.83	0.77	0.78
	2	0.19	0.16	0.04	0.10	1.19	0.96	0.82	0.84
	1	0.03	0.05	0.07	0.07	1.43	1.29	1.68	1.22
3	3	1.13	1.04	1.19	1.07	3.65	3.37	3.08	2.46
	2	0.41	0.64	1.13	1.55	2.49	3.81	3.96	3.93
	1	0.05	0.04	0.27	1.23	1.59	1.48	2.82	5.77
4	3	0.04	0.02	0.04	0.02	0.61	0.44	0.49	0.49
	2	0.02	0.03	0.02	0.03	0.34	0.70	0.61	0.45
	1	0.01	0.02	0.03	0.02	0.52	0.61	0.42	0.62

Table 4.3.4: Contamination predictive value for various scenarios in the divergent trajectories simulations.

		Slope			
Model	Separation	0	0.25	0.5	0.75
1	3	0.745	0.715	0.681	0.650
	2	0.854	0.825	0.795	0.734
	1	0.986	0.951	0.888	0.836
2	3	0.886	0.896	0.920	0.887
	2	0.864	0.856	0.951	0.890
	1	0.977	0.960	0.963	0.943
3	3	0.764	0.764	0.722	0.697
	2	0.858	0.857	0.778	0.718
	1	0.971	0.977	0.914	0.825
4	3	0.938	0.965	0.924	0.953
	2	0.958	0.960	0.970	0.929
	1	0.985	0.965	0.927	0.964

trajectory models have an average number of groups greater than two in almost all scenarios. After outlier removal the average number of groups reduced in all scenarios, but by varying degrees. The average number of groups reduced the most for model 1, where the number of groups estimated was higher when outliers were present. After outlier removal for model 1, the average number of groups was close to the correct number of 2 in the scenarios except with the separation of 1 and slope ≤ 0.25. This corresponds with the results in Table 4.3.3 as these were the scenarios where at least 15 trajectories were identified on average. As discussed earlier, for models 2 and 4, the average number of groups did not change greatly. The changes for model 3 were smaller than those observed for model 1, but greater than those for models 2 and 4. For model 3, the average number of groups after outlier removal remained around 3 even in scenarios with large separation of group means.

Table 4.3.6 shows the average relative CCR before and after outlier removal for the different divergent trajectories simulations. For model 1, where the largest number of outlying trajectories were identified, the relative CCR increased after outlier removal when the separation and slope were sufficiently large. This was the case for all simulations except those with a separation of 2 and a slope of 0 or those with a separation of 1 and a slope ≤ 0.5. The largest increase in the relative CCR was 0.05. However when the two group means were close together the relative CCR decreased by between 0.057 and 0.158. These correspond to the same scenarios identified in Section 4.2.3 where the relative CCR was higher after the introduction of outliers. In those scenarios the effect of the outliers was to find two groups, when

Table 4.3.5: Average number of groups estimated for various scenarios before and after outlier removal in divergent trajectories simulations.

		Slope							
		Before Outlier Removal				After Outlier Removal			
Model	Separation	0	0.25	0.5	0.75	0	0.25	0.5	0.75
1	3	5.18	5.48	5.53	5.63	2.15	2.20	2.32	2.46
	2	4.77	4.92	5.08	5.28	1.81	2.18	2.14	2.27
	1	2.67	3.19	4.19	4.42	1.39	1.41	1.72	2.09
2	3	3.20	3.21	3.24	3.22	3.03	3.04	3.05	3.03
	2	2.30	2.95	3.19	3.21	2.22	2.81	3.03	3.04
	1	1.98	1.98	2.25	2.98	1.93	1.94	2.14	2.87
3	3	3.50	3.55	3.50	3.44	2.87	2.89	2.94	2.94
	2	2.50	3.28	3.51	3.56	2.20	2.67	2.86	2.87
	1	2.02	2.01	2.51	3.45	1.91	1.93	2.17	2.65
4	3	3.90	3.82	3.78	3.73	3.89	3.83	3.77	3.71
	2	3.79	3.88	3.91	3.95	3.77	3.85	3.90	3.93
	1	3.04	3.30	3.59	3.75	3.00	3.24	3.57	3.73

in the absence of contamination a single group had been the best fitting. Therefore it is not surprising that the CCR has decreased after application of the outlier identification algorithm in these cases, as a single group model becomes more likely again. For models 2, 3 and 4, the average relative CCR did not change greatly after outlier removal. The differences ranged from a decrease of 0.016 to an increase of 0.009. This was expected as so few points were identified by the algorithm for these models.

Table 4.3.7 shows the average number of contamination trajectories assigned to a substantive group for various scenarios before and after outlier removal in the divergent trajectories simulations. Before outlier removal, the number of contamination trajectories assigned to a substantive group was high for model 1, particularly for

Table 4.3.6: Average relative CCR for various scenarios before and after outlier removal in divergent trajectories simulations.

		Slope							
		Before Outlier Removal				After Outlier Removal			
Model	Separation	0	0.25	0.5	0.75	0	0.25	0.5	0.75
1	3	0.934	0.932	0.935	0.934	0.976	0.971	0.965	0.959
	2	0.907	0.929	0.940	0.934	0.850	0.980	0.982	0.973
	1	0.705	0.748	0.879	0.946	0.644	0.590	0.787	0.985
2	3	0.987	0.989	0.991	0.994	0.990	0.991	0.992	0.995
	2	0.655	0.908	0.985	0.986	0.644	0.899	0.987	0.989
	1	0.641	0.584	0.654	0.944	0.641	0.583	0.637	0.938
3	3	0.986	0.986	0.989	0.992	0.988	0.990	0.988	0.991
	2	0.673	0.913	0.981	0.984	0.674	0.917	0.986	0.986
	1	0.643	0.588	0.685	0.954	0.641	0.584	0.677	0.963
4	3	0.841	0.852	0.856	0.870	0.839	0.845	0.853	0.865
	2	0.808	0.829	0.831	0.824	0.806	0.828	0.831	0.822
	1	0.869	0.808	0.833	0.845	0.866	0.806	0.832	0.844

low values of separation and slope. Even when the separation and slope were large, more than a third of contamination points were assigned to a substantive group. After outlier removal, the number of contamination trajectories assigned to a substantive group reduced considerably for model 1, particularly when the slope and separation were large. Reductions were also observed for models 2 and 3, although these changes were smaller than those observed for model 1. For model 4 the number of contamination trajectories assigned to a substantive group did not change greatly after outlier removal, as very few trajectories were removed. These results should be viewed in light of the number of contamination trajectories less than the 99th percentile of Mahalanobis distance from the closest substantive group mean, shown in Table 4.3.8. For each scenario this number was less than 3.08 on average,

and reduced with increasing separation and slope. Although small, the number of contamination trajectories assigned to a substantive group may not be expected to be reduced below these numbers even after outlier removal.

Table 4.3.7: Average number of contamination trajectories assigned to a substantive group for various scenarios before and after outlier removal in divergent trajectories simulations.

		Slope							
		Before Outlier Removal				After Outlier Removal			
Model	Separation	0	0.25	0.5	0.75	0	0.25	0.5	0.75
1	3	9.08	8.12	7.90	7.66	3.08	2.52	2.43	2.33
	2	10.3	9.99	9.42	8.90	4.49	4.19	3.43	3.24
	1	14.5	14.0	11.9	11.5	5.94	5.64	5.22	4.79
2	3	4.43	4.16	4.00	3.63	3.28	2.91	2.83	2.62
	2	4.57	4.70	4.46	4.44	3.82	3.67	3.29	3.24
	1	4.78	5.14	5.31	5.76	4.32	4.54	4.32	4.03
3	3	4.69	4.15	3.97	3.59	3.27	2.88	2.74	2.58
	2	4.76	5.23	4.97	4.76	3.98	3.79	3.36	3.24
	1	4.76	5.12	5.55	6.97	4.39	4.61	4.37	4.47
4	3	2.41	2.18	2.09	2.03	2.40	2.16	2.09	2.01
	2	2.63	2.69	2.43	2.34	2.67	2.70	2.44	2.35
	1	3.03	3.07	2.93	2.85	3.16	3.21	2.98	2.91

Table 4.3.8: Average number of contamination trajectories less than the 99th percentile of Mahalanobis distance from the closest substantive group mean, before outlier removal in divergent trajectories simulations.

	Slope			
Separation	0	0.25	0.5	0.75
3	1.99	1.75	1.58	1.41
2	2.49	2.35	2.10	1.95
1	3.08	2.89	2.67	2.56

Figures 4.3.9 to 4.3.16 show the Mahalanobis distance percentiles of all trajectories from their assigned group mean in the divergent trajectories simulations, before and after outlier removal, for models 1 to 4, and with either separation 3 and slope 0.75, or separation 1 and slope 0. The plots for the remaining values of the separation and slope are shown in Appendix A. The plots with separation 3 and slope 0.75 are for the scenarios when the substantive groups are furthest apart, whereas those for separation 1 and slope 0 are when the substantive groups are closest together. The scenarios with other values of separation and slope appear somewhere between these two extremes.

For model 1, when the substantive groups were furthest apart, most of the contamination trajectories were identified by the algorithm, regardless of whether they were assigned to a substantive group or an outlier group initially. However, when the substantive groups were closest together, a high proportion of the contamination trajectories with high Mahalanobis distance percentile remained unidentified by the algorithm. Many of these were allocated to an outlier group, which appeared to have only high Mahalanobis distance percentile values as the covariance between groups was constrained to be equal. The distribution of the Mahalanobis distance percentiles for substantive trajectories also had a peak for the highest values, as the substantive groups were generally combined together, so appeared more like a single distribution with a heavy tail.

For models 2, 3 and 4, the Mahalanobis distance percentiles do not change greatly after outlier removal, with relatively few trajectories being identified in these cases.

For these models there were also little differences with separation or slope.

In a small number of cases for models 1 and 3, Mahalanobis distances of trajectories could not be included in the plots as the means and covariance matrices could not be estimated. These were due to small numbers of observations in one or more of the groups. In each scenario this occurred for at most 1.8% of the trajectories to be plotted.

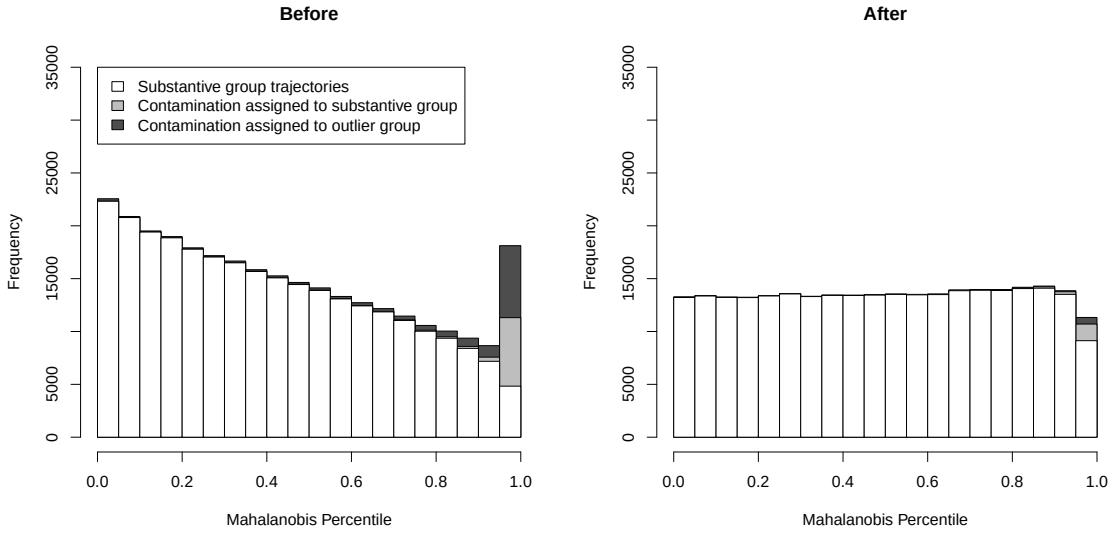


Figure 4.3.9: Mahalanobis distance percentiles of trajectories from their assigned group mean, for model 1 with separation 3 and slope 0.75, before and after outlier removal.

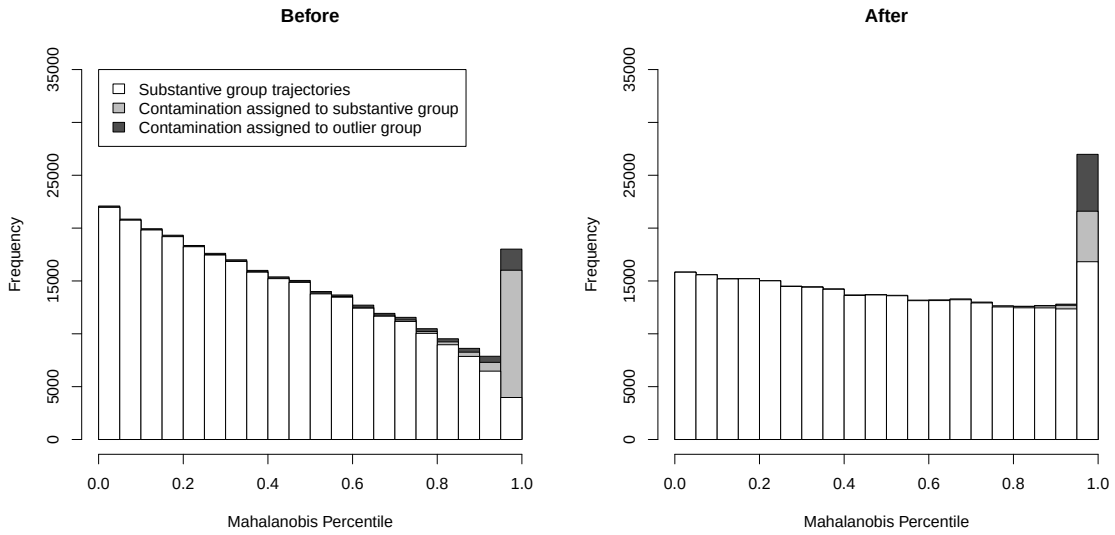


Figure 4.3.10: Mahalanobis distance percentiles of trajectories from their assigned group mean, for model 1 with separation 1 and slope 0, before and after outlier removal.

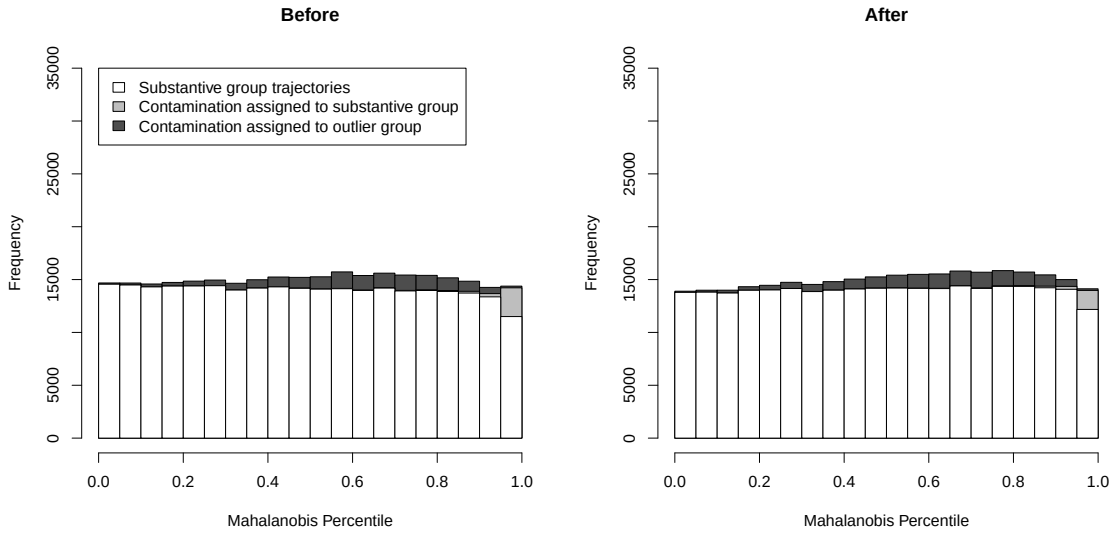


Figure 4.3.11: Mahalanobis distance percentiles of trajectories from their assigned group mean, for model 2 with separation 3 and slope 0.75, before and after outlier removal.

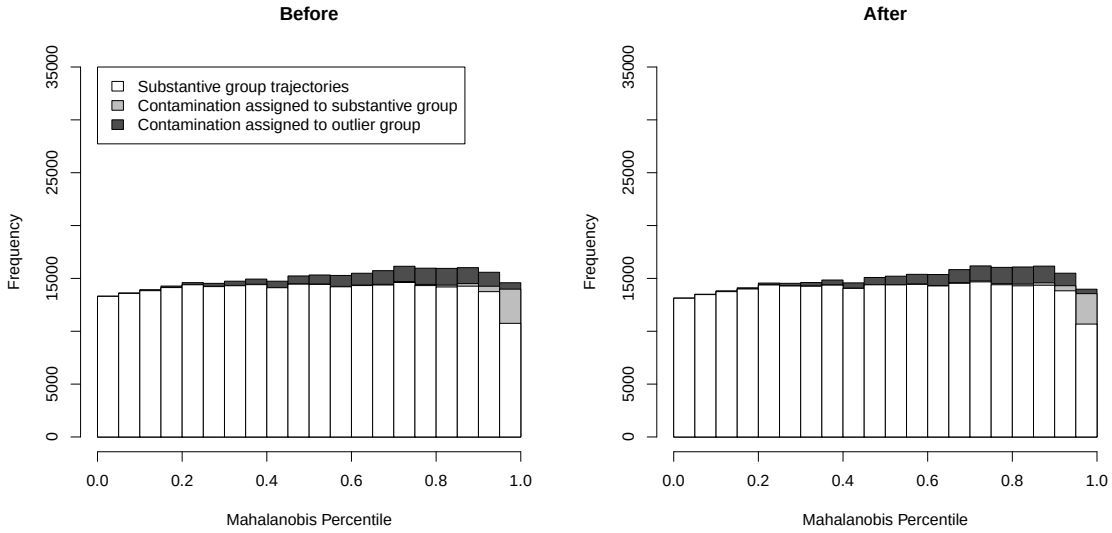


Figure 4.3.12: Mahalanobis distance percentiles of trajectories from their assigned group mean, for model 2 with separation 1 and slope 0, before and after outlier removal.

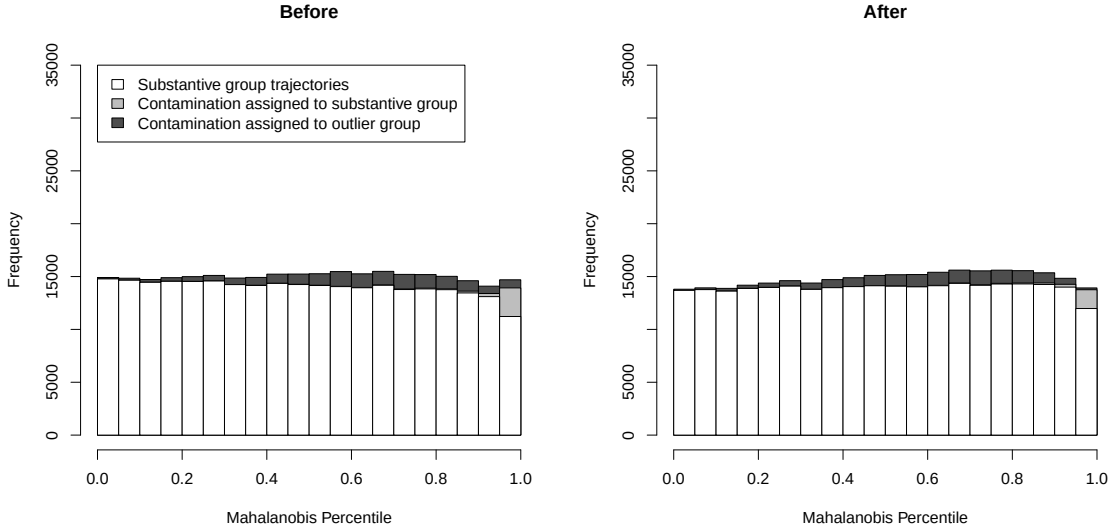


Figure 4.3.13: Mahalanobis distance percentiles of trajectories from their assigned group mean, for model 3 with separation 3 and slope 0.75, before and after outlier removal.

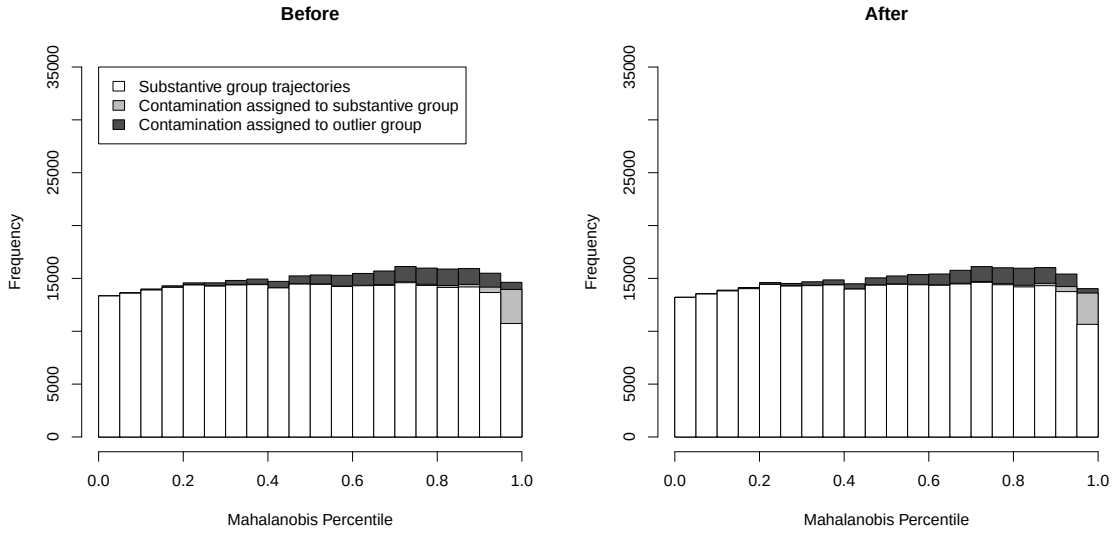


Figure 4.3.14: Mahalanobis distance percentiles of trajectories from their assigned group mean, for model 3 with separation 1 and slope 0, before and after outlier removal.

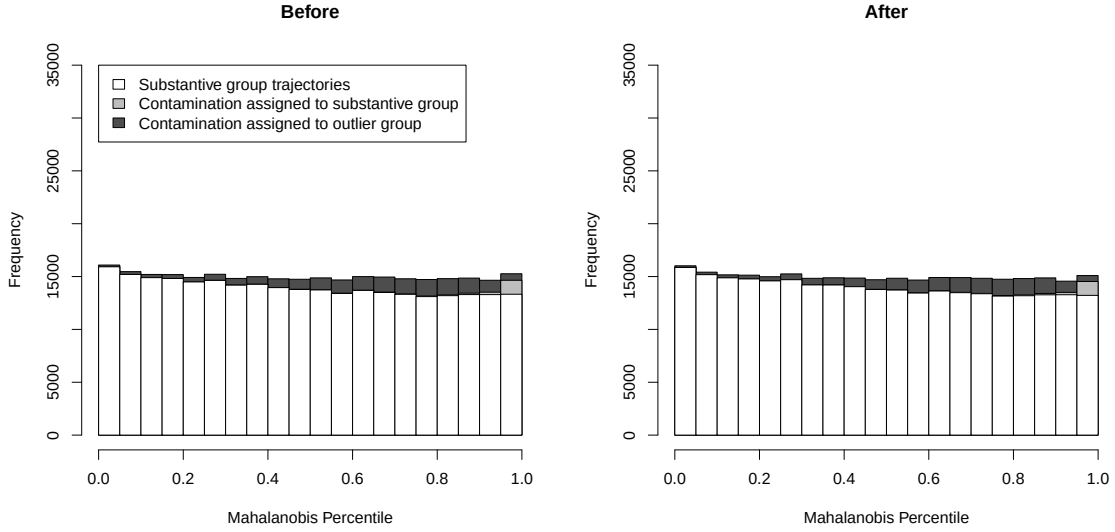


Figure 4.3.15: Mahalanobis distance percentiles of trajectories from their assigned group mean, for model 4 with separation 3 and slope 0.75, before and after outlier removal.

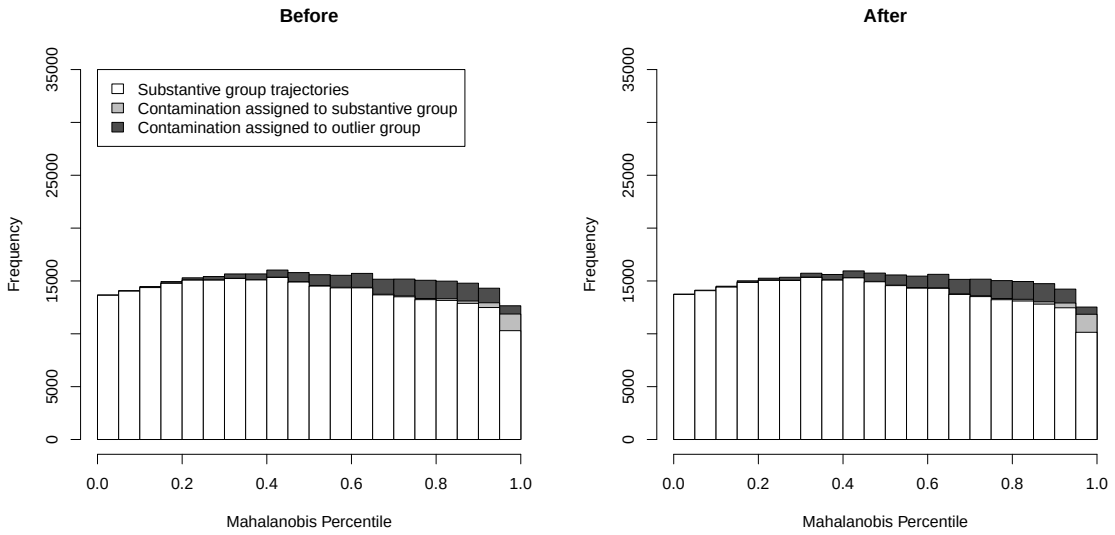


Figure 4.3.16: Mahalanobis distance percentiles of trajectories from their assigned group mean, for model 4 with separation 1 and slope 0, before and after outlier removal.

Results for outlier removal from crossed trajectories simulations

Table 4.3.9 shows the average number of trajectories identified by the outlier identification algorithm for the crossed trajectories simulations, according to whether they were substantive or contamination trajectories. The pattern of the number of trajectories identified was similar to that seen for the divergent trajectories simulations. For models 1 and 3 the number of trajectories identified increased as the slope increased, with model 1 having the greater number of trajectories identified. Models 2 and 4 had fewer trajectories identified, with the number of contamination trajectories identified highest in both cases when the slope was 0.5.

Table 4.3.10 shows the CPV for the various scenarios of the crossed trajectories simulations. As was the case for the divergent trajectories simulations, for models

Table 4.3.9: Average number of trajectories identified by the algorithm in crossed trajectories simulations, according to whether they were substantive or contamination trajectories.

	Slope							
	Substantive				Contamination			
Model	0.25	0.5	0.75	1	0.25	0.5	0.75	1
1	0.01	0.14	1.04	1.68	7.13	7.58	12.7	13.7
2	0.00	0.02	0.04	0.06	2.72	3.62	2.92	2.44
3	0.01	0.02	0.66	0.99	2.95	3.81	8.91	8.62
4	0.00	0.00	0.02	0.02	0.71	1.26	0.61	0.53

1 and 3, the CPV decreased as the slope increased. However, unlike the divergent trajectories simulations, the CPV was generally very high for all models and the lowest value was 0.891.

Table 4.3.10: Contamination predictive value for various scenarios in the crossed trajectories simulations.

	Slope			
Model	0.25	0.5	0.75	1
1	0.999	0.982	0.925	0.891
2	0.999	0.994	0.986	0.975
3	0.997	0.996	0.931	0.897
4	0.994	0.998	0.973	0.967

Table 4.3.11 shows the average number of groups estimated before and after outlier removal for the crossed trajectories simulations. For model 1, the number of groups decreased after outlier removal, similar to the divergent trajectories simulations. When the slope was 0.75 or 1, the average number of groups estimated after outlier removal was close to 2, the true number of substantive groups. Thus the additional groups created for the outliers were no longer identified after outlier

removal. For model 3, the number of trajectories also decreased for the scenarios with high slope, however the average number of groups was still at least 2.3 after outlier removal. As previously for models 2 and 4, the average number of groups did not change greatly. However, the reduction in the number of groups was greatest for a slope of 0.5, where the number of trajectories identified was highest.

Table 4.3.11: Average number of groups estimated for various scenarios before and after outlier removal in crossed trajectories simulations.

	Slope							
	Before Outlier Removal				After Outlier Removal			
Model	0.25	0.5	0.75	1	0.25	0.5	0.75	1
1	1.68	1.81	3.40	3.82	1.32	1.28	2.05	2.06
2	1.81	1.67	2.50	2.90	1.71	1.59	2.52	2.85
3	1.85	1.77	3.18	3.48	1.68	1.59	2.30	2.40
4	2.77	2.88	3.80	3.88	2.66	2.65	3.66	3.73

Table 4.3.12 shows the average relative CCR before and after outlier removal for the crossed trajectories scenarios. The relative CCR increased or remained the same for slope values other than 0.5. There were slight increases in the relative CCR for the scenarios where two groups were always identified without contamination in the previous section. When the slope was 0.5, the relative CCR decreased by 0.062 for model 1 after outlier removal. This is surprising as the group means are further apart than when the slope is 0.25, yet the performance is worse both before and after outlier removal for a slope of 0.5. In this scenario the relative CCR after outlier removal of 0.567 was also lower than with no contamination, where it was 0.604 in Section 4.2.4.

Table 4.3.12: Average relative CCR for various scenarios before and after outlier removal in crossed trajectories simulations.

	Slope							
	Before Outlier Removal				After Outlier Removal			
Model	0.25	0.5	0.75	1	0.25	0.5	0.75	1
1	0.666	0.629	0.976	0.976	0.673	0.567	0.988	0.993
2	0.670	0.557	0.889	0.987	0.670	0.556	0.895	0.987
3	0.670	0.571	0.976	0.983	0.671	0.564	0.984	0.991
4	0.536	0.696	0.826	0.821	0.548	0.690	0.844	0.846

Table 4.3.13 shows the number of contamination trajectories assigned to a substantive group for various scenarios before and after outlier removal in the crossed trajectories simulations. As for the divergent trajectories simulations the number of contamination trajectories assigned to a substantive group before outlier removal was highest for model 1, however in this case the highest value was for a slope of 0.5 rather than for lowest value of slope. As was the case for the divergent trajectories simulations, the number of contamination trajectories assigned to a substantive group reduced considerably for model 1 after outlier removal. For models 2 and 3, smaller reductions were again observed after outlier removal. Similarly for model 4, the number of contamination trajectories assigned to a substantive group did not change greatly after outlier removal. The number of contamination trajectories less than the 99th percentile of Mahalanobis distance from the closest substantive group mean were 3.70, 3.47, 3.36 and 2.99 for the increasing slope values, respectively.

Figures 4.3.17 to 4.3.24 show the Mahalanobis distance percentiles of all trajectories from their assigned group mean in the crossed trajectories simulations, before

Table 4.3.13: Number of contamination trajectories assigned to a substantive group for various scenarios before and after outlier removal in crossed trajectories simulations.

	Slope							
	Before Outlier Removal				After Outlier Removal			
Model	0.25	0.5	0.75	1	0.25	0.5	0.75	1
1	17.2	17.6	15.1	13.8	8.25	9.02	7.14	6.10
2	7.27	10.0	10.1	8.21	6.23	7.58	6.60	5.08
3	7.32	9.79	12.3	10.4	6.36	7.72	6.90	5.87
4	2.40	4.30	2.97	2.96	2.72	4.72	3.09	3.10

and after outlier removal, for models 1 to 4, and with either slope 1 or slope 0.25. The plots for the remaining values of the slope are shown in Appendix A. The plots with slope 1 are for the scenarios when the substantive group means are furthest apart, whereas those for slope 0.25 are when the substantive group means are closest together. The scenarios with other values of slope appear somewhere between these two extremes.

Comparison with Figures 4.3.9 to 4.3.16 shows that the distribution of Mahalanobis distance percentiles before and after outlier removal in the crossed trajectories simulations were very similar to the divergent trajectories simulations.

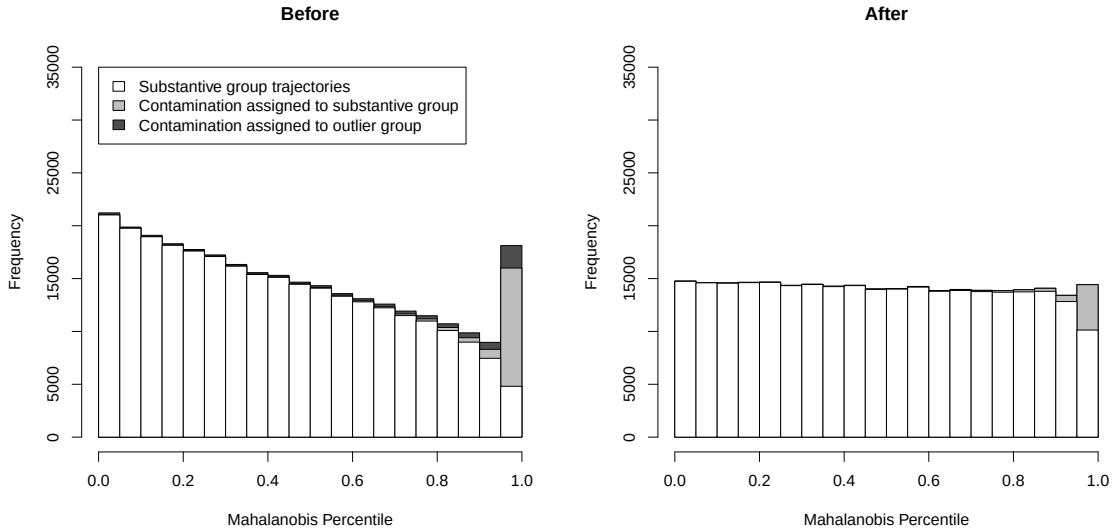


Figure 4.3.17: Mahalanobis distance percentiles of trajectories from their assigned group mean, for model 1 with slope 1, before and after outlier removal.

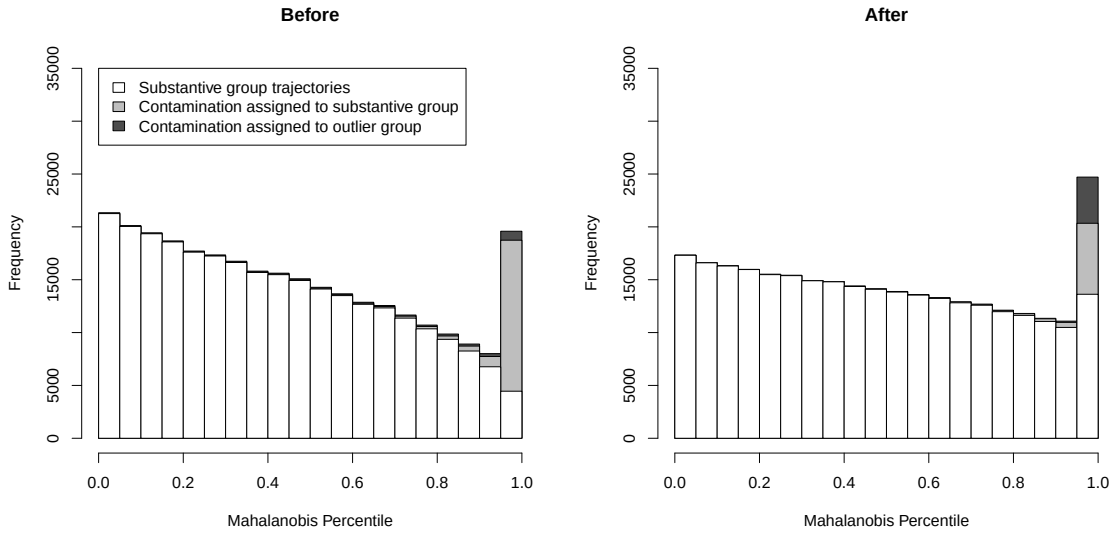


Figure 4.3.18: Mahalanobis distance percentiles of trajectories from their assigned group mean, for model 1 with slope 0.25, before and after outlier removal.

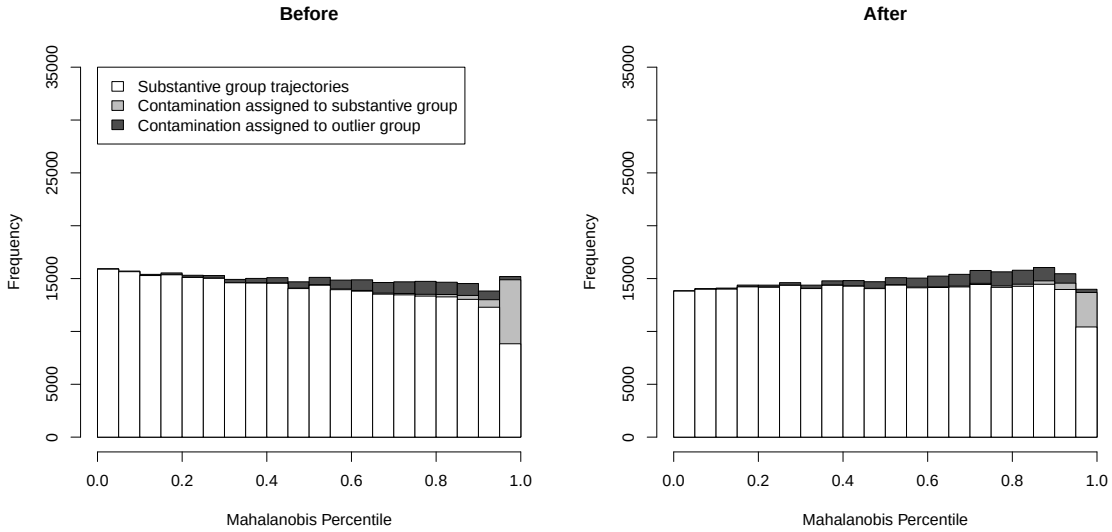


Figure 4.3.19: Mahalanobis distance percentiles of trajectories from their assigned group mean, for model 2 with slope 1, before and after outlier removal.

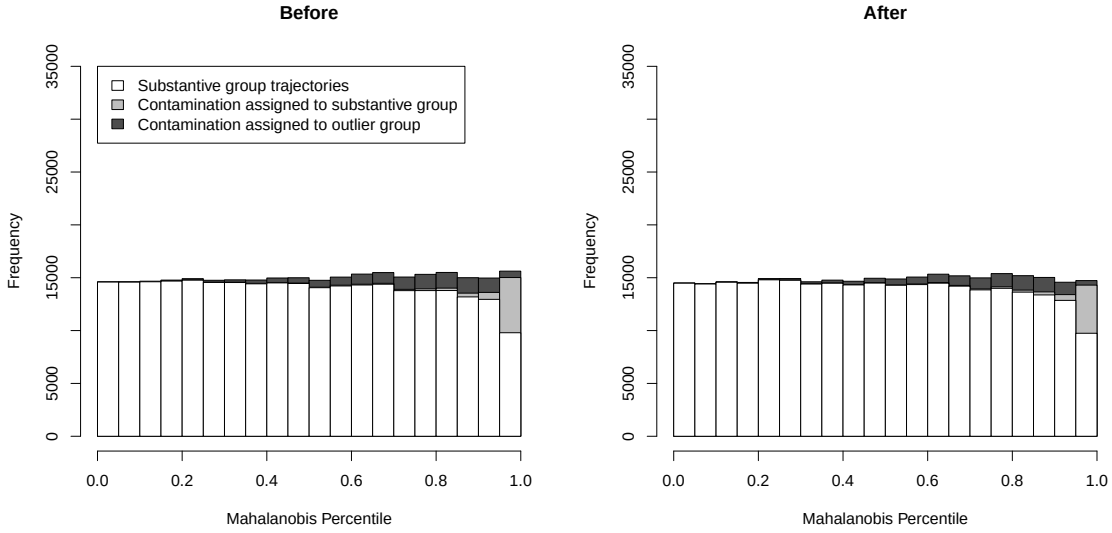


Figure 4.3.20: Mahalanobis distance percentiles of trajectories from their assigned group mean, for model 2 with slope 0.25, before and after outlier removal.

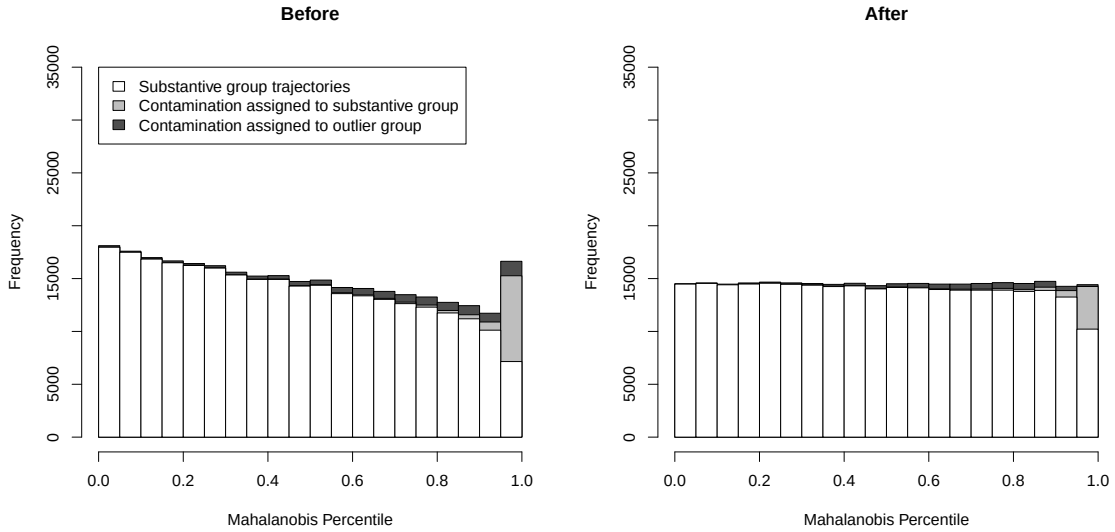


Figure 4.3.21: Mahalanobis distance percentiles of trajectories from their assigned group mean, for model 3 with slope 1, before and after outlier removal.

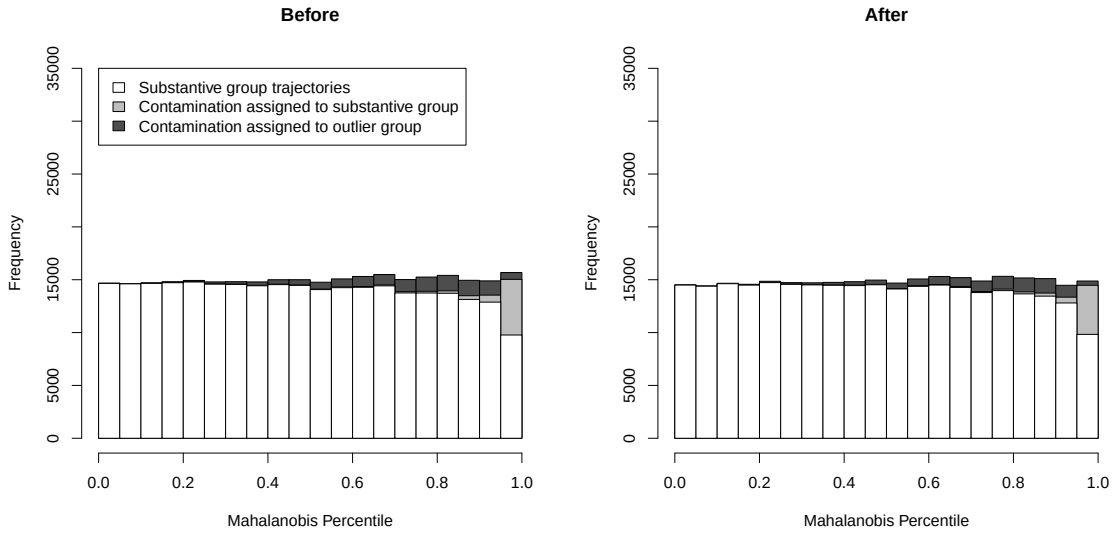


Figure 4.3.22: Mahalanobis distance percentiles of trajectories from their assigned group mean, for model 3 with slope 0.25, before and after outlier removal.

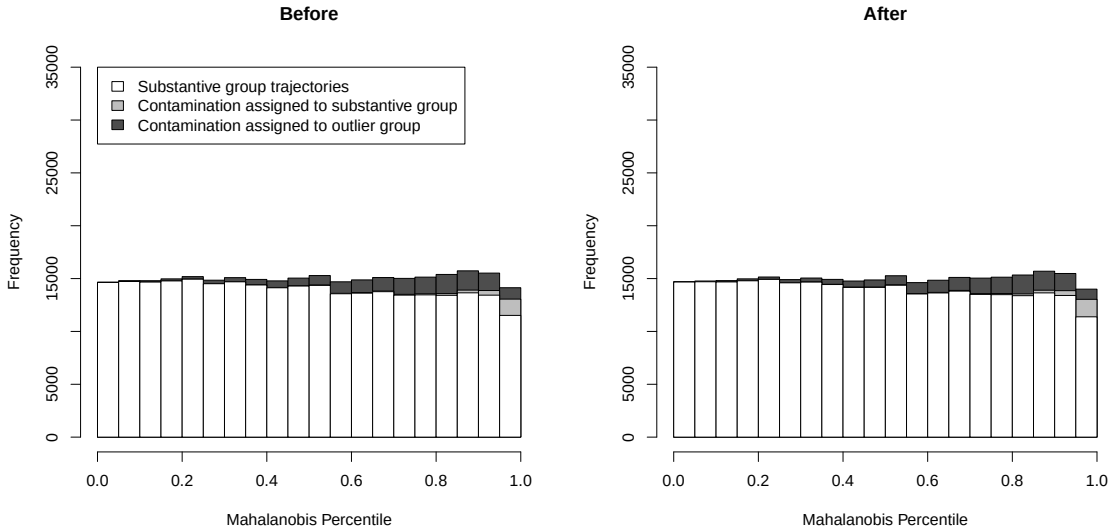


Figure 4.3.23: Mahalanobis distance percentiles of trajectories from their assigned group mean, for model 4 with slope 1, before and after outlier removal.

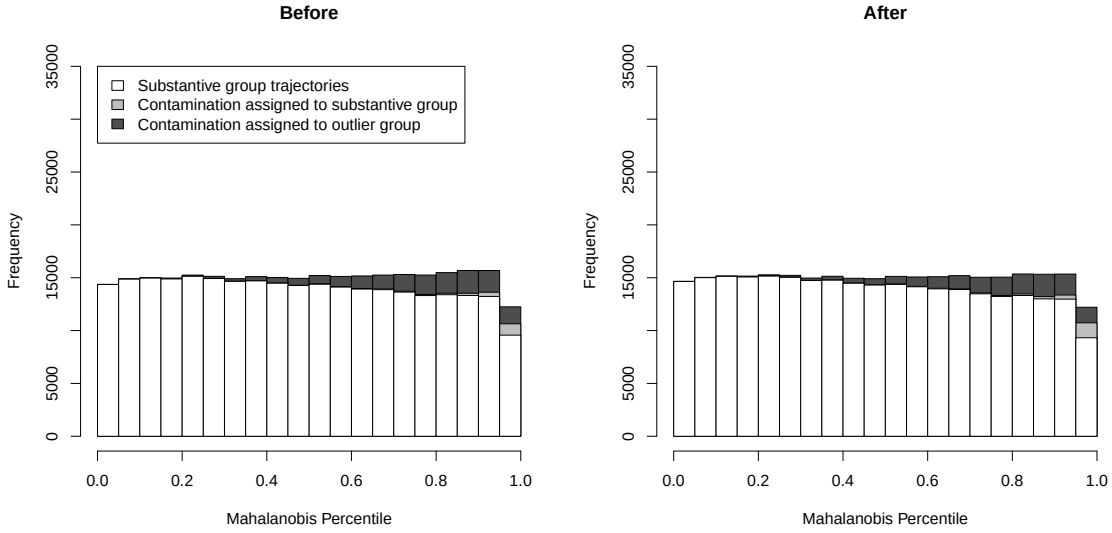


Figure 4.3.24: Mahalanobis distance percentiles of trajectories from their assigned group mean, for model 4 with slope 0.25, before and after outlier removal.

