Regression modeling with missing outcomes : competing risks and longitudinal data

Margarita Moreno Betancur

Année : 2013 N° |_|_|_|_|_|_|_|_|_|_|

UNIVERSITÉ PARIS-SUD XI

FACULTÉ DE MÉDECINE

ÉCOLE DOCTORALE DE SANTÉ PUBLIQUE

THÈSE DE DOCTORAT

Spécialité Biostatistique

Contributions aux modèles de régression avec réponses manquantes : Risques concurrents et données longitudinales

Regression modeling with missing outcomes: Competing risks and longitudinal data

Margarita Moreno Betancur

Thèse dirigée par Aurélien Latouche

Présentée et soutenue publiquement le 5 décembre 2013

Jury :

M. Joël Coste, Pr.	Président
Mme. Els Goetghebeur, Pr.	Rapporteur
M. Geert Molenberghs, Pr.	Rapporteur
M. Per Kragh Andersen, Pr.	Examinateur
Mme. Agathe Guilloux, Dr.	Examinateur
M. Aurélien Latouche, Pr.	Directeur de thèse

It's not what you know that hurts you;
it's the things you think you know, but don't.

– Scharfstein, Rotnitzky & Robins, *JASA*, 1999

Acknowledgements

Mes remerciements s'adressent en premier lieu à Aurélien Latouche, mon directeur de thèse, pour son encadrement très attentif, son soutien constant, sa disponibilité à toute heure et ses qualités humaines. Ce fut un grand plaisir de travailler avec lui sur le sujet qu'il m'a proposé et de bénéficier de son expérience dans le domaine des risques concurrents. J'ai apprécié la confiance qu'il m'a accordée pour mener ce travail vers les domaines de recherche qui m'intéressaient plus spécialement. Aurélien a énormément contribué à ma formation de chercheuse aussi en m'encourageant à présenter mes travaux, en me mettant en relation avec d'autres chercheurs et en me donnant l'opportunité de faire un séjour à l'Université de Copenhague.

Je tiens à remercier chaleureusement Michel Chavance, qui m'a introduit au domaine des données manquantes pendant mon stage de master et avec qui j'ai pu continuer de collaborer pendant ma thèse. J'ai beaucoup apprécié travailler avec lui, et nos nombreux échanges très enrichissants pour moi du point de vue statistique.

Je remercie vivement mes collaborateurs au CépiDc, notamment Grégoire Rey, de m'avoir proposé d'appliquer et développer des méthodes pour l'étude des inégalités socioéconomiques de mortalité. Je me réjouis de pouvoir continuer ces travaux au sein de leur équipe dans la suite.

Je voudrais remercier les personnes et institutions qui nous ont donné accès aux données permettant de motiver et illustrer les méthodes développées dans cette thèse. Notamment, merci à l'Eastern Cooperative Oncology Group et au Professeur Robert Gray pour les données de l'essai clinique sur le cancer du sein et à Sanofi-Aventis France

pour les données de l'essai clinique sur le trouble du maintien du sommeil.

Thank you to all the members of the PhD committee for accepting to participate in the evaluation of this dissertation. I am grateful to Professor Joël Coste for giving me the honor of presiding the committee; to my referees, Professors Els Goetghebeur and Geert Molenberghs, for the time they devoted to carefully reading this dissertation and for their valuable, constructive and kind comments; to my examiners, Professor Per Kragh Andersen and Doctor Agathe Guilloux, for evaluating my work.

Un grand merci à Pascale Tubert-Bitter de m'avoir accueillie au sein de l'équipe Biostatistique du CESP. Je suis très reconnaissante de son soutien permanent, au niveau scientifique et personnel, et ses conseils tout au long de mon stage de master et ma thèse.

Je remercie l'ensemble de l'équipe Biostatistique pour leur accueil chaleureux. Ce fut un grand plaisir d'appartenir à cette équipe. Je tiens à remercier spécialement Ghislaine Breton pour son aide et sa patience dans mes démarches administratives. Je remercie aussi l'ensemble de l'équipe Épidémiologie Respiratoire et Environnementale du CESP avec qui ce fut formidable de partager le quotidien de nos bureaux.

Mes remerciements s'adressent ensuite à tous mes collègues du CESP avec qui j'ai pu partager des moments sympathiques pendant ces années. Je remercie particulièrement Helena, Dorota, Juliette, Fanny, Elsa, Orianne, Margaux, Annabelle, Marta, Emilie, Jonathan et Yves.

A special thank you to Professor Per Kragh Andersen and all the members of the Section of Biostatistics of the University of Copenhagen for a warm welcome during my 5-month research visit. I am grateful for having had the opportunity to attend the seminars and meetings organized by the team, for being able to present my own work and for some very enlightening conversations, especially with Per Kragh Andersen and Thomas Gerds. This stay greatly contributed to my growth as a researcher and a statistician.

Je remercie l'Université Paris-Sud XI qui a financé les trois premiers années de ce travail. Je remercie l'équipe Biostatistique d'avoir financé les derniers trois mois de cette thèse et les congrès et séminaires auxquels j'ai participé au long de ces années.

Je remercie l'Université Pierre et Marie Curie de m'avoir donné l'opportunité d'effectuer ma mission d'enseignement au sein de l'UFR de Mathématiques. Ce fut une expérience très enrichissante. Un merci spécial à Vincent Humilière et Elodie Duprat pour leur encadrement.

Quisiera agradecer a toda mi familia y a mis amigos por su apoyo incondicional, sin quienes ninguno de mis logros sería posible. Agradezco especialmente a mis padres, que siempre han creído en mí y me han proporcionado todo lo necesario para que logre mis metas. También agradezco a mi hermana Ana María quien ha abierto caminos para mí y desde siempre ha sido un modelo y una guía esencial. Gracias a mis hermanitos, Natalia y Daniel, y a Patricia, Ricardo, Gabor, Olivia y la familia Dugué por la alegría y el apoyo que me brindan. Estoy muy feliz de poder compartir este logro con ustedes.