4.3.5 Discussion of outlier identification

Our study demonstrates the difficulty of correctly identifying outlier trajectories in the group-based trajectory modelling setting, as changes in the number of groups and model structure may arise from the inclusion of even a small percentage of contaminating trajectories.

In these simulation studies it was found that the presence of outliers can have surprising results. In particular, outliers can improve the classification of substantive trajectories in certain circumstances. This phenomenon persisted when investigating the effects of removing the identified outlying trajectories in these situations, as outlier removal resulted in the relative CCR decreasing. This emphasises that the use of a performance measure based on correct classification of the two simulated groups will provide limited information if the groups cannot be identified as distinct before the introduction of outliers.

From my simulation studies, I found a significant improvement in the CCR through removing outliers identified by my algorithm, in scenarios mostly involving model 1. This model has a covariance structure constrained to be equal between groups, and matches the covariance structure of the two simulated groups. This covariance structure also made it less likely that a single group with only outliers would be estimated. As a result, the algorithm was able to identify the outliers when the groups were well separated. This occurred for scenarios in which the correct number of groups could be identified before outliers were introduced. Where the group

means were not well separated, removal of outliers identified by the algorithm did not improve the relative CCR. Although the CCR was observed to decrease in some situations following outlier removal due to the substantive groups being combined together, this highlights the fragility of the original group structure proposed.

Removal of outliers identified by my algorithm was also observed to reduce the number of contamination trajectories assigned to a substantive group for models 1, 2 and 3, particularly for model 1 when it was high before outlier removal. This can be expected to reduce the impact of these contamination trajectories on the estimated parameters for the substantive groups. Future research could investigate how the sensitivity of my algorithm changes under different significance levels.

The average number of contamination trajectories less than the 99th percentile of Mahalanobis distance from the closest substantive group mean was less than 4 in all scenarios, so would have only contributed minimally to the number of contamination trajectories assigned to a substantive group. These trajectories do not appear as outliers, so cannot be expected to be identified as such.

Plots of the Mahalanobis distance percentiles of trajectories from their assigned group means illustrated the ability of the algorithm to identify contamination trajectories for model 1, particularly when the substantive groups were far apart. In contrast, these plots showed little changes for models 2, 3 and 4.

As the number of outlying trajectories identified increased, the CPV was observed to decrease. Therefore, in the circumstances where the algorithm identifies at least three outlying trajectories, it is also more likely to identify substantive group

trajectories as outliers with a non-negligible frequency.

Under model 2 or 4, the algorithm identified very few outliers as additional groups could be estimated containing only outlier trajectories. This resulted in the contamination points not appearing as outliers relative to the groups estimated for them, similar to the phenomenon of 'masking' in regression outlier identification.[72] Model 3 had results that were between those of models 1 and 2, as this model is defined as the choice between those two models according to the BIC. When model 1 was more appropriate initially, this resulted in better performance, as opposed to when model 2 was chosen initially. These results would likely have been different if the two simulated groups had different covariance structures, as model 1 would not have fit well in this case.

A potential consideration in using the outlier removal method is the computational time required, as the group-based model needs to be run multiple times for each step of removing a trajectory. However, for small to moderate sized datasets this is unlikely to be a limiting factor in the analysis.

4.4 Conclusion

In this chapter I assessed the impact of outliers on divergent and crossed group-based trajectory models. The presence of outliers tended to lead to an increased number of groups in all situations, and a reduction in the CCR when the group means were well separated. I developed and described an algorithm for identifying these outlying

trajectories, and evaluated its performance on the simulated divergent and crossed trajectories datasets.

Assessing the effectiveness of outlier removal in group-based trajectory modelling is complicated by the flexibility of these models, in terms of the covariance structures available and the potential for new groups to be created. The outlier identification algorithm performed well under certain model assumptions and where the simulated groups were well separated, but was less effective when the model was more flexibly defined or the group means were close together.

Based on these results, I recommend the application of my outlier identification algorithm in similar settings and a comparison of results with and without outliers removed as part of sensitivity analyses to determine the effect that the detected outliers may have. After potential outliers have been identified, any other information on these observations can also be inspected for further indication of a systematic error.

Chapter 5

Performance of methods for estimating the effect of covariates on group membership probabilities in group-based trajectory models

5.1 Preface

This chapter contains the second article contributing to this thesis which has been published in a peer reviewed journal:

- Davies CE, Giles LC and Glonek GFV. Performance of methods for estimating the effect of covariates on group membership probabilities in group-based

trajectory models. *Statistical Methods in Medical Research.* Prepublished January 18, 2017.

It addresses the fourth aim of this thesis by comparing the performance of methods for estimating the effect of covariates on group membership probabilities in group-based trajectory models.

The various methods available for estimating the effect of covariates on group membership probabilities in group-based trajectory models were described in Section 2.6.2. No study has considered all of these of methods, nor compared how the inclusion of additional covariates affects resulting bias. This article addresses this important gap in the literature and recommends which of the methods should be used.

Supplementary tables 1, 2 and 3 referred to in the article have been reproduced in Appendix B, or are available from:
<http://journals.sagepub.com/doi/suppl/10.1177/0962280216689580>.

5.2 Statement of authorship

Statement of Authorship

Title of Paper	Performance of methods for estimating the effect of covariates on group membership probabilities in group-based trajectory models.
Publication Status	☑ Published ☐ Accepted for Publication ☐ Submitted for Publication ☐ Unpublished and Unsubmitted w ork w ritten in manuscript style
Publication Details	Davies CE, Giles LC and Glonek GFV. Performance of methods for estimating the effect of covariates on group membership probabilities in group-based trajectory models. Statistical Methods in Medical Research. Prepublished January 18, 2017; DOI: 10.1177/0962280216689580. <http://journals.sagepub.com/doi/10.1177/0962280216689580> The final, definitive version of this paper has been published in Statistical Methods in Medical Research, 2017 by SAGE Publications Ltd, All rights reserved. Copyright © 2017 Davies CE, Giles LC and Glonek GFV. Reprinted by permission of SAGE Publications.

Principal Author

Name of Principal Author (Candidate)	Christopher E Davies		
Contribution to the Paper	Designed the study, simulated the data, performed all analyses, interpreted the results, drafted the manuscript and acted as corresponding author.		
Overall percentage (%)	80		
Certification:	This paper reports on original research I conducted during the period of my Higher Degree by Research candidature and is not subject to any obligations or contractual agreements with a third party that would constrain its inclusion in this thesis. I am the primary author of this paper.		
Signature		Date	06/12/17

Co-Author Contributions

By signing the Statement of Authorship, each author certifies that:

i. the candidate's stated contribution to the publication is accurate (as detailed above);
ii. permission is granted for the candidate in include the publication in the thesis; and
iii. the sum of all co-author contributions is equal to 100% less the candidate's stated contribution.

Name of Co-Author	Lynne C Giles		
Contribution to the Paper	Contributed to the design of the study and interpretation of the results, and reviewed the manuscript.		
Signature		Date	06/12/17

Name of Co-Author	Gary FV Glonek		
Contribution to the Paper	Contributed to the design of the study and interpretation of the results, and reviewed the manuscript.		
Signature		Date	06/12/17

5.3 Article

Article

SMMR
STATISTICAL METHODS IN MEDICAL RESEARCH

Statistical Methods in Medical Research
0(0) 1–15

DOI: 10.1177/0962280216689580
journals.sagepub.com/home/smm
SAGE

Performance of methods for estimating the effect of covariates on group membership probabilities in group-based trajectory models

Christopher E Davies,[1,2,3] Lynne C Giles[2] and Gary FV Glonek[1]

Abstract

One purpose of a longitudinal study is to gain insight of how characteristics at earlier points in time can impact on subsequent outcomes. Typically, the outcome variable varies over time and the data for each individual can be used to form a discrete path of measurements, that is a trajectory. Group-based trajectory modelling methods seek to identify subgroups of individuals within a population with trajectories that are more similar to each other than to trajectories in distinct groups. An approach to modelling the influence of covariates measured at earlier time points in the group-based setting is to consider models wherein these covariates affect the group membership probabilities. Models in which prior covariates impact the trajectories directly are also possible but are not considered here. In the present study, we compared six different methods for estimating the effect of covariates on the group membership probabilities, which have different approaches to account for the uncertainty in the group membership assignment. We found that when investigating the effect of one or several covariates on a group-based trajectory model, the full likelihood approach minimized the bias in the estimate of the covariate effect. In this '1-step' approach, the estimation of the effect of covariates and the trajectory model are carried out simultaneously. Of the '3-step' approaches, where the effect of the covariates is assessed subsequent to the estimation of the group-based trajectory model, only Vermunt's improved 3 step resulted in bias estimates similar in size to the full likelihood approach. The remaining methods considered resulted in considerably higher bias in the covariate effect estimates and should not be used. In addition to the bias empirically demonstrated for the probability regression approach, we have shown analytically that it is biased in general.

Keywords

covariates, group-based trajectory modelling, mixture models, longitudinal data, simulation

1 Introduction

Longitudinal studies can provide an understanding of how characteristics at earlier points in time can impact on subsequent outcomes. Typically, the outcome variable varies over time and the data for each individual can be used to form a discrete path of measurements, that is a trajectory. Group-based trajectory modelling methods seek to identify subgroups of individuals within a population with trajectories that are more similar to each other than to trajectories in distinct groups.

One approach to modelling the influence of covariates measured at earlier time points in the group-based setting is to consider models wherein these covariates affect the group membership probabilities. Models in which prior covariates impact the trajectories directly are also possible but are not considered here. The effect of the covariates on the group membership probabilities is modelled using multinomial logistic regression.

[1]School of Mathematical Sciences, The University of Adelaide, Adelaide, Australia
[2]School of Public Health, The University of Adelaide, Adelaide, Australia
[3]Australia and New Zealand Dialysis and Transplant Registry, South Australian Health and Medical Research Institute, Adelaide, Australia

Corresponding author:
Christopher E Davies, School of Mathematical Sciences, The University of Adelaide, Adelaide, SA 5005, Australia.
Email: chris.davies@adelaide.edu.au

A number of ways to estimate the multinomial logistic regression on the covariates have been proposed, with different approaches to account for the uncertainty in the group membership assignment. Here we consider six different methods and compare their bias in estimating the effects of covariates in a simulation study. The six different methods are 1-step (1S),[1–3] 3-step (3S),[1,3] pseudo class 3-step (PC3S),[4,5] improved 3-step (I3S),[1,3] probability regression (PR)[4] and probability weighted 3-step (PW3S).[4]

Bolck et al.[3] compared the performance of the 1S, 3S and the I3S methods in their study introducing the I3S method and demonstrated both analytically and through simulation that the classification error in the second step of the 3S method leads to attenuation of parameter estimates. This was confirmed by Vermunt[1] in a study extending the I3S method, in which the extension was also compared with the other methods presented in Bolck et al. In a simulation study, Clark and Muthén[4] compared all methods considered here except for I3S and found that the 1S method performed best. The studies by Bolck et al.[3] and Clark and Muthén[4] considered only a single continuous covariate, while Vermunt[1] used three discrete numeric covariates but did not compare with results for a single covariate. However, no study has considered all of these methods, nor compared how the inclusion of additional covariates affects resulting bias.

In the present study, the six estimation methods were compared in scenarios with varying odds ratio of the covariate relationship, difficulty of the classification problem, and with one or two covariates that were either continuous or binary. We also demonstrate analytically the bias of the PR method.

The group-based trajectory model specifications and the methods for estimating the effect of covariates are elaborated on in Section 2. The simulations are described in Section 3, while Section 4 presents the results of the simulations and the analytical treatment of the PR method. The implications of our findings are discussed in Section 5.

2 Methods

2.1 Group-based trajectory models with covariates

Latent class growth analysis (LCGA) and growth mixture modelling (GMM) are the most frequently used methods for modelling group-based trajectories.[6] Finite mixture modelling is the basis for both methods, which means that for K groups the marginal probability distribution of a randomly chosen trajectory is modelled by

$$P(\boldsymbol{y}) = \sum_{k=1}^{K} \pi_k P^k(\boldsymbol{y})$$

where $P^k(\boldsymbol{y})$ is the conditional distribution of the trajectory, $\boldsymbol{y}$, given the individual is in group k (which we denote as $G=k$), and π_k is the group membership probability ($\pi_k = P(G=k)$) such that $\pi_k > 0$ for $k = 1, \ldots, K$ and $\sum_{k=1}^{K} \pi_k = 1$.

Finite mixture modelling requires the specification of the number of groups, K. In practice, K is rarely clear from the data. There has been considerable discussion in the literature around methods to estimate the appropriate number of groups, with the results of studies comparing methods inconsistent.[7] For simplicity, we assume that the correct number of groups is known, as has been done in similar studies.[1,4,5]

In a group-based model without covariates, π_k is modelled as

$$P(G=k) = \pi_k = \frac{e^{\theta_k}}{\sum_{k=1}^{K} e^{\theta_k}}$$

This ensures $0 < \pi_k < 1$ and $\sum_{k=1}^{K} \pi_k = 1$, while the θ_k are free to vary without restriction. As a result of the summation constraint, only $K-1$ estimates of θ_k are required, and by convention $\theta_1 = 0$.

The effect of a covariate vector $\boldsymbol{x} = (x_1, \ldots, x_p)$ can be incorporated through the multinomial logistic regression model

$$P(G=k|\boldsymbol{x}) = \frac{e^{\theta_{0k} + \boldsymbol{x}^T \boldsymbol{\theta}_k}}{\sum_{l=1}^{K} e^{\theta_{0l} + \boldsymbol{x}^T \boldsymbol{\theta}_l}}$$

where $\boldsymbol{\theta}_k = (\theta_{1k}, \ldots, \theta_{pk})$. The effect of x_1 in group k relative to group 1 is estimated by θ_{1k}, with $e^{\theta_{1k}}$ providing an estimate of the odds ratio.

2.2 Group-based trajectory model specification

A multivariate Gaussian distribution is assumed for each of the $P^k(\boldsymbol{y})$, as is commonplace for group-based models of continuous outcomes. In what follows, the set of time points of observation is taken to be the same for all subjects, so that for subject i in group k

$$\boldsymbol{y_i} \sim MVN(\boldsymbol{\mu}^k, \Sigma^k) \tag{1}$$

where $\boldsymbol{y_i}$ are the responses for subject i, $i = 1, \ldots, n$, and $\boldsymbol{\mu}^k$ and Σ^k are the mean vector and covariance matrix for group k, respectively. Constraints placed on $\boldsymbol{\mu}^k$ and Σ^k determine whether the method is LCGA or GMM.

For simplicity, we describe linear LCGA and GMM models here; however, extension to higher order polynomials is straightforward. A more detailed description of the different types of these models is provided by Davies et al.[8]

2.2.1 LCGA

The LCGA model[9] can be defined by taking $\boldsymbol{\mu}^k = \alpha^k + \beta^k \boldsymbol{t}$ and Σ^k as diagonal, where $\boldsymbol{t}$ is the vector of time points of observation. Conditional independence is implied for all LCGA models with the assumption that Σ^k is diagonal.

2.2.2 GMM

The mean trajectories of LCGA are extended in GMM by the inclusion of random effects to allow for the variation of trajectories of individuals.[10,11] A GMM model can be specified by taking

$$\boldsymbol{y}_i^k = \alpha^k + a_i^k + \beta^k \boldsymbol{t}_i + b_i^k \boldsymbol{t}_i + \epsilon_i^k$$

where $\epsilon_i^k \sim MVN(0, \boldsymbol{R}^k)$ with $\boldsymbol{R}^k$ diagonal, and $(a_i^k, b_i^k)' \sim MVN(0, \boldsymbol{D}^k)$. In our notation, this corresponds to expressing the GMM as

$$\boldsymbol{\mu}^k = \alpha^k + \beta^k \boldsymbol{t} \quad \text{and}$$
$$\Sigma^k = \boldsymbol{R}^k + \boldsymbol{Z}\boldsymbol{D}^k\boldsymbol{Z}'$$

where $\boldsymbol{Z} = [\boldsymbol{1}\ \boldsymbol{x}]$.

Non-zero covariances between measurements in the same group are implied by the GMM model and therefore, unlike LCGA models, conditional independence does not follow. While $\boldsymbol{R}^k$ is assumed to be diagonal, the random effects allow for dependence over time for individuals within the same group.

2.3 Methods for estimating the effect of covariates on group membership probabilities

2.3.1 1S

In the 1S method,[1–3] estimation of the effect of covariates and the trajectory model are carried out simultaneously. That is to say, full maximum likelihood is used for all parameters of the group-based model ($\boldsymbol{\mu}^k$, Σ^k) and the θs of the multinomial logistic regression. While this is a reasonable approach from an estimation perspective, there are various reasons why researchers have preferred other methods in which the group-based model is estimated first and the effect of the covariate is assessed subsequently. As argued by Vermunt,[1] this method may be impractical if there is a large number of covariates to consider, in combination with a complex group-based trajectory model that would need to be re-estimated for each covariate. The 1S method also complicates model building decisions, may not fit with the logic of researchers wishing to decide on groups first and assumes the group-based model has not yet been constructed. According to Vermunt 'in many applications, it is more natural to use a stepwise approach and, moreover, ... sometimes it is the only reasonable way to proceed'. The remaining methods considered in the present study are such stepwise approaches, where the effect of the covariate is estimated subsequent to the estimation of the group-based model.

2.3.2 3S

In the 3S method,[1,3] the LCGA or GMM trajectory model is first estimated without inclusion of covariates. In the second step, each individual is allocated to groups according to the maximum estimated posterior probability

$$m_i = \arg\max_k \hat{P}(G_i = k | \boldsymbol{y}_i)$$

In the third step, a multinomial logistic regression model, as in Section 2.1, is estimated using maximum likelihood, with m_i as the response instead of G_i.

2.3.3 PC3S

The PC3S method[4,5] is similar to the 3S method; however at the second step, individuals are allocated randomly to groups J times according to the posterior probabilities

$$P(m_{ij} = k) = \hat{P}(G_i = k|\boldsymbol{y}_i)$$

for $j = 1, \ldots, J$. As in the 3S method, multinomial logistic regressions with m_{ij} as the response are estimated with maximum likelihood to obtain J estimates of the covariate effect, with their average providing the PC3S estimate.

2.3.4 I3S

The first two steps of the I3S method[1,3] are the same as for the 3S method. The third step differs as it takes into account the misclassification error in the second step when individuals were allocated to m_i according to the maximum estimated posterior probability. This is achieved through a latent class model where the estimated classification errors are treated as known errors of classification. To estimate $\hat{P}(M = k|G = l)$, the classification error in M, we first calculate the classification uncertainty rate for M

$$\hat{P}(G = l\,|M = k) = \frac{1}{n_k} \sum_{m_i = k} \hat{P}(G_i = l\,|\boldsymbol{y}_i)$$

where n_k is the number of observations classified in group k by the most likely class variable M. The classification uncertainty rate is the average of the posterior probabilities for each of the classes among each of the allocated classes. Bayes Theorem is then used to calculate the classification error in M

$$\hat{P}(M = k|G = l) = \frac{\hat{P}(G = l\,|M = k)n_k}{\sum_{s=1}^{K} \hat{P}(G = l\,|M = s)n_s}$$

These classification errors allow M to be treated as an imperfect measurement of G in a latent class model, with a regression of G on x, and in this way the measurement error in M is taken into account. This involves maximizing the log likelihood function

$$\sum_{i=1}^{n} \sum_{k=1}^{K} \log \sum_{l=1}^{K} P(G_i = l\,|\boldsymbol{x}_i)P(M = k|G = l)$$

This approach was first described by Vermunt[1] and expanded on improvements made to the 3S method by Bolck et al.[3] More details are provided in Vermunt's paper.[1]

2.3.5 PR

In the PR method,[4] the posterior probabilities from step one of the 3S method are transformed using the logit function, and the transformed probabilities are subsequently used in a linear regression on the covariate x. That is, the following linear regression is performed

$$\log \frac{\hat{P}(G_i = k|\boldsymbol{y}_i)}{1 - \hat{P}(G_i = k|\boldsymbol{y}_i)} = \gamma_{0k} + \gamma_k x_i$$

For two groups, the effect of x in group k relative to group 1 is estimated by γ_k, with e^{γ_k} providing an estimate of the odds ratio.

2.3.6 PW3S

The PW3S method[4] is based on the 3S method; however, the posterior probabilities of membership of the chosen class, $\hat{P}(m_i = k|\boldsymbol{y}_i)$, are used as weights in the multinomial logistic regression of step three, thereby accounting for the differing certainty in the assignments.

3 Simulations

The performance of the six methods for estimating the effect of covariates was assessed on a range of simulated datasets.

3.1 Motivating example

The simulations are motivated by the childhood behaviour data from the Generation 1 Study, a prospective longitudinal cohort study of women and their children living in Adelaide, South Australia. The women were recruited in pregnancy between 1998 and 2000.[12] Data were collected from a total of 536 children at ages 2, 31.5, 5 and 91.5 years; 354 contributed data at all four time points. Childhood behaviour was measured in the Generation 1 Study using the Achenbach Child Behavior Checklist,[13] which records the parent's views on specific behavioural, emotional and social issues for their child. Total and sub-scale scores can be derived as the sum of the ordinal items, with higher scores suggesting worse behaviour.

We focus here on the externalizing sub-scale derived from the subset of questions relating to behaviours such as aggression and bullying. A number of antenatal covariates are of interest including family socio-economic index at birth (a continuous covariate), gender (a common binary covariate) and history of family violence (a rare binary covariate). Figure 1(a) presents the externalizing data from these four time points of the Generation 1 Study, with (b) and (c) showing the groups identified from a two-group GMM model fit to these data. Each of plots (b) and (c) has a randomly selected 25 trajectories joined for clarity of individual trajectories, with all of those selected trajectories shown in (a).

3.2 Specification of simulations

In the present study, datasets of 500 observations were simulated to have properties approximating those for the externalizing behaviour data from the Generation 1 Study. Datasets were simulated using a two-group linear GMM, with 350 in the low group and 150 in the high group, similar to the proportions observed in the Generation 1 Study. The simulation model had a common random effects covariance matrix between groups, but residual variances free to vary between groups and also over time. This data generation model was chosen according to the minimum BIC fit to the Generation 1 data of the two-group linear LCGA and GMM models described in Section 2.2 (model fit shown in Figure 1). For simplicity, only two groups are considered here. The covariance structure, including the residual covariance matrix and the random effects covariance matrix, and the group means were based on the parameter estimates obtained from the fit to the Generation 1 data. In applying the various methods, the group-based trajectory model was estimated using a GMM with the same constraints as for the data generation model. The time points were standardized to improve the convergence of the group-based models, and the same four time points were used for each subject, i.e. for $i = 1, \ldots, n_k$ $\boldsymbol{t}_i = (-1, -0.5, 0.1, 1.3)'$.

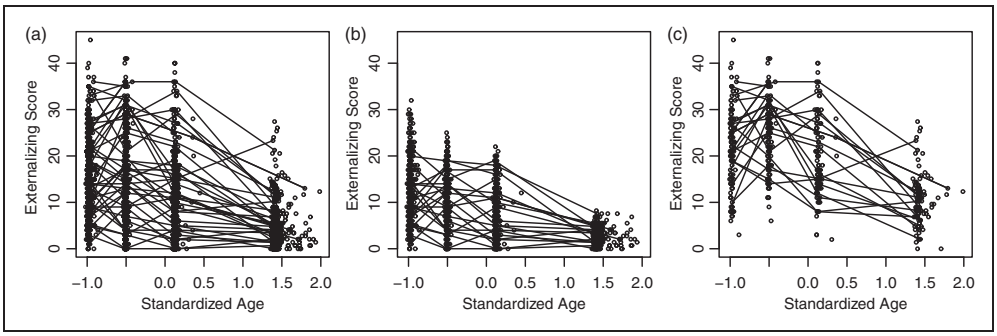


Figure 1. (a) Externalizing scores versus standardized age in the Generation 1 Study. (b) The 'low' group from a GMM model fitted to these data. (c) The 'high' group from a GMM model fitted to these data. Each of plots (b) and (c) has a randomly selected 25 trajectories joined for clarity of individual trajectories, with all of those selected trajectories shown in (a).
GMM: growth mixture modelling.

The simulations were produced across a range of scenarios to provide various dimensions on which the methods could be compared. These dimensions were the separation of the two groups, the number and type of the covariates, and the odds ratio of the covariate relationship(s). The separation between the mean trajectories of the groups was initially varied to provide different levels of classification difficulty. This was achieved by altering the y-intercept of the high group mean trajectory so that the vertical separation between the two groups took values 6, 8, 10, 12 or 14, similar to the separation level in the Generation 1 data of approximately 10. Two hundred datasets were simulated for each separation level.

Following simulation of the datasets with varying separation levels, the simulated covariates were appended to these datasets to maintain consistency of the outcome data between covariate types and odds ratios. As the covariates were simulated separately from the outcome variables, there are no direct effects of the covariates on the latent trajectories and the covariates are conditionally independent of the outcomes, given the group.

The single covariate scenarios had either a continuous covariate, a 'common' binary covariate with a prevalence of 50% in the population, or a 'rare' binary covariate with a prevalence of 10% in the population. The scenarios with two covariates had one continuous covariate, while the other covariate was either rare or common binary.

The odds ratio of each covariate relationship in the single covariate scenarios took values 0.25, 0.5, 1, 2 and 4. All combinations of covariate odds ratios from the single covariate scenarios were considered for the two covariate scenarios. In order to achieve the different odds ratios of the covariate relationships, the covariate values in the two groups were sampled retrospectively from different distributions. Normal distributions with different means but unit standard deviation were used for the continuous covariate, whereas Bernoulli distributions with different probabilities of success were used for the binary variables. For example, to simulate a continuous predictor with an odds ratio of 2, 350 observations were sampled from $N(\log_e 2, 1)$ for the low group and 150 observations were sampled from $N(0, 1)$ for the high group. This approach takes advantage of the relationship between logistic regression and discriminant analysis.[14]

3.3 Assessing estimation performance

For each of the 200 datasets corresponding to a scenario, the effect of the covariate was estimated by each of the methods in Section 2.3. The estimation methods were then compared using the average standardized bias in the parameter estimate for the effect of the covariate on the group membership probability. This bias was standardized according to its standard error so as to enable comparison across scenarios, and the standardization was carried out for each dataset so as to account for the correlation between estimates and their standard errors. The coverage probability of the 95% confidence interval for the covariate estimate was also calculated to allow comparison of the methods.

Hypothesis testing was also considered for each of the single covariate scenarios, with the null hypothesis that the effect of the covariate was zero, and $\alpha = 0.05$ the nominal level of significance. P-values were calculated using t-tests for the PR method and Wald tests for the remaining methods. The power and type I error for each method was estimated to assess the performance of the methods under hypothesis testing.

3.4 Validity of estimates

When using the methods there were some cases of estimation difficulties. To mirror good practice in using the methods, the estimates in these cases were not used. First, when applying the first step of the 3S methods, that is estimation of the group-based trajectory model without covariates, it was found that questionable estimates were produced in 0.4% (4/1000) of the datasets simulated without covariates. The estimates were invalid due to non-positive definiteness of the random effect covariance matrix. These datasets were discarded and re-simulated until the estimates for the group-based trajectory model were found to be valid.

Despite all datasets producing valid estimates for the group-based trajectory model without covariates (including those re-simulated), questionable or invalid estimates were produced for the 1S method in 0.5% (306/65000) of the combinations of datasets with covariates included. Some of these estimates were invalid due to convergence issues that occurred when the rare binary covariate had a large odds ratio or due to model non-identification, but some were also questionable due to non-positive definiteness of the random effect covariance matrix. These simulations were well distributed across the scenarios and we excluded these cases from the estimation of the results.

For the rare binary covariate in scenarios with a large odds ratio, it sometimes occurred that one of the estimated groups contained zero or few observations with the positive response for the covariate. We decided

to exclude simulations where one of the groups had fewer than five positive response observations, as in these cases an analyst should be aware that the estimates were unreliable. As a result of this, a large proportion of the rare binary simulations with an odds ratio of 4 were excluded (in the worst case, 63.5% of the simulations within a particular scenario). We therefore decided to only consider results for rare binary covariates with odds ratio up to 2. Once simulations with invalid or questionable estimates and those with fewer than five positive binary response observations were excluded, at least 83.5% of simulations were used for each of the scenarios. In total 1.3% (762/59000) of simulations were excluded for the 1S method and 0.6% (349/59000) of simulations were excluded for the 3S methods.

3.5 Analysis software

R[15] was used for the data simulations and Mplus[11] was used via R with the MplusAutomation[16] package for estimation of the GMM models. Code used for the simulations showing the methods and population parameters used is available on request. Estimation of mixture models can be sensitive to starting points.[11] Therefore, 200 random starts were used for each model estimation in Mplus, with the maximum likelihood solution being chosen to reduce the chance of finding local optima.

4 Results

4.1 Simulation results

This section presents the results of the simulations for each method, in terms of their average standardized bias, that is the difference between the covariate effect estimate and the true parameter divided by the standard error of the estimate. The average standardized biases for all odds ratios of single continuous and common binary covariates are shown in Tables 1 and 2, respectively. Tables 1 and 2 also contain the average estimates, standard errors and coverage probability of the 95% confidence interval. The results for a single rare binary covariate are shown in Supplementary Table 1. Here we discuss the results in terms of the standardized bias only as the coverage can be observed to be high when the bias is low and vice versa. Under the assumption of normally distributed estimates, a standardized bias of 1.0 corresponds to a coverage probability of 0.83, whereas a standardized bias of 2.0 corresponds to a coverage probability of 0.48.

Figure 2 presents the average standardized bias of the covariate effect estimates for the simulations with a single covariate and an odds ratio of 2. The average standardized biases for an odds ratio of 4 were similar to those for an odds ratio of 2 except of a greater magnitude. For odds ratios of 0.5 and 0.25, the results were symmetrical with those obtained for odds ratios 2 and 4, that is roughly the same bias was observed in the opposite direction. The average standardized biases for an odds ratio of 1 were all very close to zero as all methods were approximately unbiased when there was no true effect of the covariate.

In Figure 2 the same trends were observed for the two types of binary covariates as for the continuous covariate; however, lower levels of bias were observed for the binary covariates than the continuous covariate, and the rare binary covariate was less biased than the common binary covariate. The lower levels of standardized bias for the binary covariates were due to the increased variability of these estimates. The median standard error of the continuous estimates was 0.125, whereas it was 0.223 and 0.394 for common and rare binary estimates, respectively.

The best performing were the 1S and I3S methods, which showed negligible levels of bias for all levels of separation and sizes of odds ratio. For the 3S, PC3S and PW3S methods, the standardized bias decreased as separation increased, and for these methods and the PR method the bias increased with greater odds ratio for the covariate. The use of probability weighting did reduce the bias of PW3S slightly compared to 3S, whereas the PC3S had greater levels of bias than the 3S. The PR approach displayed the worst levels of bias, with bias increasing as the separation of the groups increased. This was surprising as high separation could be expected to reduce uncertainty. Furthermore, the estimate of the covariate was biased positively, so a larger estimate was obtained than the true value. The reason for the different behaviour of the PR method is explained in Section 4.2. As the 1S and I3S methods performed the best, code to conduct these methods in Mplus has been provided in Appendix 1.

The results for multiple logistic regression with one continuous and one binary covariate (the latter either rare or common) show that the bias for each covariate was very similar to the bias that would be observed if only that covariate were included. A relationship between the two covariates was seen for the 3S, PC3S and PW3S methods, where the bias present in the binary covariate increased if there was a strong relationship of the continuous

Table 1. Average estimates (Est), standard errors (SE), standardized bias (Bias) and coverage probability of the 95% confidence interval (CP) in covariate estimates for a single continuous covariate.

		Separation																	
		6				8				10				12				14			
Method	OR	Est	SE	Bias	CP	Est	SE	Bias	CP	Est	SE	Bias	CP	Est	SE	Bias	CP	Est	SE	Bias	CP
1S	0.25	−1.43	0.19	−0.10	0.97	−1.39	0.17	0.07	0.93	−1.40	0.16	0.03	0.94	−1.41	0.15	−0.09	0.96	−1.40	0.15	−0.02	0.95
	0.5	−0.72	0.14	−0.12	0.96	−0.71	0.13	−0.07	0.96	−0.69	0.12	0.06	0.92	−0.71	0.12	−0.08	0.95	−0.69	0.11	0.05	0.97
	1	0.01	0.12	0.08	0.94	−0.01	0.12	−0.07	0.96	−0.01	0.11	−0.06	0.94	0.02	0.10	0.18	0.94	−0.01	0.10	−0.12	0.93
	2	0.71	0.14	0.06	0.96	0.69	0.13	−0.05	0.95	0.71	0.12	0.12	0.93	0.69	0.11	−0.03	0.95	0.70	0.11	0.05	0.95
	4	1.41	0.19	0.01	0.95	1.40	0.17	−0.02	0.97	1.40	0.16	−0.02	0.95	1.40	0.15	−0.01	0.98	1.41	0.15	0.08	0.94
3S	0.25	−0.96	0.11	3.77	0.09	−1.04	0.12	3.02	0.19	−1.15	0.12	2.01	0.50	−1.24	0.13	1.19	0.77	−1.30	0.13	0.71	0.89
	0.5	−0.56	0.11	1.30	0.73	−0.59	0.11	1.05	0.84	−0.61	0.11	0.79	0.86	−0.65	0.11	0.42	0.93	−0.66	0.11	0.32	0.92
	1	0.01	0.10	0.10	0.95	0.00	0.10	−0.05	0.97	0.00	0.10	−0.04	0.96	0.02	0.10	0.18	0.96	−0.01	0.10	−0.12	0.94
	2	0.54	0.11	−1.44	0.71	0.57	0.11	−1.16	0.79	0.63	0.11	−0.64	0.89	0.64	0.11	−0.50	0.93	0.67	0.11	−0.23	0.94
	4	0.95	0.11	−3.88	0.07	1.05	0.12	−2.89	0.20	1.15	0.12	−2.03	0.51	1.23	0.13	−1.27	0.74	1.31	0.13	−0.65	0.87
PC3S	0.25	−0.75	0.12	5.16	0.00	−0.87	0.13	4.11	0.00	−1.02	0.13	2.85	0.17	−1.15	0.14	1.80	0.57	−1.24	0.14	1.10	0.84
	0.5	−0.45	0.12	2.10	0.41	−0.51	0.12	1.64	0.64	−0.56	0.11	1.24	0.79	−0.62	0.11	0.71	0.92	−0.64	0.11	0.49	0.93
	1	0.01	0.11	0.06	1.00	−0.01	0.11	−0.06	0.99	0.00	0.11	−0.05	0.98	0.02	0.10	0.16	0.98	−0.01	0.10	−0.11	0.96
	2	0.44	0.12	−2.15	0.39	0.50	0.12	−1.74	0.63	0.57	0.11	−1.08	0.83	0.61	0.11	−0.79	0.90	0.65	0.11	−0.41	0.94
	4	0.75	0.12	−5.20	0.00	0.88	0.13	−4.03	0.02	1.01	0.13	−2.87	0.15	1.14	0.14	−1.88	0.53	1.25	0.14	−1.04	0.82
I3S	0.25	−1.40	0.22	0.09	0.98	−1.37	0.19	0.23	0.93	−1.39	0.18	0.12	0.95	−1.41	0.16	−0.03	0.96	−1.40	0.15	0.03	0.95
	0.5	−0.71	0.15	−0.06	0.97	−0.70	0.14	0.00	0.95	−0.69	0.13	0.09	0.94	−0.70	0.12	−0.03	0.94	−0.69	0.11	0.07	0.97
	1	0.01	0.13	0.10	0.96	−0.01	0.12	−0.04	0.97	0.00	0.11	−0.04	0.95	0.02	0.10	0.18	0.95	−0.01	0.10	−0.12	0.94
	2	0.69	0.15	−0.05	0.94	0.69	0.13	−0.11	0.96	0.71	0.13	0.05	0.93	0.69	0.12	−0.04	0.96	0.70	0.11	0.03	0.95
	4	1.37	0.22	−0.20	0.95	1.39	0.20	−0.11	0.96	1.38	0.17	−0.13	0.96	1.39	0.16	−0.09	0.97	1.41	0.15	0.02	0.92
PR	0.25	−1.82	0.21	−2.02	0.49	−2.39	0.26	−3.81	0.06	−3.22	0.32	−5.69	0.00	−4.18	0.39	−7.23	0.00	−5.34	0.46	−8.60	0.00
	0.5	−1.18	0.25	−1.92	0.52	−1.55	0.31	−2.74	0.20	−2.01	0.39	−3.41	0.07	−2.68	0.47	−4.24	0.02	−3.36	0.57	−4.73	0.00
	1	0.01	0.27	0.05	0.95	0.00	0.34	0.00	0.94	−0.05	0.42	−0.13	0.93	0.05	0.51	0.10	0.94	−0.08	0.61	−0.13	0.92
	2	1.15	0.25	1.80	0.61	1.53	0.31	2.67	0.21	2.08	0.39	3.57	0.05	2.64	0.47	4.17	0.02	3.41	0.56	4.85	0.00
	4	1.81	0.21	1.95	0.51	2.40	0.26	3.82	0.04	3.20	0.32	5.63	0.00	4.16	0.39	7.15	0.00	5.34	0.46	8.62	0.00
PW3S	0.25	−1.03	0.12	2.97	0.20	−1.10	0.13	2.39	0.37	−1.19	0.13	1.58	0.66	−1.27	0.13	0.91	0.83	−1.32	0.13	0.55	0.92
	0.5	−0.58	0.11	1.01	0.83	−0.61	0.11	0.81	0.90	−0.63	0.11	0.63	0.90	−0.66	0.11	0.30	0.94	−0.67	0.11	0.26	0.94
	1	0.01	0.11	0.09	0.95	0.00	0.11	−0.05	0.97	−0.01	0.10	−0.05	0.95	0.02	0.10	0.18	0.96	−0.01	0.10	−0.12	0.94
	2	0.57	0.11	−1.12	0.82	0.60	0.11	−0.92	0.85	0.65	0.11	−0.48	0.90	0.65	0.11	−0.39	0.95	0.68	0.11	−0.16	0.95
	4	1.01	0.12	−3.08	0.16	1.11	0.13	−2.29	0.38	1.19	0.13	−1.59	0.65	1.26	0.13	−0.99	0.83	1.33	0.14	−0.50	0.89

1S: 1-step; 3S: 3-step; PC3S: pseudo class 3-step; I3S: improved 3-step; PR: probability regression; PW3S: probability weighted 3-step.

Table 2. Average estimates (Est), standard errors (SE), standardized bias (Bias) and coverage probability of the 95% confidence interval (CP) in covariate estimates for a single common binary covariate.

		Separation																			
		6				8				10				12				14			
Method	OR	Est	SE	Bias	CP	Est	SE	Bias	CP	Est	SE	Bias	CP	Est	SE	Bias	CP	Est	SE	Bias	CP
	0.25	−1.41	0.27	−0.03	0.93	−1.39	0.25	0.02	0.97	−1.42	0.24	−0.09	0.96	−1.39	0.23	0.02	0.95	−1.38	0.22	0.06	0.96
	0.5	−0.69	0.25	0.03	0.95	−0.70	0.24	−0.02	0.94	−0.72	0.22	−0.08	0.98	−0.70	0.21	−0.04	0.96	−0.70	0.21	−0.02	0.94
1S	1	−0.02	0.24	−0.08	0.93	0.01	0.23	0.04	0.95	−0.02	0.22	−0.11	0.94	−0.01	0.21	−0.03	0.94	0.00	0.20	0.01	0.97
	2	0.69	0.25	−0.02	0.97	0.71	0.23	0.05	0.97	0.68	0.22	−0.07	0.93	0.69	0.21	−0.03	0.95	0.70	0.21	0.03	0.97
	4	1.41	0.27	0.02	0.93	1.42	0.25	0.08	0.95	1.41	0.24	0.08	0.96	1.41	0.23	0.09	0.95	1.40	0.22	0.01	0.96
	0.25	−1.11	0.22	1.29	0.72	−1.17	0.22	1.00	0.82	−1.27	0.22	0.57	0.91	−1.30	0.22	0.43	0.94	−1.32	0.21	0.32	0.95
	0.5	−0.55	0.21	0.72	0.88	−0.59	0.21	0.51	0.91	−0.65	0.20	0.24	0.97	−0.66	0.20	0.19	0.95	−0.68	0.20	0.10	0.95
3S	1	−0.02	0.21	−0.10	0.92	0.01	0.20	0.05	0.94	−0.02	0.20	−0.12	0.95	0.00	0.20	−0.01	0.94	0.00	0.20	0.00	0.97
	2	0.56	0.21	−0.65	0.91	0.61	0.21	−0.43	0.94	0.61	0.20	−0.40	0.95	0.64	0.20	−0.26	0.94	0.67	0.20	−0.11	0.97
	4	1.11	0.22	−1.29	0.71	1.20	0.22	−0.88	0.88	1.27	0.22	−0.57	0.92	1.32	0.22	−0.35	0.96	1.34	0.22	−0.23	0.95
	0.25	−0.90	0.24	2.05	0.48	−1.01	0.23	1.62	0.67	−1.15	0.23	1.05	0.86	−1.22	0.22	0.75	0.92	−1.29	0.22	0.48	0.94
	0.5	−0.45	0.23	1.05	0.91	−0.52	0.22	0.79	0.93	−0.59	0.22	0.48	0.98	−0.63	0.21	0.33	0.97	−0.66	0.21	0.19	0.96
PC3S	1	−0.01	0.23	−0.06	0.99	0.01	0.22	0.04	1.00	−0.02	0.21	−0.09	0.99	−0.01	0.21	−0.02	0.95	0.00	0.20	0.01	0.98
	2	0.46	0.23	−1.04	0.91	0.53	0.22	−0.75	0.94	0.56	0.22	−0.62	0.94	0.61	0.21	−0.39	0.96	0.66	0.21	−0.18	0.98
	4	0.90	0.24	−2.05	0.45	1.03	0.23	−1.54	0.75	1.15	0.23	−1.05	0.88	1.25	0.22	−0.64	0.96	1.30	0.22	−0.41	0.94
	0.25	−1.38	0.28	0.08	0.95	−1.38	0.26	0.06	0.98	−1.41	0.25	−0.05	0.96	−1.39	0.23	0.03	0.96	−1.37	0.22	0.09	0.96
	0.5	−0.67	0.26	0.11	0.95	−0.69	0.24	0.04	0.94	−0.71	0.23	−0.08	0.97	−0.70	0.22	−0.02	0.95	−0.70	0.21	−0.01	0.95
I3S	1	−0.03	0.25	−0.10	0.92	0.01	0.23	0.05	0.94	−0.03	0.22	−0.12	0.95	0.00	0.21	−0.01	0.94	0.00	0.20	0.00	0.98
	2	0.69	0.26	−0.04	0.95	0.71	0.24	0.04	0.97	0.68	0.23	−0.08	0.94	0.68	0.22	−0.06	0.94	0.70	0.21	0.00	0.97
	4	1.38	0.28	−0.08	0.96	1.41	0.26	0.06	0.96	1.41	0.25	0.06	0.97	1.41	0.23	0.05	0.97	1.39	0.22	0.00	0.96
	0.25	−2.40	0.53	−1.91	0.51	−3.20	0.65	−2.76	0.21	−4.22	0.81	−3.50	0.06	−5.45	0.98	−4.12	0.01	−7.00	1.18	−4.74	0.01
	0.5	−1.22	0.53	−0.99	0.84	−1.67	0.67	−1.46	0.69	−2.26	0.83	−1.89	0.57	−2.89	1.01	−2.18	0.40	−3.69	1.21	−2.48	0.31
PR	1	−0.02	0.54	−0.03	0.95	0.01	0.67	0.00	0.95	−0.02	0.83	−0.03	0.95	−0.01	1.01	−0.02	0.95	0.03	1.22	0.02	0.95
	2	1.24	0.53	1.01	0.80	1.66	0.67	1.44	0.72	2.13	0.83	1.74	0.59	2.90	1.01	2.18	0.39	3.74	1.21	2.52	0.32
	4	2.39	0.53	1.90	0.55	3.19	0.65	2.76	0.18	4.32	0.81	3.60	0.05	5.65	0.98	4.35	0.01	7.11	1.18	4.85	0.00
	0.25	−1.16	0.23	1.00	0.80	−1.22	0.23	0.78	0.91	−1.30	0.22	0.42	0.94	−1.32	0.22	0.33	0.95	−1.34	0.22	0.26	0.95
	0.5	−0.57	0.22	0.56	0.92	−0.61	0.21	0.38	0.92	−0.66	0.21	0.18	0.98	−0.67	0.21	0.13	0.96	−0.68	0.20	0.07	0.95
PW3S	1	−0.02	0.22	−0.08	0.94	0.01	0.21	0.05	0.95	−0.02	0.20	−0.11	0.96	0.00	0.20	−0.02	0.94	0.00	0.20	0.00	0.97
	2	0.58	0.22	−0.52	0.93	0.63	0.21	−0.32	0.95	0.63	0.21	−0.34	0.96	0.65	0.20	−0.21	0.95	0.68	0.20	−0.08	0.97
	4	1.16	0.23	−1.00	0.81	1.24	0.23	−0.67	0.92	1.30	0.22	−0.42	0.93	1.34	0.22	−0.24	0.97	1.36	0.22	−0.17	0.96

1S: 1-step; 3S: 3-step; PC3S: pseudo class 3-step; I3S: improved 3-step; PR: probability regression; PW3S: probability weighted 3-step.

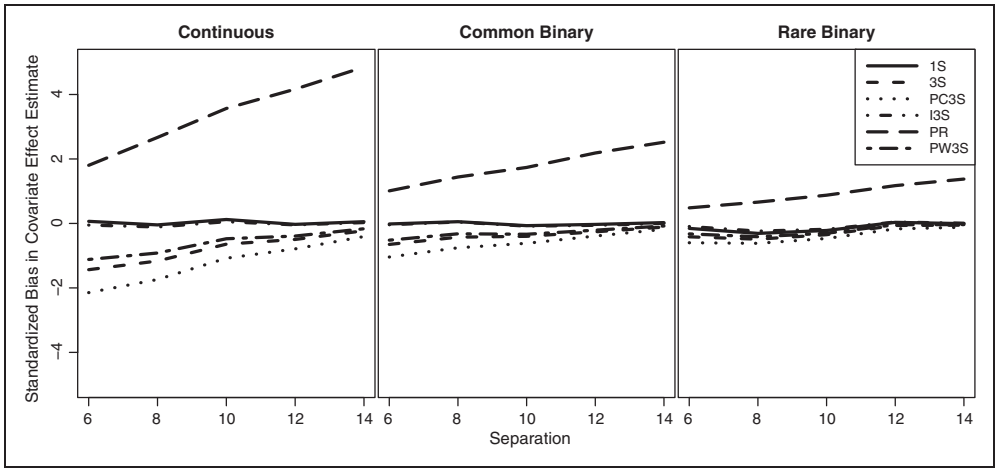


Figure 2. Average standardized bias in covariate effect estimates for single covariates with an odds ratio of 2.

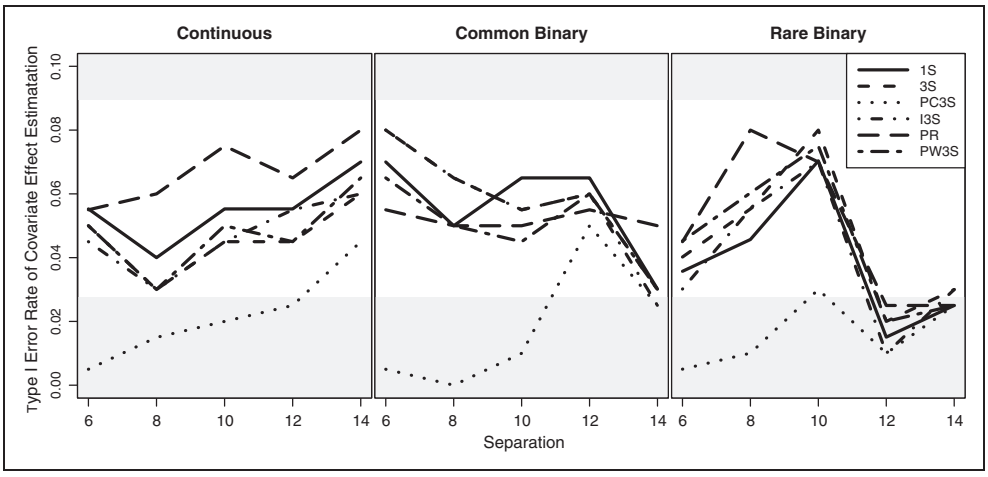


Figure 3. Estimated type I error rates for hypothesis tests of the covariate effect for single covariates. Lines entering the grey areas have 95% confidence intervals which do not contain the nominal type I error rate of 0.05.

covariate in addition to a strong relationship for the binary covariate. Conversely, for the PR approach, the bias decreased for the binary covariate as the odds ratio of the relationship with the continuous covariate increased. The results for models with two covariates are given in Supplementary Table 2.

Figure 3 presents the estimated type I error rates for hypothesis tests of single covariate effects. On average the type I errors for the scenarios conducted under the null hypothesis were close to the α level, with the exception of the PC3S method which had an average type I error of 0.019. Lines entering the grey areas of Figure 3 have 95% confidence intervals which do not contain the nominal type I error of 0.05. The correlation that can be observed between the type I errors, for the different methods, is due to sampling variability as the different methods were applied to the same simulated samples. This sampling variability is why all methods had an estimated type I error rate significantly below 0.05 for the rare binary simulation with separation 12.

The estimated power of the single covariate models for the 1S and I3S methods is shown in Supplementary Table 3. The power is only shown for these two methods as they were free from significant bias. In each of the continuous scenarios, the power was estimated to be 1.00. Due to the increased variability of the binary estimates described earlier, the power was lower in these scenarios, and in particular for the rare binary covariates. When it

was estimated to be lower than 1.00, the power tended to increase as the separation increased. The power was generally similar for the 1S and I3S methods. For the rare binary scenarios with odds ratio of 2, the power was not symmetrical with those scenarios with odds ratio of 0.5. In these cases the power was lower than expected, possibly due to the exclusion of simulations with fewer than five positive response observations.

4.2 Analysis of PR

This section details why the PR method was observed to perform poorly, with bias in the opposite direction to the remainder of the methods that also increased with separation. This is counter-intuitive, as bias should decrease with the classification difficulty, as separation increases. We demonstrate that the estimator of the effect of the covariate is unreliable for this method, as it is almost always biased, and that the bias increases as the separation increases.

Consider a two-group mixture model ($G \in \{1, 2\}$) for a multivariate outcome $\boldsymbol{Y}$ such that $\boldsymbol{Y}|G = 1 \sim MVN(\boldsymbol{\mu}_1, \Sigma)$ with probability π and $\boldsymbol{Y}|G = 2 \sim MVN(\boldsymbol{\mu}_2, \Sigma)$ with probability $1 - \pi$. We assume homogeneous variance matrices here to simplify the calculations. The general case for non-homogeneous variance matrices is provided in Appendix 1. Consider also a univariate covariate X with $X|G = 1 \sim N(\gamma, 1)$ and $X|G = 2 \sim N(0, 1)$, independently of $\boldsymbol{Y}$. Here the outcome in a logistic regression of G on X, logit $P(G = 1|x)$, can be simplified as follows

$$\begin{aligned}
\text{logit}P(G = 1|x) &= \log \frac{P(G = 1|x)}{P(G = 2|x)} \\
&= \log \frac{\pi f(x|G = 1)}{(1 - \pi) f(x|G = 2)} \\
&= \log \frac{\pi}{1 - \pi} - \frac{1}{2}(x - \gamma)^2 + \frac{1}{2}x^2 \\
&= \log \frac{\pi}{1 - \pi} - \frac{\gamma^2}{2} + \gamma x
\end{aligned}$$

Thus, the true value for the effect of the covariate, given by the coefficient of x in the logistic regression, is γ, or the mean of X in group 1.

In the PR method, the logit of the posterior probabilities is used as the outcome in a linear regression on the covariate

$$\text{logit } P(G = 1|\boldsymbol{y}) = \beta_0 + \beta x$$

Using the same argument as above

$$\text{logit } P(G = 1|\boldsymbol{y}) = \log \frac{\pi}{1 - \pi} + \left(\boldsymbol{y} - \frac{\boldsymbol{\mu}_1 + \boldsymbol{\mu}_2}{2}\right)^T \Sigma^{-1}(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)$$

Thus, in the PR simple linear regression the slope parameter is

$$\beta = \frac{\text{cov}(\text{logit } P(G = 1|\boldsymbol{Y}), X)}{\text{var}(X)} = \frac{\text{cov}(\boldsymbol{Y}^T \Sigma^{-1}(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2), X)}{\text{var}(X)} = \frac{\text{cov}(\boldsymbol{Y}, X)^T \Sigma^{-1}(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)}{\text{var}(X)}$$

where $\text{cov}(\boldsymbol{Y}, X) = [\text{cov}(Y_1, X), \text{cov}(Y_2, X), \ldots, \text{cov}(Y_p, X)]^T$ and p is the dimension of $\boldsymbol{Y}$. Now, using the facts that $E(X) = \gamma\pi$, $E(\boldsymbol{Y}) = \pi\boldsymbol{\mu}_1 + (1 - \pi)\boldsymbol{\mu}_2$, $E(X\boldsymbol{Y}) = \pi\boldsymbol{\mu}_1\gamma$ and $E(X^2) = \gamma^2\pi + 1$, we have

$$\beta = \frac{(E(X\boldsymbol{Y}) - E(X)E(\boldsymbol{Y}))^T \Sigma^{-1}(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)}{E(X^2) - E(X)^2} = \frac{\gamma\pi(1 - \pi)(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)^T \Sigma^{-1}(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)}{\gamma^2\pi(1 - \pi) + 1}$$

Thus, the estimate for the effect of x will only be unbiased if by coincidence it happens that

$$\pi(1 - \pi)(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)^T \Sigma^{-1}(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2) = \gamma^2\pi(1 - \pi) + 1$$

Therefore in general, the PR method provides a biased estimator of the effect of the covariate on the group membership probabilities. A similar expression for the slope can be obtained in the case of $\Sigma_1 \neq \Sigma_2$ and details are provided in Appendix 1.

These calculations provide an explanation why the PR simulation estimates were found to differ so greatly from the known parameter. In the scenarios with homogeneous variances, the bias for the PR method increases with the size of the squared Mahalanobis distance between the group means, $(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)^T \Sigma^{-1} (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)$, which increases as the separation between the groups increases. This is very different behaviour from the situation with the other methods, where large separation generally implies good performance.

5 Discussion

The results of these simulation studies confirm that, when possible, the 1S method should be used to estimate the effect of a covariate, as it displayed negligible levels of bias. When it is not desired to use the 1S method, the I3S method should be used as it was the only method to provide inconsequential bias among the methods that allow the group-based model to be estimated before including covariates. As the 1S and I3S methods are those we recommend, code to conduct these methods in Mplus has been provided in Appendix 1. The other 3S methods, including the 3S method and those that attempt to improve on it, should not be used as they had greater levels of bias than the I3S method. In addition to displaying high levels of bias, PR was demonstrated to result in a biased estimator for the effect of the covariate on the group membership probability and should also not be used. These results are in keeping with previous studies[1,3,4]; however in the case of PR, we have provided an explanation for the bias observed previously.

The 3S estimates appear to be biased downwards due to misclassification of trajectories on the boundary of the two groups. This misclassification results in attenuation of any effect of the covariate that exists, causing the estimate of the odds ratio to be biased downwards towards 1. The misclassification decreases as separation increases, and thus the bias of the method reduces. The bias of the 3S method can be corrected by accounting for the uncertainty in these misclassified trajectories with the I3S method.

This study considered the bias in estimates for the effect of multiple covariates and found that the presence of two covariates did not greatly change the results expected with one covariate. The only situation where a difference was observed was for two very strongly related covariates. As far as we are aware, no previous study has compared the effect of using multiple covariates with that of using individual covariates.

One of the reasons that has been suggested for why the 1S approach is not desirable when there are a large number of covariates to consider is that it can be computationally intensive to re-estimate the complex group-based trajectory model for each covariate or combination of covariates.[1] While this is a problem that would have been intractable for some models in the past, it becomes more feasible with improvements in computing and technological advances.

In this study, we have only considered data simulated from two groups and with no direct effects of the covariates on the trajectories. Although we anticipate the results of our study could generalize more broadly, future research is needed to investigate results with three or more groups. The presence of direct effects has been explored by Vermunt[1] in his paper and is also an avenue worthy of further investigation. Finally, although our simulations used complete data at all time points, these methods can be used in datasets with incomplete data. There is no reason to expect our conclusions would be altered in the case of incomplete data.

In summary, when investigating the effect of one or several covariates on a group-based trajectory model, the 1S or I3S methods should be used to minimize the bias that can result from misclassification error. The other 3S methods (3S, PC3S, PW3S and PR) resulted in considerably higher bias in the covariate effect estimates and should not be used. The PR approach especially should be avoided as it has been demonstrated analytically to be inconsistent.

Acknowledgements

The support of the Generation 1 Study Chief Investigators, Prof Vivienne Moore and Prof Michael Davies, in allowing access to data from that study is gratefully acknowledged.

Declaration of conflicting interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: Christopher Davies received financial support for this research through a postgraduate scholarship from the University of Adelaide. Data collection for the Generation 1 study was supported by grants from the Faculty of Health Sciences at the University of Adelaide, Dairy Research and Development Corporation, Channel 7 Children's Research Foundation Grant 13745, and National Health and Medical Research Council (Grants 465455 and 465437 and Australian Based Public Health Training Fellowship 627033 to LCG).

Supplementary material

Supplementary material for this paper can be found at http://journals.sagepub.com/doi/suppl/10.1177/0962280216689580

References

1. Vermunt JK. Latent class modeling with covariates: two improved three-step approaches. *Polit Anal* 2010; **18**: 450–469.
2. Dayton CM and Macready GB. Concomitant-variable latent-class models. *J Am Stat Assoc* 1988; **83**: 173–178.
3. Bolck A, Croon M and Hagenaars J. Estimating latent structure models with categorical variables: one-step versus three-step estimators. *Polit Anal* 2004; **12**: 3–27.
4. Clark SL and Muthén B. Relating latent class analysis results to variables not included in the analysis, www.statmodel.com/download/relatinglca.pdf (2009, accessed 21 September 2016).
5. Asparouhov T and Muthén B. Auxiliary variables in mixture modeling: three-step approaches using Mplus. *Struct Equ Modeling* 2014; **21**: 329–341.
6. Jung T and Wickrama KAS. An introduction to latent class growth analysis and growth mixture modeling. *Soc Pers Psychol Compass* 2008; **2**: 302–317.
7. Oliveira-Brochado A and Martins FV. Assessing the number of components in mixture models: a review, https://ideas.repec.org/p/por/fepwps/194.html (2005, accessed 21 September 2016).
8. Davies CE, Glonek GFV and Giles LC. The impact of covariance misspecification in group-based trajectory models for longitudinal data with non-stationary covariance structure. *Stat Methods Med Res*. Prepublished August 17, 2015; DOI: 10.1177/0962280215598806.
9. Nagin DS and Land KC. Age, criminal careers, and population heterogeneity: specification and estimation of a nonparametric, mixed poisson model. *Criminology* 1993; **31**: 327–362.
10. Muthén B and Muthén LK. Integrating person-centered and variable-centered analyses: growth mixture modeling with latent trajectory classes. *Alcohol Clin Exp Res* 2000; **24**: 882–891.
11. Muthén LK and Muthén BO. *Mplus user's guide*, 6th ed. Los Angeles, CA: Muthén & Muthén, 1998–2010.
12. Moore VM, Davies MJ, Willson KJ, et al. Dietary composition of pregnant women is related to size of the baby at birth. *J Nutr* 2004; **134**: 1820–1826.
13. Achenbach TM and Rescorla LA. *ASEBA school-age forms & profiles*. Burlington, VT: Aseba, 2001.
14. Hastie T, Tibshirani R and Friedman J. *The elements of statistical learning*, 2nd ed. New York: Springer, 2009.
15. R Core Team. *R: a language and environment for statistical computing*. R Foundation for Statistical Computing: Vienna, Austria, 2014.
16. Hallquist M and Wiley J. *MplusAutomation: automating Mplus model estimation and interpretation (Version 0.6-3)*, 2014 https://CRAN.R-project.org/package=MplusAutomation.
17. Mathai AM and Provost SB. *Quadratic forms in random variables: theory and applications*. New York: Marcel Dekker Inc., 1992.

Appendix 1

Extension of analysis for PR

Here we extend the argument in Section 4.2 to the general case where the two groups do not have the same covariance matrix.

Consider a two-group mixture model ($G \in \{1, 2\}$) for a multivariate outcome $\boldsymbol{Y}$ such that $\boldsymbol{Y}|G = 1 \sim MVN(\boldsymbol{\mu}_1, \Sigma_1)$ with probability π and $\boldsymbol{Y}|G = 2 \sim MVN(\boldsymbol{\mu}_2, \Sigma_2)$ with probability $1 - \pi$. Consider also a univariate covariate X with $X|G = 1 \sim N(\gamma, 1)$ and $X|G = 2 \sim N(0, 1)$, independently of $\boldsymbol{Y}$, given G. As detailed in Section 4.2 the true value for the effect of the covariate, given by the coefficient of x in the logistic regression, is γ, or the mean of X in group 1.

In the PR method the logit of the posterior probabilities is used as the outcome in a linear regression on the covariate

$$\text{logit } P(G=1|\boldsymbol{y}) = \beta_0 + \beta x$$

Using the same argument as above

$$\begin{aligned}\text{logit}P(G=1|\boldsymbol{y}) &= \log\frac{\pi}{1-\pi} + \frac{1}{2}\log\frac{|\Sigma_2|}{|\Sigma_1|} - \frac{1}{2}(\boldsymbol{y}-\boldsymbol{\mu}_1)^T\Sigma_1^{-1}(\boldsymbol{y}-\boldsymbol{\mu}_1) + \frac{1}{2}(\boldsymbol{y}-\boldsymbol{\mu}_2)^T\Sigma_2^{-1}(\boldsymbol{y}-\boldsymbol{\mu}_2)\\ &= \text{constant} + \frac{1}{2}\boldsymbol{y}^T(\Sigma_2^{-1}-\Sigma_1^{-1})\boldsymbol{y} + \boldsymbol{y}^T(\Sigma_1^{-1}\boldsymbol{\mu}_1 - \Sigma_2^{-1}\boldsymbol{\mu}_2)\end{aligned}$$

Thus, in the PR simple linear regression the theoretical estimate of the slope is given by

$$\begin{aligned}\hat{\beta} &= \frac{\text{cov}(\text{logit } P(G=1|\boldsymbol{Y}), X)}{\text{var}(X)}\\ &= \frac{\frac{1}{2}\text{cov}(\boldsymbol{Y}^T(\Sigma_2^{-1}-\Sigma_1^{-1})\boldsymbol{Y}, X) + \text{cov}(\boldsymbol{Y}^T(\Sigma_1^{-1}\boldsymbol{\mu}_1 - \Sigma_2^{-1}\boldsymbol{\mu}_2), X)}{\text{var}(X)}\\ &= \frac{\frac{1}{2}\text{cov}(\boldsymbol{Y}^T\boldsymbol{S}\boldsymbol{Y}, X) + \text{cov}(\boldsymbol{Y}, X)^T(\Sigma_1^{-1}\boldsymbol{\mu}_1 - \Sigma_2^{-1}\boldsymbol{\mu}_2)}{\text{var}(X)}\end{aligned}$$

where $\boldsymbol{S} = \Sigma_2^{-1} - \Sigma_1^{-1}$, $\text{cov}(\boldsymbol{Y}, X) = [\text{cov}(Y_1, X), \text{cov}(Y_2, X), \ldots, \text{cov}(Y_p, X)]^T$ and p is the dimension of $\boldsymbol{Y}$.

Next we use the facts that $E(X) = \pi\gamma$, $E(\boldsymbol{Y}) = \pi\boldsymbol{\mu}_1 + (1-\pi)\boldsymbol{\mu}_2$, $E(X\boldsymbol{Y}) = \pi\boldsymbol{\mu}_1\gamma$, $E(X^2) = \pi\gamma^2 + 1$, $E(\boldsymbol{Y}^T\boldsymbol{S}\boldsymbol{Y}) = \pi[\text{tr}(\boldsymbol{S}\Sigma_1) + \boldsymbol{\mu}_1^T\boldsymbol{S}\boldsymbol{\mu}_1] + (1-\pi)[\text{tr}(\boldsymbol{S}\Sigma_2) + \boldsymbol{\mu}_2^T\boldsymbol{S}\boldsymbol{\mu}_2]$ and $E(X\boldsymbol{Y}^T\boldsymbol{S}\boldsymbol{Y}) = \pi\gamma[\text{tr}(\boldsymbol{S}\Sigma_1) + \boldsymbol{\mu}_1^T\boldsymbol{S}\boldsymbol{\mu}_1]$. The last two of these expressions use the known identity for the expected value of a quadratic form in random variables.[17] The theoretical estimate of the slope in the PR simple linear regression is

$$\begin{aligned}\hat{\beta} &= \frac{\frac{1}{2}[E(X\boldsymbol{Y}^T\boldsymbol{S}\boldsymbol{Y}) - E(X)E(\boldsymbol{Y}^T\boldsymbol{S}\boldsymbol{Y})] + [E(X\boldsymbol{Y}) - E(X)E(\boldsymbol{Y})]^T(\Sigma_1^{-1}\boldsymbol{\mu}_1 - \Sigma_2^{-1}\boldsymbol{\mu}_2)}{E(X^2) - E(X)^2}\\ &= \frac{\gamma\pi(1-\pi)\left[\frac{1}{2}\{\text{tr}(\boldsymbol{S}\Sigma_1 - \boldsymbol{S}\Sigma_2) + \boldsymbol{\mu}_1^T\boldsymbol{S}\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2^T\boldsymbol{S}\boldsymbol{\mu}_2\} + (\boldsymbol{\mu}_1-\boldsymbol{\mu}_2)^T(\Sigma_1^{-1}\boldsymbol{\mu}_1 - \Sigma_2^{-1}\boldsymbol{\mu}_2)\right]}{\gamma^2\pi(1-\pi)+1}\\ &= \frac{\gamma\pi(1-\pi)\left[\frac{1}{2}\text{tr}(\Sigma_2^{-1}\Sigma_1 + \Sigma_1^{-1}\Sigma_2) - p + (\boldsymbol{\mu}_1-\boldsymbol{\mu}_2)^T\left(\frac{\Sigma_1^{-1}+\Sigma_2^{-1}}{2}\right)(\boldsymbol{\mu}_1-\boldsymbol{\mu}_2)\right]}{\gamma^2\pi(1-\pi)+1}\end{aligned}$$

Thus, the estimate for the effect of x will only be correct if by coincidence it happens that

$$\pi(1-\pi)\left[\frac{1}{2}\text{tr}(\Sigma_2^{-1}\Sigma_1 + \Sigma_1^{-1}\Sigma_2) - p + (\boldsymbol{\mu}_1-\boldsymbol{\mu}_2)^T\left(\frac{\Sigma_1^{-1}+\Sigma_2^{-1}}{2}\right)(\boldsymbol{\mu}_1-\boldsymbol{\mu}_2)\right] = \gamma^2\pi(1-\pi)+1$$

Therefore in general, the PR method provides an inconsistent estimate of the effect of the covariate on the group membership probabilities.

Examples of Mplus code for 1S and I3S methods

```
TITLE: Example code for a GMM with 1S method
DATA: FILE IS data.csv;
VARIABLE: NAMES ARE y1-y4 x;
CLASSES = c (2);
ANALYSIS: TYPE = MIXTURE;
STARTS = 200 20;
MODEL:
%OVERALL%
is | y1@0 y2@1 y3@2 y4@3;
C#1 on x;
SAVEDATA: FILE IS output.csv;
SAVE = CPROB;

TITLE: Example code for a GMM with I3S method
DATA: FILE IS data.csv;
VARIABLE: NAMES ARE y1-y4 x;
CLASSES = c (2);
AUXILIARY = x(R3STEP);
ANALYSIS: TYPE = MIXTURE;
STARTS = 200 20;
MODEL:
%OVERALL%
is | y1@0 y2@1 y3@2 y4@3;
SAVEDATA: FILE IS output.csv;
SAVE = CPROB;
```

Chapter 6

Summary and conclusions

In this thesis I have examined the statistical properties of group-based trajectory models. Specifically, I have addressed four aims relating to covariance assumptions, outliers and the use of covariates in these models. Externalising behaviour data from the Generation 1 Study was used to demonstrate fitting of these models and was the motivating example on which the suite of simulation studies was based. In this final chapter, major findings are summarised, limitations of this work are described and future directions for research are discussed.

6.1 Major findings and contributions

6.1.1 Effect of covariance misspecification in group-based trajectory models

The first aim of this thesis was to investigate the effect of covariance misspecification on misclassification of trajectories in group-based trajectory models, including data with non-stationary covariance structure. In Chapter 3 the simulation study I undertook to achieve this aim was described. It was found that covariance misspecification can result in large reductions in the CCR. However, if model selection was made from a broad range of models using the BIC, a high CCR could be achieved without knowing the true model. Therefore it is recommended that in group-based trajectory applications, the range of models to be considered should have the flexibility to account for different covariance structures across time points, and especially between groups. After investigating between these more complex models and simpler candidates, the model with the best fit to the particular data set should be chosen.

6.1.2 Impact of outliers on group-based trajectory models

The second aim of this thesis was to explore the impact of outliers on group-based trajectory models. In Chapter 4 this aim was investigated through two simulation studies. From these studies I found that introducing outliers into group-based tra-

jectory datasets tended to increase the number of groups estimated and to reduce the CCR provided the group means were well separated. However, surprisingly, in particular scenarios when the simulated group means were close together, the presence of outliers could lead to improvements in CCR. Even more so than in other statistical areas, the possibility of outliers in the group-based trajectory modelling setting should not be ignored due to their large and unpredictable impacts.

6.1.3 Outlier detection and removal for group-based trajectory models

The third aim of this thesis was to develop an algorithm to identify outliers in the group-based trajectory modelling context, and to determine its effectiveness. In Chapter 4 this aim was addressed through the development of an algorithm based on changes in the BIC. The effectiveness of the algorithm was tested on the same datasets that were used to address aim 2. The difficulty of identifying outlying trajectories in the group-based trajectory modelling setting was observed, as changes in the number of groups and model structure can appear with the presence of even a small percentage of noisy trajectories. The algorithm performed well under certain model assumptions and where the simulated groups were well separated, but was less effective when the model was more flexible or the group means were close together. As established through addressing aim 2, outliers can have significant impacts on group-based trajectory models. Good practice should be to check for outliers and to

consider sensitivity analysis to understand their impact. The algorithm developed to address this aim can be used as a method for doing so.

6.1.4 Methods for estimating the effects of covariates on group membership probabilities

The final aim of this thesis was to compare the performance of methods that estimate the effects of covariates on the group membership probabilities in group-based trajectory models. In Chapter 5, the performance of six methods from the literature was compared through a simulation study. I found that when investigating the effect of one or several covariates on a group-based trajectory model, use of the 1S or I3S method minimised the bias that can result from misclassification error. The other 3S methods (3S, PC3S, PW3S and PR) resulted in considerably higher bias in the covariate effect estimates. In the case of the probability regression approach, the performance was found to be very poor, even in 'easy problems'. I derived an expression for the bias to verify these results analytically. Use of the 1S or I3S method is recommended to estimate the effects of covariates in the group-based trajectory modelling setting. This conclusion was supported by a recent article, published subsequently to mine, in which the 1S method was also found to perform better than the I3S method in certain circumstances.[75]

6.2 Limitations and future directions

The major conclusions outlined in the previous section are subject to various limitations and these are discussed in the substantive chapters of the thesis. In what follows, I address some general limitations of this thesis and identify potential future directions for research.

As a consequence of using simulation, the studies undertaken as part of this thesis were necessarily restricted to cover only some of the many possible scenarios that could be of interest. Most importantly, the simulation studies in Chapters 3 and 5 assumed the number of groups was assumed known and fixed at two. Future research in these areas is needed to consider the impact of three or more groups, and also situations in which the number of groups is not known. A Bayesian approach could be used in the case of an unknown number of groups, with the number of groups as a parameter. The complexity resulting from the number of groups being free to vary was observed in Chapter 4, and this needs careful consideration in future studies of group-based trajectory models.

The impact of covariates on the group membership probabilities was considered in Chapter 5, extending the work of Bolck et al., Vermunt and Clark and Muthn.[57–59] However explicit effects of covariates on the mean trajectories themselves were not considered. This further extension has been examined recently by Kim et al, who found it was important to at least consider the possibility of these direct effects.[76]

This thesis was motivated by applying group-based trajectory models to the

externalising behaviour data from the Generation 1 Study. In Chapter 3, the non-stationary covariance structure of the externalising behaviour data motivated the development of Aim 1. In Chapter 4, the outlier identification algorithm was applied to these data to illustrate the method, but also to investigate the impact of any outliers identified. With the removal of the three trajectories identified as outliers, the best fitting model changed from four groups to five groups. This shows that in practical settings, outliers can have a marked influence on model selection and estimation. With these estimated groups, researchers can better understand the various courses that externalising behaviour can take through childhood. The estimation of these groups also enables research into relationships with covariates from early childhood, using the methods recommended in Chapter 5, or prediction of subsequent outcomes in adolescence or adulthood.

One phenomenon encountered during the research for this thesis was the occurrence of invalid estimates from standard software used for fitting group-based trajectory models. In some simulated scenarios, the relevant group-based model could not be estimated or the estimates that resulted were invalid due to non positive-definite matrices. This was particularly the case when a rare-binary covariate was used in Chapter 5 with a large odds ratio. Although this was a rare occurrence and did not affect the conclusions reached, researchers using these models should be aware of the possibility that invalid estimates can be generated, and all model output should be checked closely.

In terms of computational resources, the simulation studies I undertook for Chap-

ter 4 were found to take impractical lengths of time on a standard desktop computer. However, access to a high performance computer allowed the simulations to be completed in a reasonable time. In practice, analysts wanting to conduct simulation studies may have limited access to such facilities and so algorithms that improve performance in the estimation of group-based trajectory models will make an important contribution and warrants future research.

Extensive simulation studies were used in each of Chapters 3, 4 and 5 to investigate the properties of group-based trajectory models. A possible future direction is to limit the classes of models being considered and to derive analytical results that demonstrate the simulation results observed.

6.3 Concluding remarks

The findings in this thesis serve to guide statisticians and applied researchers in their use of group-based trajectory models. When investigating the effect of covariates on a group-based trajectory model, the 1S or I3S method should be used to minimise the bias that can result from misclassification error. The use of my outlier identification algorithm is recommended as part of sensitivity analyses to understand the potential impact of outliers and to better understand the underlying structures in the data. In general, the results of fitting a group-based model to any dataset are highly dependent on the assumptions made regarding the covariance when determining the specific model to be used. Researchers should consider a wide range of models, and

bearing in mind the assumptions they make, carefully choose that which fits best with the data.

Appendix A

A.1 Mahalanobis distance percentile plots for divergent trajectories simulations

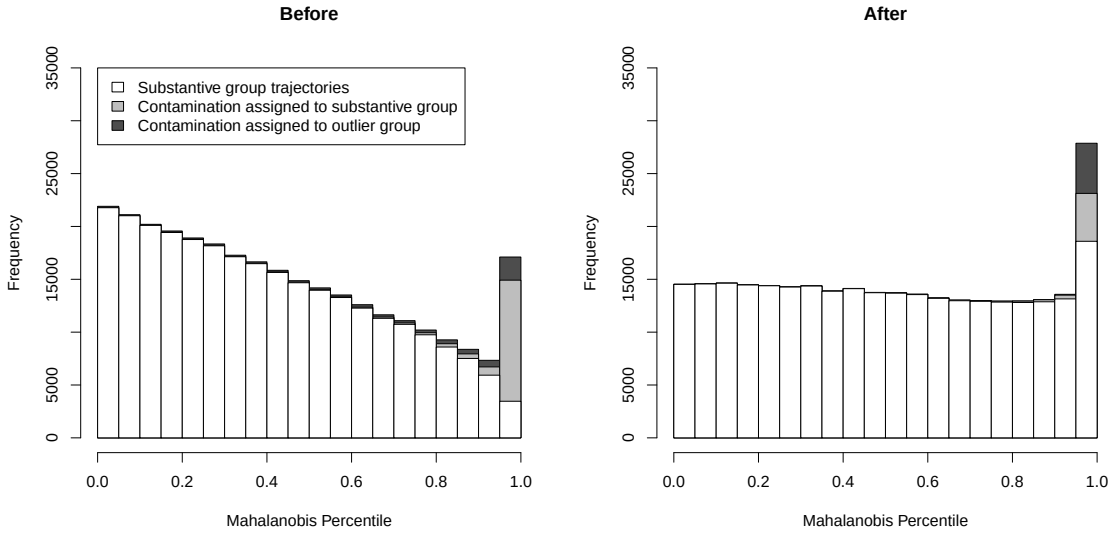


Figure A.1.1: Mahalanobis distance percentiles of trajectories from their assigned group mean, for model 1 with separation 1 and slope 0.25, before and after outlier removal.

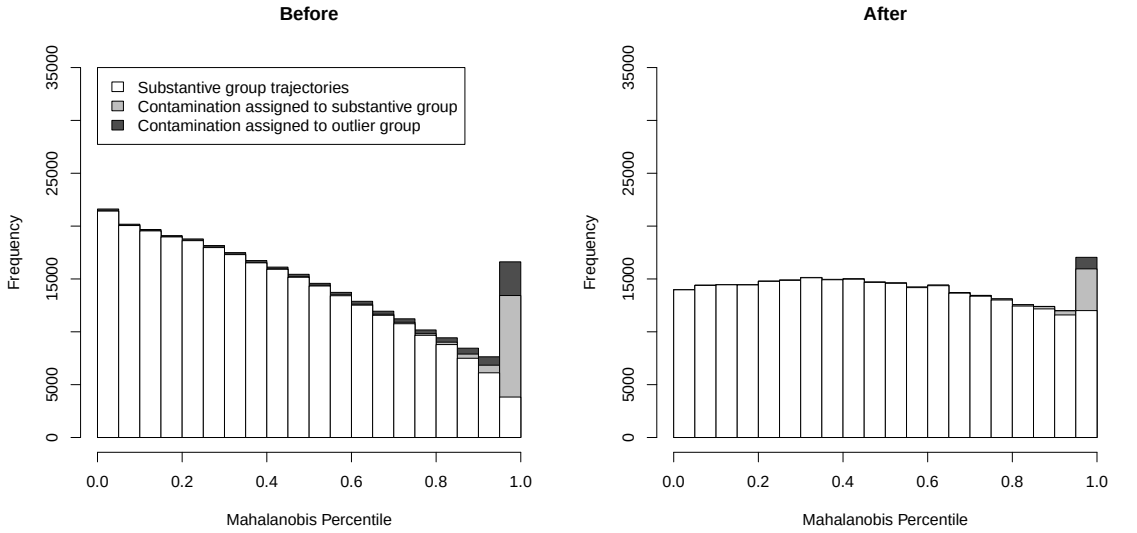


Figure A.1.2: Mahalanobis distance percentiles of trajectories from their assigned group mean, for model 1 with separation 1 and slope 0.5, before and after outlier removal.

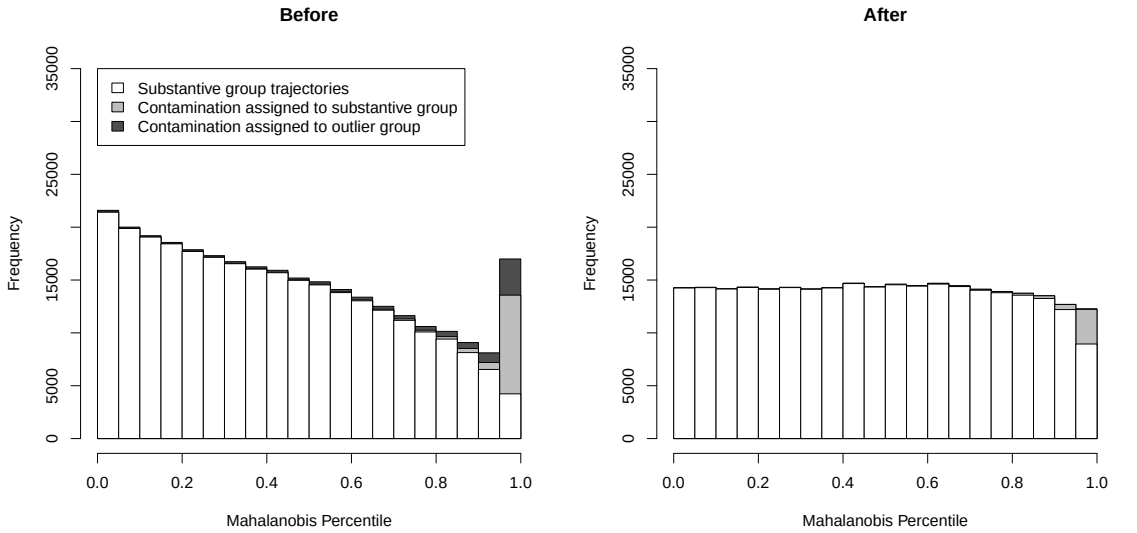


Figure A.1.3: Mahalanobis distance percentiles of trajectories from their assigned group mean, for model 1 with separation 1 and slope 0.75, before and after outlier removal.

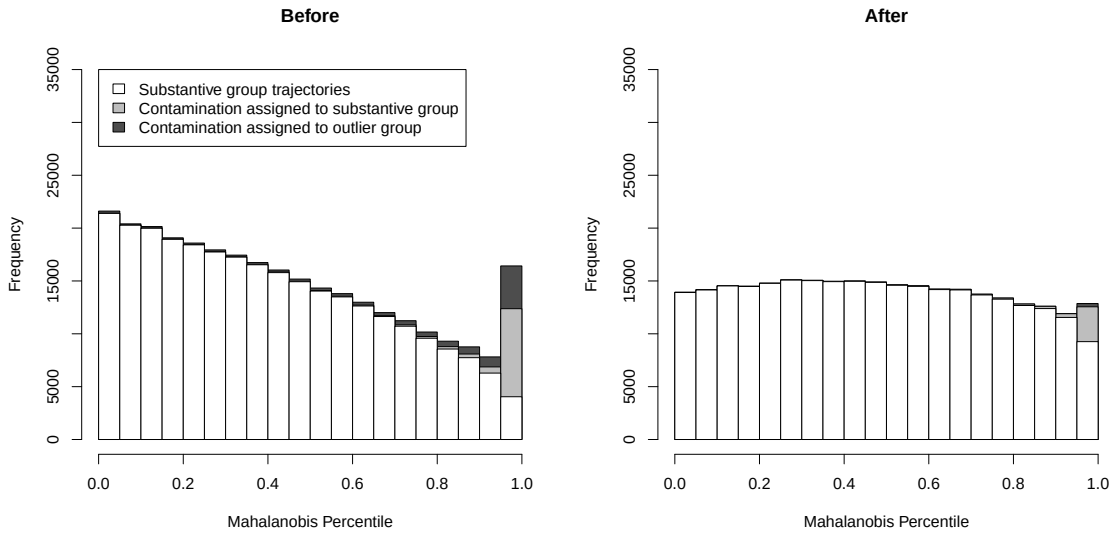


Figure A.1.4: Mahalanobis distance percentiles of trajectories from their assigned group mean, for model 1 with separation 2 and slope 0, before and after outlier removal.

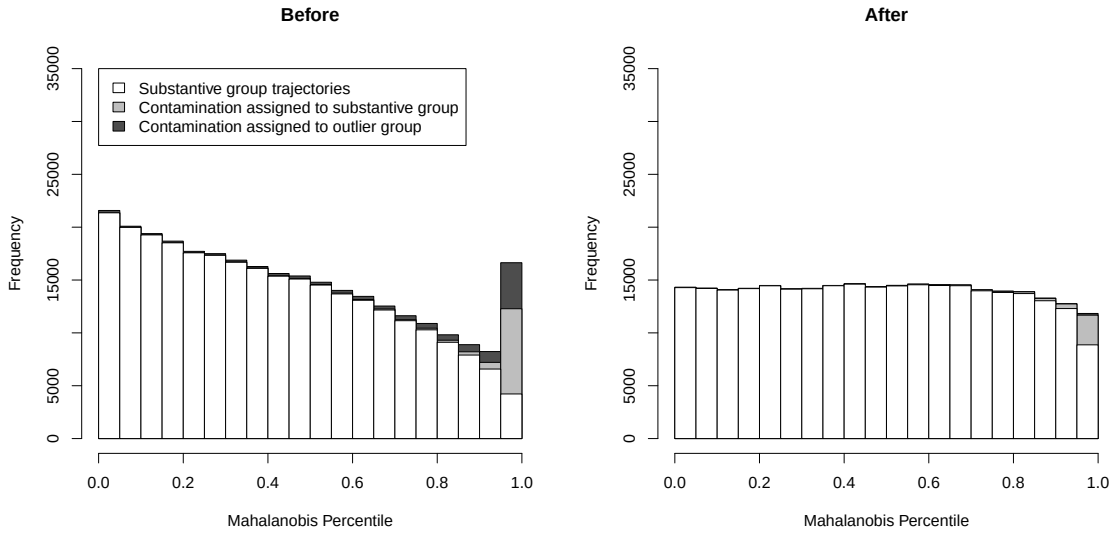


Figure A.1.5: Mahalanobis distance percentiles of trajectories from their assigned group mean, for model 1 with separation 2 and slope 0.25, before and after outlier removal.

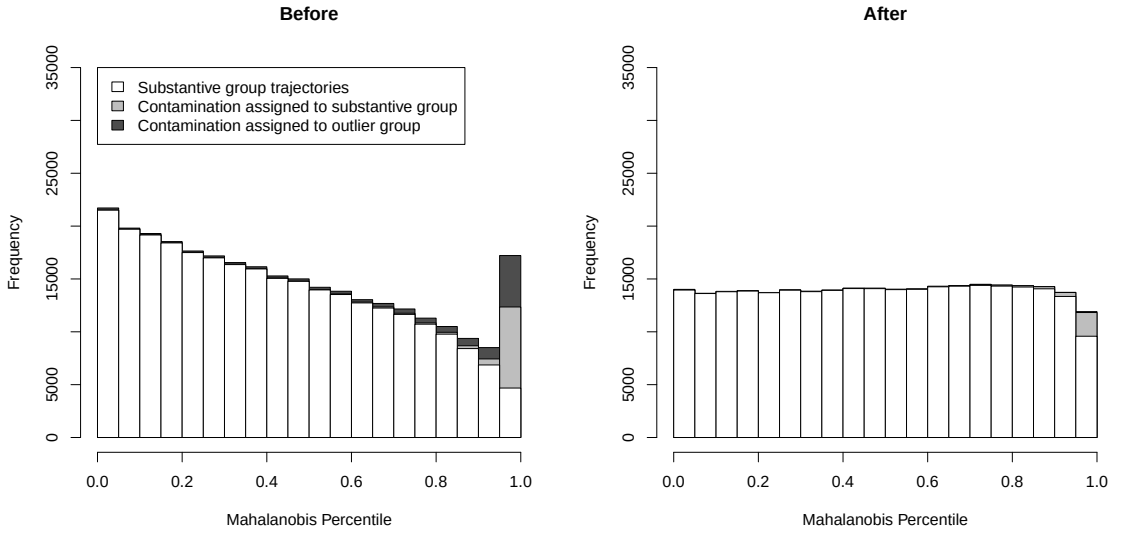


Figure A.1.6: Mahalanobis distance percentiles of trajectories from their assigned group mean, for model 1 with separation 2 and slope 0.5, before and after outlier removal.

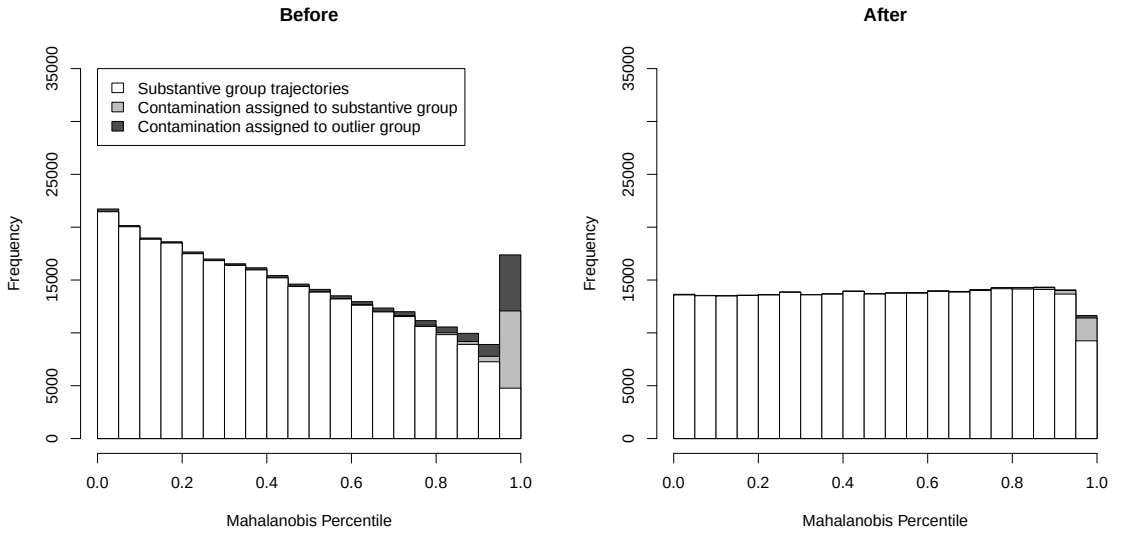


Figure A.1.7: Mahalanobis distance percentiles of trajectories from their assigned group mean, for model 1 with separation 2 and slope 0.75, before and after outlier removal.

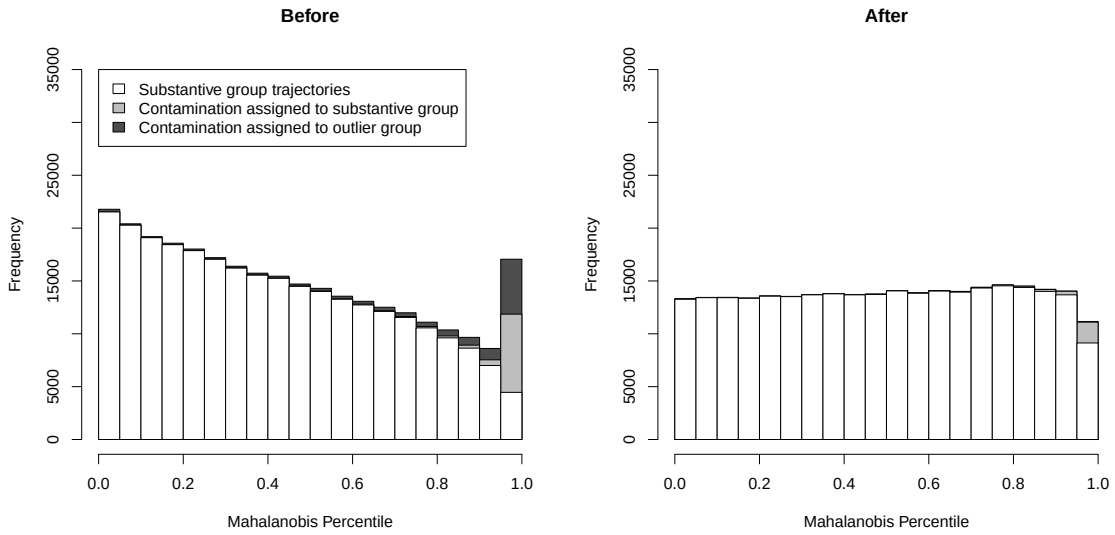


Figure A.1.8: Mahalanobis distance percentiles of trajectories from their assigned group mean, for model 1 with separation 3 and slope 0, before and after outlier removal.

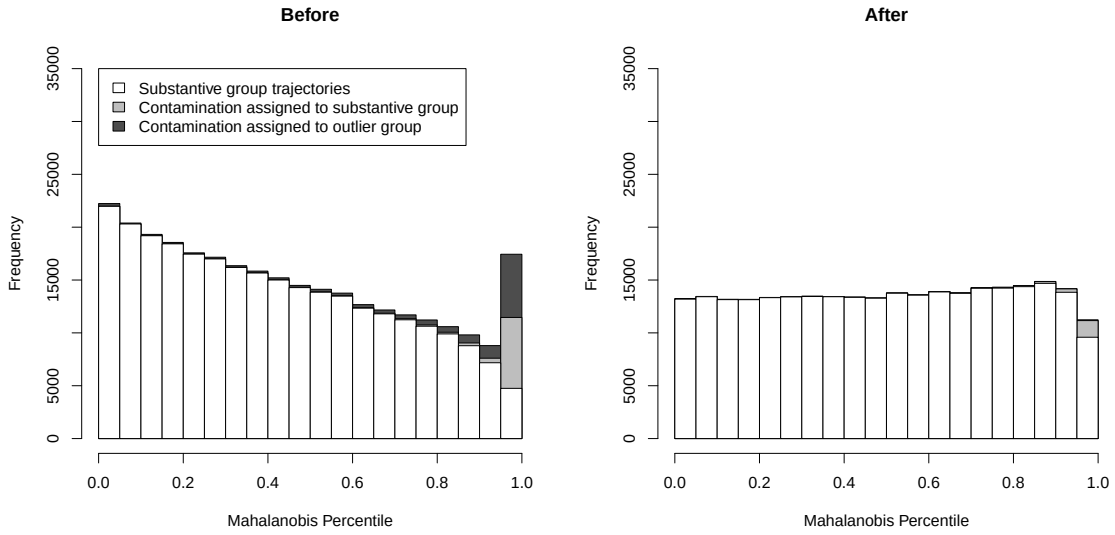


Figure A.1.9: Mahalanobis distance percentiles of trajectories from their assigned group mean, for model 1 with separation 3 and slope 0.25, before and after outlier removal.