Enfin, je voudrais remercier Pierre-Antoine pour son écoute attentive et ses conseils avisés, qu'il m'a toujours apportés avec une grande générosité. Il a été beaucoup plus qu'une compagnie et un soutien inconditionnels dans ce chemin.

Abstract

Missing data are a common occurrence in medical studies. In regression modeling, missing outcomes limit our capability to draw inferences about the covariate effects of medical interest, which are those describing the distribution of the entire set of planned outcomes. In addition to losing precision, the validity of any method used to draw inferences from the observed data will require that some assumption about the mechanism leading to missing outcomes holds. Rubin (1976) called the missingness mechanism MAR (for *missing at random*) if the probability of an outcome being missing does not depend on missing outcomes when conditioning on the observed data, and MNAR (for *missing not at random*) otherwise. This distinction has important implications regarding the modeling requirements to draw valid inferences from the available data, but generally it is not possible to assess from these data whether the missingness mechanism is MAR or MNAR. Hence, sensitivity analyses should be routinely performed to assess the robustness of inferences to assumptions about the missingness mechanism.

In the field of incomplete multivariate data, in which the outcomes are gathered in a vector $(Y_1, \ldots, Y_J)$ for which some components may be missing, MAR methods are widely available and increasingly used, and several MNAR modeling strategies have also been proposed. On the other hand, although some sensitivity analysis methodology has been developed, this is still an active area of research. The first aim of this dissertation was to develop a sensitivity analysis approach for continuous longitudinal data with drop-outs, that is, continuous outcomes that are ordered in time and completely observed for each individual up to a certain time-point, at which the individual drops-

out so that all the subsequent outcomes are missing. The proposed approach consists in assessing the inferences obtained across a family of MNAR pattern-mixture models indexed by a so-called sensitivity parameter that quantifies the departure from MAR. The approach was prompted by a randomized clinical trial investigating the benefits of a treatment for sleep-maintenance insomnia, from which 22% of the individuals had dropped-out before the study end.

The second aim was to build on the existing theory for incomplete multivariate data to develop methods for competing risks data with missing causes of failure. The competing risks model is an extension of the standard survival analysis model in which failures from different causes are distinguished. Strategies for modeling competing risks functionals, such as the cause-specific hazards (CSH) and the cumulative incidence function (CIF), generally assume that the cause of failure is known for all patients, but this is not always the case. Some methods for regression with missing causes under the MAR assumption have already been proposed, especially for semi-parametric modeling of the CSH. But other useful models have received little attention, and MNAR modeling and sensitivity analysis approaches have never been considered in this setting. We propose a general framework for semi-parametric regression modeling of the CIF under MAR using inverse probability weighting and multiple imputation ideas. Also under MAR, we propose a direct likelihood approach for parametric regression modeling of the CSH and the CIF. Furthermore, we consider MNAR pattern-mixture models in the context of sensitivity analyses. In the competing risks literature, a starting point for methodological developments for handling missing causes was a stage II breast cancer randomized clinical trial in which 23% of the deceased women had missing cause of death. We use these data to illustrate the practical value of the proposed approaches.

Keywords: Missing data; longitudinal data; competing risks; regression; missing outcomes; drop-out; missing cause of failure; multiple imputation; inverse probability weighting; direct likelihood; pattern-mixture model; sensitivity analysis; linear mixed model; cumulative incidence function; cause-specific hazard; pseudo-values.

Résumé

Les données manquantes sont fréquentes dans les études médicales. Dans les modèles de régression, les réponses manquantes limitent notre capacité à faire des inférences sur les effets des covariables décrivant la distribution de la totalité des réponses prévues sur laquelle porte l'intérêt médical. Outre la perte de précision, toute inférence statistique requière qu'une hypothèse sur le mécanisme de manquement soit vérifiée. Rubin (1976) a appelé le mécanisme de manquement MAR (pour les sigles en anglais de *manquant au hasard*) si la probabilité qu'une réponse soit manquante ne dépend pas des réponses manquantes conditionnellement aux données observées, et MNAR (pour les sigles en anglais de *manquant non au hasard*) autrement. Cette distinction a des implications importantes pour la modélisation, mais en général il n'est pas possible de déterminer si le mécanisme de manquement est MAR ou MNAR à partir des données disponibles. Par conséquent, il est indispensable d'effectuer des analyses de sensibilité pour évaluer la robustesse des inférences aux hypothèses de manquement.

Pour les données multivariées incomplètes, c'est-à-dire, lorsque l'intérêt porte sur un vecteur de réponses $(Y_1, \ldots, Y_J)$ dont certaines composantes peuvent être manquantes, plusieurs méthodes de modélisation sous l'hypothèse MAR et, dans une moindre mesure, sous l'hypothèse MNAR ont été proposées. En revanche, le développement de méthodes pour effectuer des analyses de sensibilité est un domaine actif de recherche. Le premier objectif de cette thèse était de développer une méthode d'analyse de sensibilité pour les données longitudinales continues avec des sorties d'étude, c'est-à-dire, pour les réponses continues, ordonnées dans le temps, qui sont complètement observées pour chaque in-

dividu jusqu'à la fin de l'étude ou jusqu'à ce qu'il sorte définitivement de l'étude. Dans l'approche proposée, on évalue les inférences obtenues à partir d'une famille de modèles MNAR dits « de mélange de profils », indexés par un paramètre qui quantifie le départ par rapport à l'hypothèse MAR. La méthode a été motivée par un essai clinique étudiant un traitement pour le trouble du maintien du sommeil, durant lequel 22% des individus sont sortis de l'étude avant la fin.