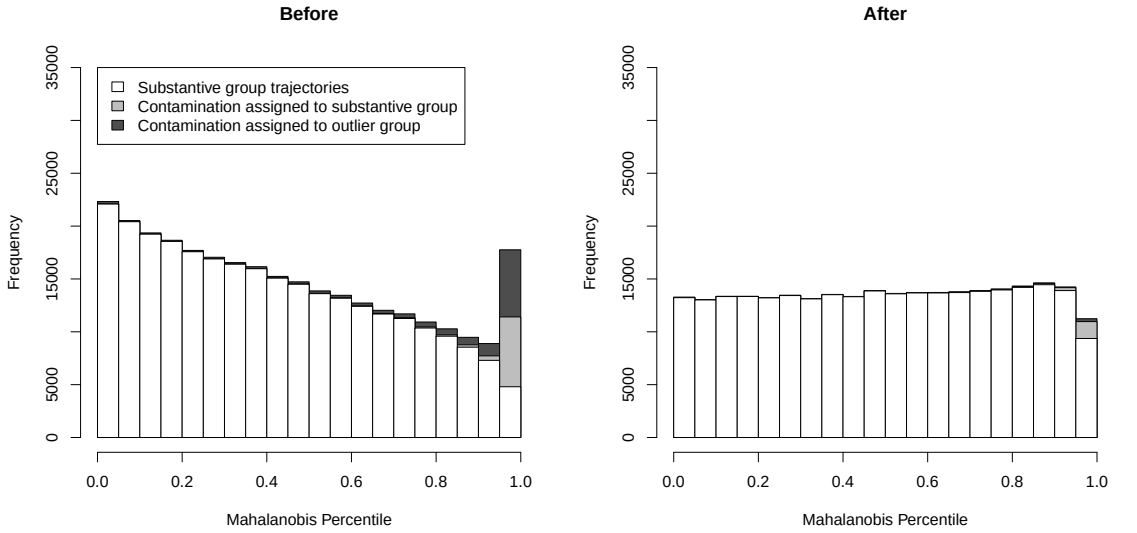


Figure A.1.10: Mahalanobis distance percentiles of trajectories from their assigned group mean, for model 1 with separation 3 and slope 0.5, before and after outlier removal.

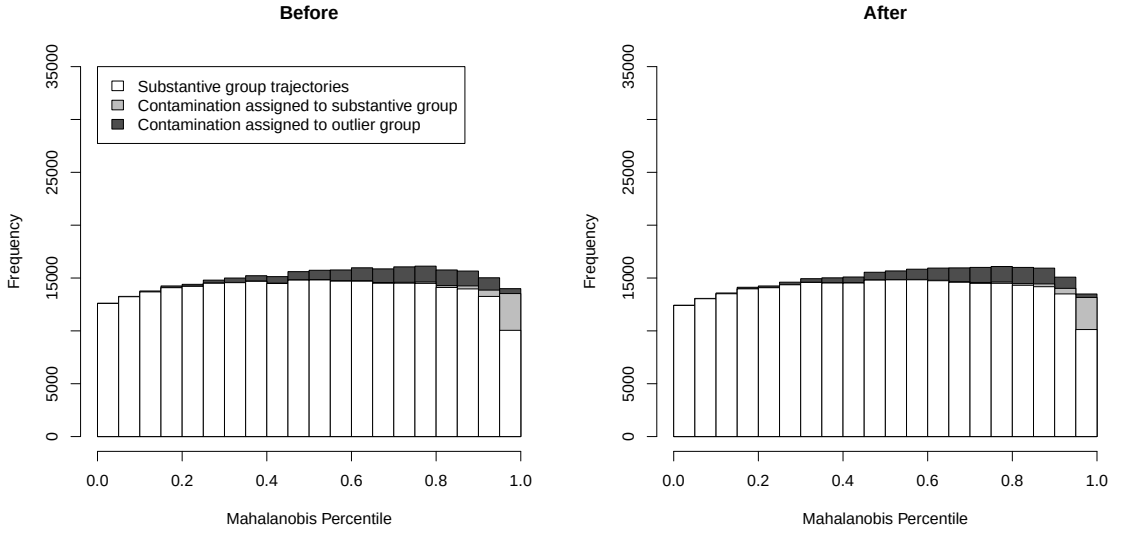


Figure A.1.11: Mahalanobis distance percentiles of trajectories from their assigned group mean, for model 2 with separation 1 and slope 0.25, before and after outlier removal.

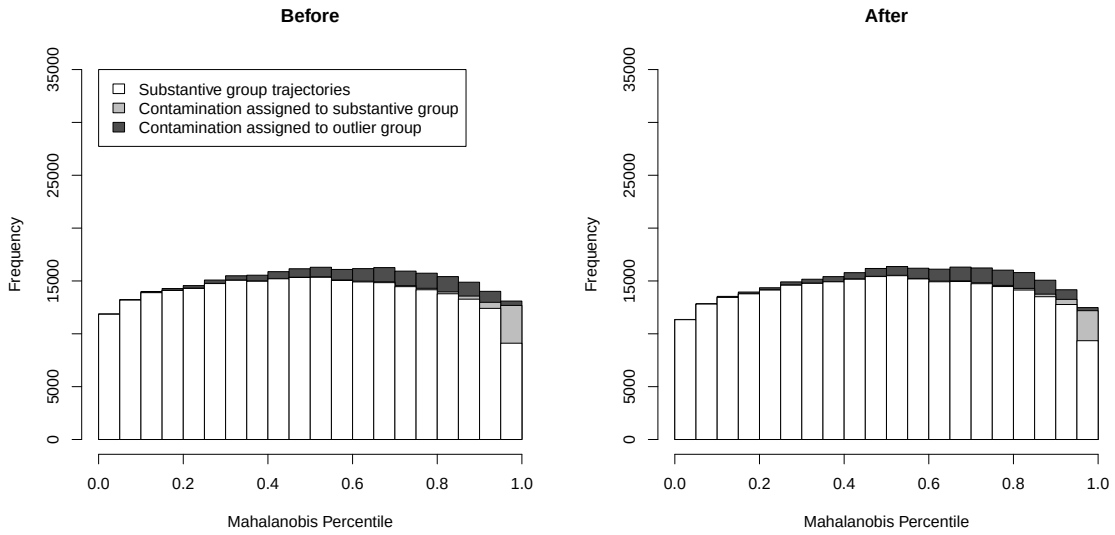


Figure A.1.12: Mahalanobis distance percentiles of trajectories from their assigned group mean, for model 2 with separation 1 and slope 0.5, before and after outlier removal.

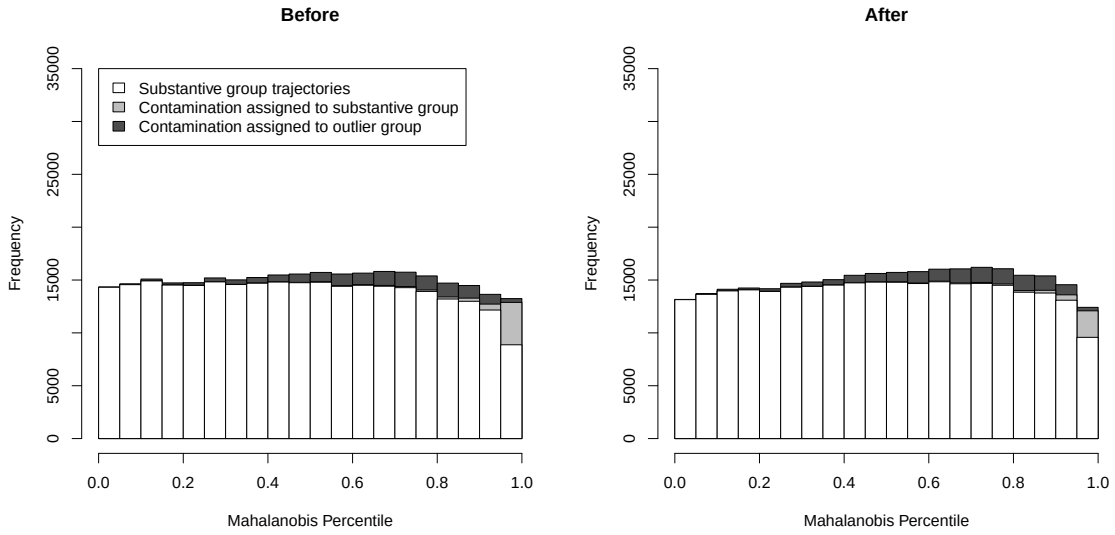


Figure A.1.13: Mahalanobis distance percentiles of trajectories from their assigned group mean, for model 2 with separation 1 and slope 0.75, before and after outlier removal.

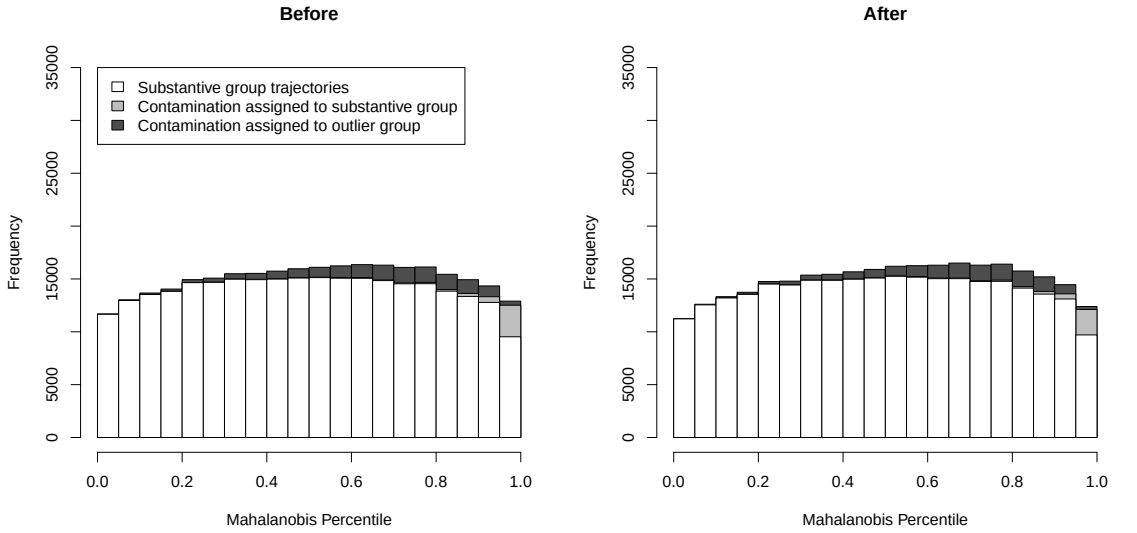


Figure A.1.14: Mahalanobis distance percentiles of trajectories from their assigned group mean, for model 2 with separation 2 and slope 0, before and after outlier removal.

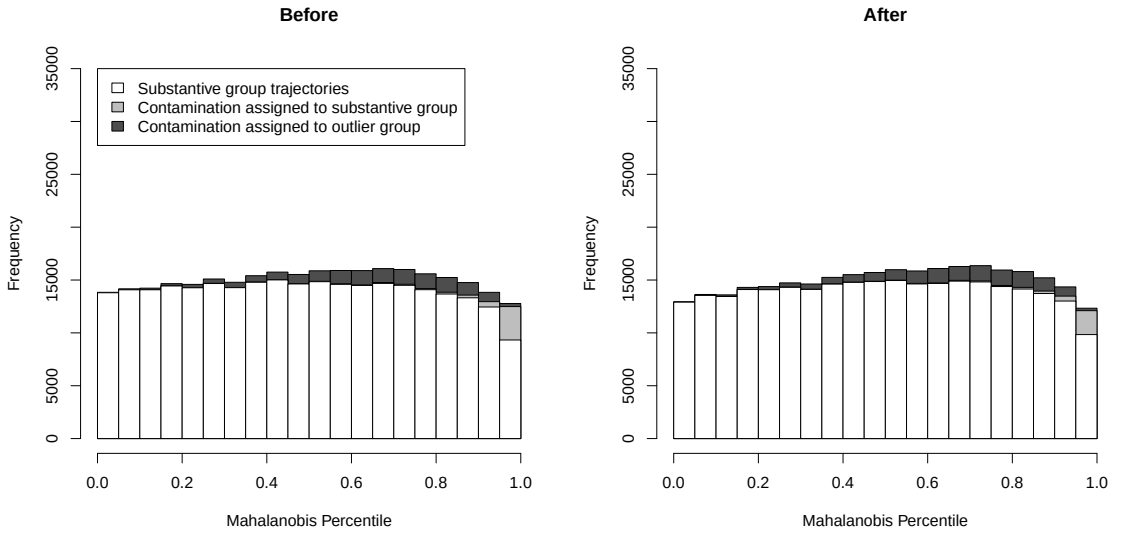


Figure A.1.15: Mahalanobis distance percentiles of trajectories from their assigned group mean, for model 2 with separation 2 and slope 0.25, before and after outlier removal.

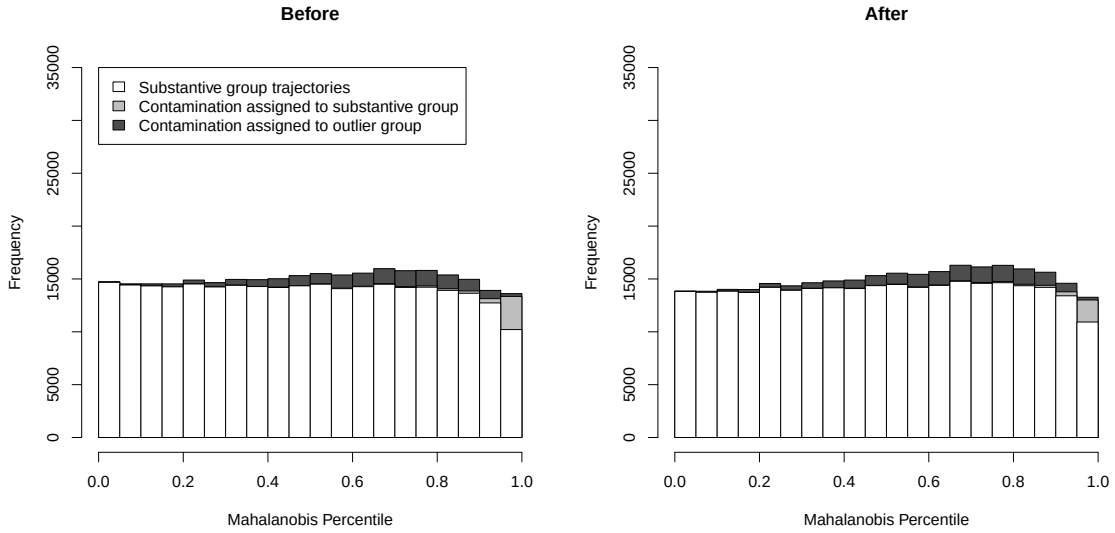


Figure A.1.16: Mahalanobis distance percentiles of trajectories from their assigned group mean, for model 2 with separation 2 and slope 0.5, before and after outlier removal.

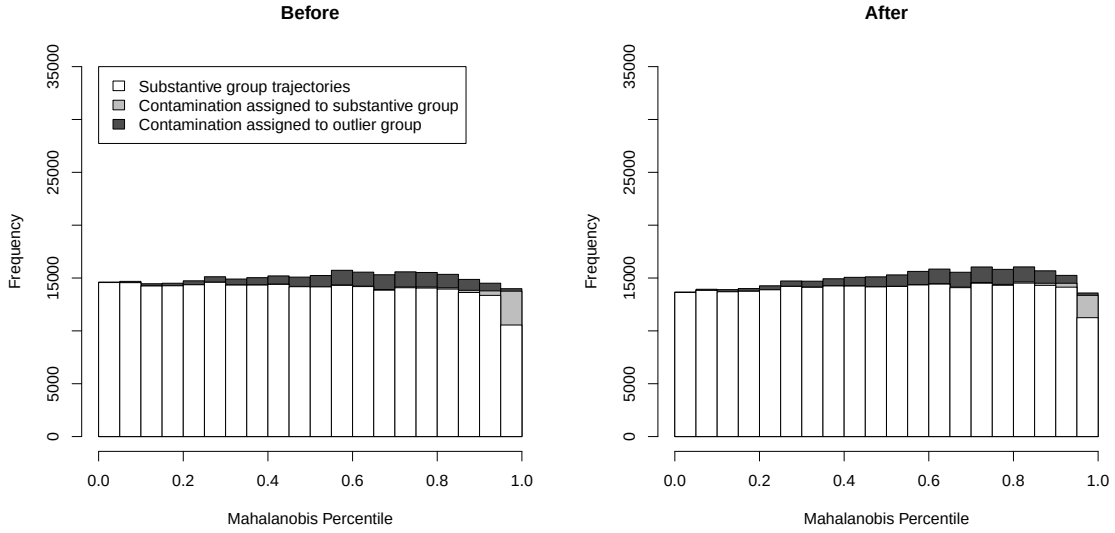


Figure A.1.17: Mahalanobis distance percentiles of trajectories from their assigned group mean, for model 2 with separation 2 and slope 0.75, before and after outlier removal.

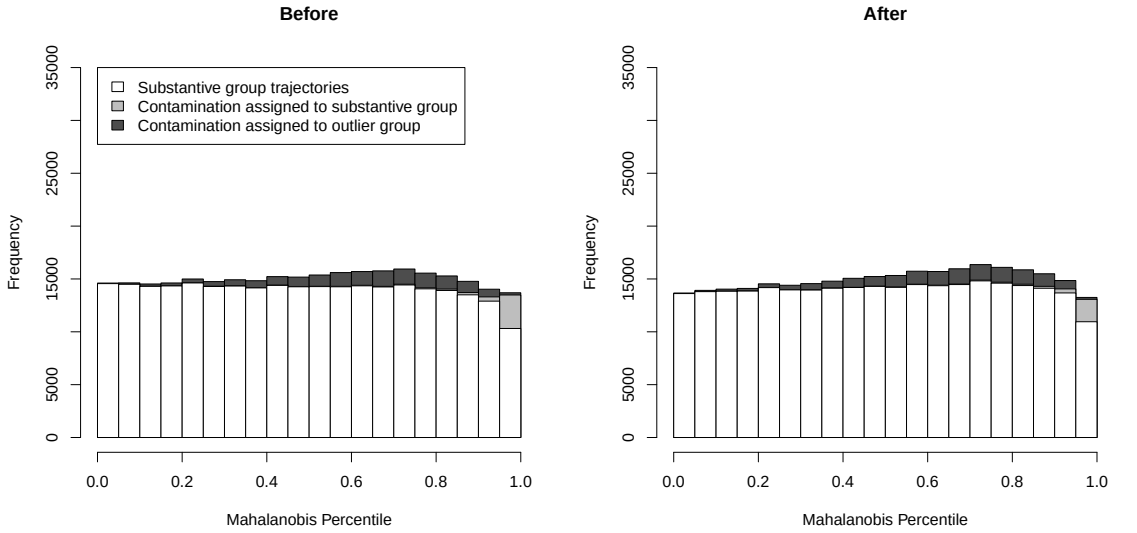


Figure A.1.18: Mahalanobis distance percentiles of trajectories from their assigned group mean, for model 2 with separation 3 and slope 0, before and after outlier removal.

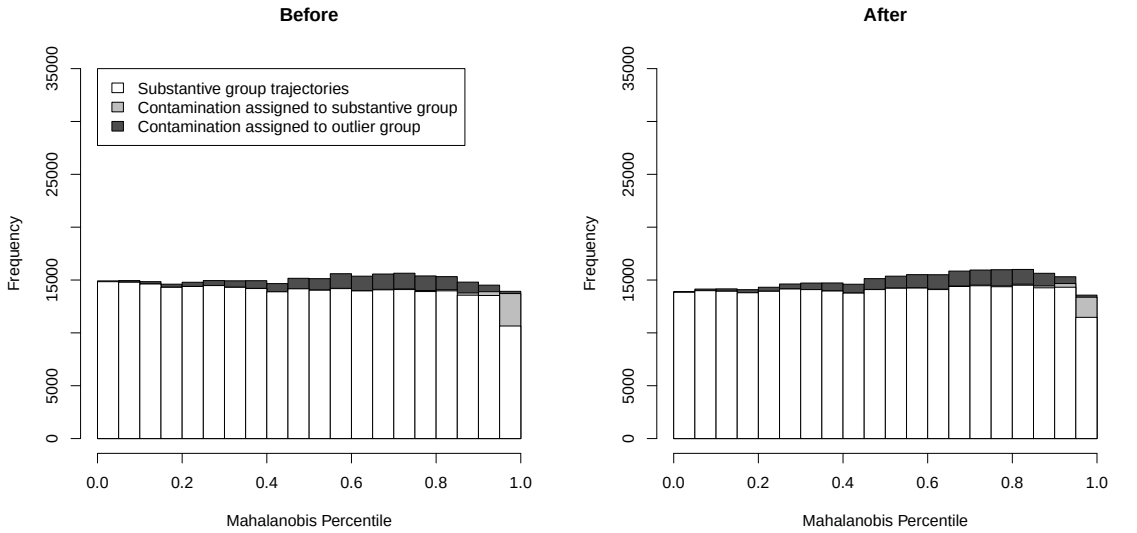


Figure A.1.19: Mahalanobis distance percentiles of trajectories from their assigned group mean, for model 2 with separation 3 and slope 0.25, before and after outlier removal.

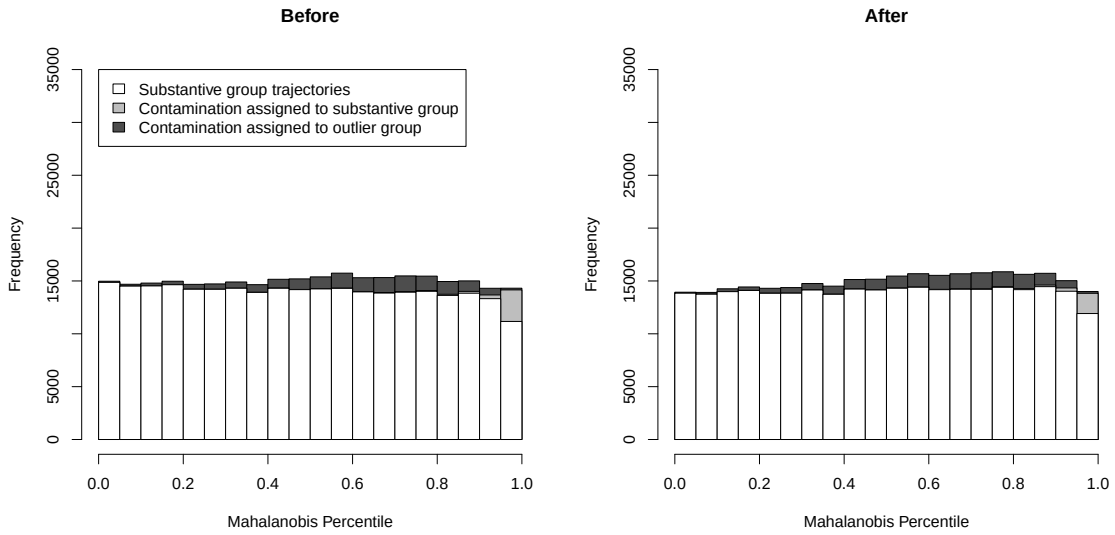


Figure A.1.20: Mahalanobis distance percentiles of trajectories from their assigned group mean, for model 2 with separation 3 and slope 0.5, before and after outlier removal.

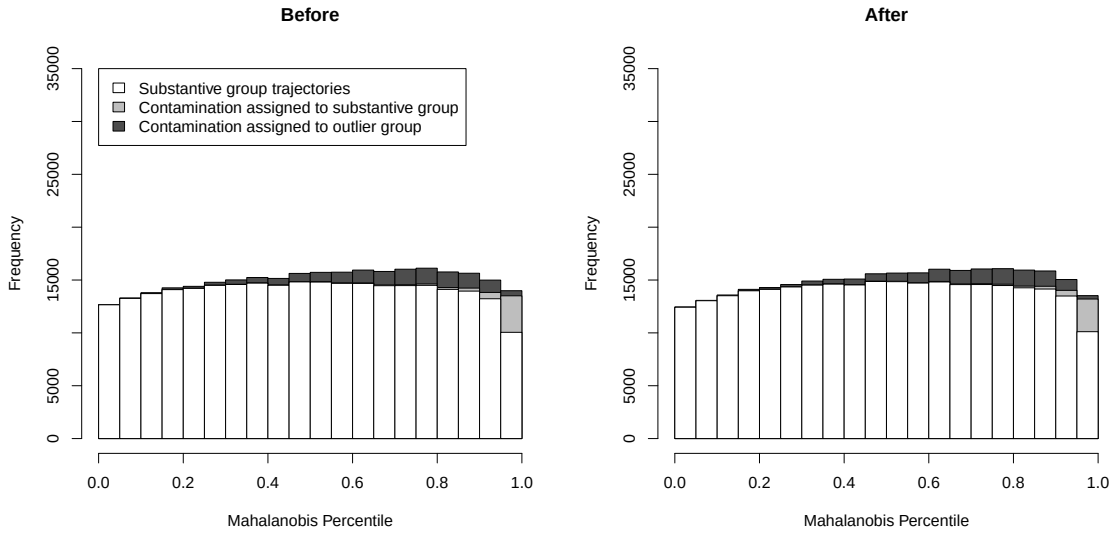


Figure A.1.21: Mahalanobis distance percentiles of trajectories from their assigned group mean, for model 3 with separation 1 and slope 0.25, before and after outlier removal.

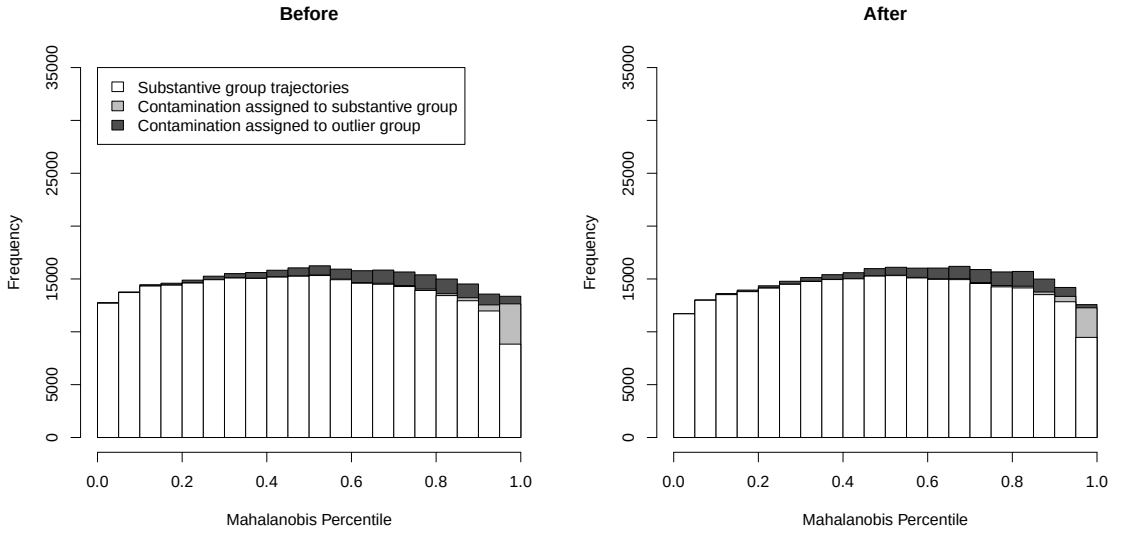


Figure A.1.22: Mahalanobis distance percentiles of trajectories from their assigned group mean, for model 3 with separation 1 and slope 0.5, before and after outlier removal.

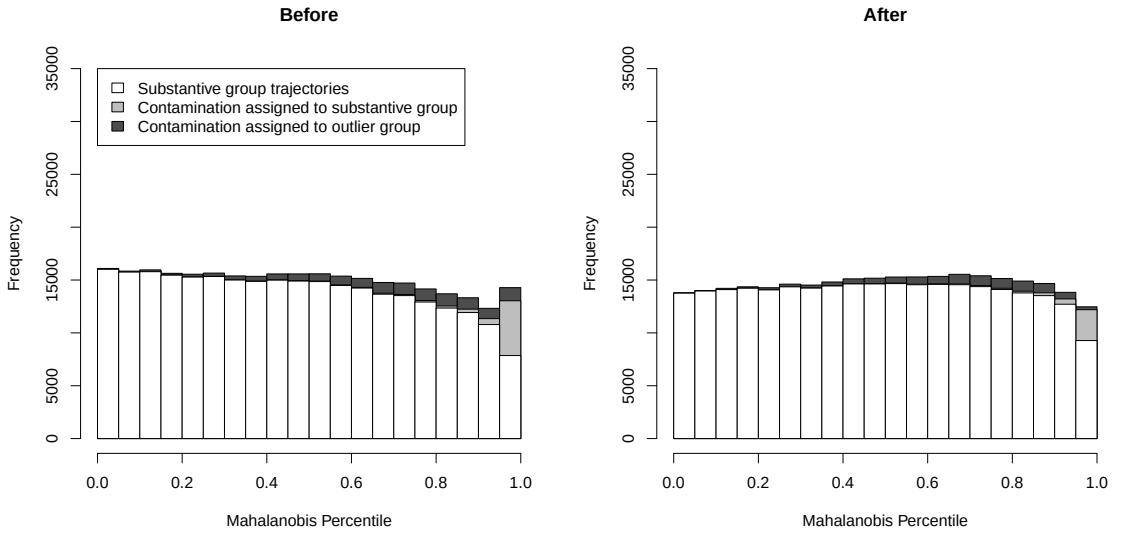


Figure A.1.23: Mahalanobis distance percentiles of trajectories from their assigned group mean, for model 3 with separation 1 and slope 0.75, before and after outlier removal.

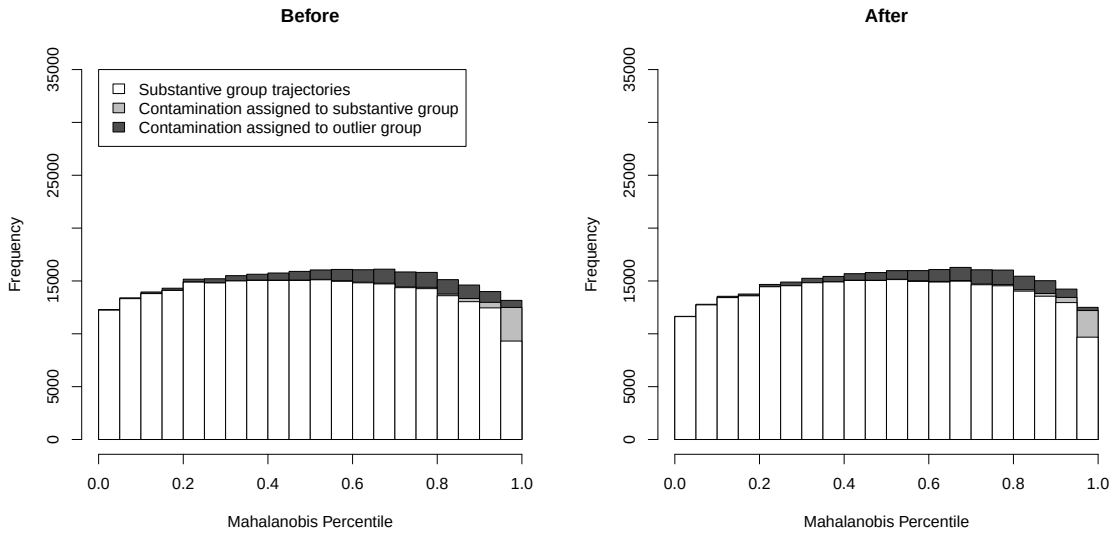


Figure A.1.24: Mahalanobis distance percentiles of trajectories from their assigned group mean, for model 3 with separation 2 and slope 0, before and after outlier removal.

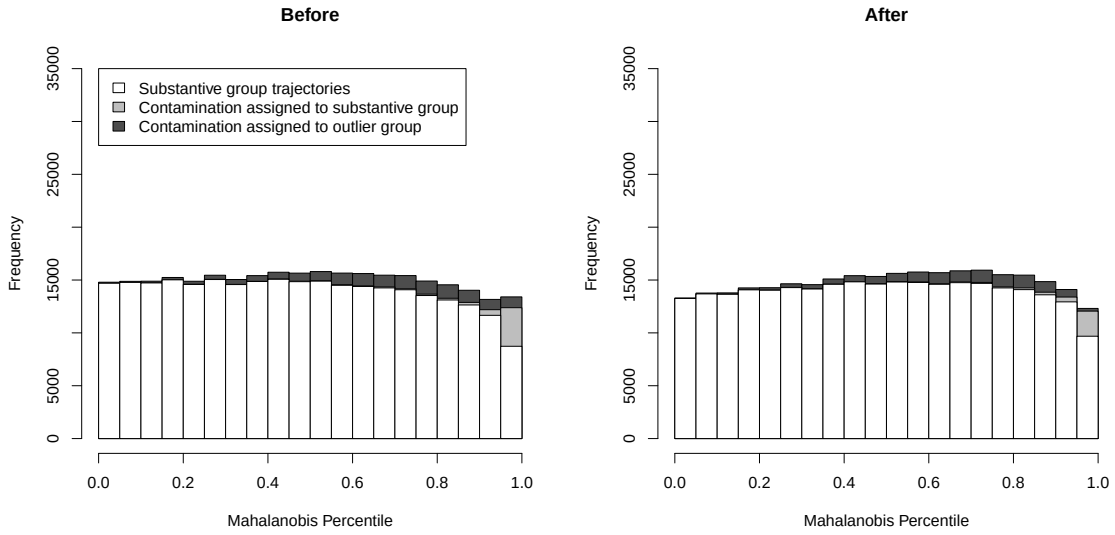


Figure A.1.25: Mahalanobis distance percentiles of trajectories from their assigned group mean, for model 3 with separation 2 and slope 0.25, before and after outlier removal.

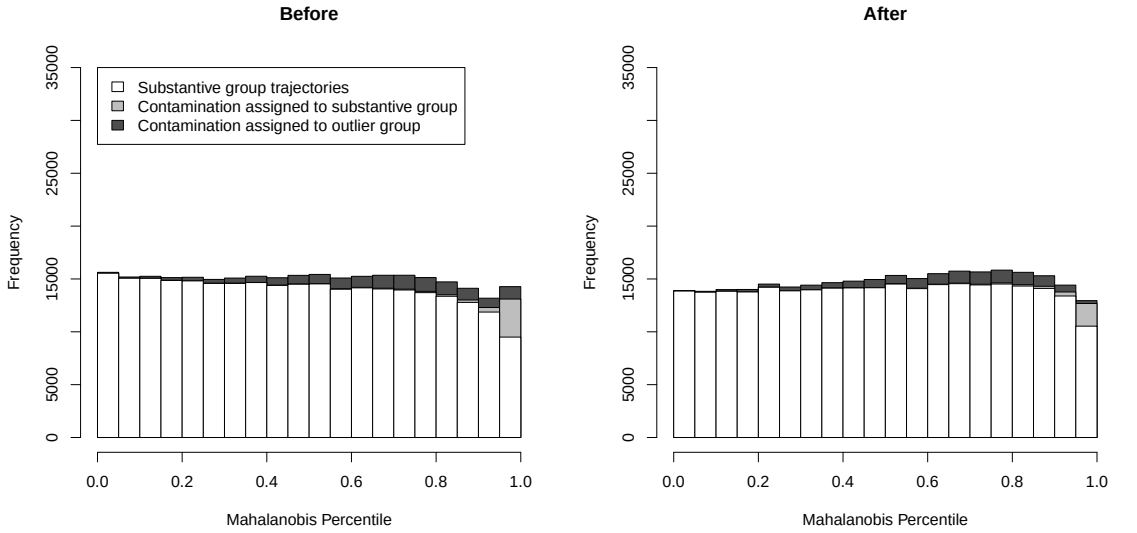


Figure A.1.26: Mahalanobis distance percentiles of trajectories from their assigned group mean, for model 3 with separation 2 and slope 0.5, before and after outlier removal.

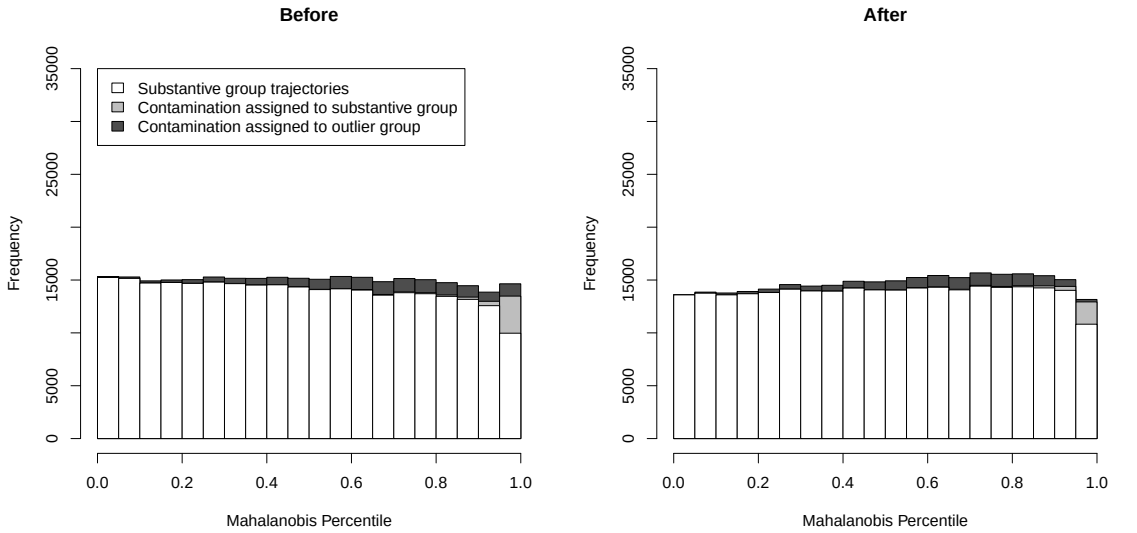


Figure A.1.27: Mahalanobis distance percentiles of trajectories from their assigned group mean, for model 3 with separation 2 and slope 0.75, before and after outlier removal.

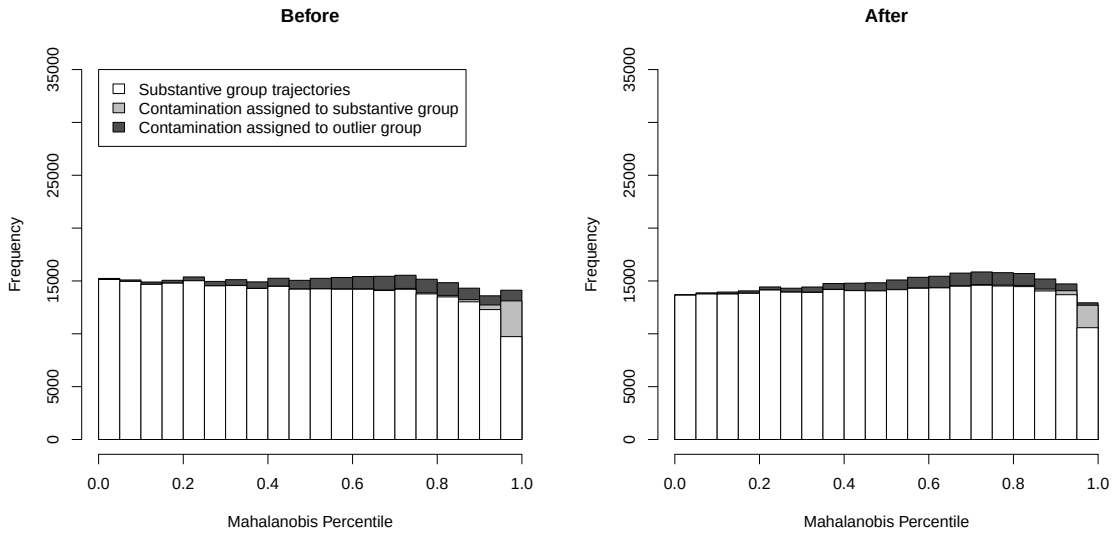


Figure A.1.28: Mahalanobis distance percentiles of trajectories from their assigned group mean, for model 3 with separation 3 and slope 0, before and after outlier removal.

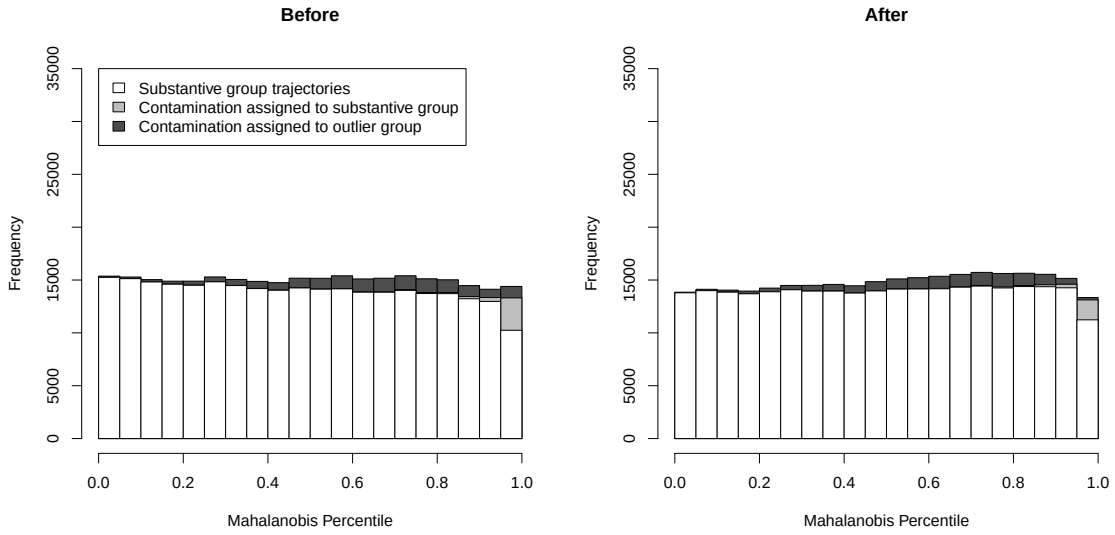


Figure A.1.29: Mahalanobis distance percentiles of trajectories from their assigned group mean, for model 3 with separation 3 and slope 0.25, before and after outlier removal.

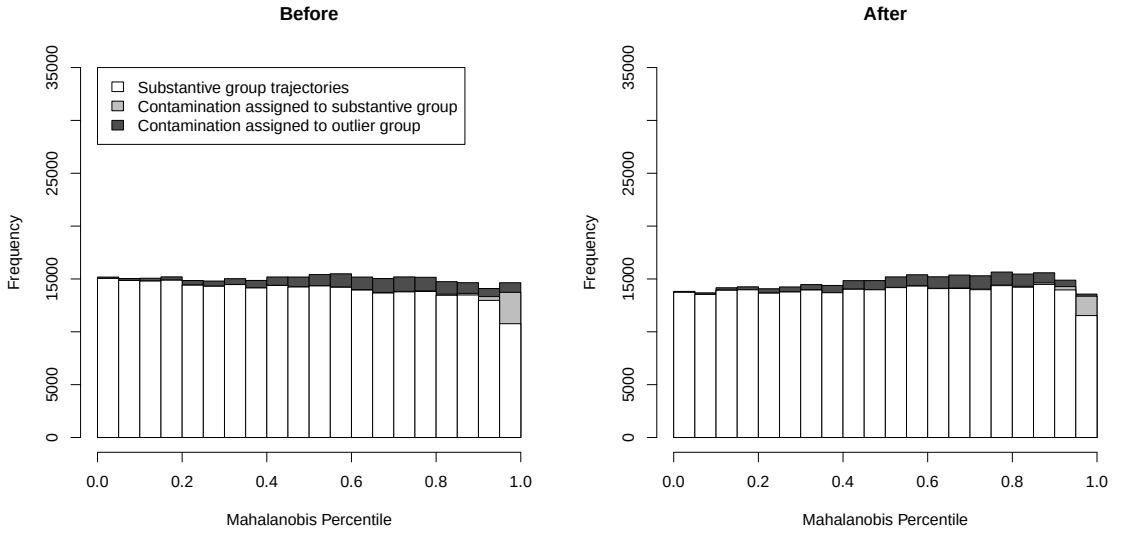


Figure A.1.30: Mahalanobis distance percentiles of trajectories from their assigned group mean, for model 3 with separation 3 and slope 0.5, before and after outlier removal.

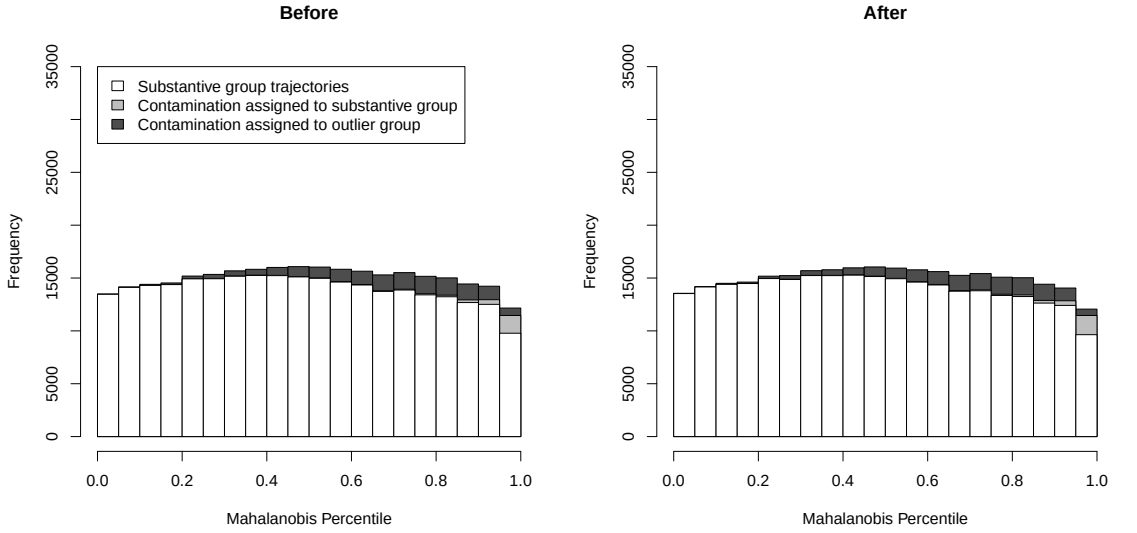


Figure A.1.31: Mahalanobis distance percentiles of trajectories from their assigned group mean, for model 4 with separation 1 and slope 0.25, before and after outlier removal.

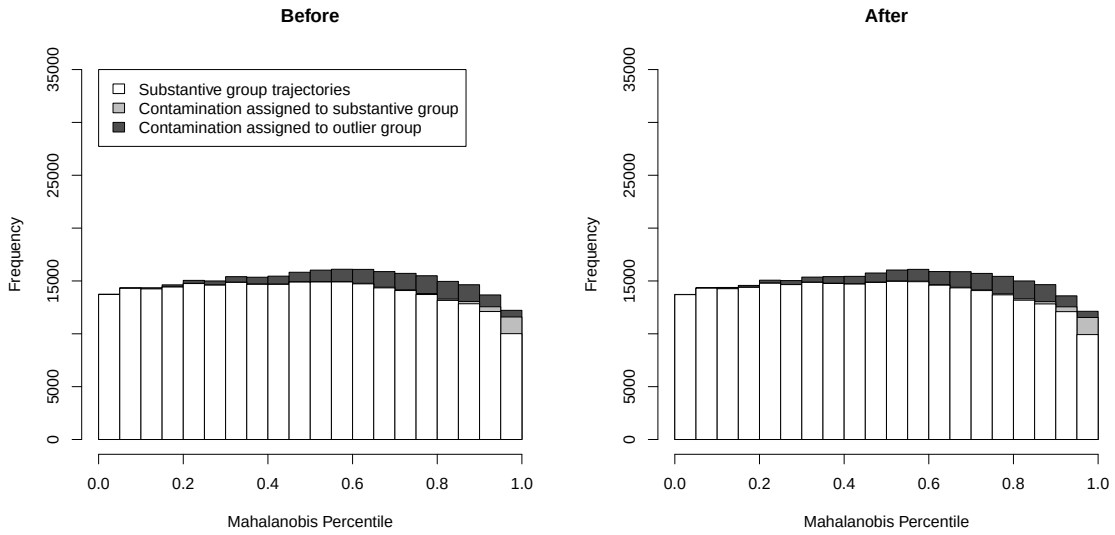


Figure A.1.32: Mahalanobis distance percentiles of trajectories from their assigned group mean, for model 4 with separation 1 and slope 0.5, before and after outlier removal.

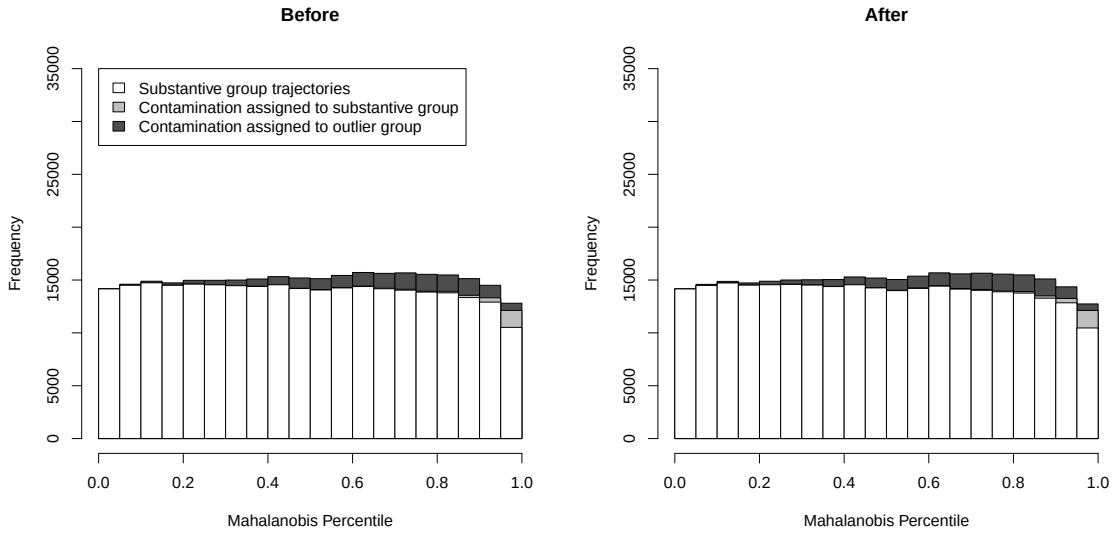


Figure A.1.33: Mahalanobis distance percentiles of trajectories from their assigned group mean, for model 4 with separation 1 and slope 0.75, before and after outlier removal.

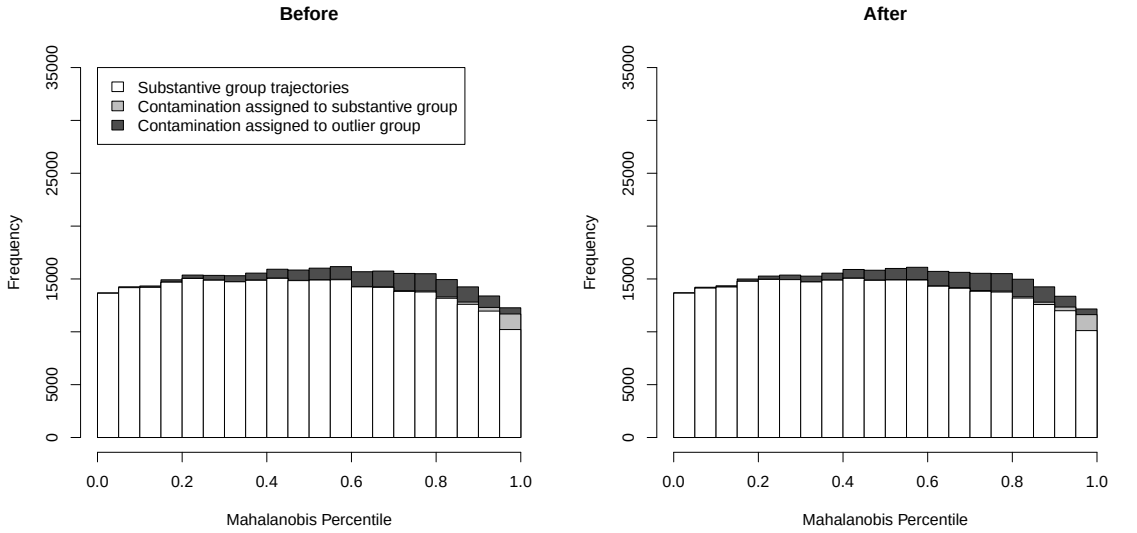


Figure A.1.34: Mahalanobis distance percentiles of trajectories from their assigned group mean, for model 4 with separation 2 and slope 0, before and after outlier removal.

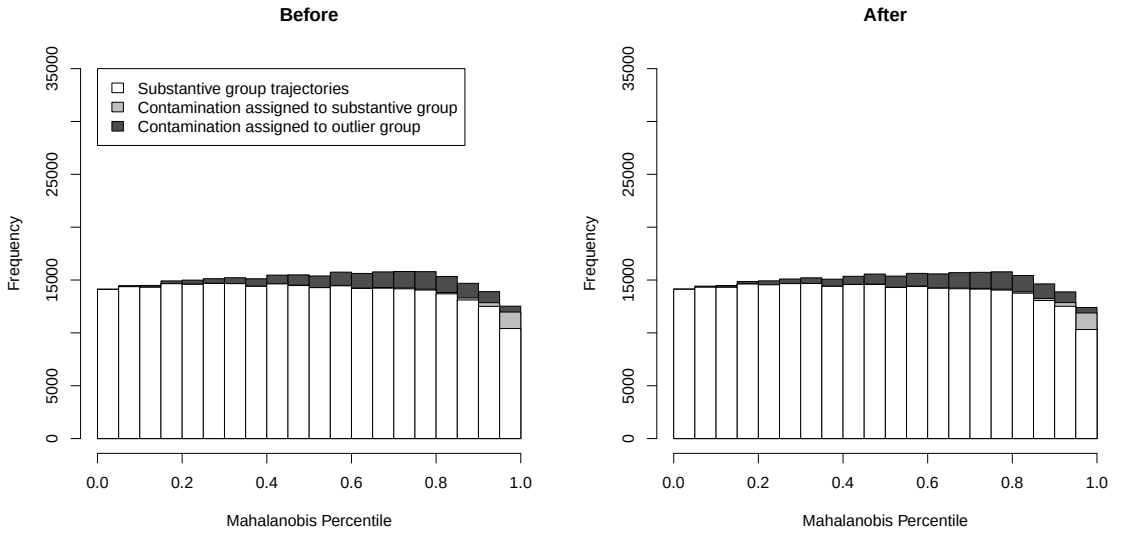


Figure A.1.35: Mahalanobis distance percentiles of trajectories from their assigned group mean, for model 4 with separation 2 and slope 0.25, before and after outlier removal.

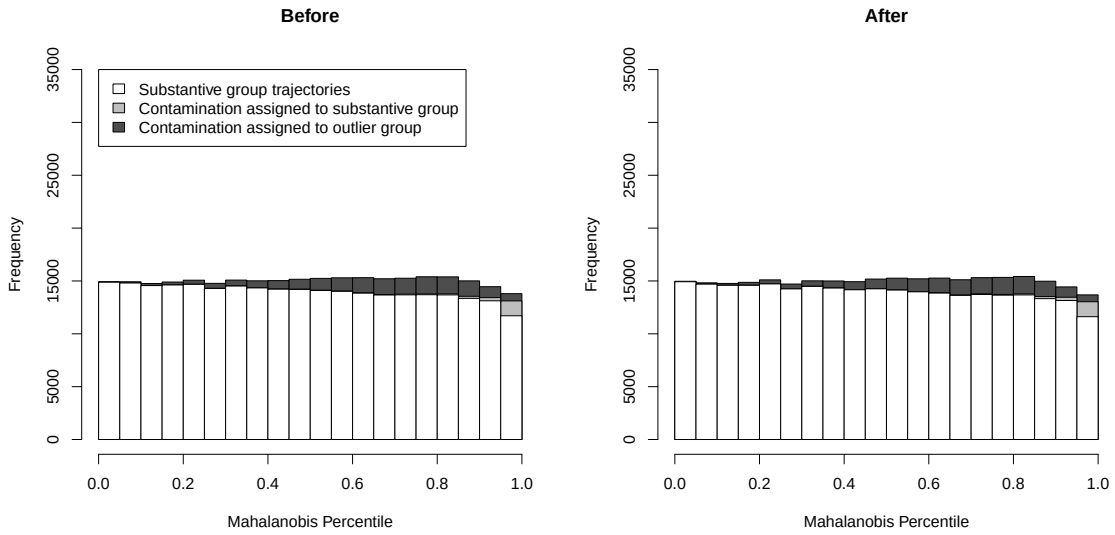


Figure A.1.36: Mahalanobis distance percentiles of trajectories from their assigned group mean, for model 4 with separation 2 and slope 0.5, before and after outlier removal.

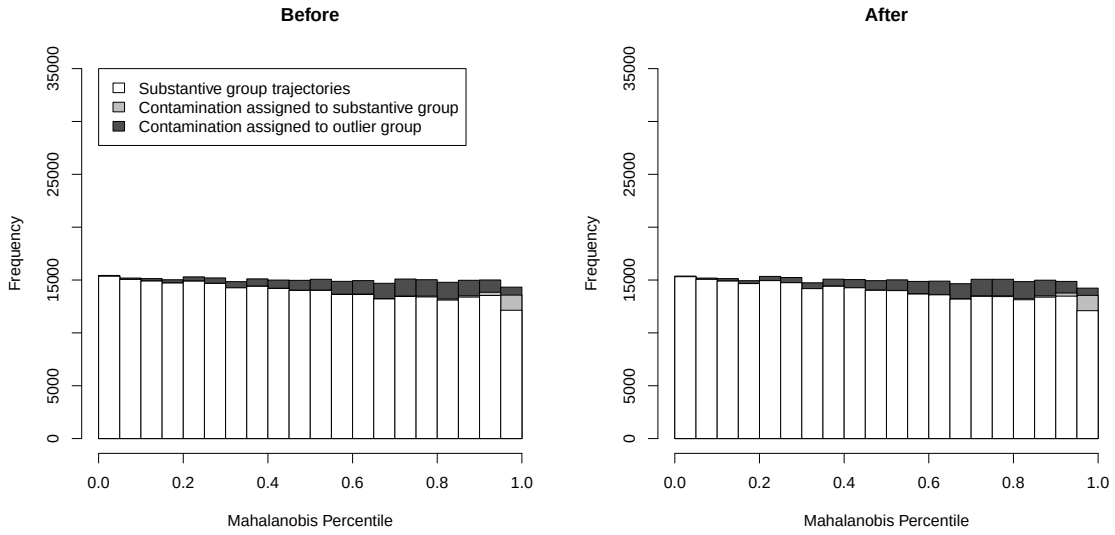


Figure A.1.37: Mahalanobis distance percentiles of trajectories from their assigned group mean, for model 4 with separation 2 and slope 0.75, before and after outlier removal.

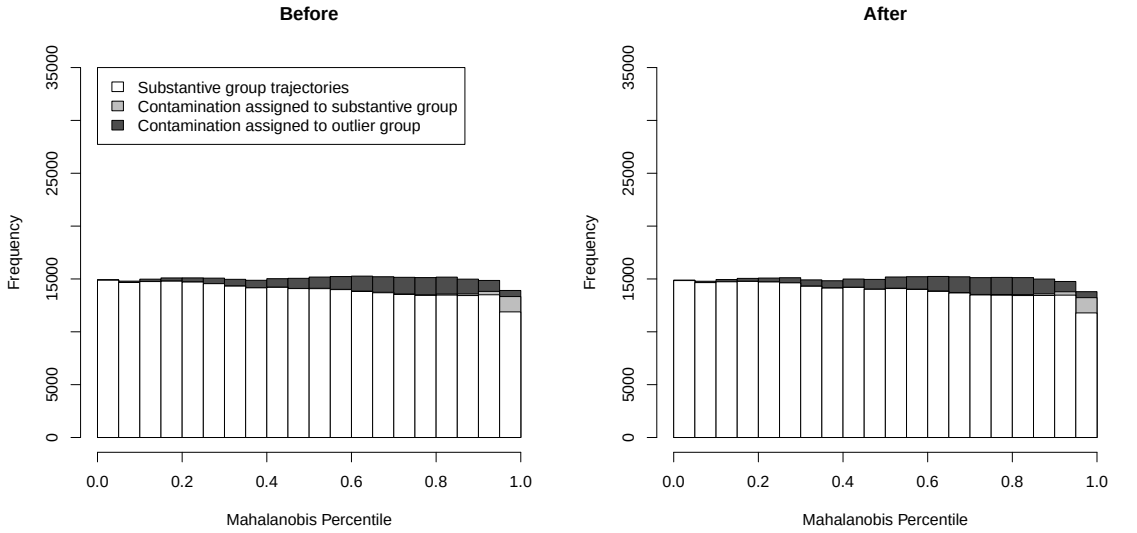


Figure A.1.38: Mahalanobis distance percentiles of trajectories from their assigned group mean, for model 4 with separation 3 and slope 0, before and after outlier removal.

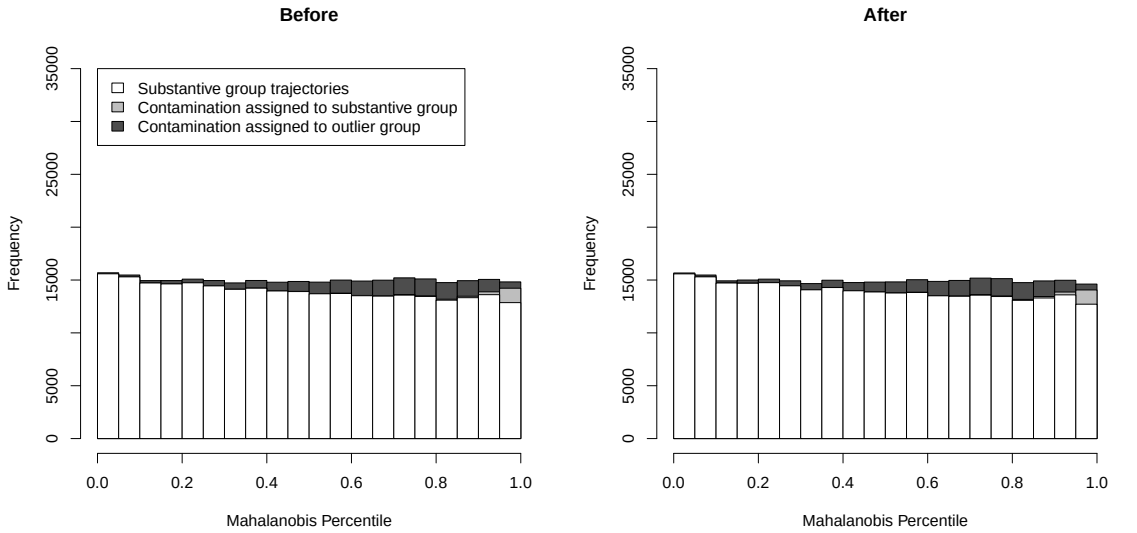


Figure A.1.39: Mahalanobis distance percentiles of trajectories from their assigned group mean, for model 4 with separation 3 and slope 0.25, before and after outlier removal.

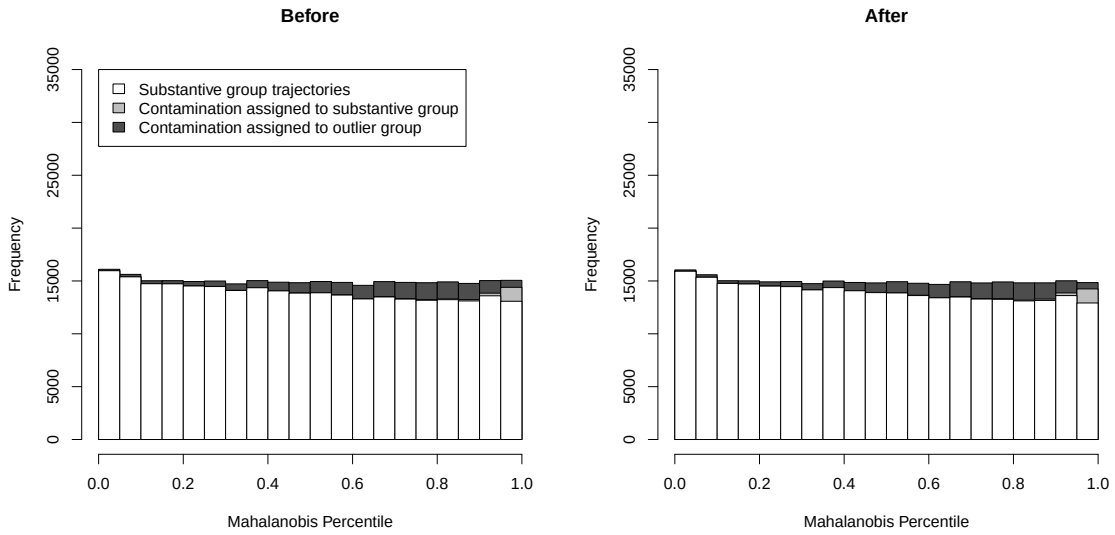


Figure A.1.40: Mahalanobis distance percentiles of trajectories from their assigned group mean, for model 4 with separation 3 and slope 0.5, before and after outlier removal.

A.2 Mahalanobis distance percentile plots for crossed trajectories simulations

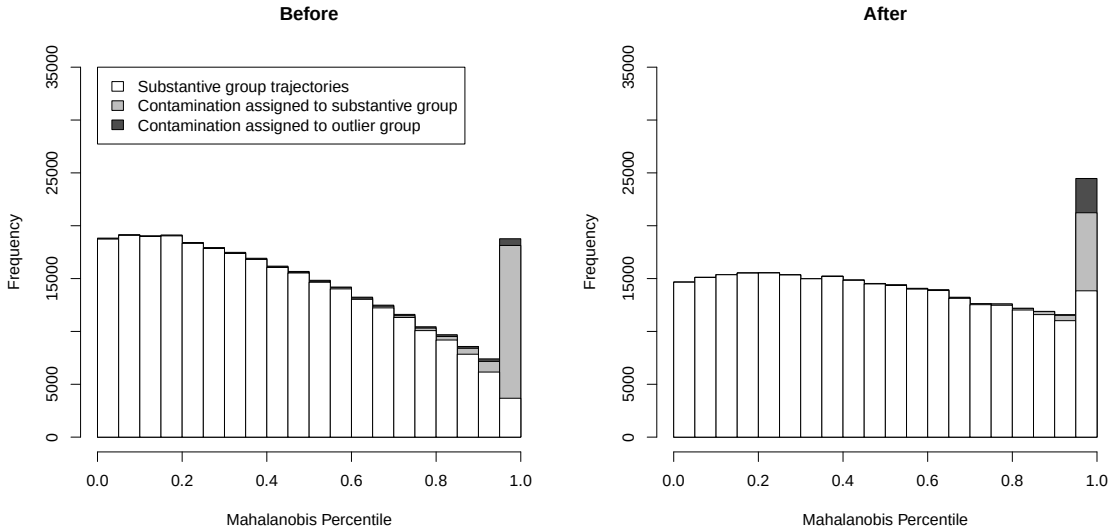


Figure A.2.1: Mahalanobis distance percentiles of trajectories from their assigned group mean, for model 1 with slope 0.5, before and after outlier removal.

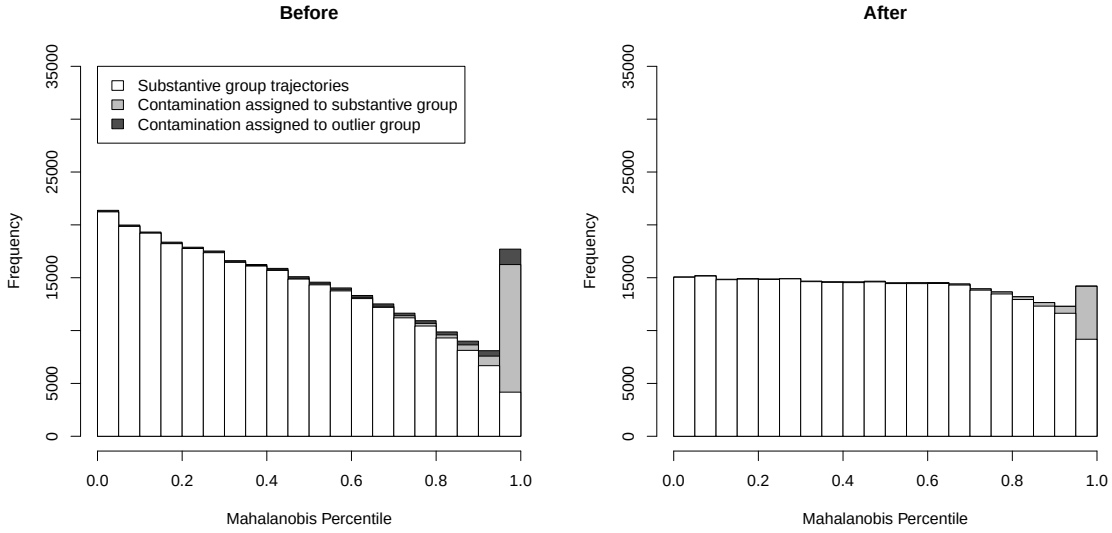


Figure A.2.2: Mahalanobis distance percentiles of trajectories from their assigned group mean, for model 1 with slope 0.75, before and after outlier removal.

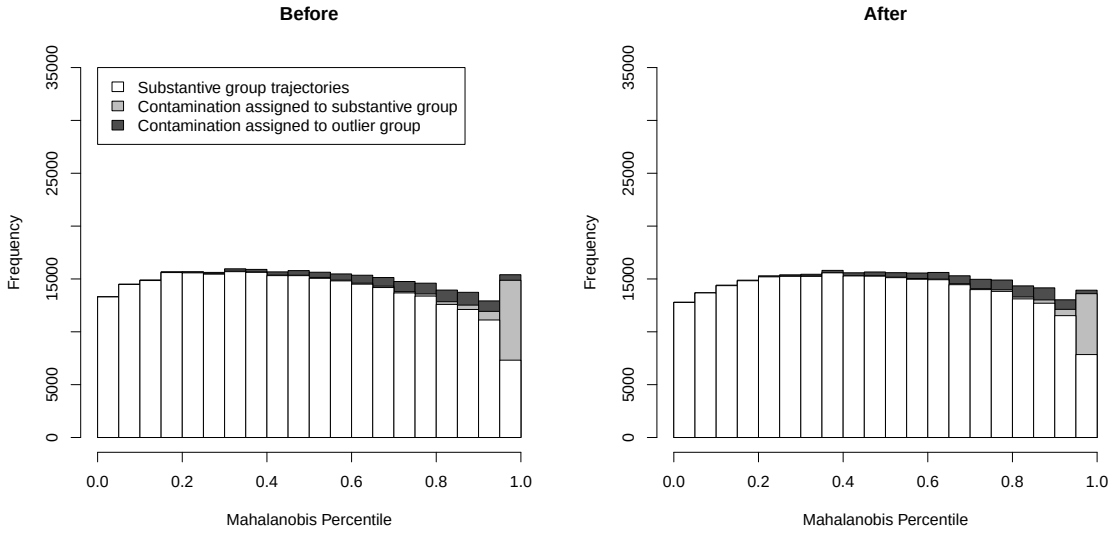


Figure A.2.3: Mahalanobis distance percentiles of trajectories from their assigned group mean, for model 2 with slope 0.5, before and after outlier removal.

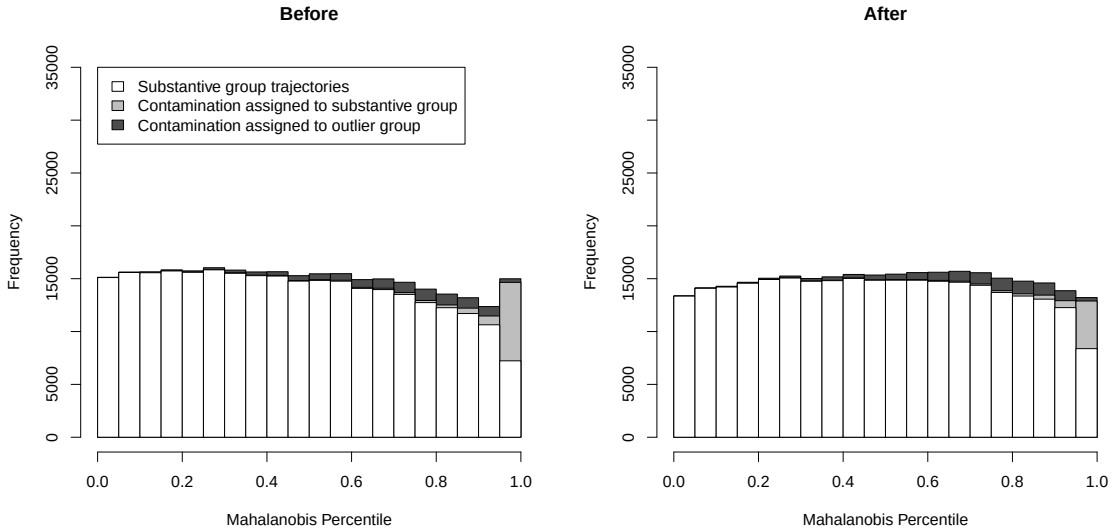


Figure A.2.4: Mahalanobis distance percentiles of trajectories from their assigned group mean, for model 2 with slope 0.75, before and after outlier removal.

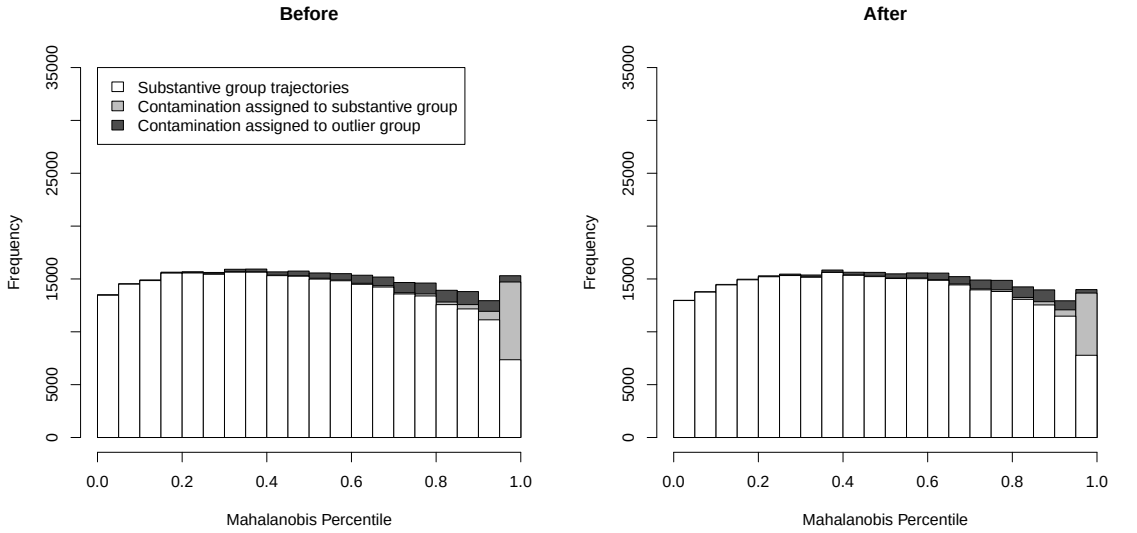


Figure A.2.5: Mahalanobis distance percentiles of trajectories from their assigned group mean, for model 3 with slope 0.5, before and after outlier removal.

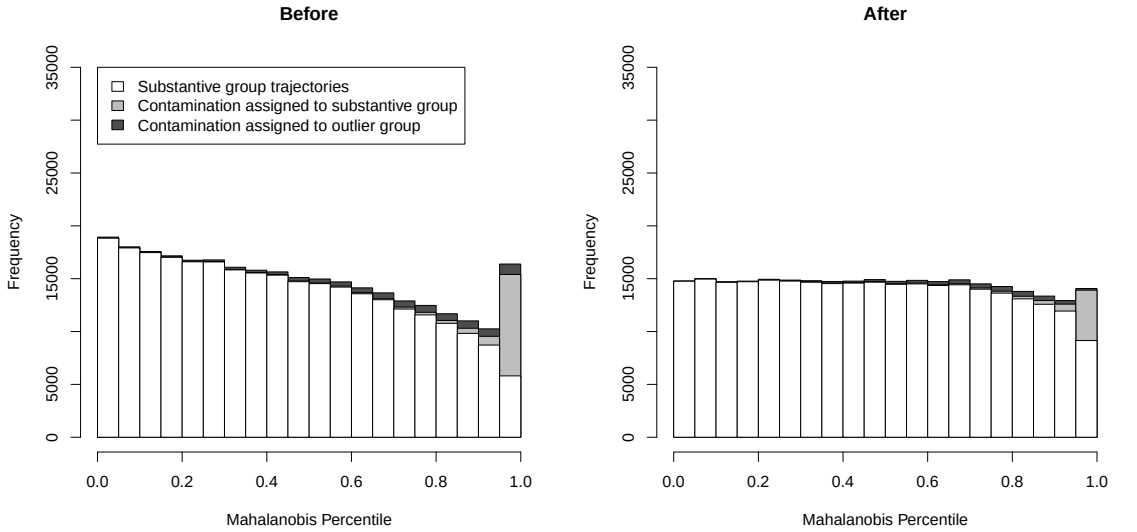


Figure A.2.6: Mahalanobis distance percentiles of trajectories from their assigned group mean, for model 3 with slope 0.75, before and after outlier removal.

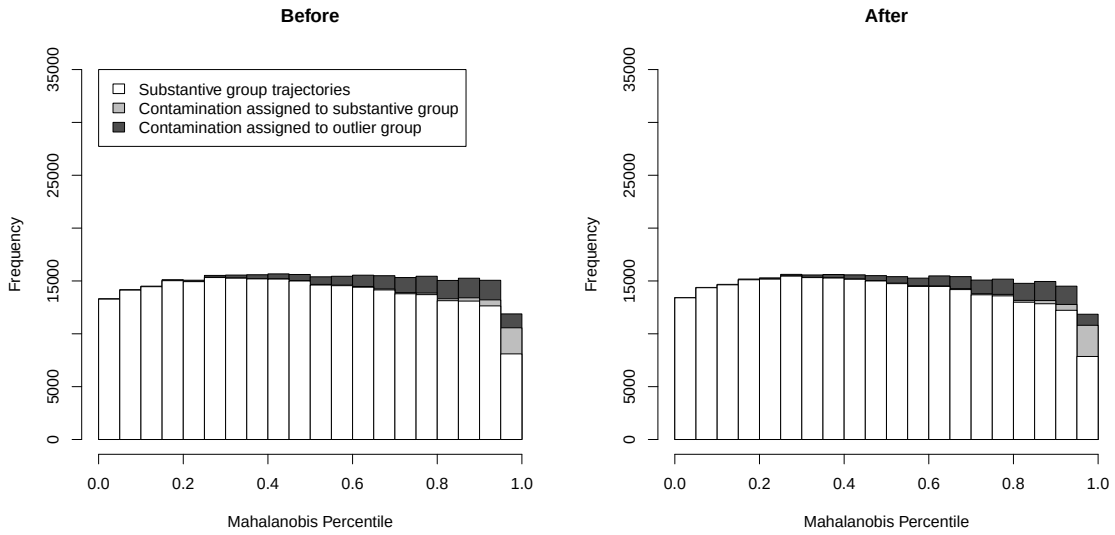


Figure A.2.7: Mahalanobis distance percentiles of trajectories from their assigned group mean, for model 4 with slope 0.5, before and after outlier removal.

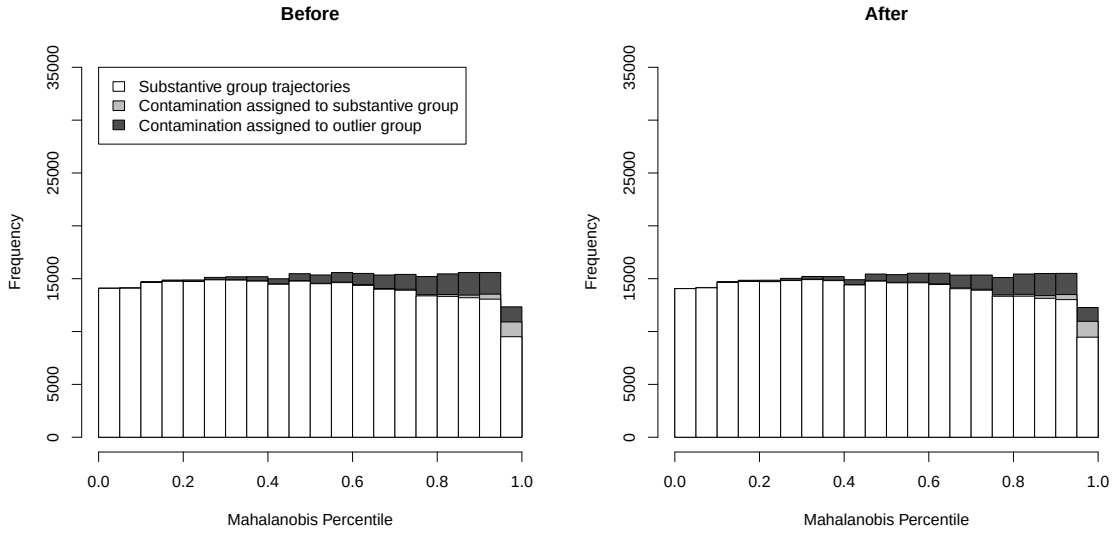


Figure A.2.8: Mahalanobis distance percentiles of trajectories from their assigned group mean, for model 4 with slope 0.75, before and after outlier removal.

Appendix B

B.1 Supplementary Table 1

Supplementary Table 1: Average estimates (Est), standard errors (SE), standardized bias (Bias) and coverage probability of the 95% confidence interval (CP) in covariate estimates for a single rare binary covariate. Results for simulations with OR of 4 are not presented for reasons described in Section 3.4.

		Separation																			
		6				8				10				12				14			
Method	OR	Est	SE	Bias	CP	Est	SE	Bias	CP	Est	SE	Bias	CP	Est	SE	Bias	CP	Est	SE	Bias	CP
1S	0.25	-1.39	0.40	0.02	0.97	-1.36	0.37	0.11	0.98	-1.37	0.35	0.06	0.96	-1.40	0.33	0.00	0.94	-1.42	0.32	-0.08	0.96
	0.5	-0.69	0.39	0.01	0.98	-0.69	0.37	-0.02	0.93	-0.67	0.34	0.04	0.97	-0.74	0.33	-0.14	0.97	-0.67	0.32	0.06	0.97
	1	0.00	0.41	-0.04	0.96	0.04	0.39	0.02	0.95	0.01	0.37	-0.05	0.93	0.00	0.35	-0.05	0.98	-0.03	0.34	-0.12	0.98
	2	0.65	0.48	-0.15	0.98	0.59	0.44	-0.31	0.97	0.64	0.42	-0.22	0.95	0.74	0.41	0.03	0.97	0.72	0.40	-0.01	0.97
	4																				
3S	0.25	-1.08	0.31	1.00	0.82	-1.12	0.31	0.88	0.88	-1.21	0.31	0.60	0.91	-1.29	0.31	0.33	0.92	-1.36	0.31	0.08	0.94
	0.5	-0.56	0.32	0.39	0.97	-0.59	0.32	0.30	0.94	-0.61	0.31	0.26	0.98	-0.69	0.31	0.00	0.98	-0.65	0.31	0.12	0.96
	1	0.00	0.35	-0.05	0.96	0.03	0.34	0.02	0.94	0.02	0.34	-0.02	0.92	0.01	0.34	-0.01	0.98	-0.02	0.33	-0.11	0.97
	2	0.56	0.40	-0.42	0.92	0.53	0.38	-0.49	0.94	0.59	0.39	-0.35	0.94	0.70	0.39	-0.06	0.96	0.70	0.38	-0.05	0.97
	4																				
PC3S	0.25	-0.93	0.35	1.30	0.80	-1.02	0.34	1.09	0.87	-1.14	0.33	0.76	0.91	-1.25	0.32	0.46	0.92	-1.33	0.32	0.19	0.95
	0.5	-0.46	0.36	0.62	0.99	-0.52	0.35	0.47	0.96	-0.57	0.33	0.36	0.98	-0.66	0.32	0.09	0.99	-0.63	0.32	0.18	0.97
	1	0.01	0.38	0.00	0.99	0.03	0.37	0.03	0.99	0.01	0.36	-0.04	0.97	0.00	0.35	-0.03	0.99	-0.02	0.34	-0.10	0.98
	2	0.45	0.43	-0.60	0.97	0.46	0.42	-0.62	0.95	0.53	0.41	-0.47	0.95	0.65	0.40	-0.17	0.97	0.68	0.39	-0.12	0.98
	4																				
I3S	0.25	-1.39	0.44	0.10	0.95	-1.33	0.39	0.18	0.97	-1.35	0.36	0.12	0.95	-1.39	0.34	0.04	0.94	-1.42	0.33	-0.08	0.96
	0.5	-0.69	0.41	0.00	0.98	-0.69	0.38	0.00	0.95	-0.67	0.35	0.05	0.98	-0.73	0.33	-0.13	0.98	-0.67	0.32	0.05	0.96
	1	0.00	0.43	-0.05	0.97	0.04	0.40	0.01	0.94	0.02	0.38	-0.02	0.93	0.01	0.36	-0.01	0.99	-0.02	0.34	-0.11	0.97
	2	0.70	0.51	-0.10	0.97	0.63	0.46	-0.24	0.95	0.66	0.44	-0.18	0.95	0.75	0.42	0.04	0.97	0.73	0.40	0.01	0.97
	4																				
PR	0.25	-2.83	0.89	-1.61	0.59	-3.62	1.11	-2.01	0.52	-5.02	1.37	-2.66	0.29	-6.55	1.66	-3.13	0.19	-8.41	2.01	-3.52	0.09
	0.5	-1.38	0.90	-0.77	0.87	-1.98	1.12	-1.17	0.75	-2.50	1.38	-1.30	0.77	-3.47	1.69	-1.64	0.60	-3.87	2.06	-1.56	0.63
	1	-0.01	0.90	0.00	0.96	0.13	1.11	0.13	0.92	-0.09	1.40	-0.06	0.93	-0.01	1.70	-0.01	0.98	-0.23	2.05	-0.12	0.98
	2	1.12	0.90	0.48	0.98	1.42	1.12	0.66	0.95	1.92	1.40	0.88	0.89	2.68	1.69	1.17	0.85	3.48	2.02	1.38	0.74
	4																				
PW3S	0.25	-1.12	0.33	0.81	0.88	-1.16	0.32	0.72	0.91	-1.24	0.31	0.48	0.93	-1.31	0.31	0.26	0.93	-1.37	0.31	0.06	0.95
	0.5	-0.58	0.34	0.32	0.97	-0.60	0.33	0.24	0.94	-0.62	0.32	0.21	0.98	-0.70	0.31	-0.03	0.98	-0.65	0.31	0.11	0.97
	1	0.01	0.36	-0.03	0.95	0.03	0.35	0.02	0.94	0.01	0.35	-0.03	0.93	0.01	0.34	-0.02	0.98	-0.02	0.33	-0.11	0.98
	2	0.59	0.42	-0.32	0.95	0.56	0.40	-0.41	0.95	0.61	0.40	-0.30	0.95	0.71	0.39	-0.04	0.97	0.71	0.39	-0.04	0.97
	4																				

B.2 Supplementary Table 2

Supplementary Table 2: Average estimates (Est), standard errors (SE), standardized bias (Bias) and coverage probability of the 95% confidence interval (CP) in covariate estimates for two covariates simulations. C-CB refers to the simulations with a continuous and a common binary covariate. C-RB refers to the simulations with a continuous and a rare binary covariate. Results for simulations with a rare binary covariate with OR of 4 are not presented for reasons described in Section 3.4.

			Continuous in C-CB																			
			Separation																			
			6				8				10				12				14			
Method	Binary OR	Continuous OR	Est	SE	Bias	CP	Est	SE	Bias	CP	Est	SE	Bias	CP	Est	SE	Bias	CP	Est	SE	Bias	CP
		0.25	-1.41	0.20	0.02	0.94	-1.41	0.18	0.00	0.95	-1.42	0.17	-0.11	0.95	-1.42	0.16	-0.13	0.96	-1.41	0.15	-0.11	0.97
		0.5	-0.72	0.15	-0.16	1.00	-0.70	0.14	-0.01	0.96	-0.71	0.13	-0.06	0.92	-0.71	0.12	-0.13	0.95	-0.70	0.12	-0.01	0.93
	0.25	1	0.00	0.13	0.02	0.94	0.00	0.12	-0.02	0.95	0.02	0.11	0.15	0.94	-0.03	0.11	-0.27	0.95	0.01	0.11	0.09	0.96
		2	0.73	0.15	0.17	0.95	0.72	0.14	0.11	0.94	0.69	0.13	-0.03	0.97	0.71	0.12	0.12	0.94	0.69	0.12	-0.07	0.98
		4	1.43	0.20	0.13	0.95	1.43	0.18	0.13	0.96	1.41	0.17	0.06	0.98	1.41	0.16	0.06	0.97	1.42	0.15	0.15	0.94
		0.25	-1.44	0.19	-0.19	0.96	-1.42	0.18	-0.09	0.96	-1.42	0.16	-0.10	0.98	-1.43	0.15	-0.18	0.92	-1.41	0.15	-0.05	0.97
		0.5	-0.70	0.14	0.01	0.96	-0.70	0.13	0.03	0.94	-0.71	0.12	-0.12	0.95	-0.70	0.12	-0.03	0.97	-0.70	0.11	-0.02	0.97
	0.5	1	-0.01	0.12	-0.09	0.93	0.01	0.12	0.06	0.97	0.00	0.11	0.03	0.96	0.01	0.11	0.09	0.97	0.00	0.10	0.03	0.96
		2	0.72	0.14	0.10	0.96	0.72	0.13	0.17	0.94	0.70	0.12	0.01	0.95	0.71	0.12	0.10	0.96	0.70	0.11	-0.02	0.96
		4	1.43	0.20	0.12	0.97	1.41	0.18	0.06	0.96	1.43	0.17	0.18	0.97	1.41	0.15	0.07	0.95	1.39	0.15	-0.02	0.95
		0.25	-1.43	0.19	-0.09	0.95	-1.41	0.18	-0.04	0.96	-1.40	0.16	0.03	0.96	-1.41	0.15	-0.10	0.97	-1.41	0.15	-0.09	0.97
		0.5	-0.71	0.14	-0.09	0.96	-0.71	0.13	-0.09	0.95	-0.70	0.12	-0.03	0.98	-0.71	0.12	-0.09	0.93	-0.70	0.11	-0.05	0.91
1S	1	1	0.00	0.12	0.02	0.96	0.00	0.12	-0.04	0.94	-0.01	0.11	-0.07	0.95	0.00	0.10	0.01	0.95	0.00	0.10	-0.03	0.93
		2	0.70	0.14	0.02	0.94	0.70	0.13	0.02	0.94	0.70	0.12	0.02	0.97	0.70	0.12	-0.02	0.97	0.69	0.11	-0.03	0.94
		4	1.43	0.19	0.14	0.97	1.39	0.17	-0.06	0.96	1.41	0.16	0.03	0.94	1.41	0.15	0.06	0.94	1.41	0.15	0.11	0.96
		0.25	-1.41	0.19	-0.04	0.96	-1.45	0.18	-0.26	0.95	-1.40	0.16	0.01	0.94	-1.39	0.15	0.04	0.93	-1.39	0.15	0.03	0.96
		0.5	-0.70	0.14	-0.01	0.96	-0.70	0.13	-0.02	0.96	-0.69	0.12	0.08	0.96	-0.71	0.12	-0.07	0.93	-0.71	0.11	-0.10	0.92
	2	1	0.01	0.12	0.10	0.93	0.00	0.12	0.03	0.97	0.00	0.11	-0.02	0.94	0.01	0.11	0.07	0.96	-0.01	0.10	-0.07	0.96
		2	0.70	0.14	-0.04	0.97	0.71	0.13	0.04	0.93	0.71	0.12	0.08	0.96	0.70	0.12	0.05	0.96	0.70	0.11	0.02	0.97
		4	1.41	0.19	0.04	0.95	1.41	0.18	0.04	0.98	1.41	0.16	0.08	0.93	1.43	0.15	0.18	0.96	1.41	0.15	0.05	0.94
		0.25	-1.44	0.20	-0.14	0.96	-1.43	0.19	-0.12	0.95	-1.40	0.17	-0.02	0.98	-1.40	0.16	-0.04	0.97	-1.39	0.15	0.04	0.96
		0.5	-0.71	0.15	-0.07	0.97	-0.71	0.14	-0.05	0.95	-0.70	0.13	0.03	0.95	-0.69	0.12	0.04	0.95	-0.71	0.12	-0.10	0.91
	4	1	-0.02	0.13	-0.13	0.97	0.00	0.12	-0.04	0.97	0.00	0.11	0.02	0.92	0.01	0.11	0.05	0.91	0.00	0.11	0.02	0.95
		2	0.72	0.15	0.11	0.95	0.72	0.14	0.15	0.94	0.69	0.13	-0.06	0.93	0.71	0.12	0.08	0.92	0.69	0.12	-0.06	0.94
		4	1.44	0.20	0.18	0.94	1.41	0.18	0.04	0.95	1.44	0.17	0.23	0.92	1.40	0.16	0.04	0.96	1.40	0.15	0.00	0.96

Supplementary Table 2: Average estimates (Est), standard errors (SE), standardized bias (Bias) and coverage probability of the 95% confidence interval (CP) in covariate estimates for two covariates simulations. C-CB refers to the simulations with a continuous and a common binary covariate. C-RB refers to the simulations with a continuous and a rare binary covariate. Results for simulations with a rare binary covariate with OR of 4 are not presented for reasons described in Section 3.4.