Le second objectif était de développer des méthodes pour la modélisation de risques concurrents avec des causes d'évènement manquantes en s'appuyant sur la théorie existante pour les données multivariées incomplètes. Les risques concurrents apparaissent comme une extension du modèle standard de l'analyse de survie où l'on distingue le type d'évènement ou la cause l'ayant entrainé. Les méthodes pour modéliser le risque cause-spécifique et la fonction d'incidence cumulée supposent en général que la cause d'évènement est connue pour tous les individus, ce qui n'est pas toujours le cas. Certains auteurs ont proposé des méthodes de régression gérant les causes manquantes sous l'hypothèse MAR, notamment pour la modélisation semi-paramétrique du risque. Mais d'autres modèles n'ont pas été considérés, de même que la modélisation sous MNAR et les analyses de sensibilité. Nous proposons des estimateurs pondérés et une approche par imputation multiple pour la modélisation semi-paramétrique de l'incidence cumulée sous l'hypothèse MAR. En outre, nous étudions une approche par maximum de vraisemblance pour la modélisation paramétrique du risque et de l'incidence sous MAR. Enfin, nous considérons des modèles de mélange de profils dans le contexte des analyses de sensibilité. Un essai clinique étudiant un traitement pour le cancer du sein de stade II avec 23% des causes de décès manquantes sert à illustrer les méthodes proposées.

Mots clés: Données manquantes; données longitudinales; risques concurrents ; régression ; réponses manquantes ; sorties d'étude ; cause d'évènement manquante ; imputation multiple ; estimateurs pondérés ; maximum de vraisemblance ; modèle de mélange de profils ; analyse de sensibilité ; modèle linéaire mixte ; fonction d'incidence cumulée ; risque cause-spécifique ; pseudo-valeurs.

Scientific production

Published manuscripts

Moreno-Betancur M, Latouche A. Regression modeling of the cumulative incidence function with missing causes of failure using pseudo-values. *Statistics in Medicine* 2013; **32**(18):3206-3223. DOI: 10.1002/sim.5755.

Moreno-Betancur M, Chavance M. Sensitivity analysis of incomplete longitudinal data departing from the missing at random assumption: Methodology and application in a clinical trial with drop-outs. *Statistical Methods in Medical Research* [Epub ahead of print on May 22, 2013]. DOI: 10.1177/0962280213490014.

Unpublished manuscripts

Moreno-Betancur M, Rey G, Latouche A. Competing risks regression with missing causes of death: Assessing the sensitivity of inferences to missing data assumptions (*in preparation*).

Moreno-Betancur M, Latouche A, Rey G. A structured framework for estimation of the relative index and slope index of inequality in studies of socioeconomic gradients in event rates and risks (*in preparation*).

Conference presentations

Moreno-Betancur M, Latouche A**. Regression modelling with missing causes of death: Assessing the sensitivity of inferences to missingness assumptions, *6th International Conference of the ERCIM WG on Computational and Methodological Statistics*, London, England, December 2013.

Moreno-Betancur M*, Rey G, Latouche A. Competing risks regression with missing causes of death: Assessing the sensitivity of inferences to missing data assumptions,

4th Conference of the International Biometric Society Channel Network, St Andrews, Scotland, July 2013.

Moreno-Betancur M*, Latouche A. Regression modeling of the cumulative incidence function with missing causes of failure, *Joint Statistical Meetings of the American Statistical Association*, San Diego, USA, July 2012.

Moreno-Betancur M**, Latouche A. Regression modeling of the cumulative incidence function with missing causes of failure, *International Conference on Statistical Models and Methods for Reliability and Survival Analyses and their Validation (S2MRSA)*, Bordeaux, France, July 2012.

Moreno-Betancur M, Latouche A**. Regression modeling of the cumulative incidence function with missing causes of failure using pseudo-observations, *10th Annual ASA CT Chapter Mini-Conference at Yale University*, Connecticut, USA, March 2012.

Moreno-Betancur M, Latouche A**. Regression modeling of the cumulative incidence function with missing causes of failure using pseudo-observations, *58. Biometrisches Kolloquium*, Berlin, Germany, March 2012.

Moreno-Betancur M*, Latouche A. Additive model for the cumulative incidence function when some causes of failure are missing, *3rd Conference of the International Biometric Society Channel Network*, Bordeaux, France, April 2011.

Moreno-Betancur M*, Chavance M. Imputation multiple des réponses manquantes après sortie d'étude, *8ème Journées de Jeunes Chercheurs en Biométrie de la Société Française de Biométrie*, Paris, France, November 2010.

Seminars

Moreno-Betancur M**, Rey G, Latouche A. Différentiels de mortalité par suicide avec prise en compte des données manquantes, *Séminaire scientifique du Centre d'épidémiologie sur les causes médicales de décès, Inserm-CépiDc*, Paris, France, November 2013.

Moreno-Betancur M**, Latouche A. Modélisation de la fonction d'incidence cumulée avec des causes d'événement manquantes, *Séminaire du Département de Biostatistique et d'Epidémiologie, Institut Gustave Roussy*, Paris, France, March 2013.

Moreno-Betancur M**, Latouche A. Modélisation de la fonction d'incidence cumulée avec des causes d'événement manquantes, *Séminaire des doctorants du Laboratoire de Statistique Théorique et Appliquée, Université Pierre et Marie Curie*, Paris, France, December 2012.

Moreno-Betancur M*, Latouche A. Measuring socioeconomic inequalities in cause-specific mortality in France, *Department of Biostatistics, University of Copenhagen*, Copenhagen, Denmark, June 2012.