			Binary in C-CB																			
			Separation																			
			6				8				10				12				14			
Method	Binary OR	Continuous OR	Est	SE	Bias	CP	Est	SE	Bias	CP	Est	SE	Bias	CP	Est	SE	Bias	CP	Est	SE	Bias	CP
		0.25	-1.40	0.34	0.00	0.97	-1.38	0.31	0.06	0.95	-1.45	0.30	-0.18	0.96	-1.43	0.28	-0.12	0.93	-1.42	0.27	-0.09	0.96
		0.5	-1.43	0.29	-0.09	0.95	-1.41	0.27	-0.03	0.94	-1.44	0.25	-0.16	0.94	-1.43	0.24	-0.13	0.98	-1.38	0.23	0.04	0.95
	0.25	1	-1.42	0.27	-0.07	0.97	-1.38	0.25	0.05	0.96	-1.39	0.24	0.04	0.96	-1.42	0.23	-0.12	0.97	-1.39	0.22	0.00	0.98
		2	-1.41	0.29	-0.03	0.95	-1.41	0.27	-0.06	0.96	-1.40	0.25	-0.04	0.95	-1.39	0.24	0.03	0.93	-1.40	0.23	-0.04	0.96
		4	-1.40	0.34	0.02	0.98	-1.41	0.31	-0.03	0.93	-1.38	0.29	0.06	0.95	-1.40	0.28	0.00	0.93	-1.44	0.27	-0.16	0.96
		0.25	-0.76	0.32	-0.18	0.96	-0.74	0.29	-0.14	0.98	-0.72	0.27	-0.07	0.96	-0.73	0.26	-0.13	0.94	-0.68	0.25	0.07	0.94
		0.5	-0.70	0.27	0.01	0.94	-0.71	0.25	-0.04	0.96	-0.72	0.24	-0.10	0.98	-0.69	0.22	0.04	0.94	-0.70	0.22	0.00	0.92
	0.5	1	-0.71	0.25	-0.03	0.94	-0.70	0.24	-0.03	0.93	-0.71	0.22	-0.08	0.96	-0.71	0.21	-0.08	0.94	-0.70	0.21	-0.02	0.93
		2	-0.72	0.27	-0.06	0.96	-0.70	0.25	-0.01	0.95	-0.69	0.23	0.03	0.96	-0.69	0.22	0.02	0.95	-0.67	0.22	0.13	0.93
		4	-0.75	0.31	-0.16	0.94	-0.68	0.29	0.06	0.96	-0.70	0.27	-0.01	0.96	-0.68	0.26	0.08	0.95	-0.71	0.25	-0.04	0.96
		0.25	0.01	0.31	0.03	0.96	0.00	0.28	-0.01	0.95	0.00	0.27	0.02	0.95	0.00	0.26	-0.01	0.94	0.02	0.25	0.08	0.97
		0.5	0.00	0.26	0.01	0.95	-0.01	0.24	-0.03	0.97	-0.05	0.23	-0.20	0.91	-0.01	0.22	-0.06	0.92	0.02	0.21	0.08	0.94
1S	1	1	-0.01	0.24	-0.02	0.96	-0.02	0.23	-0.09	0.93	-0.01	0.22	-0.07	0.94	0.01	0.21	0.07	0.96	0.02	0.20	0.08	0.93
		2	0.01	0.26	0.04	0.96	0.01	0.24	0.06	0.96	0.01	0.23	0.05	0.95	0.00	0.22	0.02	0.96	-0.01	0.21	-0.06	0.98
		4	0.01	0.31	0.03	0.98	0.02	0.28	0.06	0.98	0.01	0.27	0.05	0.96	-0.01	0.25	-0.04	0.97	0.02	0.25	0.08	0.96
		0.25	0.71	0.31	0.03	0.96	0.70	0.29	0.01	0.94	0.72	0.27	0.09	0.94	0.68	0.26	-0.07	0.96	0.65	0.25	-0.18	0.94
		0.5	0.71	0.27	0.03	0.95	0.69	0.25	-0.04	0.96	0.70	0.23	0.03	0.96	0.71	0.22	0.06	0.95	0.69	0.22	-0.01	0.95
	2	1	0.72	0.25	0.10	0.96	0.70	0.24	-0.01	0.96	0.72	0.22	0.10	0.97	0.69	0.21	-0.03	0.96	0.72	0.21	0.11	0.95
		2	0.70	0.27	0.02	0.96	0.69	0.25	-0.01	0.97	0.70	0.24	0.03	0.95	0.70	0.23	0.03	0.96	0.70	0.22	0.02	0.97
		4	0.71	0.31	0.03	0.97	0.69	0.29	-0.01	0.96	0.70	0.27	0.01	0.94	0.72	0.26	0.09	0.97	0.73	0.25	0.10	0.97
		0.25	1.40	0.34	0.00	0.96	1.38	0.31	-0.05	0.93	1.44	0.29	0.16	0.93	1.44	0.28	0.16	0.94	1.39	0.27	-0.02	0.96
		0.5	1.41	0.29	0.04	0.96	1.43	0.27	0.13	0.97	1.36	0.25	-0.13	0.94	1.42	0.24	0.10	0.96	1.39	0.23	0.00	0.96
	4	1	1.41	0.27	0.02	0.95	1.40	0.25	0.03	0.95	1.43	0.24	0.12	0.95	1.40	0.23	0.01	0.95	1.41	0.22	0.07	0.95
		2	1.41	0.29	0.04	0.97	1.42	0.27	0.09	0.96	1.42	0.25	0.11	0.95	1.41	0.24	0.07	0.97	1.40	0.23	0.03	0.97
		4	1.40	0.34	-0.02	0.95	1.42	0.31	0.06	0.96	1.41	0.29	0.04	0.93	1.41	0.28	0.05	0.98	1.40	0.27	0.02	0.97

Supplementary Table 2: Average estimates (Est), standard errors (SE), standardized bias (Bias) and coverage probability of the 95% confidence interval (CP) in covariate estimates for two covariates simulations. C-CB refers to the simulations with a continuous and a common binary covariate. C-RB refers to the simulations with a continuous and a rare binary covariate. Results for simulations with a rare binary covariate with OR of 4 are not presented for reasons described in Section 3.4.

			Continuous in C-RB																			
			Separation																			
			6				8				10				12				14			
Method	Binary OR	Continuous OR	Est	SE	Bias	CP	Est	SE	Bias	CP	Est	SE	Bias	CP	Est	SE	Bias	CP	Est	SE	Bias	CP
		0.25	-1.42	0.19	-0.08	0.96	-1.42	0.18	-0.07	0.96	-1.41	0.16	-0.04	0.97	-1.40	0.15	-0.04	0.96	-1.40	0.15	-0.03	0.96
		0.5	-0.72	0.14	-0.15	0.96	-0.72	0.13	-0.16	0.97	-0.71	0.12	-0.08	0.94	-0.70	0.12	0.04	0.93	-0.70	0.11	-0.02	0.97
	0.25	1	0.00	0.13	-0.03	0.96	0.01	0.12	0.10	0.95	0.00	0.11	-0.02	0.95	0.00	0.11	0.01	0.93	0.00	0.10	-0.01	0.97
		2	0.70	0.14	-0.03	0.93	0.70	0.13	0.01	0.96	0.71	0.12	0.06	0.95	0.70	0.12	0.00	0.95	0.71	0.11	0.08	0.96
		4	1.43	0.20	0.10	0.98	1.42	0.18	0.08	0.95	1.40	0.16	0.01	0.97	1.41	0.15	0.04	0.96	1.39	0.15	-0.03	0.95
		0.25	-1.44	0.19	-0.21	0.97	-1.43	0.18	-0.16	0.97	-1.42	0.16	-0.09	0.95	-1.42	0.15	-0.13	0.94	-1.40	0.15	-0.04	0.96
		0.5	-0.75	0.14	-0.32	0.95	-0.70	0.13	-0.01	0.94	-0.71	0.12	-0.05	0.95	-0.70	0.12	0.02	0.99	-0.71	0.11	-0.11	0.95
	0.5	1	0.00	0.12	-0.01	0.95	0.01	0.12	0.05	0.94	-0.01	0.11	-0.08	0.95	0.01	0.11	0.13	0.97	0.00	0.10	0.02	0.92
		2	0.70	0.14	0.01	0.95	0.72	0.13	0.13	0.96	0.70	0.12	0.05	0.96	0.70	0.12	0.04	0.94	0.70	0.11	0.05	0.97
		4	1.42	0.19	0.02	0.94	1.42	0.18	0.07	0.95	1.41	0.16	0.04	0.93	1.42	0.15	0.16	0.96	1.41	0.15	0.12	0.96
		0.25	-1.42	0.19	-0.06	0.97	-1.43	0.18	-0.15	0.97	-1.40	0.16	-0.01	0.95	-1.42	0.15	-0.13	0.95	-1.40	0.15	-0.01	0.97
		0.5	-0.71	0.14	-0.08	0.96	-0.71	0.13	-0.10	0.94	-0.69	0.12	0.11	0.95	-0.70	0.12	-0.02	0.95	-0.70	0.11	0.00	0.95
1S	1	1	0.00	0.12	0.03	0.96	0.00	0.11	-0.04	0.95	0.01	0.11	0.07	0.95	0.00	0.10	-0.02	0.97	0.00	0.10	-0.02	0.97
		2	0.71	0.14	0.09	0.94	0.69	0.13	-0.05	0.96	0.72	0.12	0.16	0.96	0.71	0.12	0.08	0.95	0.69	0.11	-0.08	0.94
		4	1.42	0.19	0.05	0.96	1.41	0.17	0.01	0.95	1.42	0.16	0.10	0.96	1.41	0.15	0.05	0.95	1.39	0.14	-0.04	0.96
		0.25	-1.44	0.19	-0.17	0.98	-1.39	0.17	0.07	0.96	-1.40	0.16	0.01	0.95	-1.41	0.15	-0.08	0.96	-1.40	0.15	-0.01	0.93
		0.5	-0.70	0.14	-0.01	0.96	-0.72	0.13	-0.17	0.93	-0.70	0.12	-0.03	0.95	-0.71	0.12	-0.12	0.93	-0.69	0.11	0.03	0.95
	2	1	-0.01	0.12	-0.07	0.98	0.01	0.12	0.05	0.92	0.00	0.11	0.03	0.96	-0.01	0.10	-0.04	0.96	-0.01	0.10	-0.08	0.95
		2	0.70	0.14	-0.03	0.92	0.69	0.13	-0.05	0.97	0.70	0.12	0.00	0.95	0.70	0.12	0.02	0.97	0.69	0.11	-0.02	0.93
		4	1.42	0.19	0.10	0.95	1.40	0.17	-0.05	0.93	1.43	0.16	0.20	0.97	1.40	0.15	-0.01	0.97	1.40	0.14	0.01	0.94
		0.25	-1.40	0.19	0.05	0.93	-1.43	0.18	-0.16	0.94	-1.39	0.16	0.07	0.94	-1.42	0.15	-0.11	0.97	-1.40	0.15	-0.01	0.95
		0.5	-0.72	0.14	-0.11	0.97	-0.69	0.13	0.07	0.98	-0.72	0.13	-0.12	0.93	-0.69	0.12	0.06	0.96	-0.71	0.11	-0.09	0.95
	4	1	0.03	0.12	0.23	0.92	-0.01	0.11	-0.12	0.97	0.00	0.11	0.01	0.98	-0.01	0.11	-0.10	0.96	0.00	0.10	-0.04	0.96
		2	0.72	0.14	0.12	0.95	0.73	0.13	0.22	0.94	0.70	0.12	0.03	0.97	0.69	0.12	-0.07	0.94	0.70	0.11	0.05	0.94
		4	1.43	0.19	0.12	0.96	1.39	0.17	-0.06	0.97	1.40	0.16	0.00	0.95	1.40	0.15	0.00	0.92	1.44	0.15	0.26	0.95

Supplementary Table 2: Average estimates (Est), standard errors (SE), standardized bias (Bias) and coverage probability of the 95% confidence interval (CP) in covariate estimates for two covariates simulations. C-CB refers to the simulations with a continuous and a common binary covariate. C-RB refers to the simulations with a continuous and a rare binary covariate. Results for simulations with a rare binary covariate with OR of 4 are not presented for reasons described in Section 3.4.

			Binary in C-RB																			
			Separation																			
			6				8				10				12				14			
Method	Binary OR	Continuous OR	Est	SE	Bias	CP	Est	SE	Bias	CP	Est	SE	Bias	CP	Est	SE	Bias	CP	Est	SE	Bias	CP
		0.25	-1.48	0.51	-0.15	0.97	-1.42	0.47	-0.03	0.95	-1.38	0.43	0.05	0.94	-1.37	0.41	0.07	0.96	-1.43	0.40	-0.07	0.95
		0.5	-1.44	0.43	-0.07	0.96	-1.42	0.40	-0.06	0.95	-1.41	0.37	-0.03	0.97	-1.45	0.36	-0.13	0.94	-1.39	0.34	0.00	0.95
	0.25	1	-1.36	0.40	0.11	0.96	-1.42	0.37	-0.05	0.96	-1.43	0.35	-0.08	0.96	-1.44	0.33	-0.13	0.95	-1.42	0.32	-0.07	0.96
		2	-1.36	0.43	0.13	0.94	-1.37	0.40	0.06	0.97	-1.39	0.37	0.00	0.96	-1.38	0.36	0.04	0.95	-1.41	0.34	-0.05	0.95
		4	-1.54	0.52	-0.22	0.96	-1.37	0.47	0.06	0.98	-1.44	0.44	-0.07	0.93	-1.44	0.41	-0.09	0.93	-1.40	0.40	-0.02	0.99
		0.25	-0.70	0.49	-0.02	0.94	-0.64	0.46	0.10	0.97	-0.64	0.43	0.10	0.97	-0.72	0.41	-0.06	0.93	-0.68	0.39	0.02	0.95
		0.5	-0.66	0.42	0.06	0.96	-0.67	0.38	0.04	0.94	-0.68	0.36	0.02	0.94	-0.74	0.34	-0.13	0.95	-0.69	0.34	-0.01	0.96
	0.5	1	-0.69	0.39	-0.02	0.97	-0.71	0.36	-0.06	0.96	-0.71	0.34	-0.06	0.97	-0.70	0.33	-0.03	0.95	-0.67	0.32	0.06	0.96
		2	-0.71	0.41	-0.03	0.97	-0.63	0.39	0.15	0.96	-0.72	0.36	-0.08	0.94	-0.71	0.35	-0.05	0.95	-0.71	0.34	-0.05	0.96
		4	-0.69	0.50	0.01	0.95	-0.66	0.45	0.07	0.95	-0.76	0.43	-0.17	0.96	-0.72	0.41	-0.07	0.94	-0.72	0.40	-0.06	0.93
		0.25	0.04	0.50	0.02	0.93	0.00	0.47	-0.03	0.94	0.00	0.45	-0.07	0.94	-0.02	0.43	-0.10	0.94	0.00	0.41	-0.03	0.92
		0.5	0.02	0.44	0.01	0.97	-0.01	0.41	-0.08	0.95	0.01	0.38	-0.03	0.95	0.01	0.37	-0.02	0.94	-0.01	0.36	-0.06	0.95
1S	1	1	-0.03	0.41	-0.13	0.96	-0.01	0.39	-0.09	0.97	0.05	0.37	0.07	0.94	0.04	0.35	0.04	0.97	0.03	0.34	0.03	0.98
		2	0.00	0.44	-0.06	0.96	0.03	0.41	0.03	0.97	-0.04	0.39	-0.17	0.96	0.03	0.37	0.05	0.98	0.00	0.36	-0.03	0.95
		4	0.02	0.51	0.00	0.96	0.02	0.47	-0.01	0.98	0.00	0.44	-0.05	0.96	-0.01	0.43	-0.05	0.95	0.00	0.41	-0.04	0.96
		0.25	0.63	0.56	-0.15	0.97	0.70	0.53	-0.02	0.97	0.67	0.50	-0.10	0.95	0.61	0.48	-0.24	0.93	0.65	0.46	-0.17	0.94
		0.5	0.64	0.50	-0.18	0.96	0.62	0.47	-0.22	0.95	0.65	0.44	-0.18	0.95	0.63	0.42	-0.22	0.95	0.67	0.41	-0.13	0.97
	2	1	0.63	0.47	-0.21	0.98	0.61	0.44	-0.26	0.95	0.67	0.43	-0.12	0.97	0.65	0.40	-0.19	0.97	0.74	0.40	0.02	0.98
		2	0.54	0.49	-0.37	0.93	0.63	0.46	-0.18	0.97	0.66	0.44	-0.15	0.98	0.68	0.42	-0.12	0.96	0.66	0.41	-0.15	0.98
		4	0.56	0.56	-0.29	0.96	0.66	0.53	-0.13	0.98	0.67	0.50	-0.10	0.96	0.71	0.48	-0.01	0.97	0.71	0.47	-0.03	0.98
		0.25																				
		0.5																				
	4	1																				
		2																				
		4																				

Supplementary Table 2: Average estimates (Est), standard errors (SE), standardized bias (Bias) and coverage probability of the 95% confidence interval (CP) in covariate estimates for two covariates simulations. C-CB refers to the simulations with a continuous and a common binary covariate. C-RB refers to the simulations with a continuous and a rare binary covariate. Results for simulations with a rare binary covariate with OR of 4 are not presented for reasons described in Section 3.4.

			Continuous in C-CB																			
			Separation																			
			6				8				10				12				14			
Method	Binary OR	Continuous OR	Est	SE	Bias	CP	Est	SE	Bias	CP	Est	SE	Bias	CP	Est	SE	Bias	CP	Est	SE	Bias	CP
		0.25	-0.92	0.12	4.09	0.06	-1.01	0.12	3.20	0.14	-1.14	0.13	2.03	0.46	-1.23	0.13	1.29	0.72	-1.30	0.14	0.67	0.89
		0.5	-0.53	0.11	1.53	0.64	-0.56	0.11	1.21	0.76	-0.61	0.11	0.76	0.87	-0.65	0.11	0.43	0.94	-0.67	0.11	0.28	0.94
	0.25	1	0.00	0.11	0.02	0.96	0.00	0.10	-0.04	0.95	0.02	0.10	0.15	0.91	-0.03	0.10	-0.27	0.93	0.01	0.10	0.09	0.95
		2	0.53	0.11	-1.49	0.66	0.57	0.11	-1.13	0.79	0.60	0.11	-0.87	0.86	0.65	0.11	-0.39	0.94	0.66	0.11	-0.37	0.94
		4	0.93	0.12	-4.02	0.06	1.04	0.12	-2.90	0.23	1.12	0.13	-2.14	0.45	1.22	0.13	-1.33	0.75	1.31	0.14	-0.66	0.88
		0.25	-0.97	0.12	3.68	0.08	-1.04	0.12	2.97	0.21	-1.16	0.13	1.91	0.53	-1.26	0.13	1.01	0.80	-1.30	0.13	0.69	0.88
		0.5	-0.54	0.11	1.45	0.67	-0.57	0.11	1.16	0.78	-0.62	0.11	0.68	0.88	-0.64	0.11	0.50	0.92	-0.67	0.11	0.27	0.96
	0.5	1	-0.01	0.10	-0.12	0.94	0.00	0.10	0.04	0.97	0.00	0.10	0.04	0.95	0.01	0.10	0.08	0.96	0.00	0.10	0.02	0.97
		2	0.54	0.11	-1.42	0.70	0.59	0.11	-1.01	0.82	0.62	0.11	-0.76	0.92	0.65	0.11	-0.41	0.94	0.67	0.11	-0.30	0.93
		4	0.95	0.12	-3.81	0.06	1.04	0.12	-2.98	0.18	1.16	0.13	-1.89	0.55	1.24	0.13	-1.19	0.78	1.30	0.13	-0.72	0.89
		0.25	-0.97	0.12	3.69	0.09	-1.05	0.12	2.89	0.26	-1.15	0.12	2.01	0.50	-1.24	0.13	1.16	0.81	-1.31	0.13	0.67	0.87
		0.5	-0.55	0.11	1.41	0.70	-0.59	0.11	0.98	0.83	-0.62	0.11	0.68	0.90	-0.65	0.11	0.41	0.92	-0.67	0.11	0.23	0.94
3S	1	1	0.00	0.10	0.02	0.94	0.00	0.10	-0.03	0.97	-0.01	0.10	-0.08	0.94	0.00	0.10	-0.01	0.94	0.00	0.10	-0.02	0.95
		2	0.54	0.11	-1.47	0.70	0.58	0.11	-1.08	0.81	0.62	0.11	-0.75	0.88	0.64	0.11	-0.50	0.93	0.67	0.11	-0.30	0.93
		4	0.97	0.12	-3.69	0.11	1.04	0.12	-3.05	0.19	1.16	0.12	-1.95	0.47	1.24	0.13	-1.24	0.73	1.32	0.13	-0.59	0.92
		0.25	-0.95	0.12	3.86	0.06	-1.07	0.12	2.72	0.29	-1.14	0.12	2.07	0.48	-1.23	0.13	1.32	0.70	-1.30	0.13	0.76	0.86
		0.5	-0.53	0.11	1.52	0.70	-0.58	0.11	1.12	0.82	-0.60	0.11	0.88	0.88	-0.65	0.11	0.44	0.93	-0.67	0.11	0.21	0.91
	2	1	0.01	0.10	0.08	0.92	0.00	0.10	0.03	0.97	0.00	0.10	0.00	0.93	0.01	0.10	0.07	0.96	-0.01	0.10	-0.06	0.95
		2	0.53	0.11	-1.55	0.64	0.58	0.11	-1.08	0.81	0.62	0.11	-0.69	0.89	0.65	0.11	-0.43	0.95	0.67	0.11	-0.28	0.95
		4	0.95	0.11	-3.92	0.06	1.04	0.12	-2.95	0.20	1.15	0.13	-1.94	0.52	1.25	0.13	-1.11	0.78	1.30	0.13	-0.74	0.84
		0.25	-0.93	0.12	4.04	0.05	-1.03	0.12	3.00	0.19	-1.12	0.13	2.15	0.45	-1.22	0.13	1.34	0.73	-1.28	0.14	0.82	0.85
		0.5	-0.53	0.11	1.53	0.67	-0.57	0.11	1.20	0.75	-0.60	0.11	0.86	0.84	-0.63	0.11	0.61	0.94	-0.67	0.11	0.21	0.93
	4	1	-0.02	0.11	-0.14	0.96	-0.01	0.11	-0.05	0.97	0.00	0.10	0.02	0.93	0.00	0.10	0.04	0.92	0.00	0.10	0.01	0.95
		2	0.53	0.11	-1.58	0.62	0.57	0.11	-1.14	0.77	0.60	0.11	-0.91	0.88	0.65	0.11	-0.47	0.90	0.66	0.11	-0.37	0.92
		4	0.93	0.12	-3.97	0.05	1.02	0.12	-3.08	0.18	1.15	0.13	-1.92	0.52	1.22	0.13	-1.34	0.69	1.29	0.14	-0.81	0.86

Supplementary Table 2: Average estimates (Est), standard errors (SE), standardized bias (Bias) and coverage probability of the 95% confidence interval (CP) in covariate estimates for two covariates simulations. C-CB refers to the simulations with a continuous and a common binary covariate. C-RB refers to the simulations with a continuous and a rare binary covariate. Results for simulations with a rare binary covariate with OR of 4 are not presented for reasons described in Section 3.4.

			Binary in C-CB																			
			Separation																			
			6				8				10				12				14			
Method	Binary OR	Continuous OR	Est	SE	Bias	CP	Est	SE	Bias	CP	Est	SE	Bias	CP	Est	SE	Bias	CP	Est	SE	Bias	CP
		0.25	-0.95	0.24	1.83	0.50	-1.02	0.24	1.53	0.67	-1.17	0.25	0.89	0.88	-1.25	0.25	0.58	0.93	-1.31	0.25	0.32	0.94
		0.5	-1.07	0.23	1.42	0.71	-1.16	0.22	1.05	0.81	-1.26	0.23	0.61	0.91	-1.31	0.22	0.38	0.95	-1.32	0.22	0.32	0.94
	0.25	1	-1.13	0.22	1.21	0.77	-1.17	0.22	1.04	0.82	-1.24	0.22	0.73	0.86	-1.32	0.22	0.33	0.94	-1.34	0.22	0.25	0.96
		2	-1.07	0.23	1.44	0.72	-1.16	0.22	1.06	0.83	-1.23	0.22	0.74	0.89	-1.27	0.22	0.55	0.89	-1.34	0.22	0.24	0.95
		4	-0.93	0.24	1.93	0.51	-1.05	0.24	1.41	0.73	-1.11	0.25	1.15	0.82	-1.22	0.25	0.70	0.85	-1.33	0.25	0.24	0.94
		0.25	-0.50	0.23	0.84	0.89	-0.54	0.23	0.68	0.90	-0.57	0.23	0.53	0.95	-0.65	0.24	0.21	0.92	-0.62	0.24	0.30	0.92
		0.5	-0.53	0.22	0.78	0.88	-0.58	0.21	0.54	0.93	-0.63	0.21	0.30	0.97	-0.63	0.21	0.29	0.93	-0.67	0.21	0.14	0.94
	0.5	1	-0.56	0.21	0.67	0.89	-0.60	0.21	0.47	0.90	-0.64	0.20	0.27	0.97	-0.67	0.20	0.15	0.92	-0.67	0.20	0.11	0.93
		2	-0.55	0.22	0.66	0.91	-0.58	0.21	0.54	0.90	-0.61	0.21	0.40	0.97	-0.64	0.21	0.28	0.96	-0.64	0.21	0.26	0.92
		4	-0.51	0.23	0.80	0.87	-0.50	0.23	0.84	0.87	-0.57	0.24	0.55	0.91	-0.59	0.24	0.45	0.92	-0.66	0.24	0.17	0.98
		0.25	0.00	0.23	0.02	0.96	0.00	0.23	0.02	0.96	0.00	0.23	0.00	0.92	0.00	0.23	-0.01	0.93	0.02	0.23	0.09	0.96
		0.5	0.00	0.21	0.00	0.96	-0.01	0.21	-0.03	0.96	-0.04	0.21	-0.19	0.92	-0.01	0.21	-0.06	0.92	0.02	0.21	0.08	0.95
3S	1	1	0.00	0.21	0.01	0.95	-0.01	0.20	-0.07	0.93	-0.01	0.20	-0.06	0.93	0.02	0.20	0.08	0.95	0.02	0.20	0.08	0.92
		2	0.01	0.21	0.04	0.97	0.01	0.21	0.04	0.94	0.01	0.21	0.05	0.95	0.00	0.21	0.01	0.96	-0.01	0.21	-0.06	0.98
		4	0.03	0.23	0.13	0.95	0.01	0.23	0.05	0.95	0.01	0.23	0.04	0.97	-0.01	0.23	-0.04	0.96	0.02	0.24	0.08	0.96
		0.25	0.49	0.23	-0.90	0.86	0.52	0.23	-0.76	0.89	0.60	0.23	-0.40	0.92	0.60	0.24	-0.42	0.94	0.60	0.24	-0.39	0.94
		0.5	0.55	0.22	-0.70	0.89	0.56	0.21	-0.63	0.87	0.61	0.21	-0.38	0.93	0.66	0.21	-0.19	0.97	0.67	0.21	-0.14	0.95
	2	1	0.58	0.21	-0.57	0.91	0.60	0.21	-0.49	0.91	0.65	0.20	-0.22	0.95	0.65	0.20	-0.22	0.94	0.70	0.20	0.00	0.97
		2	0.54	0.22	-0.70	0.89	0.57	0.21	-0.58	0.92	0.62	0.21	-0.36	0.93	0.65	0.21	-0.22	0.94	0.67	0.21	-0.12	0.96
		4	0.48	0.23	-0.94	0.85	0.52	0.23	-0.74	0.89	0.58	0.24	-0.50	0.91	0.64	0.24	-0.26	0.93	0.68	0.24	-0.09	0.96
		0.25	0.93	0.24	-1.92	0.49	1.02	0.24	-1.53	0.65	1.19	0.25	-0.83	0.87	1.27	0.25	-0.51	0.90	1.29	0.25	-0.40	0.92
		0.5	1.07	0.23	-1.42	0.66	1.17	0.22	-0.99	0.85	1.20	0.22	-0.88	0.84	1.31	0.22	-0.39	0.94	1.32	0.22	-0.32	0.96
	4	1	1.11	0.22	-1.28	0.73	1.18	0.22	-0.97	0.83	1.28	0.22	-0.53	0.93	1.29	0.22	-0.46	0.92	1.35	0.22	-0.19	0.95
		2	1.06	0.23	-1.47	0.69	1.16	0.22	-1.04	0.82	1.24	0.22	-0.70	0.91	1.29	0.22	-0.45	0.94	1.34	0.22	-0.24	0.96
		4	0.94	0.24	-1.89	0.55	1.04	0.24	-1.47	0.72	1.15	0.25	-0.98	0.81	1.23	0.25	-0.67	0.91	1.29	0.25	-0.40	0.94

Supplementary Table 2: Average estimates (Est), standard errors (SE), standardized bias (Bias) and coverage probability of the 95% confidence interval (CP) in covariate estimates for two covariates simulations. C-CB refers to the simulations with a continuous and a common binary covariate. C-RB refers to the simulations with a continuous and a rare binary covariate. Results for simulations with a rare binary covariate with OR of 4 are not presented for reasons described in Section 3.4.

			Continuous in C-RB																			
			Separation																			
			6				8				10				12				14			
Method	Binary OR	Continuous OR	Est	SE	Bias	CP	Est	SE	Bias	CP	Est	SE	Bias	CP	Est	SE	Bias	CP	Est	SE	Bias	CP
		0.25	-0.95	0.12	3.87	0.07	-1.04	0.12	3.02	0.19	-1.15	0.13	1.99	0.46	-1.23	0.13	1.26	0.74	-1.30	0.13	0.70	0.90
		0.5	-0.55	0.11	1.38	0.70	-0.59	0.11	1.01	0.86	-0.62	0.11	0.69	0.88	-0.64	0.11	0.53	0.92	-0.67	0.11	0.26	0.95
	0.25	1	0.00	0.10	-0.03	0.94	0.01	0.10	0.09	0.96	0.00	0.10	-0.02	0.96	0.00	0.10	-0.01	0.95	0.00	0.10	-0.02	0.97
		2	0.53	0.11	-1.56	0.64	0.57	0.11	-1.16	0.80	0.62	0.11	-0.73	0.88	0.64	0.11	-0.54	0.90	0.67	0.11	-0.23	0.94
		4	0.94	0.12	-3.96	0.03	1.04	0.12	-2.92	0.21	1.14	0.12	-2.07	0.50	1.23	0.13	-1.30	0.69	1.29	0.13	-0.84	0.81
		0.25	-0.96	0.11	3.76	0.09	-1.06	0.12	2.79	0.25	-1.16	0.12	1.88	0.53	-1.24	0.13	1.18	0.74	-1.30	0.13	0.71	0.85
		0.5	-0.57	0.11	1.14	0.79	-0.57	0.11	1.15	0.77	-0.62	0.11	0.72	0.85	-0.65	0.11	0.48	0.94	-0.68	0.11	0.19	0.94
	0.5	1	0.00	0.10	0.01	0.96	0.00	0.10	0.05	0.95	-0.01	0.10	-0.10	0.97	0.01	0.10	0.13	0.97	0.00	0.10	0.00	0.94
		2	0.54	0.11	-1.48	0.68	0.58	0.11	-1.06	0.84	0.62	0.11	-0.72	0.91	0.65	0.11	-0.44	0.93	0.67	0.11	-0.24	0.95
		4	0.95	0.11	-3.92	0.09	1.05	0.12	-2.95	0.18	1.15	0.12	-1.99	0.50	1.25	0.13	-1.17	0.77	1.32	0.13	-0.57	0.93
		0.25	-0.96	0.11	3.83	0.07	-1.05	0.12	2.86	0.19	-1.15	0.12	1.95	0.51	-1.25	0.13	1.13	0.80	-1.30	0.13	0.69	0.87
		0.5	-0.54	0.11	1.44	0.70	-0.59	0.11	1.01	0.80	-0.61	0.11	0.85	0.86	-0.65	0.11	0.43	0.94	-0.67	0.11	0.29	0.95
3S	1	1	0.00	0.10	0.03	0.96	0.00	0.10	-0.04	0.96	0.01	0.10	0.06	0.96	0.00	0.10	-0.02	0.97	0.00	0.10	-0.03	0.97
		2	0.55	0.11	-1.34	0.74	0.57	0.11	-1.16	0.80	0.63	0.11	-0.61	0.91	0.65	0.11	-0.42	0.93	0.66	0.11	-0.34	0.93
		4	0.96	0.11	-3.80	0.07	1.06	0.12	-2.87	0.27	1.16	0.12	-1.86	0.53	1.24	0.13	-1.21	0.78	1.29	0.13	-0.76	0.88
		0.25	-0.98	0.12	3.63	0.09	-1.03	0.12	3.10	0.20	-1.15	0.12	1.98	0.51	-1.24	0.13	1.19	0.78	-1.30	0.13	0.71	0.85
		0.5	-0.55	0.11	1.40	0.68	-0.59	0.11	0.99	0.85	-0.62	0.11	0.76	0.89	-0.66	0.11	0.35	0.89	-0.67	0.11	0.30	0.94
	2	1	0.00	0.10	-0.04	0.97	0.01	0.10	0.05	0.96	0.00	0.10	-0.02	0.94	0.00	0.10	-0.04	0.97	-0.01	0.10	-0.09	0.96
		2	0.54	0.11	-1.46	0.70	0.57	0.11	-1.18	0.76	0.62	0.11	-0.74	0.87	0.65	0.11	-0.45	0.92	0.66	0.11	-0.33	0.94
		4	0.95	0.11	-3.89	0.07	1.04	0.12	-3.00	0.21	1.16	0.13	-1.84	0.56	1.24	0.13	-1.24	0.73	1.30	0.13	-0.72	0.89
		0.25	-0.94	0.11	4.02	0.07	-1.04	0.12	2.99	0.20	-1.14	0.12	2.09	0.42	-1.23	0.13	1.26	0.80	-1.30	0.13	0.74	0.84
		0.5	-0.54	0.11	1.46	0.63	-0.57	0.11	1.21	0.78	-0.62	0.11	0.72	0.88	-0.64	0.11	0.53	0.95	-0.68	0.11	0.16	0.96
	4	1	0.01	0.10	0.07	0.95	-0.01	0.10	-0.09	0.95	0.01	0.10	0.07	0.96	-0.01	0.10	-0.08	0.95	0.00	0.10	-0.05	0.97
		2	0.54	0.11	-1.44	0.66	0.59	0.11	-1.03	0.84	0.61	0.11	-0.80	0.90	0.64	0.11	-0.49	0.91	0.67	0.11	-0.24	0.92
		4	0.96	0.11	-3.85	0.06	1.02	0.12	-3.17	0.15	1.14	0.12	-2.09	0.45	1.24	0.13	-1.25	0.77	1.33	0.14	-0.52	0.92

Supplementary Table 2: Average estimates (Est), standard errors (SE), standardized bias (Bias) and coverage probability of the 95% confidence interval (CP) in covariate estimates for two covariates simulations. C-CB refers to the simulations with a continuous and a common binary covariate. C-RB refers to the simulations with a continuous and a rare binary covariate. Results for simulations with a rare binary covariate with OR of 4 are not presented for reasons described in Section 3.4.

			Binary in C-RB																			
			Separation																			
			6				8				10				12				14			
Method	Binary OR	Continuous OR	Est	SE	Bias	CP	Est	SE	Bias	CP	Est	SE	Bias	CP	Est	SE	Bias	CP	Est	SE	Bias	CP
		0.25	-0.98	0.35	1.18	0.79	-1.02	0.35	1.05	0.82	-1.09	0.36	0.83	0.85	-1.20	0.37	0.53	0.92	-1.32	0.37	0.21	0.94
		0.5	-1.06	0.32	1.02	0.85	-1.14	0.32	0.77	0.87	-1.20	0.32	0.58	0.93	-1.32	0.33	0.23	0.95	-1.33	0.33	0.20	0.94
	0.25	1	-1.06	0.31	1.06	0.80	-1.18	0.31	0.69	0.89	-1.27	0.31	0.40	0.94	-1.33	0.31	0.21	0.96	-1.36	0.31	0.12	0.96
		2	-0.99	0.32	1.22	0.74	-1.10	0.32	0.88	0.86	-1.22	0.32	0.52	0.93	-1.27	0.33	0.38	0.94	-1.35	0.33	0.14	0.94
		4	-0.97	0.35	1.21	0.80	-0.98	0.36	1.15	0.84	-1.14	0.36	0.72	0.90	-1.24	0.36	0.42	0.92	-1.29	0.38	0.27	0.98
		0.25	-0.46	0.36	0.61	0.90	-0.46	0.36	0.63	0.95	-0.52	0.37	0.44	0.97	-0.62	0.37	0.19	0.96	-0.63	0.37	0.15	0.94
		0.5	-0.49	0.33	0.58	0.91	-0.57	0.33	0.37	0.95	-0.60	0.33	0.27	0.95	-0.68	0.32	0.02	0.95	-0.65	0.33	0.10	0.95
	0.5	1	-0.55	0.32	0.40	0.94	-0.60	0.31	0.29	0.97	-0.64	0.31	0.16	0.98	-0.65	0.31	0.11	0.96	-0.64	0.31	0.14	0.96
		2	-0.53	0.33	0.47	0.94	-0.52	0.33	0.49	0.93	-0.63	0.33	0.16	0.93	-0.65	0.33	0.11	0.97	-0.67	0.33	0.04	0.98
		4	-0.46	0.36	0.64	0.90	-0.47	0.36	0.59	0.92	-0.61	0.37	0.21	0.95	-0.63	0.37	0.16	0.95	-0.67	0.38	0.07	0.92
		0.25	0.02	0.38	0.00	0.95	0.01	0.39	-0.01	0.97	0.00	0.39	-0.06	0.93	-0.02	0.39	-0.11	0.94	0.00	0.39	-0.04	0.92
		0.5	0.03	0.36	0.03	0.98	-0.01	0.35	-0.07	0.95	0.00	0.35	-0.06	0.97	0.01	0.35	-0.02	0.95	0.00	0.35	-0.05	0.96
3S	1	1	-0.01	0.35	-0.10	0.94	0.00	0.35	-0.07	0.96	0.03	0.34	0.02	0.96	0.03	0.34	0.04	0.97	0.03	0.33	0.05	0.98
		2	0.02	0.36	-0.03	0.95	0.04	0.35	0.05	0.96	-0.04	0.35	-0.18	0.97	0.03	0.35	0.05	0.99	0.00	0.35	-0.04	0.94
		4	-0.02	0.39	-0.10	0.96	-0.01	0.39	-0.06	0.96	0.00	0.39	-0.04	0.95	0.00	0.39	-0.03	0.97	0.00	0.39	-0.05	0.96
		0.25	0.45	0.43	-0.63	0.90	0.53	0.44	-0.44	0.94	0.58	0.44	-0.32	0.95	0.53	0.43	-0.45	0.93	0.60	0.44	-0.28	0.94
		0.5	0.51	0.40	-0.53	0.91	0.53	0.40	-0.48	0.91	0.61	0.40	-0.29	0.95	0.59	0.40	-0.34	0.94	0.63	0.40	-0.22	0.96
	2	1	0.54	0.39	-0.47	0.92	0.57	0.38	-0.43	0.92	0.64	0.39	-0.22	0.96	0.60	0.38	-0.31	0.94	0.71	0.39	-0.04	0.97
		2	0.50	0.40	-0.59	0.88	0.56	0.40	-0.40	0.95	0.63	0.40	-0.25	0.97	0.64	0.40	-0.22	0.96	0.62	0.40	-0.25	0.97
		4	0.43	0.42	-0.69	0.91	0.52	0.43	-0.46	0.93	0.57	0.44	-0.35	0.95	0.66	0.45	-0.14	0.96	0.67	0.45	-0.12	0.96
		0.25																				
		0.5																				
	4	1																				
		2																				
		4																				

Supplementary Table 2: Average estimates (Est), standard errors (SE), standardized bias (Bias) and coverage probability of the 95% confidence interval (CP) in covariate estimates for two covariates simulations. C-CB refers to the simulations with a continuous and a common binary covariate. C-RB refers to the simulations with a continuous and a rare binary covariate. Results for simulations with a rare binary covariate with OR of 4 are not presented for reasons described in Section 3.4.

			Continuous in C-CB																			
			Separation																			
			6				8				10				12				14			
Method	Binary OR	Continuous OR	Est	SE	Bias	CP	Est	SE	Bias	CP	Est	SE	Bias	CP	Est	SE	Bias	CP	Est	SE	Bias	CP
		0.25	-0.71	0.12	5.43	0.00	-0.84	0.13	4.29	0.00	-0.99	0.14	2.95	0.16	-1.13	0.14	1.90	0.53	-1.24	0.14	1.10	0.81
		0.5	-0.43	0.12	2.24	0.37	-0.48	0.12	1.80	0.58	-0.55	0.12	1.23	0.79	-0.61	0.12	0.75	0.92	-0.64	0.12	0.49	0.94
	0.25	1	0.00	0.12	0.02	1.00	0.00	0.11	-0.01	1.00	0.01	0.11	0.12	0.97	-0.03	0.11	-0.24	0.97	0.01	0.11	0.09	0.98
		2	0.43	0.12	-2.20	0.37	0.49	0.12	-1.74	0.60	0.54	0.12	-1.32	0.79	0.61	0.12	-0.74	0.94	0.63	0.12	-0.55	0.93
		4	0.72	0.13	-5.39	0.00	0.85	0.13	-4.12	0.01	0.99	0.14	-3.00	0.12	1.12	0.14	-1.94	0.53	1.24	0.14	-1.09	0.84
		0.25	-0.75	0.13	5.11	0.00	-0.87	0.13	4.11	0.01	-1.02	0.13	2.80	0.18	-1.16	0.14	1.68	0.61	-1.25	0.14	1.08	0.80
		0.5	-0.43	0.12	2.23	0.35	-0.49	0.12	1.77	0.58	-0.57	0.12	1.11	0.84	-0.61	0.11	0.78	0.90	-0.65	0.11	0.44	0.96
	0.5	1	-0.01	0.12	-0.07	1.00	0.01	0.11	0.05	0.98	0.00	0.11	0.04	0.99	0.01	0.10	0.08	0.98	0.00	0.10	0.03	0.97
		2	0.44	0.12	-2.15	0.43	0.51	0.12	-1.60	0.67	0.56	0.12	-1.20	0.83	0.62	0.11	-0.71	0.93	0.64	0.11	-0.48	0.93
		4	0.74	0.12	-5.21	0.00	0.86	0.13	-4.12	0.00	1.02	0.13	-2.78	0.22	1.14	0.14	-1.83	0.55	1.24	0.14	-1.13	0.81
		0.25	-0.75	0.12	5.14	0.00	-0.88	0.13	4.03	0.01	-1.01	0.13	2.88	0.16	-1.15	0.14	1.78	0.59	-1.25	0.14	1.05	0.82
		0.5	-0.45	0.12	2.12	0.42	-0.51	0.12	1.61	0.64	-0.56	0.11	1.15	0.84	-0.62	0.11	0.70	0.91	-0.65	0.11	0.41	0.92
PC3S	1	1	0.00	0.11	0.00	0.98	0.00	0.11	-0.02	0.99	-0.01	0.11	-0.06	0.97	0.00	0.10	0.01	0.98	0.00	0.10	-0.02	0.95
		2	0.44	0.12	-2.18	0.40	0.50	0.12	-1.69	0.63	0.56	0.11	-1.17	0.82	0.61	0.11	-0.80	0.90	0.65	0.11	-0.47	0.94
		4	0.76	0.12	-5.09	0.00	0.87	0.13	-4.13	0.01	1.02	0.13	-2.81	0.20	1.15	0.14	-1.82	0.56	1.26	0.14	-1.00	0.88
		0.25	-0.74	0.12	5.22	0.00	-0.89	0.13	3.90	0.02	-1.01	0.13	2.89	0.17	-1.13	0.14	1.93	0.55	-1.24	0.14	1.14	0.81
		0.5	-0.43	0.12	2.23	0.38	-0.50	0.12	1.72	0.63	-0.55	0.11	1.28	0.80	-0.61	0.11	0.73	0.91	-0.66	0.11	0.38	0.92
	2	1	0.01	0.11	0.07	0.99	0.00	0.11	0.03	1.00	0.00	0.11	-0.02	0.97	0.01	0.10	0.07	0.98	-0.01	0.10	-0.06	0.97
		2	0.43	0.12	-2.26	0.35	0.50	0.12	-1.67	0.62	0.56	0.12	-1.15	0.85	0.61	0.11	-0.75	0.91	0.65	0.11	-0.46	0.94
		4	0.74	0.12	-5.26	0.00	0.87	0.13	-4.08	0.00	1.01	0.13	-2.82	0.15	1.15	0.14	-1.74	0.61	1.24	0.14	-1.12	0.80
		0.25	-0.72	0.13	5.37	0.00	-0.85	0.13	4.15	0.01	-0.98	0.14	3.02	0.12	-1.12	0.14	1.99	0.50	-1.22	0.14	1.23	0.78
		0.5	-0.42	0.12	2.26	0.33	-0.48	0.12	1.80	0.57	-0.54	0.12	1.30	0.81	-0.59	0.12	0.89	0.91	-0.65	0.12	0.41	0.93
	4	1	-0.01	0.12	-0.09	1.00	0.00	0.11	-0.02	1.00	0.00	0.11	0.03	0.98	0.01	0.11	0.05	0.95	0.00	0.11	0.01	0.95
		2	0.42	0.12	-2.27	0.33	0.49	0.12	-1.73	0.62	0.54	0.12	-1.37	0.75	0.61	0.12	-0.78	0.89	0.63	0.12	-0.56	0.93
		4	0.73	0.13	-5.29	0.00	0.84	0.13	-4.24	0.00	1.01	0.14	-2.84	0.20	1.12	0.14	-2.00	0.45	1.22	0.14	-1.21	0.81

Supplementary Table 2: Average estimates (Est), standard errors (SE), standardized bias (Bias) and coverage probability of the 95% confidence interval (CP) in covariate estimates for two covariates simulations. C-CB refers to the simulations with a continuous and a common binary covariate. C-RB refers to the simulations with a continuous and a rare binary covariate. Results for simulations with a rare binary covariate with OR of 4 are not presented for reasons described in Section 3.4.

			Binary in C-CB																			
			Separation																			
			6				8				10				12				14			
Method	Binary OR	Continuous OR	Est	SE	Bias	CP	Est	SE	Bias	CP	Est	SE	Bias	CP	Est	SE	Bias	CP	Est	SE	Bias	CP
		0.25	-0.73	0.26	2.56	0.19	-0.84	0.26	2.14	0.42	-1.03	0.26	1.39	0.73	-1.15	0.26	0.95	0.88	-1.25	0.26	0.56	0.94
		0.5	-0.86	0.25	2.18	0.41	-0.98	0.24	1.69	0.61	-1.13	0.24	1.09	0.85	-1.23	0.23	0.71	0.92	-1.27	0.23	0.52	0.95
	0.25	1	-0.90	0.24	2.03	0.47	-1.01	0.23	1.64	0.68	-1.13	0.23	1.15	0.82	-1.25	0.22	0.62	0.93	-1.30	0.22	0.42	0.96
		2	-0.85	0.24	2.20	0.36	-0.98	0.24	1.71	0.62	-1.10	0.24	1.21	0.84	-1.20	0.23	0.85	0.87	-1.29	0.23	0.43	0.95
		4	-0.72	0.26	2.61	0.15	-0.86	0.26	2.06	0.47	-0.98	0.26	1.58	0.70	-1.12	0.26	1.04	0.81	-1.27	0.26	0.49	0.93
		0.25	-0.39	0.25	1.21	0.87	-0.45	0.25	0.98	0.92	-0.51	0.25	0.75	0.96	-0.60	0.25	0.40	0.94	-0.60	0.25	0.39	0.95
		0.5	-0.43	0.24	1.12	0.88	-0.50	0.23	0.85	0.93	-0.57	0.23	0.54	0.97	-0.60	0.22	0.43	0.95	-0.65	0.22	0.23	0.94
	0.5	1	-0.46	0.23	1.00	0.89	-0.52	0.22	0.78	0.92	-0.59	0.22	0.48	0.98	-0.63	0.21	0.29	0.96	-0.66	0.21	0.19	0.94
		2	-0.45	0.24	1.05	0.90	-0.50	0.23	0.87	0.92	-0.55	0.22	0.64	0.96	-0.61	0.22	0.41	0.97	-0.62	0.22	0.36	0.93
		4	-0.39	0.25	1.20	0.86	-0.42	0.25	1.12	0.87	-0.50	0.25	0.79	0.94	-0.55	0.25	0.59	0.95	-0.63	0.24	0.27	0.97
		0.25	0.01	0.25	0.02	1.00	0.00	0.25	0.00	1.00	0.00	0.24	0.01	0.98	0.00	0.24	-0.01	0.96	0.02	0.24	0.10	0.98
		0.5	0.00	0.23	0.02	1.00	-0.01	0.23	-0.03	0.99	-0.04	0.22	-0.17	0.96	-0.01	0.22	-0.05	0.95	0.02	0.21	0.07	0.95
PC3S	1	1	-0.01	0.23	-0.02	1.00	-0.01	0.22	-0.06	0.98	-0.01	0.21	-0.06	0.97	0.01	0.21	0.06	0.98	0.01	0.20	0.07	0.94
		2	0.01	0.23	0.03	1.00	0.01	0.23	0.05	0.99	0.01	0.22	0.05	0.99	0.00	0.22	0.01	0.97	-0.01	0.21	-0.06	0.98
		4	0.01	0.25	0.06	1.00	0.01	0.25	0.05	0.99	0.01	0.24	0.05	0.99	-0.01	0.24	-0.03	0.98	0.02	0.24	0.08	0.98
		0.25	0.37	0.25	-1.28	0.84	0.43	0.25	-1.05	0.87	0.52	0.25	-0.70	0.94	0.55	0.25	-0.59	0.94	0.58	0.24	-0.48	0.95
		0.5	0.44	0.24	-1.08	0.89	0.49	0.23	-0.89	0.89	0.56	0.22	-0.59	0.96	0.62	0.22	-0.35	0.96	0.65	0.22	-0.23	0.95
	2	1	0.48	0.23	-0.96	0.93	0.52	0.22	-0.81	0.95	0.59	0.22	-0.46	0.97	0.61	0.21	-0.39	0.96	0.67	0.21	-0.11	0.98
		2	0.44	0.24	-1.09	0.91	0.49	0.23	-0.88	0.94	0.56	0.22	-0.58	0.94	0.61	0.22	-0.38	0.97	0.65	0.22	-0.22	0.98
		4	0.37	0.25	-1.29	0.83	0.43	0.25	-1.05	0.91	0.51	0.25	-0.75	0.93	0.59	0.25	-0.45	0.94	0.64	0.25	-0.23	0.97
		0.25	0.72	0.26	-2.60	0.18	0.84	0.26	-2.13	0.44	1.03	0.26	-1.38	0.79	1.16	0.26	-0.92	0.89	1.23	0.26	-0.63	0.91
		0.5	0.85	0.24	-2.20	0.38	1.00	0.24	-1.63	0.67	1.07	0.24	-1.35	0.74	1.22	0.23	-0.72	0.93	1.27	0.23	-0.51	0.96
	4	1	0.90	0.24	-2.05	0.46	1.02	0.23	-1.60	0.69	1.16	0.23	-1.01	0.87	1.23	0.22	-0.73	0.91	1.31	0.22	-0.37	0.96
		2	0.85	0.25	-2.19	0.37	0.98	0.24	-1.69	0.64	1.12	0.24	-1.15	0.82	1.21	0.23	-0.76	0.93	1.29	0.23	-0.44	0.97
		4	0.72	0.26	-2.59	0.17	0.86	0.26	-2.07	0.48	1.00	0.26	-1.51	0.70	1.13	0.26	-1.02	0.88	1.23	0.26	-0.63	0.92

Supplementary Table 2: Average estimates (Est), standard errors (SE), standardized bias (Bias) and coverage probability of the 95% confidence interval (CP) in covariate estimates for two covariates simulations. C-CB refers to the simulations with a continuous and a common binary covariate. C-RB refers to the simulations with a continuous and a rare binary covariate. Results for simulations with a rare binary covariate with OR of 4 are not presented for reasons described in Section 3.4.