Moreno-Betancur M**, Latouche A. Regression modeling of the cumulative incidence function with missing causes of failure, *Seminar of the Department of Biostatistics, University of Copenhagen*, Copenhagen, Denmark, March 2012.

Moreno-Betancur M**, Chavance M. Imputation multiple des réponses manquantes après sortie d'étude : Application à un essai clinique du trouble du maintien du sommeil, *Sanofi-Aventis*, Paris, France, November 2010.

* speaker (contributed), ** speaker (invited)

Contents

Some notational conventions

- Boldface will be used for vectors and matrices.
- For a matrix or vector $\mathbf{X}$, $\mathbf{X}'$ denotes the transpose of $\mathbf{X}$.
- $I(\cdot)$ is the indicator function, i.e. for a logical proposition q, $I(q) = 1$ if q is true and $I(q) = 0$ otherwise.
- $\xrightarrow{p}$ denotes convergence in probability.
- $o_p(1)$ denotes a sequence of random variables that converges to zero in probability.

For $\mathbf{X}$ and $\mathbf{Y}$ random vectors:

- $\mathbf{X}_i$ denotes the realization of $\mathbf{X}$ by individual i.
- $f(\mathbf{x}|\cdot)$ denotes the (conditional) joint density of $\mathbf{X}$ evaluated at $\mathbf{x}$. Where necessary to avoid ambiguity, the notation $f(\mathbf{x} = \cdot|\cdot)$ will be used instead.
- $f(\mathbf{x}|\mathbf{y})$ is short for $f(\mathbf{x}|\mathbf{Y} = \mathbf{y})$.

List of Acronyms

AJ	Approximate jackknife
CC	Complete case
CCD	Complete censored data
CI	Confidence interval
CIF	Cumulative incidence function
CP	Coverage probability
CSH	Cause-specific hazard
ECOG	Eastern Cooperative Oncology Group
ER	Estrogen-receptor
ES	Extra state
FEELC	Feeling of sleepiness
GEE	Generalized estimating equations
GR	Goetghebeur and Ryan
IPCW	Inverse probability of censoring weighting
IPW	Inverse probability weighting
IPWpv	Inverse probability weighted pseudo-values
LMM	Linear mixed model
LT	Lu and Tsiatis
MAR	Missing at random
MB	Mean bias
MCAR	Missing completely at random
MI	Multiple imputation
MNAR	Missing not at random
MRB	Mean relative bias
MSE	Mean squared error
NAW	Number of nocturnal awakenings
PDS	Permanent demographic sample
PMM	Pattern-mixture model
RRE	Root relative efficiency
SD	Standard deviation
SE	Standard error
SLREF	Sleep refreshing quality
SMI	Sleep-maintenance insomnia
SOL	Sleep onset latency
TST	Total sleep time
WASO	Wake-time after sleep onset
WGEE	Weighted generalized estimating equations

List of Tables

List of Figures

Synthèse
(extended summary in French)

Introduction

La modélisation par régression avec des réponses manquantes

Les données manquantes sont fréquentes dans les études médicales, notamment dans celles impliquant des êtres humains. Par exemple, lorsqu'on demande aux participants d'une étude de remplir un questionnaire, certains d'entre eux laisseront souvent quelques items sans réponse. Dans les études médicales, il est commun de suivre un groupe d'individus au cours du temps, par exemple pour étudier l'évolution d'un marqueur biologique mesuré à plusieurs reprises ou le temps jusqu'à la survenue d'un évènement tel que le décès. L'observation prolongée des individus accentue l'occurrence de données manquantes. En effet, le recueil d'information se détériore avec le temps, entre autres parce que quelques individus n'assistent pas aux rendez-vous planifiés et d'autres sortent des études complètement par des raisons migratoires, médicales ou autres. Il est difficile d'éviter ces problèmes entièrement quelque soit le type d'étude, même lorsqu'il s'agit d'un essai clinique soigneusement planifié et suivant un protocole strict.

Dans la recherche médicale, la modélisation par régression est souvent utilisée pour étudier l'effet des covariables (ex. facteurs pronostiques, expositions) sur des réponses d'intérêt médical (ex. le niveau d'un marqueur biologique, le délai de survie). Dans ces études, les individus sont recrutés suivant des critères prédéfinis, et les réponses

qui seront mesurées pour chacun d'entre eux, ainsi que les dates et horaires de ces mesures, sont déterminés à l'avance. L'intérêt médical porte sur les effets des covariables décrivant la distribution de la totalité des réponses prévues. Or, lorsque quelques réponses manquent pour certains individus, les réponses observées et les covariables représentent toute l'information disponible pour faire des inférences (c.-à-d. des estimations ponctuelles, des intervalles de confiance, des tests d'hypothèse) sur ces paramètres. Outre l'évidente perte de précision entrainée par l'information manquante, toute méthode pour faire des inférences à partir des données disponibles requière qu'une hypothèse sur le mécanisme de manquement soit vérifiée.

Rubin (1976) a proposé une taxonomie des mécanismes de manquement qui est clé pour comprendre la problématique de l'analyse statistique de données incomplètes. Dans le contexte des réponses manquantes, le mécanisme de manquement est appelé MCAR (pour les sigles en anglais de *manquant complètement au hasard*) si la probabilité qu'une réponse soit manquante est constante. Le mécanisme de manquement est appelé MAR (pour les sigles en anglais de *manquant au hasard*) si la probabilité qu'une réponse soit manquante ne dépende pas des réponses manquantes conditionnellement aux données observées (ex. des covariables, des réponses mesurées à d'autres moments). Enfin, le mécanisme de manquement est appelé MNAR (pour les sigles en anglais de *manquant non au hasard*) si la probabilité qu'une réponse soit manquante dépend des réponses manquantes même lorsqu'on conditionne par les données observées.

La taxonomie de Rubin a des implications importantes concernant les besoins de modélisation pour obtenir des inférences valides à partir des données disponibles. Sous l'hypothèse d'un mécanisme MCAR, les réponses observées sont un échantillon aléatoire des réponses prévues. Par conséquent, dans plusieurs contextes, un mécanisme MCAR garantit que l'on puisse obtenir des inférences, certes imprécises, mais tout de même valides en effectuant une analyse dite des « cas complets » (CC), c'est à dire, une analyse excluant les individus avec des réponses manquantes. Sous MAR, les réponses observées et manquantes ont la même distribution conditionnellement au reste des données observées. Ceci implique que des inférences valides peuvent être obtenues à partir

des données disponibles sans avoir à faire des hypothèses supplémentaires sur la distribution des réponses manquantes. Sous MNAR, les distributions conditionnelles des réponses observées et manquantes sont différentes, et il est alors nécessaire de faire des hypothèses sur cette dernière, que ce soit de façon explicite ou implicite, pour résoudre les problèmes d'identifiabilité.

La taxonomie de Rubin relève des différences importantes entre les diverses méthodes existantes pour gérer les réponses manquantes. Or, comme Molenberghs et al. (2008) l'a démontré dans un contexte assez général, il n'est pas possible d'évaluer à partir des données observées si le mécanisme de manquement est MAR ou MNAR. En effet, sauf dans certaines situations où le schéma de l'étude prévoit le manquement de certaines données, les hypothèses de manquement requises par quelconque stratégie de modélisation avec des données manquantes ne sont pas vérifiables. C'est pourquoi il est indispensable d'effectuer des *analyses de sensibilité*, pour évaluer la robustesse des inférences obtenues dans une analyse principale aux écarts par rapport aux hypothèses de manquement sous-jacentes.

Objectifs de la thèse

L'analyse CC et d'autres méthodes ad-hoc étaient souvent utilisées auparavant, et elles sont toujours utilisées mais dans une moindre mesure. En effet, une meilleure compréhension des fortes hypothèses requises par ces méthodes, irréalistes dans la plupart des cas, a stimulé le développement, l'implémentation et l'utilisation de méthodes reposant sur des hypothèses moins fortes. Ceci est en particulier vrai dans le contexte des données multivariées incomplètes, c'est-à-dire, lorsque l'intérêt porte sur un vecteur de réponses $(Y_1, \ldots, Y_J)$ dont certaines composantes peuvent être manquantes. Pour ce type de données, plusieurs méthodes de modélisation reposant sur l'hypothèse MAR sont utilisées couramment et diverses méthodes sous MNAR ont été proposées aussi. En revanche, même si quelques méthodes pour effectuer des analyses de sensibilité ont été proposées, celui-ci est encore un domaine actif de recherche.

Le premier objectif de cette thèse était de développer une méthode d'analyse de sensibilité pour les données longitudinales continues avec des sorties d'étude. Ces données sont un type particulier de données multivariées incomplètes, où les réponses sont continues et ordonnées dans le temps. De plus, pour chaque individu, on observe soit toutes les réponses, soit toutes les réponses jusqu'à un certain moment, auquel l'individu sort de l'étude et ne revient jamais, de façon à ce que le reste de ses réponses sont manquantes.

Le second objectif de la thèse était de développer des méthodes pour la modélisation de risques concurrents avec des causes d'évènement manquantes en s'appuyant sur la théorie existante pour les données multivariées incomplètes. Le modèle de risques concurrents est une extension du modèle standard de l'analyse de survie où l'on distingue le type d'évènement ou la cause l'ayant entrainé. Certains auteurs ont proposé des méthodes de régression pour risques concurrents avec des causes d'évènement manquantes sous l'hypothèse MAR, notamment pour la modélisation semi-paramétrique du risque cause-spécifique. Mais d'autres modèles n'ont pas été considérés, de même que la modélisation sous MNAR et les analyses de sensibilité. Notre but était donc de combler certains de ces vides dans la littérature.

Modélisation de données longitudinales avec des sorties d'étude

Contexte et état de la question

On parle de *données longitudinales* lorsqu'une variable réponse d'intérêt est mesurée pour chaque individu à plusieurs reprises au cours du temps, donnant lieu à une séquence de mesures répétées ordonnées dans le temps. Ici, nous nous intéressons aux variables réponse continues et, pour faciliter la notation, nous nous concentrons sur le cas où tous les individus ont le même nombre J de mesures prévues, ayant lieu aux mêmes dates. On dénote par $\mathbf{Y}_i = (Y_{i1}, \ldots, Y_{iJ})$ le vecteur de réponses de l'individu i $(i = 1, \ldots, n)$.

L'objectif d'une étude longitudinale est d'étudier l'évolution intra-individu de la variable réponse au cours du temps et de déterminer les facteurs qui influencent cette évolution.

Un exemple d'étude longitudinale est l'essai clinique randomisé qui a motivé les développements de cette partie de la thèse. Dans cet essai, on étudiait un traitement pour le trouble du maintien du sommeil. Les 962 patients recrutés dans cet essai ont eu à remplir quotidiennement un questionnaire pendant la durée de l'étude qui contenait des scores quantifiant la qualité du sommeil. Les scores étaient : la durée de réveil après l'induction du sommeil (WASO); le nombre de réveils (NAW); la qualité rafraichissante du sommeil (SLREF), prenant des valeurs de un (excellent) à quatre (pauvre); la sensation de somnolence (FEELC), prenant des valeurs de zéro (très somnolent) à neuf (pas du tout somnolent); le temps total de sommeil (TST); et la latence à l'endormissement (SOL), qui est la durée entre la fermeture des yeux et l'entrée effective en sommeil. Une décroissance dans les scores WASO, NAW, SLREF ou SOL, ou une croissance dans les scores FEELC ou TST indiquerait une amélioration dans la qualité du sommeil du patient. Les scores quotidiens n'étaient pas disponibles, seuls leurs moyennes sur six périodes communes à tous les individus. Donc, toutes les variables réponse considérées étaient continues.