			Continuous in C-RB																			
			Separation																			
			6				8				10				12				14			
Method	Binary OR	Continuous OR	Est	SE	Bias	CP	Est	SE	Bias	CP	Est	SE	Bias	CP	Est	SE	Bias	CP	Est	SE	Bias	CP
		0.25	-0.74	0.12	5.24	0.00	-0.86	0.13	4.13	0.00	-1.01	0.13	2.86	0.14	-1.14	0.14	1.87	0.51	-1.24	0.14	1.12	0.83
		0.5	-0.44	0.12	2.14	0.39	-0.50	0.12	1.64	0.67	-0.56	0.12	1.16	0.84	-0.60	0.11	0.84	0.89	-0.64	0.11	0.46	0.97
	0.25	1	0.00	0.12	-0.02	1.00	0.01	0.11	0.08	0.99	0.00	0.11	-0.02	1.00	0.00	0.11	0.00	0.96	0.00	0.10	-0.01	0.97
		2	0.43	0.12	-2.27	0.34	0.49	0.12	-1.74	0.60	0.56	0.12	-1.16	0.83	0.60	0.11	-0.83	0.88	0.65	0.11	-0.41	0.96
		4	0.73	0.12	-5.30	0.00	0.86	0.13	-4.08	0.01	1.00	0.13	-2.93	0.16	1.13	0.14	-1.92	0.51	1.23	0.14	-1.20	0.76
		0.25	-0.75	0.12	5.12	0.00	-0.88	0.13	3.94	0.00	-1.02	0.13	2.76	0.17	-1.15	0.14	1.80	0.55	-1.24	0.14	1.10	0.79
		0.5	-0.46	0.12	1.99	0.49	-0.50	0.12	1.72	0.61	-0.56	0.11	1.16	0.84	-0.61	0.11	0.80	0.92	-0.66	0.11	0.36	0.93
	0.5	1	0.00	0.11	-0.02	1.00	0.00	0.11	0.04	0.98	-0.01	0.11	-0.09	0.98	0.01	0.10	0.12	0.98	0.00	0.10	0.01	0.94
		2	0.44	0.12	-2.21	0.40	0.51	0.12	-1.63	0.66	0.56	0.11	-1.16	0.84	0.61	0.11	-0.73	0.90	0.65	0.11	-0.42	0.96
		4	0.74	0.12	-5.22	0.00	0.87	0.13	-4.06	0.00	1.01	0.13	-2.86	0.19	1.15	0.14	-1.78	0.57	1.25	0.14	-1.00	0.88
		0.25	-0.75	0.12	5.17	0.00	-0.88	0.13	3.99	0.01	-1.02	0.13	2.82	0.18	-1.15	0.14	1.76	0.63	-1.24	0.14	1.09	0.82
		0.5	-0.44	0.12	2.15	0.44	-0.51	0.12	1.63	0.65	-0.55	0.11	1.27	0.77	-0.61	0.11	0.76	0.92	-0.65	0.11	0.45	0.95
PC3S	1	1	0.00	0.11	0.02	1.00	0.00	0.11	-0.04	0.99	0.01	0.11	0.06	0.98	0.00	0.10	-0.02	0.99	0.00	0.10	-0.03	0.97
		2	0.45	0.12	-2.11	0.42	0.50	0.12	-1.74	0.62	0.58	0.11	-1.05	0.88	0.62	0.11	-0.71	0.92	0.64	0.11	-0.52	0.93
		4	0.75	0.12	-5.17	0.00	0.88	0.13	-4.03	0.02	1.03	0.13	-2.73	0.22	1.15	0.14	-1.82	0.58	1.24	0.14	-1.14	0.83
		0.25	-0.76	0.13	5.04	0.00	-0.86	0.13	4.16	0.01	-1.02	0.13	2.81	0.16	-1.15	0.14	1.82	0.60	-1.25	0.14	1.09	0.80
		0.5	-0.44	0.12	2.16	0.40	-0.51	0.12	1.59	0.70	-0.56	0.11	1.17	0.86	-0.62	0.11	0.67	0.88	-0.64	0.11	0.48	0.93
	2	1	-0.01	0.11	-0.05	1.00	0.01	0.11	0.05	0.97	0.00	0.11	0.01	0.98	0.00	0.10	-0.04	0.98	-0.01	0.10	-0.08	0.96
		2	0.44	0.12	-2.20	0.37	0.49	0.12	-1.75	0.63	0.56	0.11	-1.16	0.83	0.61	0.11	-0.75	0.90	0.64	0.11	-0.48	0.96
		4	0.75	0.12	-5.19	0.00	0.87	0.13	-4.11	0.01	1.03	0.13	-2.71	0.21	1.14	0.14	-1.86	0.58	1.24	0.14	-1.11	0.83
		0.25	-0.74	0.12	5.25	0.00	-0.88	0.13	4.03	0.02	-1.00	0.13	2.95	0.15	-1.14	0.14	1.86	0.50	-1.24	0.14	1.13	0.77
		0.5	-0.44	0.12	2.18	0.42	-0.49	0.12	1.81	0.56	-0.56	0.12	1.15	0.85	-0.60	0.11	0.81	0.92	-0.66	0.11	0.35	0.96
	4	1	0.01	0.11	0.05	0.99	-0.01	0.11	-0.06	0.99	0.01	0.11	0.05	0.98	-0.01	0.10	-0.08	0.98	-0.01	0.10	-0.07	0.97
		2	0.44	0.12	-2.14	0.41	0.51	0.12	-1.61	0.68	0.55	0.11	-1.24	0.81	0.61	0.11	-0.80	0.89	0.65	0.11	-0.42	0.92
		4	0.74	0.12	-5.25	0.00	0.85	0.13	-4.24	0.00	1.01	0.13	-2.89	0.20	1.14	0.14	-1.86	0.55	1.26	0.14	-0.93	0.86

Supplementary Table 2: Average estimates (Est), standard errors (SE), standardized bias (Bias) and coverage probability of the 95% confidence interval (CP) in covariate estimates for two covariates simulations. C-CB refers to the simulations with a continuous and a common binary covariate. C-RB refers to the simulations with a continuous and a rare binary covariate. Results for simulations with a rare binary covariate with OR of 4 are not presented for reasons described in Section 3.4.

			Binary in C-RB																			
			Separation																			
			6				8				10				12				14			
Method	Binary OR	Continuous OR	Est	SE	Bias	CP	Est	SE	Bias	CP	Est	SE	Bias	CP	Est	SE	Bias	CP	Est	SE	Bias	CP
		0.25	-0.80	0.38	1.52	0.71	-0.89	0.38	1.31	0.77	-0.99	0.38	1.03	0.86	-1.11	0.38	0.73	0.93	-1.26	0.38	0.36	0.95
		0.5	-0.90	0.36	1.37	0.82	-1.02	0.35	1.04	0.88	-1.12	0.35	0.78	0.93	-1.27	0.34	0.37	0.96	-1.29	0.33	0.29	0.95
	0.25	1	-0.91	0.35	1.37	0.80	-1.07	0.34	0.94	0.92	-1.19	0.33	0.61	0.93	-1.29	0.32	0.33	0.96	-1.33	0.32	0.20	0.96
		2	-0.84	0.36	1.51	0.71	-0.99	0.35	1.12	0.86	-1.13	0.35	0.75	0.94	-1.21	0.34	0.54	0.95	-1.31	0.34	0.23	0.96
		4	-0.80	0.38	1.53	0.73	-0.84	0.39	1.41	0.74	-1.03	0.39	0.93	0.88	-1.16	0.38	0.60	0.94	-1.24	0.39	0.40	0.98
		0.25	-0.38	0.39	0.79	0.97	-0.40	0.39	0.73	0.98	-0.47	0.39	0.55	0.98	-0.58	0.39	0.29	0.99	-0.60	0.38	0.23	0.95
		0.5	-0.42	0.37	0.73	0.95	-0.49	0.36	0.55	0.96	-0.55	0.35	0.39	0.97	-0.65	0.34	0.12	0.97	-0.64	0.34	0.13	0.97
	0.5	1	-0.47	0.36	0.61	0.98	-0.53	0.34	0.45	0.97	-0.59	0.33	0.29	0.99	-0.63	0.33	0.20	0.97	-0.63	0.32	0.18	0.97
		2	-0.44	0.37	0.66	0.98	-0.46	0.36	0.63	0.97	-0.58	0.35	0.31	0.97	-0.62	0.34	0.21	0.97	-0.66	0.33	0.10	0.99
		4	-0.37	0.40	0.80	0.96	-0.40	0.39	0.72	0.97	-0.55	0.39	0.35	0.99	-0.58	0.39	0.29	0.98	-0.64	0.39	0.14	0.94
		0.25	0.02	0.41	0.02	1.00	0.01	0.42	-0.01	1.00	0.00	0.41	-0.05	0.99	-0.02	0.41	-0.10	0.97	0.00	0.40	-0.04	0.96
		0.5	0.03	0.40	0.04	1.00	-0.01	0.38	-0.06	0.98	0.00	0.37	-0.03	0.98	0.01	0.37	-0.02	0.97	0.00	0.36	-0.05	0.98
PC3S	1	1	-0.01	0.38	-0.08	0.99	0.00	0.38	-0.05	1.00	0.04	0.36	0.06	0.99	0.03	0.35	0.04	0.98	0.03	0.34	0.03	0.99
		2	0.01	0.40	-0.01	0.99	0.02	0.38	0.03	1.00	-0.04	0.37	-0.14	0.98	0.03	0.37	0.04	1.00	0.00	0.36	-0.04	0.96
		4	0.00	0.42	-0.02	1.00	0.00	0.42	-0.03	1.00	0.00	0.41	-0.04	0.98	0.00	0.41	-0.03	0.98	0.00	0.40	-0.04	0.98
		0.25	0.35	0.46	-0.78	0.96	0.43	0.47	-0.60	0.97	0.50	0.46	-0.46	0.95	0.49	0.45	-0.51	0.95	0.57	0.45	-0.34	0.95
		0.5	0.42	0.44	-0.67	0.96	0.46	0.43	-0.59	0.96	0.54	0.42	-0.43	0.95	0.56	0.42	-0.40	0.95	0.62	0.41	-0.25	0.98
	2	1	0.44	0.43	-0.64	0.97	0.48	0.41	-0.58	0.95	0.57	0.41	-0.35	0.97	0.57	0.40	-0.37	0.97	0.69	0.40	-0.10	0.97
		2	0.39	0.43	-0.75	0.95	0.47	0.43	-0.56	0.96	0.57	0.42	-0.37	0.98	0.60	0.41	-0.30	0.96	0.60	0.41	-0.28	0.98
		4	0.31	0.46	-0.88	0.94	0.43	0.46	-0.61	0.97	0.49	0.46	-0.48	0.96	0.60	0.47	-0.25	0.96	0.64	0.46	-0.18	0.98
		0.25																				
		0.5																				
	4	1																				
		2																				
		4																				

Supplementary Table 2: Average estimates (Est), standard errors (SE), standardized bias (Bias) and coverage probability of the 95% confidence interval (CP) in covariate estimates for two covariates simulations. C-CB refers to the simulations with a continuous and a common binary covariate. C-RB refers to the simulations with a continuous and a rare binary covariate. Results for simulations with a rare binary covariate with OR of 4 are not presented for reasons described in Section 3.4.

			Continuous in C-CB																			
			Separation																			
			6				8				10				12				14			
Method	Binary OR	Continuous OR	Est	SE	Bias	CP	Est	SE	Bias	CP	Est	SE	Bias	CP	Est	SE	Bias	CP	Est	SE	Bias	CP
		0.25	-1.39	0.23	0.15	0.92	-1.39	0.21	0.12	0.96	-1.41	0.19	-0.02	0.95	-1.41	0.18	-0.05	0.95	-1.41	0.16	-0.09	0.98
		0.5	-0.70	0.16	0.01	1.00	-0.70	0.15	0.02	0.95	-0.71	0.13	-0.04	0.91	-0.71	0.13	-0.08	0.96	-0.70	0.12	-0.01	0.94
	0.25	1	0.00	0.14	0.02	0.95	0.00	0.13	-0.04	0.95	0.02	0.12	0.16	0.91	-0.03	0.11	-0.27	0.94	0.01	0.11	0.09	0.95
		2	0.71	0.16	0.03	0.97	0.71	0.15	0.04	0.98	0.69	0.13	-0.08	0.95	0.71	0.13	0.11	0.95	0.69	0.12	-0.08	0.98
		4	1.41	0.23	-0.06	0.96	1.42	0.21	0.04	0.94	1.40	0.19	-0.04	0.97	1.40	0.17	-0.03	0.97	1.42	0.16	0.09	0.95
		0.25	-1.42	0.22	-0.04	0.96	-1.40	0.20	0.04	0.96	-1.40	0.18	-0.01	0.98	-1.43	0.17	-0.16	0.94	-1.40	0.15	0.00	0.95
		0.5	-0.70	0.15	0.05	0.97	-0.69	0.14	0.07	0.96	-0.71	0.13	-0.05	0.94	-0.69	0.12	0.03	0.96	-0.70	0.11	0.01	0.96
	0.5	1	-0.02	0.13	-0.12	0.94	0.01	0.12	0.04	0.96	0.00	0.11	0.04	0.96	0.01	0.11	0.08	0.96	0.00	0.10	0.02	0.97
		2	0.70	0.15	-0.01	0.96	0.71	0.14	0.07	0.93	0.70	0.13	-0.04	0.95	0.71	0.12	0.06	0.95	0.69	0.12	-0.04	0.95
		4	1.41	0.23	-0.03	0.97	1.40	0.20	-0.08	0.96	1.42	0.18	0.05	0.94	1.41	0.16	0.02	0.97	1.39	0.15	-0.03	0.94
		0.25	-1.40	0.22	0.06	0.97	-1.40	0.20	0.05	0.94	-1.39	0.18	0.11	0.95	-1.41	0.16	-0.03	0.97	-1.40	0.15	-0.03	0.95
		0.5	-0.70	0.15	0.04	0.95	-0.71	0.14	-0.07	0.94	-0.70	0.13	-0.04	0.98	-0.70	0.12	-0.04	0.95	-0.70	0.11	-0.03	0.94
I3S	1	1	0.00	0.13	0.02	0.96	0.00	0.12	-0.03	0.97	-0.01	0.11	-0.07	0.94	0.00	0.11	-0.01	0.94	0.00	0.10	-0.02	0.95
		2	0.69	0.15	-0.09	0.93	0.70	0.14	-0.02	0.95	0.70	0.13	-0.04	0.96	0.69	0.12	-0.04	0.95	0.69	0.11	-0.04	0.95
		4	1.41	0.22	-0.03	0.96	1.37	0.19	-0.21	0.95	1.40	0.18	-0.05	0.93	1.40	0.16	-0.04	0.93	1.41	0.15	0.08	0.95
		0.25	-1.39	0.22	0.17	0.95	-1.43	0.20	-0.10	0.96	-1.39	0.18	0.10	0.96	-1.39	0.16	0.10	0.93	-1.39	0.15	0.07	0.96
		0.5	-0.69	0.15	0.10	0.94	-0.70	0.14	0.04	0.95	-0.68	0.13	0.15	0.96	-0.70	0.12	-0.03	0.93	-0.71	0.12	-0.05	0.92
	2	1	0.01	0.13	0.08	0.90	0.00	0.12	0.03	0.98	0.00	0.11	0.00	0.94	0.01	0.11	0.07	0.96	-0.01	0.10	-0.06	0.95
		2	0.69	0.15	-0.14	0.96	0.70	0.14	-0.01	0.93	0.70	0.13	0.03	0.97	0.70	0.12	0.05	0.97	0.70	0.12	-0.01	0.97
		4	1.39	0.23	-0.13	0.94	1.39	0.20	-0.08	0.98	1.41	0.18	0.01	0.95	1.42	0.17	0.13	0.96	1.40	0.16	-0.01	0.94
		0.25	-1.42	0.24	0.04	0.96	-1.42	0.21	0.01	0.96	-1.39	0.19	0.08	0.97	-1.40	0.17	0.01	0.95	-1.39	0.16	0.08	0.95
		0.5	-0.70	0.16	0.02	0.95	-0.70	0.14	0.02	0.96	-0.69	0.13	0.08	0.95	-0.69	0.13	0.08	0.96	-0.71	0.12	-0.08	0.94
	4	1	-0.02	0.14	-0.14	0.97	-0.01	0.13	-0.05	0.97	0.00	0.12	0.02	0.93	0.00	0.11	0.04	0.92	0.00	0.11	0.01	0.94
		2	0.70	0.16	-0.05	0.93	0.71	0.15	0.02	0.95	0.69	0.13	-0.08	0.95	0.71	0.13	0.04	0.94	0.69	0.12	-0.07	0.94
		4	1.43	0.24	0.02	0.96	1.40	0.21	-0.08	0.95	1.44	0.19	0.12	0.93	1.40	0.17	-0.02	0.96	1.39	0.16	-0.06	0.96

Supplementary Table 2: Average estimates (Est), standard errors (SE), standardized bias (Bias) and coverage probability of the 95% confidence interval (CP) in covariate estimates for two covariates simulations. C-CB refers to the simulations with a continuous and a common binary covariate. C-RB refers to the simulations with a continuous and a rare binary covariate. Results for simulations with a rare binary covariate with OR of 4 are not presented for reasons described in Section 3.4.

			Binary in C-CB																			
			Separation																			
			6				8				10				12				14			
Method	Binary OR	Continuous OR	Est	SE	Bias	CP	Est	SE	Bias	CP	Est	SE	Bias	CP	Est	SE	Bias	CP	Est	SE	Bias	CP
		0.25	-1.39	0.37	0.06	0.97	-1.37	0.34	0.10	0.95	-1.44	0.31	-0.10	0.95	-1.43	0.29	-0.09	0.94	-1.41	0.28	-0.06	0.94
		0.5	-1.39	0.31	0.05	0.94	-1.41	0.28	-0.02	0.96	-1.43	0.26	-0.11	0.95	-1.42	0.25	-0.09	0.98	-1.38	0.24	0.06	0.95
	0.25	1	-1.41	0.29	-0.01	0.97	-1.37	0.26	0.10	0.96	-1.37	0.25	0.10	0.95	-1.41	0.23	-0.07	0.97	-1.39	0.22	0.02	0.99
		2	-1.39	0.31	0.06	0.96	-1.41	0.28	-0.02	0.95	-1.39	0.26	0.02	0.97	-1.38	0.25	0.08	0.93	-1.40	0.24	-0.03	0.95
		4	-1.37	0.37	0.12	0.99	-1.40	0.34	0.01	0.94	-1.36	0.31	0.14	0.93	-1.39	0.29	0.05	0.94	-1.44	0.28	-0.13	0.96
		0.25	-0.74	0.34	-0.10	0.96	-0.72	0.31	-0.06	0.98	-0.70	0.29	0.01	0.97	-0.73	0.27	-0.12	0.92	-0.67	0.26	0.11	0.95
		0.5	-0.68	0.28	0.07	0.94	-0.70	0.26	0.00	0.96	-0.71	0.24	-0.06	0.97	-0.68	0.23	0.06	0.95	-0.70	0.22	0.01	0.93
	0.5	1	-0.69	0.26	0.06	0.95	-0.70	0.24	0.00	0.93	-0.71	0.23	-0.05	0.97	-0.71	0.22	-0.06	0.94	-0.70	0.21	-0.01	0.93
		2	-0.71	0.28	-0.03	0.96	-0.70	0.26	0.00	0.95	-0.69	0.24	0.04	0.96	-0.69	0.23	0.04	0.96	-0.67	0.22	0.13	0.94
		4	-0.74	0.34	-0.10	0.96	-0.66	0.31	0.13	0.98	-0.69	0.29	0.03	0.96	-0.67	0.27	0.12	0.96	-0.70	0.26	-0.02	0.97
		0.25	0.01	0.33	0.02	0.97	0.00	0.30	0.01	0.97	0.00	0.28	-0.01	0.94	0.00	0.26	-0.02	0.93	0.02	0.25	0.08	0.96
		0.5	0.00	0.27	-0.01	0.96	-0.01	0.25	-0.02	0.96	-0.05	0.24	-0.19	0.92	-0.01	0.22	-0.06	0.92	0.02	0.22	0.08	0.95
I3S	1	1	0.00	0.25	0.01	0.95	-0.02	0.24	-0.07	0.93	-0.01	0.22	-0.06	0.93	0.02	0.21	0.08	0.95	0.02	0.20	0.08	0.92
		2	0.01	0.27	0.03	0.96	0.01	0.25	0.04	0.94	0.01	0.23	0.04	0.95	0.00	0.22	0.01	0.96	-0.01	0.22	-0.06	0.98
		4	0.03	0.33	0.11	0.97	0.01	0.30	0.04	0.95	0.02	0.28	0.05	0.95	-0.01	0.26	-0.03	0.96	0.02	0.25	0.09	0.97
		0.25	0.71	0.34	0.00	0.97	0.69	0.31	-0.04	0.94	0.73	0.29	0.10	0.93	0.67	0.27	-0.09	0.94	0.65	0.26	-0.20	0.95
		0.5	0.70	0.28	0.00	0.95	0.68	0.26	-0.09	0.95	0.69	0.24	-0.02	0.94	0.71	0.23	0.04	0.98	0.69	0.22	-0.01	0.95
	2	1	0.71	0.26	0.04	0.94	0.69	0.24	-0.02	0.97	0.72	0.23	0.09	0.95	0.69	0.22	-0.02	0.96	0.72	0.21	0.12	0.96
		2	0.70	0.28	-0.01	0.94	0.69	0.26	-0.04	0.97	0.70	0.24	0.01	0.97	0.70	0.23	0.01	0.94	0.70	0.22	0.02	0.97
		4	0.70	0.34	-0.03	0.96	0.70	0.31	-0.02	0.94	0.70	0.29	0.01	0.96	0.72	0.27	0.08	0.97	0.73	0.26	0.11	0.97
		0.25	1.37	0.37	-0.12	0.96	1.37	0.34	-0.11	0.95	1.45	0.31	0.15	0.94	1.44	0.29	0.15	0.94	1.39	0.27	-0.03	0.95
		0.5	1.39	0.31	-0.05	0.96	1.42	0.28	0.07	0.97	1.36	0.26	-0.15	0.92	1.41	0.25	0.07	0.95	1.38	0.24	-0.05	0.98
	4	1	1.39	0.29	-0.06	0.95	1.39	0.26	-0.03	0.94	1.42	0.25	0.10	0.97	1.38	0.23	-0.06	0.95	1.40	0.23	0.04	0.96
		2	1.37	0.31	-0.09	0.95	1.41	0.28	0.02	0.96	1.41	0.26	0.03	0.95	1.40	0.25	0.01	0.96	1.40	0.24	0.02	0.98
		4	1.38	0.37	-0.08	0.96	1.39	0.34	-0.05	0.96	1.41	0.31	0.03	0.94	1.39	0.29	-0.01	0.99	1.39	0.28	-0.02	0.96

Supplementary Table 2: Average estimates (Est), standard errors (SE), standardized bias (Bias) and coverage probability of the 95% confidence interval (CP) in covariate estimates for two covariates simulations. C-CB refers to the simulations with a continuous and a common binary covariate. C-RB refers to the simulations with a continuous and a rare binary covariate. Results for simulations with a rare binary covariate with OR of 4 are not presented for reasons described in Section 3.4.

			Continuous in C-RB																			
			Separation																			
			6				8				10				12				14			
Method	Binary OR	Continuous OR	Est	SE	Bias	CP	Est	SE	Bias	CP	Est	SE	Bias	CP	Est	SE	Bias	CP	Est	SE	Bias	CP
		0.25	-1.40	0.22	0.10	0.97	-1.40	0.21	0.07	0.97	-1.40	0.18	0.06	0.96	-1.40	0.16	0.03	0.97	-1.40	0.16	-0.01	0.95
		0.5	-0.72	0.16	-0.07	0.95	-0.71	0.14	-0.09	0.98	-0.71	0.13	-0.06	0.97	-0.69	0.12	0.05	0.94	-0.70	0.12	-0.01	0.97
	0.25	1	0.00	0.13	-0.04	0.95	0.01	0.12	0.09	0.96	0.00	0.11	-0.02	0.97	0.00	0.11	-0.01	0.94	0.00	0.11	-0.02	0.97
		2	0.68	0.15	-0.13	0.95	0.69	0.14	-0.07	0.96	0.70	0.13	0.01	0.95	0.69	0.12	-0.06	0.95	0.70	0.12	0.05	0.96
		4	1.40	0.23	-0.09	0.98	1.41	0.21	-0.01	0.98	1.40	0.18	-0.07	0.96	1.40	0.17	-0.05	0.96	1.38	0.15	-0.11	0.92
		0.25	-1.41	0.22	0.00	0.98	-1.42	0.20	-0.05	0.96	-1.41	0.18	-0.03	0.95	-1.41	0.17	-0.05	0.94	-1.40	0.15	0.02	0.96
		0.5	-0.74	0.15	-0.23	0.94	-0.69	0.14	0.08	0.95	-0.70	0.13	0.00	0.95	-0.70	0.12	0.03	0.97	-0.71	0.12	-0.07	0.96
	0.5	1	0.00	0.13	0.01	0.96	0.01	0.12	0.05	0.96	-0.01	0.11	-0.10	0.97	0.01	0.11	0.13	0.97	0.00	0.10	0.00	0.93
		2	0.69	0.15	-0.09	0.96	0.70	0.14	0.02	0.96	0.70	0.13	0.00	0.97	0.70	0.12	0.01	0.94	0.70	0.12	0.02	0.96
		4	1.37	0.22	-0.23	0.92	1.40	0.20	-0.07	0.95	1.39	0.18	-0.08	0.95	1.41	0.17	0.05	0.96	1.41	0.16	0.10	0.96
		0.25	-1.38	0.22	0.15	0.97	-1.41	0.20	0.02	0.97	-1.40	0.18	0.06	0.93	-1.41	0.16	-0.06	0.96	-1.40	0.15	0.02	0.95
		0.5	-0.69	0.15	0.08	0.96	-0.70	0.13	-0.03	0.98	-0.68	0.13	0.14	0.95	-0.70	0.12	-0.02	0.96	-0.69	0.11	0.04	0.95
I3S	1	1	0.00	0.13	0.02	0.97	0.00	0.12	-0.04	0.96	0.01	0.11	0.06	0.96	0.00	0.11	-0.02	0.97	0.00	0.10	-0.02	0.97
		2	0.71	0.15	0.04	0.95	0.69	0.14	-0.09	0.97	0.71	0.13	0.10	0.98	0.70	0.12	0.04	0.95	0.69	0.11	-0.09	0.95
		4	1.39	0.22	-0.11	0.96	1.40	0.20	-0.09	0.95	1.40	0.18	-0.01	0.97	1.40	0.16	-0.01	0.94	1.39	0.15	-0.07	0.97
		0.25	-1.42	0.22	-0.04	0.98	-1.37	0.19	0.24	0.96	-1.39	0.18	0.10	0.93	-1.40	0.16	0.00	0.96	-1.40	0.15	0.03	0.92
		0.5	-0.70	0.15	0.01	0.96	-0.71	0.14	-0.07	0.94	-0.69	0.13	0.05	0.97	-0.71	0.12	-0.11	0.93	-0.69	0.11	0.04	0.95
	2	1	-0.01	0.13	-0.03	0.97	0.01	0.12	0.05	0.96	0.00	0.11	-0.02	0.94	0.00	0.11	-0.04	0.96	-0.01	0.10	-0.09	0.96
		2	0.70	0.15	-0.07	0.94	0.69	0.14	-0.12	0.95	0.70	0.13	-0.03	0.95	0.70	0.12	0.00	0.97	0.69	0.11	-0.07	0.95
		4	1.38	0.22	-0.16	0.94	1.39	0.20	-0.13	0.93	1.42	0.18	0.06	0.97	1.40	0.17	-0.01	0.97	1.39	0.15	-0.04	0.96
		0.25	-1.36	0.21	0.26	0.93	-1.38	0.20	0.14	0.96	-1.38	0.18	0.14	0.94	-1.40	0.17	-0.01	0.96	-1.39	0.15	0.04	0.94
		0.5	-0.70	0.15	0.03	0.97	-0.69	0.14	0.11	0.96	-0.71	0.13	-0.03	0.94	-0.69	0.12	0.07	0.96	-0.71	0.12	-0.10	0.96
	4	1	0.01	0.13	0.07	0.95	-0.01	0.12	-0.09	0.95	0.01	0.11	0.06	0.96	-0.01	0.11	-0.09	0.95	-0.01	0.10	-0.05	0.96
		2	0.70	0.15	-0.05	0.93	0.71	0.14	0.08	0.96	0.69	0.13	-0.06	0.96	0.70	0.12	-0.02	0.95	0.70	0.11	0.02	0.94
		4	1.41	0.22	-0.03	0.96	1.37	0.20	-0.24	0.94	1.40	0.18	-0.07	0.94	1.40	0.17	-0.02	0.96	1.43	0.16	0.16	0.95

Supplementary Table 2: Average estimates (Est), standard errors (SE), standardized bias (Bias) and coverage probability of the 95% confidence interval (CP) in covariate estimates for two covariates simulations. C-CB refers to the simulations with a continuous and a common binary covariate. C-RB refers to the simulations with a continuous and a rare binary covariate. Results for simulations with a rare binary covariate with OR of 4 are not presented for reasons described in Section 3.4.

			Binary in C-RB																			
			Separation																			
			6				8				10				12				14			
Method	Binary OR	Continuous OR	Est	SE	Bias	CP	Est	SE	Bias	CP	Est	SE	Bias	CP	Est	SE	Bias	CP	Est	SE	Bias	CP
		0.25	-1.48	0.57	-0.12	0.98	-1.42	0.51	0.01	0.95	-1.36	0.46	0.11	0.95	-1.37	0.43	0.07	0.94	-1.43	0.41	-0.08	0.95
		0.5	-1.43	0.47	-0.02	0.96	-1.42	0.43	-0.03	0.97	-1.39	0.39	0.04	0.98	-1.44	0.37	-0.11	0.96	-1.39	0.35	0.00	0.95
	0.25	1	-1.36	0.43	0.15	0.97	-1.41	0.40	-0.01	0.98	-1.43	0.37	-0.06	0.98	-1.43	0.34	-0.09	0.98	-1.41	0.33	-0.05	0.97
		2	-1.33	0.46	0.20	0.94	-1.37	0.43	0.09	0.99	-1.41	0.39	-0.02	0.98	-1.39	0.37	0.04	0.95	-1.42	0.35	-0.05	0.95
		4	-1.50	0.56	-0.13	0.94	-1.37	0.52	0.09	0.98	-1.41	0.47	0.00	0.97	-1.44	0.44	-0.06	0.95	-1.40	0.41	0.00	0.98
		0.25	-0.69	0.54	0.01	0.95	-0.63	0.49	0.12	0.96	-0.63	0.45	0.12	0.97	-0.71	0.43	-0.04	0.95	-0.68	0.40	0.03	0.96
		0.5	-0.65	0.45	0.10	0.97	-0.68	0.40	0.03	0.96	-0.68	0.38	0.03	0.95	-0.74	0.35	-0.13	0.95	-0.68	0.34	0.01	0.95
	0.5	1	-0.69	0.41	0.00	0.96	-0.70	0.37	-0.02	0.98	-0.71	0.35	-0.05	0.98	-0.70	0.34	-0.02	0.96	-0.67	0.32	0.07	0.96
		2	-0.69	0.44	0.01	0.97	-0.63	0.41	0.14	0.97	-0.72	0.37	-0.07	0.94	-0.70	0.36	-0.04	0.96	-0.70	0.34	-0.04	0.97
		4	-0.68	0.54	0.05	0.98	-0.64	0.49	0.11	0.96	-0.75	0.45	-0.12	0.96	-0.72	0.43	-0.06	0.93	-0.72	0.41	-0.05	0.93
		0.25	0.05	0.54	0.02	0.96	0.01	0.51	-0.01	0.96	0.01	0.47	-0.04	0.92	-0.03	0.44	-0.10	0.94	0.00	0.42	-0.04	0.93
		0.5	0.04	0.47	0.03	0.97	-0.01	0.43	-0.07	0.95	0.00	0.39	-0.06	0.97	0.01	0.38	-0.02	0.95	0.00	0.36	-0.05	0.96
I3S	1	1	-0.01	0.43	-0.10	0.96	0.00	0.41	-0.07	0.97	0.03	0.38	0.02	0.96	0.03	0.36	0.04	0.97	0.03	0.35	0.05	0.98
		2	0.02	0.46	-0.03	0.96	0.04	0.42	0.06	0.97	-0.05	0.40	-0.18	0.96	0.04	0.38	0.06	0.98	0.00	0.36	-0.04	0.95
		4	-0.02	0.55	-0.08	0.96	0.00	0.50	-0.04	0.96	0.00	0.47	-0.03	0.96	0.00	0.44	-0.03	0.96	0.00	0.42	-0.04	0.96
		0.25	0.68	0.63	-0.12	0.95	0.70	0.58	-0.05	0.98	0.70	0.53	-0.06	0.95	0.60	0.50	-0.26	0.94	0.65	0.47	-0.17	0.95
		0.5	0.66	0.53	-0.16	0.96	0.65	0.49	-0.19	0.97	0.69	0.47	-0.09	0.97	0.64	0.43	-0.21	0.95	0.66	0.41	-0.15	0.96
	2	1	0.68	0.51	-0.15	0.96	0.68	0.47	-0.17	0.96	0.71	0.44	-0.05	0.97	0.65	0.41	-0.20	0.97	0.74	0.41	0.02	0.98
		2	0.65	0.55	-0.22	0.95	0.69	0.50	-0.10	0.95	0.71	0.46	-0.06	0.98	0.69	0.43	-0.09	0.96	0.65	0.41	-0.17	0.96
		4	0.62	0.62	-0.19	0.96	0.69	0.59	-0.08	0.98	0.70	0.53	-0.07	0.96	0.74	0.50	0.04	0.97	0.72	0.49	-0.01	0.98
		0.25																				
		0.5																				
	4	1																				
		2																				
		4																				

Supplementary Table 2: Average estimates (Est), standard errors (SE), standardized bias (Bias) and coverage probability of the 95% confidence interval (CP) in covariate estimates for two covariates simulations. C-CB refers to the simulations with a continuous and a common binary covariate. C-RB refers to the simulations with a continuous and a rare binary covariate. Results for simulations with a rare binary covariate with OR of 4 are not presented for reasons described in Section 3.4.

			Continuous in C-CB																			
			Separation																			
			6				8				10				12				14			
Method	Binary OR	Continuous OR	Est	SE	Bias	CP	Est	SE	Bias	CP	Est	SE	Bias	CP	Est	SE	Bias	CP	Est	SE	Bias	CP
		0.25	-1.71	0.21	-1.46	0.66	-2.25	0.26	-3.28	0.11	-3.00	0.32	-5.00	0.00	-3.92	0.38	-6.61	0.00	-4.98	0.46	-7.90	0.00
		0.5	-1.08	0.25	-1.54	0.69	-1.41	0.31	-2.29	0.40	-1.89	0.38	-3.14	0.14	-2.46	0.46	-3.87	0.01	-3.14	0.55	-4.46	0.00
	0.25	1	0.00	0.26	0.01	0.95	0.00	0.33	-0.02	0.93	0.05	0.41	0.12	0.91	-0.10	0.49	-0.21	0.95	0.04	0.59	0.06	0.95
		2	1.10	0.25	1.61	0.65	1.43	0.31	2.39	0.37	1.86	0.38	3.07	0.15	2.47	0.46	3.87	0.02	3.11	0.55	4.40	0.00
		4	1.69	0.21	1.41	0.68	2.28	0.26	3.36	0.11	3.01	0.32	5.07	0.00	3.91	0.38	6.56	0.00	4.98	0.45	7.96	0.00
		0.25	-1.78	0.21	-1.84	0.52	-2.37	0.26	-3.71	0.08	-3.17	0.32	-5.53	0.00	-4.14	0.39	-7.15	0.00	-5.24	0.46	-8.41	0.00
		0.5	-1.10	0.25	-1.59	0.63	-1.48	0.31	-2.51	0.28	-2.01	0.39	-3.40	0.08	-2.60	0.47	-4.09	0.02	-3.32	0.56	-4.71	0.00
	0.5	1	-0.01	0.27	-0.03	0.94	0.02	0.33	0.05	0.98	0.03	0.42	0.07	0.95	0.02	0.50	0.04	0.94	0.02	0.61	0.04	0.95
		2	1.14	0.25	1.76	0.59	1.54	0.31	2.72	0.26	1.99	0.39	3.33	0.09	2.64	0.46	4.19	0.01	3.30	0.56	4.63	0.02
		4	1.75	0.21	1.66	0.60	2.37	0.26	3.70	0.06	3.17	0.32	5.53	0.00	4.08	0.39	6.98	0.00	5.22	0.46	8.37	0.00
		0.25	-1.81	0.21	-1.98	0.49	-2.38	0.26	-3.76	0.07	-3.19	0.32	-5.56	0.00	-4.17	0.38	-7.22	0.00	-5.34	0.46	-8.63	0.00
		0.5	-1.15	0.25	-1.82	0.53	-1.57	0.31	-2.80	0.20	-2.06	0.39	-3.52	0.04	-2.68	0.47	-4.24	0.01	-3.38	0.56	-4.77	0.01
PR	1	1	0.00	0.27	0.02	0.94	0.02	0.34	0.06	0.95	-0.02	0.42	-0.06	0.93	0.00	0.51	0.00	0.94	0.00	0.61	0.00	0.95
		2	1.14	0.25	1.76	0.58	1.51	0.31	2.62	0.26	2.04	0.39	3.47	0.05	2.66	0.47	4.17	0.01	3.36	0.56	4.75	0.00
		4	1.83	0.21	2.06	0.47	2.38	0.26	3.76	0.05	3.21	0.32	5.64	0.01	4.16	0.39	7.17	0.00	5.36	0.46	8.67	0.00
		0.25	-1.79	0.21	-1.84	0.56	-2.41	0.26	-3.85	0.02	-3.15	0.32	-5.45	0.00	-4.08	0.39	-6.93	0.00	-5.23	0.46	-8.36	0.00
		0.5	-1.10	0.25	-1.62	0.64	-1.47	0.31	-2.49	0.31	-1.99	0.39	-3.37	0.09	-2.61	0.47	-4.11	0.02	-3.36	0.56	-4.76	0.01
	2	1	0.00	0.27	0.01	0.95	0.01	0.33	0.01	0.94	-0.01	0.42	-0.03	0.94	0.06	0.51	0.12	0.96	-0.01	0.61	-0.02	0.95
		2	1.11	0.25	1.64	0.64	1.51	0.31	2.60	0.28	2.01	0.39	3.43	0.09	2.59	0.46	4.10	0.02	3.33	0.56	4.69	0.01
		4	1.77	0.21	1.76	0.53	2.36	0.26	3.73	0.05	3.17	0.32	5.55	0.00	4.13	0.38	7.13	0.00	5.22	0.46	8.37	0.00
		0.25	-1.72	0.21	-1.54	0.66	-2.30	0.26	-3.46	0.09	-3.01	0.32	-5.06	0.01	-3.87	0.38	-6.48	0.00	-4.95	0.46	-7.81	0.00
		0.5	-1.07	0.25	-1.51	0.68	-1.40	0.31	-2.31	0.37	-1.89	0.38	-3.13	0.13	-2.43	0.46	-3.78	0.04	-3.17	0.55	-4.49	0.00
	4	1	0.00	0.26	-0.01	0.94	0.03	0.33	0.08	0.96	0.02	0.41	0.06	0.93	0.01	0.49	0.04	0.93	0.00	0.59	0.00	0.94
		2	1.08	0.25	1.53	0.67	1.42	0.31	2.35	0.41	1.83	0.38	2.99	0.17	2.47	0.46	3.89	0.04	3.09	0.55	4.36	0.02
		4	1.71	0.21	1.46	0.67	2.25	0.26	3.25	0.14	3.03	0.32	5.13	0.01	3.89	0.38	6.54	0.00	4.95	0.46	7.84	0.00

Supplementary Table 2: Average estimates (Est), standard errors (SE), standardized bias (Bias) and coverage probability of the 95% confidence interval (CP) in covariate estimates for two covariates simulations. C-CB refers to the simulations with a continuous and a common binary covariate. C-RB refers to the simulations with a continuous and a rare binary covariate. Results for simulations with a rare binary covariate with OR of 4 are not presented for reasons described in Section 3.4.

			Binary in C-CB																			
			Separation																			
			6				8				10				12				14			
Method	Binary OR	Continuous OR	Est	SE	Bias	CP	Est	SE	Bias	CP	Est	SE	Bias	CP	Est	SE	Bias	CP	Est	SE	Bias	CP
		0.25	-1.72	0.50	-0.64	0.91	-2.31	0.62	-1.48	0.68	-3.20	0.76	-2.37	0.37	-4.10	0.91	-2.99	0.16	-5.17	1.08	-3.51	0.06
		0.5	-2.19	0.52	-1.54	0.65	-2.88	0.64	-2.32	0.36	-4.03	0.80	-3.31	0.11	-5.13	0.96	-3.90	0.03	-6.35	1.15	-4.31	0.01
	0.25	1	-2.38	0.53	-1.86	0.52	-3.17	0.66	-2.69	0.24	-4.19	0.81	-3.45	0.08	-5.58	0.98	-4.26	0.01	-7.01	1.18	-4.75	0.00
		2	-2.14	0.52	-1.44	0.71	-2.88	0.64	-2.31	0.41	-3.91	0.80	-3.16	0.12	-4.97	0.96	-3.73	0.06	-6.48	1.15	-4.42	0.01
		4	-1.72	0.50	-0.64	0.94	-2.30	0.62	-1.47	0.72	-3.07	0.76	-2.21	0.37	-3.98	0.91	-2.86	0.19	-5.24	1.07	-3.59	0.07
		0.25	-0.93	0.50	-0.46	0.94	-1.23	0.62	-0.88	0.86	-1.57	0.76	-1.16	0.79	-2.14	0.91	-1.59	0.62	-2.56	1.08	-1.72	0.56
		0.5	-1.19	0.52	-0.95	0.84	-1.55	0.65	-1.30	0.76	-2.00	0.81	-1.63	0.63	-2.59	0.98	-1.94	0.50	-3.38	1.17	-2.28	0.38
	0.5	1	-1.24	0.53	-1.03	0.84	-1.65	0.67	-1.44	0.72	-2.27	0.83	-1.90	0.50	-2.89	1.01	-2.19	0.42	-3.69	1.21	-2.48	0.28
		2	-1.12	0.52	-0.82	0.85	-1.49	0.65	-1.22	0.74	-1.96	0.81	-1.56	0.65	-2.57	0.98	-1.92	0.51	-3.25	1.18	-2.18	0.40
		4	-0.94	0.50	-0.48	0.90	-1.14	0.62	-0.72	0.90	-1.58	0.76	-1.16	0.79	-2.05	0.91	-1.49	0.67	-2.71	1.08	-1.85	0.58
		0.25	0.03	0.50	0.06	0.94	0.00	0.62	0.00	0.96	-0.05	0.76	-0.06	0.95	-0.04	0.91	-0.05	0.95	0.10	1.09	0.09	0.99
		0.5	0.02	0.53	0.03	0.96	-0.01	0.65	-0.01	0.95	-0.11	0.81	-0.13	0.96	-0.04	0.98	-0.04	0.93	0.05	1.18	0.04	0.94
PR	1	1	-0.02	0.54	-0.04	0.97	-0.04	0.67	-0.07	0.96	-0.09	0.84	-0.10	0.95	0.02	1.01	0.02	0.93	0.09	1.22	0.07	0.94
		2	0.02	0.53	0.04	0.94	-0.02	0.66	-0.02	0.94	0.08	0.81	0.10	0.96	0.06	0.98	0.06	0.97	-0.03	1.18	-0.03	0.96
		4	0.01	0.50	0.02	0.96	0.02	0.62	0.05	0.95	0.04	0.76	0.04	0.95	-0.01	0.91	-0.01	0.93	0.06	1.09	0.05	0.96
		0.25	0.90	0.50	0.39	0.95	1.18	0.62	0.78	0.85	1.66	0.76	1.25	0.78	1.96	0.91	1.39	0.70	2.49	1.09	1.65	0.63
		0.5	1.14	0.53	0.85	0.86	1.53	0.65	1.28	0.75	2.05	0.81	1.68	0.60	2.68	0.98	2.03	0.51	3.31	1.17	2.23	0.33
	2	1	1.28	0.53	1.10	0.77	1.68	0.67	1.46	0.68	2.24	0.83	1.86	0.52	2.81	1.01	2.11	0.42	3.79	1.21	2.55	0.31
		2	1.13	0.53	0.82	0.90	1.50	0.65	1.23	0.78	2.04	0.81	1.67	0.60	2.60	0.98	1.95	0.53	3.36	1.17	2.27	0.38
		4	0.87	0.50	0.34	0.92	1.21	0.62	0.82	0.88	1.60	0.76	1.20	0.79	2.12	0.91	1.57	0.66	2.72	1.08	1.86	0.57
		0.25	1.74	0.50	0.68	0.88	2.27	0.62	1.42	0.66	3.18	0.76	2.36	0.36	4.13	0.91	3.02	0.16	5.18	1.08	3.51	0.08
		0.5	2.18	0.52	1.50	0.66	2.95	0.64	2.40	0.32	3.76	0.80	2.98	0.18	5.14	0.96	3.90	0.01	6.45	1.15	4.40	0.01
	4	1	2.39	0.53	1.89	0.52	3.16	0.66	2.70	0.22	4.31	0.81	3.61	0.04	5.48	0.98	4.16	0.02	7.07	1.18	4.81	0.01
		2	2.21	0.52	1.56	0.64	2.96	0.64	2.42	0.33	3.99	0.80	3.25	0.11	5.10	0.96	3.87	0.03	6.49	1.15	4.42	0.01
		4	1.77	0.50	0.73	0.86	2.33	0.62	1.51	0.65	3.12	0.76	2.27	0.38	4.04	0.91	2.93	0.18	5.20	1.08	3.54	0.07

Supplementary Table 2: Average estimates (Est), standard errors (SE), standardized bias (Bias) and coverage probability of the 95% confidence interval (CP) in covariate estimates for two covariates simulations. C-CB refers to the simulations with a continuous and a common binary covariate. C-RB refers to the simulations with a continuous and a rare binary covariate. Results for simulations with a rare binary covariate with OR of 4 are not presented for reasons described in Section 3.4.