Pour étudier l'effet des covariables sur l'évolution de la variable réponse, des méthodes de régression prenant compte de la corrélation entre les réponses d'un même individu au cours du temps sont nécessaires. Si toutes les réponses de tous les individus sont observées, on peut utiliser par exemple le modèle linéaire mixte (Laird and Ware, 1982).

Lorsqu'il y a des sorties d'étude, le vecteur de réponses s'écrit $\mathbf{Y}_i = (\mathbf{Y}_i^{\mathcal{O}}, \mathbf{Y}_i^{\mathcal{M}})$, où $\mathbf{Y}_i^{\mathcal{O}} = (Y_{i1}, \ldots, Y_{i(U_i-1)})$ et $\mathbf{Y}_i^{\mathcal{M}} = (Y_{iU_i}, \ldots, Y_{iJ})$ sont les parties observées et manquantes de $\mathbf{Y}_i$, respectivement, et U_i dénote le moment de la première réponse manquante, avec $U_i \leq J$ pour les individus qui sont sortis de l'étude et $U_i = J+1$ pour ceux qui complètent l'étude. On appelle U_i *l'indicateur de sortie d'étude*. Plusieurs méthodes sont disponibles pour l'analyse de données longitudinales avec des sorties d'étude sous l'hypothèse MAR. Par exemple, Rubin (1976) a montré qu'une analyse par maximum de

vraisemblance utilisant toutes les données disponibles permet d'obtenir des estimations non-biaisées et d'efficacité maximale à condition que (i) le mécanisme de manquement soit MAR et (ii) les paramètres des mécanismes de réponse et manquement soient distincts. Si ces deux conditions sont vérifiées on dit que le mécanisme de manquement est *ignorable*. En effet, on peut obtenir des inférences valides tout simplement en ajustant le modèle linéaire mixte aux réponses disponibles. Deux autres approches utiles sous MAR sont l'imputation multiple (Rubin, 1987) et les équations d'estimation pondérées par l'inverse de la probabilité d'observation (Robins et al., 1995).

Sous MNAR, il faut modéliser la distribution jointe du vecteur de réponses $\mathbf{Y}$ et de l'indicateur de sortie d'étude U, dénotée par $f(\mathbf{y}, u)$. Deux approches distinguées par Little and Rubin (1987, Chapitre 11) sont les *modèles de sélection* et les *modèles de mélange de profils*. Ces derniers, d'importance pour la suite, sont fondés sur la factorisation suivante:

$$f(\mathbf{y}, u) = f(\mathbf{y}|u) \times f(u).$$

Une troisième approche est celle des *modèles à paramètres partagés* (Wu and Carroll, 1988; Wu and Bailey, 1988, 1989; Little, 1995).

Une approche pour effectuer des analyses de sensibilité consiste à évaluer la discordance entre les inférences obtenues dans l'analyse principale et celles obtenues à partir d'une famille de modèles MNAR reposant sur des hypothèses de manquement, distributionnelles ou structurelles différentes (ex. Little and Yau, 1996; Kenward, 1998; Kenward and Molenberghs, 1999; Michiels et al., 2002). Si l'analyse principale repose sur l'hypothèse MAR, une version plus structurée de cette approche consiste à considérer une famille de modèles MNAR indexés par un paramètre qui quantifie le départ par rapport à l'hypothèse MAR (Little, 1994; Rotnitzky et al., 1998; Scharfstein et al., 1999; Daniels and Hogan, 2000; Molenberghs et al., 2001a). D'autres méthodes d'analyse de sensibilité reposant sur les idées d'influence de Cook (1977, 1986) ont été proposées (Thijs et al., 2000; Verbeke et al., 2001; Molenberghs et al., 2001b; Jansen et al., 2006).

Contributions

L'approche développée pour répondre au premier objectif de la thèse est fondée sur le principe d'une famille de modèles MNAR indexés par un paramètre quantifiant le départ par rapport à MAR. On utilise des modèles de mélange de profils, considérés par certains auteurs comme les plus appropriés pour les analyses de sensibilité (Daniels and Hogan, 2000; Daniels and Wang, 2009; Hogan, 2009). Plus précisément, soit $\boldsymbol{\varphi}$ le paramètre de $f(\mathbf{y}, u)$. Le paramètre d'intérêt $\boldsymbol{\theta}$ est en général une fonction h de $\boldsymbol{\varphi}$, c'est-à-dire $\boldsymbol{\theta} = h(\boldsymbol{\varphi})$. Les modèles de mélange de profils sont particulièrement utiles pour les analyses de sensibilité car ils peuvent être paramétrés de façon à ce que $\boldsymbol{\varphi} = (\boldsymbol{\phi}, \boldsymbol{\kappa})$ où $\boldsymbol{\kappa}$ n'apparait pas dans le modèle pour les données observées, qui est donc indexé uniquement par $\boldsymbol{\phi}$. D'autre part, le modèle pour les réponses manquantes conditionnellement aux données observées, appelé le modèle d'extrapolation, est indexé par $(\boldsymbol{\phi}, \boldsymbol{\kappa})$:

$$f(\mathbf{y}^{\mathcal{O}}, \mathbf{y}^{\mathcal{M}}, u|\boldsymbol{\phi}, \boldsymbol{\kappa}) = f(\mathbf{y}^{\mathcal{O}}, u|\boldsymbol{\phi}) \times f(\mathbf{y}^{\mathcal{M}}|\mathbf{y}^{\mathcal{O}}, u, \boldsymbol{\phi}, \boldsymbol{\kappa}). \tag{1}$$

L'intérêt de cette paramétrisation découle des observations suivantes. Premièrement, pour une valeur fixe de $\boldsymbol{\phi}$, toute valeur de $\boldsymbol{\kappa}$ donne le même ajustement aux données observées. C'est-à-dire, la vraisemblance des données observées $\mathcal{L}(\boldsymbol{\phi}, \boldsymbol{\kappa}|\mathbf{y}^{\mathcal{O}}, u)$, considérée comme une fonction de $\boldsymbol{\kappa}$, est constante, impliquant que $\boldsymbol{\kappa}$ n'est pas identifiable. Deuxièmement, la vraisemblance $\mathcal{L}(\boldsymbol{\phi}, \boldsymbol{\kappa}|\mathbf{y}^{\mathcal{O}}, u)$, vue comme une fonction de $\boldsymbol{\phi}$, n'est pas constante, donc ce paramètre est identifiable. Enfin, le paramètre d'intérêt $\boldsymbol{\theta} = h(\boldsymbol{\phi}, \boldsymbol{\kappa})$ dépend en général de $\boldsymbol{\kappa}$. En conclusion, différentes valeurs de $\boldsymbol{\kappa}$ impliquent le même ajustement aux données observées, mais aussi des modèles d'extrapolation et des valeurs pour le paramètre d'intérêt différents. Un paramètre comme $\boldsymbol{\kappa}$ est appelé un *paramètre de sensibilité* car il représente la source des différences dans les inférences obtenues sous diverses hypothèses (non-vérifiables) sur le modèle d'extrapolation. Une analyse de sensibilité peut s'effectuer en faisant varier $\boldsymbol{\kappa}$ sur un ensemble de valeurs possibles, déterminé éventuellement avec l'aide d'experts, et en comparant les inférences obtenues sur cet ensemble et celles de l'analyse principale.

Nous considérons une famille de modèles de mélange de profils où l'on distingue uniquement entre les réponses observées et manquantes. Plus précisément, dénotant l'indicateur de manquement de Y_{ij} par $R_{ij} := I(U_i \leq j)$, on considère des modèles linéaires mixtes de la forme:

$$Y_{ij} = \mathbf{X}'_{ij}\boldsymbol{\beta} + \mathbf{Z}'_{ij}\mathbf{b}_i + \kappa R_{ij} + \varepsilon_{ij}, \qquad \varepsilon_{ij} \sim N(0, \sigma^2), \qquad \mathbf{b}_i \sim N(\mathbf{0}, \mathbf{G}), \qquad (2)$$

pour $j = 1, \ldots, J$ et $i = 1, \ldots, n$, où : $\mathbf{X}_{ij}$ est un p-vecteur contenant des covariables fixes mesurées au début de l'étude et des polynômes de t_j, le temps de la mesure j ; $\boldsymbol{\beta}$ est le p-vecteur des effets fixes communs à la population ; le q-vecteur $\mathbf{Z}_{ij}$ contient les covariables dans $\mathbf{X}_{ij}$ à effet aléatoire ; les erreurs résiduelles ε_{ij} sont indépendantes et identiquement distribuées (i.i.d), suivant une distribution normale de moyenne zéro et variance σ^2 ; et les effets aléatoires, représentés par les q-vecteurs $\mathbf{b}_i$, sont i.i.d. suivant une distribution normale de moyenne zéro et variance $\mathbf{G} = \mathbf{G}(\boldsymbol{\alpha})$, où $\boldsymbol{\alpha}$ est un vecteur de paramètres inconnu. Les effets aléatoires sont supposés être indépendants des erreurs résiduelles et des covariables.

La première partie du prédicteur linéaire, $\mathbf{X}'_{ij}\boldsymbol{\beta} + \mathbf{Z}'_{ij}\mathbf{b}_i$, et les paramètres de la variance, σ^2 et $\mathbf{G}$, déterminent complètement la distribution des réponses observées, pour lesquelles $R_{ij} = 0$. Les paramètres correspondants peuvent être estimés à partir des données observées par maximum vraisemblance. D'autre part, la distribution des réponses manquantes est identifiée au paramètre κ près, qui est un paramètre de sensibilité. Donc, le modèle (2) correspond à une paramétrisation comme celle représentée dans (1). Cette famille de modèles de mélange suppose que les réponses manquantes et observées ont la même distribution, sauf pour un décalage dans la valeur espérée quantifié par κ. L'hypothèse MAR est équivalente à l'hypothèse que ces deux distributions sont égales, ce qui se traduit par $\kappa = 0$. Donc, cette famille de modèles est 'centrée' en MAR, et κ quantifie le départ par rapport à cette hypothèse. En permettant que κ dépend des covariables, c'est-à-dire $\kappa = \kappa(\mathbf{X}_{ij})$, l'analyste peut intégrer des hypothèses plus détaillées concernant les différences entre les trajectoires des réponses observées et

manquantes.

La performance de cette approche pour étudier la sensibilité des inférences a été évaluée dans une étude de simulations. Pour implémenter la méthode, une approche par imputation multiple a été développée. Dans cette approche, les réponses manquantes sont imputées en tirant des valeurs tirées directement du modèle (2), à l'aide des estimations des paramètres obtenus sous MAR et le paramètre $\kappa(\mathbf{X}_{ij})$ choisi. En prenant $\kappa = 0$, la procédure d'imputation développée offre une alternative d'analyse sous MAR par imputation multiple et a été validée dans ce contexte dans une autre étude de simulations.