			Continuous in C-RB																			
			Separation																			
			6				8				10				12				14			
Method	Binary OR	Continuous OR	Est	SE	Bias	CP	Est	SE	Bias	CP	Est	SE	Bias	CP	Est	SE	Bias	CP	Est	SE	Bias	CP
		0.25	-1.77	0.21	-1.75	0.56	-2.35	0.26	-3.63	0.09	-3.12	0.32	-5.37	0.00	-4.06	0.38	-6.95	0.00	-5.17	0.46	-8.27	0.00
		0.5	-1.14	0.25	-1.79	0.56	-1.49	0.31	-2.55	0.25	-2.00	0.38	-3.42	0.09	-2.52	0.46	-3.94	0.01	-3.27	0.56	-4.64	0.00
	0.25	1	-0.01	0.27	-0.04	0.94	0.03	0.33	0.08	0.96	0.01	0.41	0.03	0.97	-0.01	0.50	-0.01	0.94	0.01	0.60	0.03	0.98
		2	1.10	0.25	1.60	0.67	1.48	0.31	2.52	0.30	1.97	0.38	3.32	0.11	2.53	0.47	3.94	0.00	3.26	0.56	4.62	0.01
		4	1.74	0.21	1.66	0.65	2.32	0.26	3.52	0.07	3.08	0.32	5.28	0.00	4.03	0.38	6.89	0.00	5.12	0.46	8.19	0.00
		0.25	-1.83	0.21	-2.04	0.48	-2.41	0.26	-3.88	0.05	-3.22	0.32	-5.69	0.00	-4.16	0.39	-7.18	0.00	-5.31	0.46	-8.56	0.00
		0.5	-1.20	0.25	-2.02	0.49	-1.52	0.31	-2.64	0.25	-2.03	0.39	-3.48	0.06	-2.60	0.47	-4.09	0.01	-3.38	0.56	-4.80	0.00
	0.5	1	0.00	0.27	0.00	0.95	0.03	0.33	0.09	0.91	-0.03	0.42	-0.07	0.97	0.08	0.51	0.17	0.96	0.00	0.61	0.00	0.96
		2	1.13	0.25	1.73	0.57	1.54	0.31	2.69	0.24	2.04	0.39	3.50	0.04	2.64	0.47	4.18	0.02	3.34	0.57	4.69	0.00
		4	1.79	0.21	1.86	0.54	2.38	0.26	3.74	0.04	3.19	0.32	5.59	0.00	4.15	0.38	7.19	0.00	5.29	0.46	8.54	0.00
		0.25	-1.81	0.21	-1.91	0.49	-2.43	0.26	-3.94	0.03	-3.23	0.32	-5.68	0.00	-4.17	0.39	-7.20	0.00	-5.31	0.46	-8.51	0.00
		0.5	-1.17	0.25	-1.88	0.53	-1.56	0.31	-2.76	0.22	-2.00	0.39	-3.35	0.09	-2.64	0.47	-4.15	0.01	-3.38	0.56	-4.77	0.01
PR	1	1	0.00	0.27	-0.01	0.95	0.02	0.34	0.05	0.95	0.05	0.42	0.11	0.95	0.01	0.51	0.03	0.96	-0.04	0.62	-0.07	0.96
		2	1.19	0.25	1.95	0.53	1.52	0.31	2.64	0.24	2.10	0.39	3.66	0.03	2.69	0.47	4.25	0.01	3.40	0.56	4.80	0.01
		4	1.82	0.21	1.98	0.46	2.42	0.26	3.91	0.02	3.23	0.32	5.69	0.00	4.17	0.39	7.23	0.00	5.33	0.46	8.61	0.00
		0.25	-1.83	0.21	-2.02	0.52	-2.38	0.26	-3.76	0.06	-3.18	0.32	-5.56	0.00	-4.16	0.39	-7.16	0.00	-5.32	0.46	-8.56	0.00
		0.5	-1.12	0.25	-1.66	0.61	-1.56	0.31	-2.78	0.19	-2.05	0.39	-3.51	0.05	-2.65	0.47	-4.19	0.01	-3.37	0.56	-4.76	0.00
	2	1	0.00	0.27	-0.01	0.96	0.03	0.34	0.10	0.91	0.01	0.42	0.02	0.94	-0.01	0.51	-0.02	0.98	-0.03	0.61	-0.05	0.93
		2	1.15	0.25	1.79	0.60	1.51	0.31	2.61	0.26	2.04	0.39	3.48	0.08	2.62	0.47	4.09	0.03	3.36	0.57	4.74	0.01
		4	1.81	0.21	1.94	0.50	2.40	0.26	3.82	0.05	3.21	0.32	5.61	0.00	4.15	0.39	7.15	0.00	5.34	0.46	8.61	0.00
		0.25	-1.79	0.21	-1.86	0.55	-2.37	0.26	-3.74	0.08	-3.16	0.32	-5.47	0.00	-4.11	0.39	-7.06	0.00	-5.21	0.46	-8.28	0.00
		0.5	-1.13	0.25	-1.74	0.61	-1.47	0.31	-2.48	0.31	-2.01	0.39	-3.41	0.07	-2.60	0.47	-4.06	0.01	-3.30	0.56	-4.67	0.00
	4	1	0.01	0.27	0.05	0.93	0.00	0.33	0.00	0.97	0.04	0.42	0.09	0.95	-0.04	0.50	-0.09	0.96	0.02	0.61	0.03	0.96
		2	1.16	0.25	1.86	0.56	1.53	0.31	2.66	0.26	2.01	0.39	3.42	0.09	2.58	0.47	4.05	0.02	3.30	0.56	4.64	0.01
		4	1.80	0.21	1.90	0.54	2.36	0.26	3.67	0.05	3.16	0.32	5.48	0.00	4.13	0.38	7.12	0.00	5.30	0.46	8.56	0.00

Supplementary Table 2: Average estimates (Est), standard errors (SE), standardized bias (Bias) and coverage probability of the 95% confidence interval (CP) in covariate estimates for two covariates simulations. C-CB refers to the simulations with a continuous and a common binary covariate. C-RB refers to the simulations with a continuous and a rare binary covariate. Results for simulations with a rare binary covariate with OR of 4 are not presented for reasons described in Section 3.4.

			Binary in C-RB																			
			Separation																			
			6				8				10				12				14			
Method	Binary OR	Continuous OR	Est	SE	Bias	CP	Est	SE	Bias	CP	Est	SE	Bias	CP	Est	SE	Bias	CP	Est	SE	Bias	CP
		0.25	-2.27	0.84	-1.05	0.79	-2.80	1.04	-1.37	0.69	-3.65	1.28	-1.79	0.57	-4.54	1.52	-2.09	0.49	-6.06	1.80	-2.61	0.30
		0.5	-2.58	0.87	-1.39	0.70	-3.58	1.09	-2.03	0.47	-4.53	1.35	-2.34	0.37	-6.13	1.63	-2.92	0.21	-7.65	1.95	-3.24	0.13
	0.25	1	-2.66	0.89	-1.44	0.67	-3.89	1.11	-2.24	0.43	-5.12	1.38	-2.69	0.24	-6.62	1.66	-3.17	0.16	-8.37	2.01	-3.47	0.10
		2	-2.53	0.87	-1.30	0.74	-3.32	1.09	-1.78	0.57	-4.62	1.35	-2.41	0.37	-5.64	1.65	-2.61	0.34	-7.68	1.97	-3.21	0.15
		4	-2.24	0.84	-1.02	0.75	-2.69	1.05	-1.23	0.72	-3.81	1.28	-1.89	0.57	-4.68	1.51	-2.19	0.44	-6.01	1.82	-2.55	0.33
		0.25	-1.00	0.84	-0.36	0.89	-1.15	1.04	-0.44	0.92	-1.67	1.28	-0.78	0.85	-2.39	1.52	-1.11	0.76	-2.86	1.83	-1.19	0.74
		0.5	-1.12	0.88	-0.49	0.91	-1.71	1.09	-0.94	0.78	-2.28	1.35	-1.18	0.74	-3.15	1.64	-1.49	0.67	-3.76	1.98	-1.54	0.60
	0.5	1	-1.45	0.90	-0.83	0.85	-1.88	1.11	-1.09	0.79	-2.64	1.40	-1.40	0.70	-3.21	1.70	-1.48	0.65	-4.03	2.04	-1.64	0.65
		2	-1.19	0.88	-0.58	0.93	-1.54	1.10	-0.77	0.85	-2.30	1.35	-1.18	0.76	-3.07	1.64	-1.47	0.67	-3.77	1.98	-1.56	0.66
		4	-0.98	0.84	-0.34	0.86	-1.21	1.04	-0.50	0.90	-1.89	1.28	-0.95	0.83	-2.31	1.53	-1.05	0.80	-3.16	1.83	-1.35	0.67
		0.25	-0.04	0.83	-0.06	0.97	0.01	1.04	0.01	0.95	-0.08	1.27	-0.06	0.96	-0.17	1.52	-0.12	0.94	-0.17	1.81	-0.08	0.94
		0.5	0.08	0.89	0.08	0.97	-0.13	1.09	-0.11	0.93	-0.01	1.36	0.00	0.97	-0.06	1.65	-0.03	0.96	-0.01	1.99	-0.02	0.98
PR	1	1	-0.07	0.90	-0.07	0.92	-0.04	1.14	-0.03	0.96	0.14	1.41	0.11	0.95	-0.06	1.70	-0.03	0.97	0.05	2.05	0.03	0.96
		2	0.02	0.88	0.02	0.97	0.02	1.09	0.02	0.96	-0.20	1.36	-0.15	0.94	0.08	1.65	0.04	0.97	-0.14	1.97	-0.07	0.95
		4	-0.04	0.84	-0.03	0.96	0.05	1.04	0.05	0.97	-0.20	1.28	-0.15	0.95	-0.08	1.53	-0.04	0.95	-0.02	1.82	-0.01	0.96
		0.25	0.80	0.84	0.14	0.99	1.04	1.05	0.33	0.98	1.38	1.28	0.54	0.93	1.64	1.53	0.61	0.92	2.13	1.80	0.80	0.90
		0.5	1.09	0.89	0.44	0.97	1.34	1.10	0.58	0.95	1.87	1.36	0.86	0.89	2.19	1.65	0.92	0.87	2.94	1.98	1.14	0.83
	2	1	1.18	0.90	0.54	0.97	1.46	1.11	0.68	0.94	1.97	1.40	0.92	0.87	2.46	1.69	1.04	0.88	3.51	2.05	1.38	0.73
		2	0.98	0.88	0.32	0.98	1.40	1.10	0.64	0.97	1.77	1.36	0.81	0.93	2.38	1.64	1.02	0.82	2.90	1.98	1.12	0.86
		4	0.72	0.84	0.03	0.98	1.06	1.04	0.33	0.99	1.40	1.28	0.55	0.97	1.95	1.54	0.82	0.91	2.39	1.82	0.93	0.89
		0.25																				
		0.5																				
	4	1																				
		2																				
		4																				

Supplementary Table 2: Average estimates (Est), standard errors (SE), standardized bias (Bias) and coverage probability of the 95% confidence interval (CP) in covariate estimates for two covariates simulations. C-CB refers to the simulations with a continuous and a common binary covariate. C-RB refers to the simulations with a continuous and a rare binary covariate. Results for simulations with a rare binary covariate with OR of 4 are not presented for reasons described in Section 3.4.

			Continuous in C-CB																			
			Separation																			
			6				8				10				12				14			
Method	Binary OR	Continuous OR	Est	SE	Bias	CP	Est	SE	Bias	CP	Est	SE	Bias	CP	Est	SE	Bias	CP	Est	SE	Bias	CP
		0.25	-0.99	0.13	3.26	0.13	-1.07	0.13	2.53	0.27	-1.18	0.13	1.58	0.66	-1.26	0.14	0.97	0.83	-1.32	0.14	0.50	0.92
		0.5	-0.56	0.12	1.17	0.81	-0.59	0.11	0.95	0.86	-0.63	0.11	0.59	0.90	-0.66	0.11	0.31	0.94	-0.67	0.11	0.22	0.94
	0.25	1	0.00	0.11	0.03	0.96	0.00	0.11	-0.03	0.94	0.02	0.11	0.15	0.93	-0.03	0.11	-0.27	0.95	0.01	0.10	0.09	0.95
		2	0.56	0.12	-1.14	0.78	0.60	0.12	-0.87	0.84	0.62	0.11	-0.69	0.89	0.67	0.11	-0.28	0.95	0.66	0.11	-0.30	0.95
		4	0.99	0.13	-3.20	0.18	1.10	0.13	-2.27	0.42	1.17	0.13	-1.68	0.60	1.26	0.14	-1.01	0.81	1.33	0.14	-0.48	0.92
		0.25	-1.03	0.13	2.90	0.22	-1.10	0.13	2.34	0.39	-1.20	0.13	1.49	0.65	-1.29	0.14	0.75	0.87	-1.32	0.14	0.53	0.92
		0.5	-0.57	0.11	1.15	0.78	-0.59	0.11	0.92	0.87	-0.64	0.11	0.51	0.92	-0.66	0.11	0.38	0.95	-0.68	0.11	0.20	0.97
	0.5	1	-0.01	0.11	-0.11	0.95	0.00	0.11	0.04	0.97	0.00	0.10	0.04	0.96	0.01	0.10	0.09	0.97	0.00	0.10	0.02	0.97
		2	0.57	0.11	-1.09	0.78	0.61	0.11	-0.75	0.88	0.63	0.11	-0.59	0.92	0.66	0.11	-0.30	0.95	0.67	0.11	-0.24	0.94
		4	1.02	0.13	-3.02	0.17	1.10	0.13	-2.34	0.37	1.21	0.13	-1.45	0.68	1.27	0.13	-0.91	0.83	1.32	0.14	-0.57	0.90
		0.25	-1.03	0.12	2.91	0.23	-1.11	0.13	2.27	0.42	-1.19	0.13	1.60	0.65	-1.28	0.13	0.89	0.87	-1.33	0.13	0.51	0.95
		0.5	-0.57	0.11	1.09	0.82	-0.61	0.11	0.74	0.90	-0.64	0.11	0.54	0.93	-0.67	0.11	0.29	0.96	-0.68	0.11	0.17	0.94
PW3S	1	1	0.00	0.11	0.01	0.95	0.00	0.11	-0.02	0.97	-0.01	0.10	-0.07	0.94	0.00	0.10	-0.01	0.94	0.00	0.10	-0.03	0.95
		2	0.57	0.11	-1.15	0.78	0.60	0.11	-0.86	0.88	0.63	0.11	-0.58	0.92	0.65	0.11	-0.39	0.94	0.67	0.11	-0.24	0.95
		4	1.03	0.12	-2.91	0.23	1.09	0.13	-2.42	0.38	1.20	0.13	-1.53	0.65	1.27	0.13	-0.95	0.86	1.34	0.14	-0.44	0.94
		0.25	-1.01	0.12	3.06	0.15	-1.12	0.13	2.10	0.46	-1.19	0.13	1.62	0.64	-1.26	0.13	1.04	0.78	-1.32	0.14	0.59	0.91
		0.5	-0.56	0.11	1.19	0.77	-0.60	0.11	0.87	0.90	-0.62	0.11	0.70	0.91	-0.66	0.11	0.32	0.93	-0.68	0.11	0.14	0.92
	2	1	0.01	0.11	0.08	0.93	0.00	0.11	0.04	0.97	0.00	0.10	-0.01	0.95	0.01	0.10	0.07	0.96	-0.01	0.10	-0.06	0.95
		2	0.56	0.11	-1.23	0.78	0.60	0.11	-0.84	0.87	0.64	0.11	-0.52	0.92	0.66	0.11	-0.33	0.96	0.67	0.11	-0.21	0.95
		4	1.01	0.12	-3.11	0.14	1.10	0.13	-2.31	0.35	1.20	0.13	-1.51	0.67	1.28	0.13	-0.83	0.87	1.32	0.14	-0.57	0.89
		0.25	-0.99	0.13	3.20	0.14	-1.09	0.13	2.34	0.34	-1.17	0.13	1.70	0.61	-1.25	0.14	1.04	0.80	-1.31	0.14	0.65	0.89
		0.5	-0.56	0.12	1.20	0.81	-0.59	0.11	0.94	0.87	-0.62	0.11	0.68	0.90	-0.64	0.11	0.48	0.96	-0.68	0.11	0.14	0.94
	4	1	-0.01	0.11	-0.13	0.97	0.00	0.11	-0.04	0.98	0.00	0.11	0.02	0.93	0.00	0.11	0.05	0.92	0.00	0.10	0.01	0.94
		2	0.56	0.12	-1.23	0.76	0.60	0.12	-0.87	0.86	0.61	0.11	-0.74	0.91	0.66	0.11	-0.34	0.91	0.66	0.11	-0.31	0.93
		4	1.00	0.13	-3.13	0.17	1.08	0.13	-2.44	0.32	1.20	0.13	-1.47	0.67	1.25	0.14	-1.04	0.80	1.31	0.14	-0.63	0.89

Supplementary Table 2: Average estimates (Est), standard errors (SE), standardized bias (Bias) and coverage probability of the 95% confidence interval (CP) in covariate estimates for two covariates simulations. C-CB refers to the simulations with a continuous and a common binary covariate. C-RB refers to the simulations with a continuous and a rare binary covariate. Results for simulations with a rare binary covariate with OR of 4 are not presented for reasons described in Section 3.4.

			Binary in C-CB																			
			Separation																			
			6				8				10				12				14			
Method	Binary OR	Continuous OR	Est	SE	Bias	CP	Est	SE	Bias	CP	Est	SE	Bias	CP	Est	SE	Bias	CP	Est	SE	Bias	CP
		0.25	-1.02	0.26	1.47	0.69	-1.08	0.25	1.25	0.77	-1.22	0.26	0.67	0.92	-1.29	0.26	0.43	0.94	-1.33	0.26	0.23	0.94
		0.5	-1.13	0.24	1.10	0.79	-1.20	0.23	0.82	0.86	-1.29	0.23	0.44	0.95	-1.33	0.23	0.26	0.95	-1.33	0.23	0.26	0.96
	0.25	1	-1.18	0.23	0.93	0.85	-1.21	0.23	0.83	0.88	-1.27	0.22	0.57	0.90	-1.34	0.22	0.22	0.96	-1.35	0.22	0.19	0.97
		2	-1.12	0.24	1.13	0.82	-1.20	0.23	0.82	0.90	-1.26	0.23	0.57	0.93	-1.30	0.23	0.43	0.91	-1.35	0.23	0.17	0.95
		4	-0.99	0.26	1.57	0.67	-1.11	0.26	1.12	0.80	-1.16	0.25	0.92	0.86	-1.26	0.25	0.54	0.89	-1.35	0.26	0.16	0.95
		0.25	-0.53	0.25	0.65	0.93	-0.57	0.24	0.52	0.95	-0.60	0.24	0.40	0.96	-0.66	0.24	0.13	0.93	-0.64	0.24	0.25	0.94
		0.5	-0.56	0.23	0.62	0.89	-0.60	0.22	0.42	0.94	-0.65	0.22	0.21	0.97	-0.65	0.21	0.23	0.94	-0.67	0.21	0.11	0.93
	0.5	1	-0.59	0.22	0.51	0.91	-0.62	0.21	0.36	0.91	-0.66	0.21	0.18	0.98	-0.68	0.21	0.09	0.93	-0.68	0.20	0.08	0.94
		2	-0.58	0.23	0.52	0.94	-0.60	0.22	0.42	0.93	-0.62	0.22	0.33	0.97	-0.65	0.21	0.22	0.96	-0.65	0.21	0.23	0.93
		4	-0.54	0.25	0.62	0.91	-0.53	0.24	0.70	0.90	-0.59	0.24	0.44	0.93	-0.61	0.24	0.36	0.95	-0.67	0.24	0.12	0.98
		0.25	0.01	0.24	0.02	0.98	0.00	0.24	0.01	0.97	0.00	0.24	0.00	0.95	0.00	0.24	-0.01	0.93	0.02	0.24	0.09	0.97
		0.5	0.00	0.22	0.00	0.96	-0.01	0.22	-0.03	0.97	-0.04	0.21	-0.20	0.93	-0.01	0.21	-0.06	0.92	0.02	0.21	0.08	0.94
PW3S	1	1	0.00	0.22	-0.01	0.97	-0.02	0.21	-0.08	0.94	-0.01	0.21	-0.06	0.94	0.01	0.20	0.07	0.96	0.02	0.20	0.08	0.93
		2	0.01	0.22	0.04	0.97	0.01	0.22	0.04	0.95	0.01	0.21	0.04	0.94	0.00	0.21	0.01	0.96	-0.01	0.21	-0.06	0.98
		4	0.02	0.24	0.10	0.96	0.01	0.24	0.05	0.97	0.01	0.24	0.04	0.97	-0.01	0.24	-0.04	0.96	0.02	0.24	0.08	0.97
		0.25	0.52	0.25	-0.73	0.90	0.55	0.24	-0.60	0.93	0.62	0.24	-0.32	0.94	0.61	0.24	-0.35	0.95	0.61	0.24	-0.34	0.95
		0.5	0.57	0.23	-0.55	0.91	0.59	0.22	-0.49	0.89	0.63	0.22	-0.29	0.95	0.67	0.21	-0.14	0.96	0.67	0.21	-0.11	0.95
	2	1	0.60	0.22	-0.43	0.95	0.61	0.21	-0.39	0.95	0.66	0.21	-0.15	0.97	0.66	0.21	-0.19	0.94	0.70	0.20	0.02	0.98
		2	0.57	0.23	-0.55	0.95	0.59	0.22	-0.46	0.95	0.64	0.22	-0.27	0.94	0.66	0.21	-0.16	0.95	0.68	0.21	-0.09	0.97
		4	0.51	0.25	-0.76	0.88	0.55	0.24	-0.59	0.92	0.60	0.24	-0.41	0.94	0.65	0.24	-0.18	0.95	0.68	0.24	-0.05	0.97
		0.25	1.00	0.26	-1.55	0.64	1.08	0.25	-1.23	0.72	1.23	0.26	-0.63	0.90	1.30	0.26	-0.37	0.92	1.31	0.26	-0.32	0.94
		0.5	1.13	0.24	-1.12	0.80	1.22	0.24	-0.75	0.91	1.23	0.23	-0.72	0.88	1.33	0.23	-0.28	0.95	1.34	0.23	-0.25	0.97
	4	1	1.16	0.23	-0.99	0.80	1.22	0.23	-0.76	0.88	1.31	0.22	-0.38	0.94	1.32	0.22	-0.35	0.93	1.36	0.22	-0.13	0.96
		2	1.12	0.24	-1.15	0.79	1.21	0.24	-0.80	0.90	1.27	0.23	-0.52	0.91	1.32	0.23	-0.33	0.96	1.35	0.23	-0.18	0.96
		4	1.00	0.26	-1.53	0.67	1.10	0.25	-1.17	0.80	1.19	0.26	-0.78	0.85	1.26	0.25	-0.51	0.97	1.32	0.26	-0.31	0.96

Supplementary Table 2: Average estimates (Est), standard errors (SE), standardized bias (Bias) and coverage probability of the 95% confidence interval (CP) in covariate estimates for two covariates simulations. C-CB refers to the simulations with a continuous and a common binary covariate. C-RB refers to the simulations with a continuous and a rare binary covariate. Results for simulations with a rare binary covariate with OR of 4 are not presented for reasons described in Section 3.4.

			Continuous in C-RB																			
			Separation																			
			6				8				10				12				14			
Method	Binary OR	Continuous OR	Est	SE	Bias	CP	Est	SE	Bias	CP	Est	SE	Bias	CP	Est	SE	Bias	CP	Est	SE	Bias	CP
		0.25	-1.01	0.13	3.06	0.18	-1.09	0.13	2.37	0.40	-1.19	0.13	1.56	0.65	-1.26	0.13	0.98	0.84	-1.32	0.14	0.54	0.92
		0.5	-0.58	0.12	1.05	0.82	-0.61	0.11	0.76	0.90	-0.64	0.11	0.53	0.91	-0.65	0.11	0.42	0.93	-0.67	0.11	0.21	0.96
	0.25	1	0.00	0.11	-0.03	0.95	0.01	0.11	0.09	0.97	0.00	0.10	-0.01	0.97	0.00	0.10	0.00	0.95	0.00	0.10	-0.01	0.97
		2	0.56	0.11	-1.23	0.76	0.59	0.11	-0.90	0.90	0.63	0.11	-0.56	0.91	0.65	0.11	-0.42	0.93	0.68	0.11	-0.16	0.96
		4	1.00	0.12	-3.14	0.16	1.10	0.13	-2.29	0.41	1.18	0.13	-1.63	0.63	1.26	0.13	-1.01	0.81	1.31	0.14	-0.66	0.84
		0.25	-1.03	0.12	2.94	0.20	-1.12	0.13	2.17	0.43	-1.20	0.13	1.45	0.69	-1.28	0.13	0.89	0.80	-1.32	0.14	0.55	0.89
		0.5	-0.60	0.11	0.84	0.86	-0.60	0.11	0.90	0.83	-0.64	0.11	0.55	0.90	-0.66	0.11	0.38	0.95	-0.68	0.11	0.12	0.94
	0.5	1	0.00	0.11	0.00	0.97	0.01	0.11	0.05	0.96	-0.01	0.10	-0.09	0.96	0.01	0.10	0.14	0.97	0.00	0.10	0.00	0.93
		2	0.56	0.11	-1.17	0.80	0.61	0.11	-0.80	0.90	0.63	0.11	-0.56	0.94	0.66	0.11	-0.34	0.94	0.68	0.11	-0.18	0.96
		4	1.01	0.12	-3.10	0.22	1.10	0.13	-2.31	0.37	1.19	0.13	-1.56	0.65	1.28	0.13	-0.88	0.84	1.34	0.14	-0.43	0.94
		0.25	-1.02	0.12	3.01	0.17	-1.11	0.13	2.22	0.37	-1.20	0.13	1.52	0.68	-1.28	0.13	0.86	0.87	-1.32	0.14	0.54	0.91
		0.5	-0.57	0.11	1.11	0.80	-0.61	0.11	0.77	0.86	-0.62	0.11	0.69	0.90	-0.66	0.11	0.34	0.94	-0.67	0.11	0.22	0.95
PW3S	1	1	0.00	0.11	0.03	0.97	0.00	0.11	-0.03	0.96	0.01	0.10	0.06	0.96	0.00	0.10	-0.02	0.97	0.00	0.10	-0.03	0.97
		2	0.58	0.11	-1.03	0.84	0.59	0.11	-0.92	0.84	0.65	0.11	-0.44	0.92	0.66	0.11	-0.30	0.95	0.67	0.11	-0.28	0.93
		4	1.02	0.12	-3.00	0.19	1.11	0.13	-2.25	0.38	1.21	0.13	-1.44	0.69	1.27	0.13	-0.94	0.83	1.31	0.13	-0.61	0.92
		0.25	-1.04	0.13	2.84	0.23	-1.09	0.13	2.46	0.32	-1.19	0.13	1.55	0.68	-1.27	0.13	0.91	0.82	-1.32	0.14	0.56	0.90
		0.5	-0.57	0.11	1.10	0.81	-0.61	0.11	0.74	0.88	-0.63	0.11	0.58	0.90	-0.67	0.11	0.24	0.93	-0.67	0.11	0.24	0.94
	2	1	-0.01	0.11	-0.05	0.97	0.01	0.11	0.06	0.95	0.00	0.10	-0.01	0.96	0.00	0.10	-0.04	0.97	-0.01	0.10	-0.09	0.96
		2	0.57	0.11	-1.15	0.80	0.59	0.11	-0.93	0.84	0.64	0.11	-0.56	0.92	0.66	0.11	-0.35	0.93	0.67	0.11	-0.26	0.94
		4	1.02	0.12	-3.07	0.17	1.10	0.13	-2.39	0.39	1.21	0.13	-1.41	0.69	1.27	0.13	-0.96	0.82	1.32	0.14	-0.56	0.91
		0.25	-1.00	0.12	3.17	0.14	-1.10	0.13	2.30	0.36	-1.18	0.13	1.66	0.63	-1.27	0.13	0.96	0.87	-1.32	0.14	0.59	0.86
		0.5	-0.57	0.11	1.12	0.80	-0.59	0.11	0.96	0.84	-0.64	0.11	0.54	0.90	-0.65	0.11	0.42	0.95	-0.69	0.11	0.10	0.97
	4	1	0.01	0.11	0.06	0.95	-0.01	0.11	-0.08	0.97	0.01	0.10	0.06	0.96	-0.01	0.10	-0.08	0.96	-0.01	0.10	-0.06	0.97
		2	0.57	0.11	-1.10	0.81	0.61	0.11	-0.76	0.92	0.63	0.11	-0.63	0.92	0.65	0.11	-0.39	0.92	0.68	0.11	-0.18	0.94
		4	1.02	0.12	-3.05	0.14	1.08	0.12	-2.51	0.27	1.18	0.13	-1.63	0.60	1.27	0.13	-0.97	0.83	1.35	0.14	-0.37	0.94

Supplementary Table 2: Average estimates (Est), standard errors (SE), standardized bias (Bias) and coverage probability of the 95% confidence interval (CP) in covariate estimates for two covariates simulations. C-CB refers to the simulations with a continuous and a common binary covariate. C-RB refers to the simulations with a continuous and a rare binary covariate. Results for simulations with a rare binary covariate with OR of 4 are not presented for reasons described in Section 3.4.

			Binary in C-RB																			
			Separation																			
			6				8				10				12				14			
Method	Binary OR	Continuous OR	Est	SE	Bias	CP	Est	SE	Bias	CP	Est	SE	Bias	CP	Est	SE	Bias	CP	Est	SE	Bias	CP
		0.25	-1.04	0.37	0.94	0.85	-1.08	0.37	0.84	0.86	-1.14	0.37	0.68	0.90	-1.23	0.37	0.44	0.95	-1.34	0.38	0.15	0.96
		0.5	-1.11	0.34	0.82	0.89	-1.19	0.33	0.60	0.90	-1.24	0.33	0.46	0.95	-1.34	0.33	0.15	0.96	-1.34	0.33	0.16	0.95
	0.25	1	-1.10	0.32	0.88	0.87	-1.22	0.32	0.54	0.93	-1.29	0.32	0.31	0.96	-1.35	0.31	0.13	0.96	-1.37	0.31	0.08	0.96
		2	-1.04	0.34	1.01	0.82	-1.14	0.34	0.72	0.90	-1.25	0.33	0.42	0.94	-1.29	0.33	0.31	0.95	-1.36	0.33	0.10	0.94
		4	-1.03	0.37	0.97	0.86	-1.03	0.37	0.96	0.88	-1.19	0.38	0.56	0.91	-1.28	0.37	0.31	0.94	-1.31	0.38	0.21	0.98
		0.25	-0.49	0.38	0.49	0.92	-0.49	0.38	0.52	0.97	-0.55	0.38	0.37	0.97	-0.64	0.38	0.14	0.96	-0.64	0.38	0.13	0.95
		0.5	-0.52	0.35	0.47	0.93	-0.58	0.34	0.31	0.96	-0.62	0.33	0.21	0.95	-0.69	0.33	-0.01	0.95	-0.66	0.33	0.07	0.96
	0.5	1	-0.58	0.34	0.31	0.96	-0.61	0.32	0.22	0.97	-0.65	0.32	0.11	0.98	-0.66	0.32	0.08	0.96	-0.65	0.31	0.12	0.96
		2	-0.55	0.35	0.37	0.97	-0.54	0.34	0.42	0.94	-0.65	0.33	0.12	0.94	-0.66	0.33	0.08	0.96	-0.68	0.33	0.02	0.98
		4	-0.48	0.38	0.53	0.94	-0.50	0.38	0.49	0.94	-0.64	0.38	0.13	0.96	-0.64	0.38	0.12	0.95	-0.68	0.38	0.04	0.93
		0.25	0.02	0.40	0.00	0.97	0.01	0.41	-0.02	0.97	0.00	0.40	-0.06	0.94	-0.02	0.40	-0.10	0.94	0.00	0.40	-0.04	0.93
		0.5	0.03	0.38	0.03	0.97	-0.01	0.37	-0.07	0.95	0.00	0.36	-0.05	0.97	0.01	0.36	-0.02	0.95	0.00	0.35	-0.05	0.96
PW3S	1	1	-0.01	0.36	-0.11	0.97	0.00	0.36	-0.07	0.97	0.04	0.35	0.04	0.96	0.03	0.34	0.04	0.97	0.03	0.34	0.04	0.98
		2	0.02	0.38	-0.03	0.97	0.03	0.37	0.04	0.97	-0.04	0.36	-0.18	0.98	0.03	0.36	0.05	0.99	0.00	0.35	-0.04	0.95
		4	-0.01	0.41	-0.08	0.96	0.00	0.40	-0.05	0.97	-0.01	0.40	-0.06	0.95	0.00	0.40	-0.04	0.97	0.00	0.40	-0.05	0.97
		0.25	0.49	0.46	-0.51	0.94	0.56	0.46	-0.36	0.97	0.59	0.45	-0.28	0.95	0.55	0.44	-0.40	0.93	0.61	0.44	-0.25	0.94
		0.5	0.55	0.43	-0.41	0.95	0.56	0.42	-0.39	0.94	0.62	0.41	-0.24	0.96	0.61	0.41	-0.29	0.94	0.64	0.40	-0.19	0.96
	2	1	0.57	0.42	-0.38	0.95	0.59	0.40	-0.35	0.93	0.65	0.40	-0.18	0.97	0.61	0.39	-0.28	0.96	0.72	0.39	-0.02	0.97
		2	0.52	0.43	-0.49	0.91	0.58	0.42	-0.33	0.96	0.64	0.41	-0.20	0.97	0.65	0.40	-0.19	0.97	0.63	0.40	-0.22	0.97
		4	0.45	0.45	-0.61	0.93	0.55	0.46	-0.38	0.96	0.59	0.45	-0.29	0.96	0.68	0.46	-0.10	0.96	0.68	0.46	-0.09	0.96
		0.25																				
		0.5																				
	4	1																				
		2																				
		4																				

B.3 Supplementary Table 3

Supplementary Table 3: Power for hypothesis tests of the covariate effect in scenarios with single covariates of different types. Results for simulations with a rare binary covariate with OR of 4 are not presented for reasons described in Section 3.4.

		Separation														
		Continuous					Common Binary					Rare Binary				
Method	OR	6	8	10	12	14	6	8	10	12	14	6	8	10	12	14
1S	0.25	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.95	0.97	0.98	0.99	1.00
	0.5	1.00	1.00	1.00	1.00	1.00	0.80	0.84	0.95	0.94	0.93	0.46	0.53	0.53	0.62	0.59
	2	1.00	1.00	1.00	1.00	1.00	0.78	0.90	0.90	0.90	0.95	0.19	0.18	0.31	0.40	0.47
	4	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00					
I3S	0.25	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.92	0.95	0.98	0.99	1.00
	0.5	1.00	1.00	1.00	1.00	1.00	0.74	0.79	0.91	0.92	0.91	0.43	0.47	0.53	0.63	0.58
	2	1.00	1.00	1.00	1.00	1.00	0.72	0.87	0.88	0.87	0.96	0.17	0.20	0.31	0.40	0.47
	4	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00					

Bibliography

[1] Moore VM, Davies MJ, Willson KJ et al. Dietary composition of pregnant women is related to size of the baby at birth. *The Journal of Nutrition* 2004; 134: 1820–1826.

[2] Whitrow MJ, Moore VM, Rumbold AR et al. Effect of supplemental folic acid in pregnancy on childhood asthma: A prospective birth cohort study. *American Journal of Epidemiology* 2009; 170: 1486–1493.

[3] Moffitt TE, Caspi A, Dickson N et al. Childhood-onset versus adolescent-onset antisocial conduct problems in males: Natural history from ages 3 to 18 years. *Development and Psychopathology* 1996; 8: 399–424.

[4] Haapasalo J and Tremblay RE. Physically aggressive boys from ages 6 to 12: Family background, parenting behavior, and prediction of delinquency. *Journal of Consulting and Clinical Psychology* 1994; 62: 1044–1052.

[5] Nagin DS and Land KC. Age, criminal careers, and population heterogene-

ity: Specification and estimation of a nonparametric, mixed poisson model. *Criminology* 1993; 31: 327–362.

[6] Nagin DS. Analyzing developmental trajectories: A semiparametric, group-based approach. *Psychological Methods* 1999; 4: 139–157.

[7] Muthén B and Shedden K. Finite mixture modeling with mixture outcomes using the EM algorithm. *Biometrics* 1999; 55: 463–469.

[8] Muthén B and Muthén LK. Integrating person-centered and variable-centered analyses: Growth mixture modeling with latent trajectory classes. *Alcoholism: Clinical and Experimental Research* 2000; 24: 882–891.

[9] Muthén LK and Muthén BO. *Mplus User's Guide.* 6th ed. Los Angeles, CA: Muthén & Muthén; 1998–2010.

[10] Nagin D and Odgers C. Group-based trajectory modeling (nearly) two decades later. *Journal of Quantitative Criminology* 2010; 26: 445–453.

[11] Giles LC, Whitrow MJ, Davies MJ et al. Growth trajectories in early childhood, their relationship with antenatal and postnatal factors, and development of obesity by age 9 years: results from an australian birth cohort study. *International Journal Of Obesity* 2015; 39: 1049–1056.

[12] Sajobi TT, Menon BK, Wang M et al. Early trajectory of stroke severity predicts long-term functional outcomes in ischemic stroke subjects. *Stroke* 2017; 48: 105–110.

[13] Madigan S, Plamondon A and Jenkins JM. Marital conflict trajectories and associations with children's disruptive behavior. *Journal of Marriage and Family* 2017; 79: 437–450.

[14] Robbers SC, Bartels M, van Oort FV et al. A twin-singleton comparison of developmental trajectories of externalizing and internalizing problems in 6- to 12-year-old children. *Twin Research and Human Genetics* 2010; 13: 79–87.

[15] Caspi A, Moffitt TE, Newman DL et al. Behavioral observations at age 3 years predict adult psychiatric disorders: Longitudinal evidence from a birth cohort. *Archives of General Psychiatry* 1996; 53: 1033–1039.

[16] Moffitt TE. Adolescence-limited and life-course-persistent antisocial behavior: A developmental taxonomy. *Psychological Review* 1993; 100: 674–701.

[17] Simonoff E, Elander J, Holmshaw J et al. Predictors of antisocial personality. *The British Journal of Psychiatry* 2004; 184: 118–127.

[18] Toumbourou JW, Williams I, Letcher P et al. Developmental trajectories of internalising behaviour in the prediction of adolescent depressive symptoms. *Australian Journal of Psychology* 2011; 63: 214–223.

[19] Nagin DS and Tremblay RE. Trajectories of boys' physical aggression, opposition, and hyperactivity on the path to physically violent and nonviolent juvenile delinquency. *Child Development* 1999; 70: 1181.

[20] O'Connor EE, Dearing E and Collins BA. Teacher-child relationship and behavior problem trajectories in elementary school. *American Educational Research Journal* 2011; 48: 120–162.

[21] Fanti KA and Henrich CC. Trajectories of pure and co-occurring internalizing and externalizing problems from age 2 to age 12: Findings from the national institute of child health and human development study of early child care. *Developmental Psychology* 2010; 46: 1159–1175.

[22] Achenbach TM and Rescorla LA. *Manual for the ASEBA School-Age Forms & Profiles*. Burlington, VT: University of Vermont, Research Center for Children, Youth & Families: Aseba Burlington; 2001.

[23] Jung T and Wickrama KAS. An introduction to latent class growth analysis and growth mixture modeling. *Social and Personality Psychology Compass* 2008; 2: 302–317.

[24] Nagin DS and Tremblay RE. Analyzing developmental trajectories of distinct but related behaviors: A group-based method. *Psychological Methods* 2001; 6: 18–34.

[25] McLachlan G and Peel D. *Finite Mixture Models*. John Wiley & Sons, Inc.; 2000.

[26] Dempster AP, Laird NM and Rubin DB. Maximum likelihood from incomplete

data via the em algorithm. *Journal of the Royal Statistical Society Series B (Methodological)* 1977; 39: 1–38.

[27] Fraley C, Raftery AE, Murphy TB et al. `mclust` version 4 for R: Normal mixture modeling for model-based clustering, classification, and density estimation. Department of Statistics: University of Washington; 2012. 597.

[28] R Core Team. R: A language and environment for statistical computing. Vienna, Austria; 2014.

[29] Oliveira-Brochado A and Martins FV. Assessing the number of components in mixture models: A review; 2005 [cited 5 September 2017]. Available from: `https://ideas.repec.org/p/por/fepwps/194.html`.

[30] Schwarz G. Estimating the dimension of a model. *The Annals of Statistics* 1978; 6: 461–464.

[31] Raftery AE. Bayesian model selection in social research. *Sociological Methodology* 1995; 25: 111–163.

[32] Nylund KL, Asparouhov T and Muthén BO. Deciding on the number of classes in latent class analysis and growth mixture modeling: A Monte Carlo simulation study. *Structural Equation Modeling: A Multidisciplinary Journal* 2007; 14: 535–569.

[33] Heggeseth BC and Jewell NP. The impact of covariance misspecification in

multivariate gaussian mixtures on estimation and inference: An application to longitudinal modeling. *Statistics in Medicine* 2013; 32: 2790–2803.

[34] Gray G. Bias in misspecified mixtures. *Biometrics* 1994; 50: pp. 457–470.

[35] Lo Y. Bias from misspecification of the component variances in a normal mixture. *Computational Statistics & Data Analysis* 2011; 55: 2739–2747.

[36] Enders CK and Tofighi D. The impact of misspecifying class-specific residual variances in growth mixture models. *Structural Equation Modeling: A Multidisciplinary Journal* 2008; 15: 75–95.

[37] Gilthorpe MS, Dahly DL, Tu YK et al. Challenges in modelling the random structure correctly in growth mixture models and the impact this has on model mixtures. *Journal of Developmental Origins of Health and Disease* 2014; 5: 197–205.

[38] Diallo TMO, Morin AJS and Lu H. Impact of misspecifications of the latent variance–covariance and residual matrices on the class enumeration accuracy of growth mixture models. *Structural Equation Modeling: A Multidisciplinary Journal* 2016; 23: 507–531.

[39] Hodge V and Austin J. A survey of outlier detection methodologies. *Artificial Intelligence Review* 2004; 22: 85–126.

[40] Dixon WJ. Processing data for outliers. *Biometrics* 1953; 9: 74–89.

[41] Barnett V and Lewis T. *Outliers in Statistical Data*. New York: Wiley; 1994.

[42] Chandola V, Banerjee A and Kumar V. Anomaly detection: A survey. *ACM Computing Surveys* 2009; 41: 15:1–15:58.

[43] Fraley C and Raftery AE. How many clusters? Which clustering method? Answers via model-based cluster analysis. *The Computer Journal* 1998; 41: 578–588.

[44] Longford NT and D'Urso P. Mixture models with an improper component. *Journal of Applied Statistics* 2011; 38: 2511–2521.

[45] Zhuang S, Huang Y, Palaniappan K et al. Gaussian mixture density modeling, decomposition, and applications. *Image Processing, IEEE Transactions on* 1996; 5: 1293–1302.

[46] Peel D and McLachlan GJ. Robust mixture modelling using the t distribution. *Statistics and Computing* 2000; 10: 339–348.

[47] Lin TI. Robust mixture modeling using multivariate skew t distributions. *Statistics and Computing* 2010; 20: 343–356.

[48] Lo K and Gottardo R. Flexible mixture modeling via the multivariate t distribution with the box-cox transformation: An alternative to the skew-t distribution. *Statistics and Computing* 2012; 22: 33–52.

[49] Forbes F and Wraith D. A new family of multivariate heavy-tailed distributions

with variable marginal amounts of tailweight: Application to robust clustering. *Statistics and Computing* 2013; p. 1–14.

[50] Lee S and McLachlan GJ. Finite mixtures of multivariate skew t-distributions: Some recent and new results. *Statistics and Computing* 2014; 24: 181–202.

[51] Roberts S and Tarassenko L. A probabilistic resource allocating network for novelty detection. *Neural Computation* 1994; 6: 270–284.

[52] Farcomeni A. Robust constrained clustering in presence of entry-wise outliers. *Technometrics* 2014; 56: 102–111.

[53] Kitagawa G. On the use of AIC for the detection of outliers. *Technometrics* 1979; 21: 193–199.

[54] Kadota K, Tominaga D, Akiyama Y et al. Detecting outlying samples in microarray data: A critical assessment of the effect of outliers on sample classification. *Chem-Bio Informatics Journal* 2003; 3: 30–45.

[55] Nagin DS. *Group-Based Modeling of Development*. Cambridge, MA, USA: Harvard University Press; 2005.

[56] Dayton CM and Macready GB. Concomitant-variable latent-class models. *Journal of the American Statistical Association* 1988; 83: 173–178.

[57] Bolck A, Croon M and Hagenaars J. Estimating latent structure models with

categorical variables: One-step versus three-step estimators. *Political Analysis* 2004; 12: 3–27.

[58] Vermunt JK. Latent class modeling with covariates: Two improved three-step approaches. *Political Analysis* 2010; 18: 450–469.

[59] Clark SL and Muthén B. Relating latent class analysis results to variables not included in the analysis; 2009 [cited 5 September 2017]. Available from: `www.statmodel.com/download/relatinglca.pdf`.

[60] Asparouhov T and Muthén B. Auxiliary variables in mixture modeling: Three-step approaches using Mplus. *Structural Equation Modeling: A Multidisciplinary Journal* 2014; 21: 329–341.

[61] Davies CE, Glonek GFV and Giles LC. The impact of covariance misspecification in group-based trajectory models for longitudinal data with non-stationary covariance structure. *Statistical Methods in Medical Research* 2017; 26: 1982–1991.

[62] Davies CE, Glonek GFV and Giles LC. Letter to the editor. *Statistical Methods in Medical Research* prepublished May 24, 2017.

[63] Bowers AJ and Sprott R. Examining the multiple trajectories associated with dropping out of high school: A growth mixture model analysis. *The Journal of Educational Research* 2012; 105: 176–195.

[64] Lanza ST and Collins LM. A mixture model of discontinuous development in heavy drinking from ages 18 to 30: The role of college enrollment. *Journal of Studies on Alcohol and Drugs* 2006; 67: 552.

[65] Hser YI, Huang D, Chou CP et al. Trajectories of heroin addiction: Growth mixture modeling results based on a 33-year follow-up study. *Evaluation Review* 2007; 31: 548–563.

[66] Redner RA and Walker HF. Mixture densities, maximum likelihood and the EM algorithm. *SIAM Review* 1984; 26: 195–239.

[67] Berger JO. *Statistical Decision Theory and Bayesian Analysis*. 2nd ed. New York: Springer; 1985.

[68] Hallquist M and Wiley J. MplusAutomation: Automating Mplus model estimation and interpretation (version 0.6-3); 2014.

[69] Fraley C and Raftery AE. Model-based clustering, discriminant analysis and density estimation. *Journal of the American Statistical Association* 2002; 97: 611–631.

[70] Hastie T, Tibshirani R and Friedman J. *The Elements of Statistical Learning*. New York: Springer; 2009.

[71] Chatfield C and Collins A. *Introduction to Multivariate Analysis*. London: Chapman and Hall; 1980.

[72] Atkinson AC. Masking unmasked. *Biometrika* 1986; 73: 533–541.

[73] Davies CE, Giles LC and Glonek GF. Performance of methods for estimating the effect of covariates on group membership probabilities in group-based trajectory models. *Statistical Methods in Medical Research* prepublished January 18, 2017.

[74] Mathai AM and Provost SB. *Quadratic Forms in Random Variables: Theory and Applications*. New York: Marcel Dekker, Inc.; 1992.

[75] Diallo TMO and Lu H. On the application of the three-step approach to growth mixture models. *Structural Equation Modeling: A Multidisciplinary Journal* 2017; 24: 714–732.

[76] Kim M, Vermunt J, Bakk Z et al. Modeling predictors of latent classes in regression mixture models. *Structural Equation Modeling: A Multidisciplinary Journal* 2016; 23: 601–614.