L'approche proposée a été appliquée l'essai du trouble du maintien du sommeil. L'objectif de cet essai était de faire des inférences sur l'effet du traitement sur chaque score, définit comme la différence espérée entre les groupes de contrôle et de traitement dans le changement du score entre le début et la fin de l'étude. Or, 22% des individus étaient sortis de l'étude avant sa fin. Dans les essais cliniques, les sorties d'étude peuvent s'expliquer par des effets indésirables, un manque d'efficacité et des violations du protocole entre autres raisons (Molenberghs and Kenward, 2007). Nous avons utilisé la méthode proposée pour étudier la sensibilité des inférences obtenues sous MAR aux écarts par rapport à cette hypothèse. Les analyses ont confirmé un effet significatif du traitement sur les scores WASO et NAW, cette conclusion restant stable sous un large éventail de modèles d'une famille comme (2) et même lorsqu'on a considéré une autre définition de l'effet du traitement. Pour les autres scores, les résultats étaient très fragiles et sensibles aux hypothèses de manquement.

Par rapport à d'autres méthodes proposées dans la littérature (ex. Daniels and Hogan, 2000; Ratitch et al., 2013), l'approche proposée a comme avantage qu'elle peut être facilement appliquée à des données avec un grand nombre de mesures répétées, ayant lieu possiblement à des occasions différentes pour chaque individu. Celles-ci sont des avantages héritées du modèle linéaire mixte. De plus, les paramètres de sensibilité dans notre approche ont une interprétation intuitive car ils caractérisent les trajectoires des individus, ce qui facilite la formulation des hypothèses sur la distribution des

réponses manquantes. Les résultats de cette partie de la thèse ont fait l'objet d'une publication (Moreno-Betancur and Chavance, 2013).

Modélisation de risques concurrents avec des causes d'évènement manquantes

Contexte et état de la question

Le modèle de *risques concurrents* est une extension du modèle standard de l'analyse de survie. Dans ce dernier, on étudie le délai de survenue d'un évènement, noté par T. Dans le modèle de risques concurrents, on distingue le type d'évènement ou la cause l'ayant entrainé, les différents types d'évènement étant mutuellement exclusifs. Les réponses d'intérêt dans ce contexte sont donc T et D, où D dénote la cause d'évènement. Sans perte de généralité, on peut supposer que D a deux catégories, une représentant la cause d'intérêt ($D = 1$) et l'autre regroupant toutes les autres causes ($D = 2$). En recherche médicale, le modèle de risques concurrents est typiquement employé lorsqu'on s'intéresse aux délais de survenue d'évènements tels que le décès par une cause spécifique, la rechute ou la réponse à un traitement. La modélisation par régression permet d'analyser l'effet des facteurs de risque ou des interventions sur la survenue d'évènements concurrents.

En général, T peut être *censuré à droite*, c'est-à-dire, pour certains individus on sait qu'ils n'ont pas subi d'évènement jusqu'à un certain moment, après lequel aucune information concernant la survenue d'un évènement n'est disponible. C'est cette particularité des délais de survenue des évènements qui a fait que l'analyse de survie se développe comme un domaine séparé de la statistique. Notamment, il existe de nombreuses méthodes pour la modélisation des risques concurrents en présence de censure à droite.

Soit $\mathbf{X}$ un p-vecteur de covariables mesurées au début de l'étude. Pour étudier l'impact de $\mathbf{X}$ sur la survenue des évènements entrainés par la cause j ($j = 1, 2$), on peut postuler des modèles de régression pour deux fonctions cause-spécifiques. La

première fonction est le *risque cause-spécifique* (CSH), définit au temps t comme le taux auquel se produisent des évènements de type j parmi les individus à risque juste avant t: $\lambda_j(t) := \lim_{h\to 0} \frac{1}{h} P(t \leq T < t+h, D = j | T \geq t)$. Pour cette fonction, il est commun de considérer des modèles à risques proportionnels pour chaque cause, tel que le modèle de Cox (Cox, 1972), ou des modèles additifs tel que le modèle de Aalen (Aalen, 1980).

La deuxième fonction est appelée *la fonction d'incidence cumulée* (CIF), définie comme la probabilité d'observer un évènement de type j avant le temps t: $F_j(t) := P(T \leq t, D = j)$. Pour cette fonction, il est commun de postuler des modèles linéaires généralisés semi-paramétriques, comme le modèle de Fine et Gray (Fine and Gray, 1999) ou le modèle additif (Klein, 2006). Une modèle paramétrique a aussi été proposé (Jeong and Fine, 2007). L'incidence cumulée d'une cause donnée dépend des deux fonctions de risque, λ_1 et λ_2. Donc, une augmentation du risque d'une cause ne reflètera pas forcement une augmentation de l'incidence cumulée de cette cause (Beyersmann et al., 2007). C'est pourquoi, les deux fonctions doivent être étudiées pour comprendre le mécanisme de risques concurrents entièrement (Andersen et al., 2012; Latouche et al., 2013).

Les méthodes pour modéliser le CSH ou la CIF supposent en général que la cause d'évènement D est connue pour tous les individus ayant subi un évènement, ce qui n'est pas toujours le cas. Par exemple, si l'évènement étudié est le décès, quelques causes de décès peuvent manquer lorsque les certificats de décès ne sont pas bien remplis ou les patients meurent sans autopsie (Andersen et al., 1996; Manola and Gray, 2011). Certains auteurs ont proposé des méthodes de régression gérant les causes d'évènement manquantes sous l'hypothèse MAR, notamment pour la modélisation semi-paramétrique du CSH (Goetghebeur and Ryan, 1995; Andersen et al., 1996; Nicolaie et al., 2011; Lu and Tsiatis, 2001; Gao and Tsiatis, 2005; Gao, 2006; Lu and Liang, 2008). Cependant, d'autres modèles fréquemment employés n'ont été que peu ou pas du tout considérés. Par exemple, la modélisation semi-paramétrique de la CIF n'a été considéré que par Bakoyannis et al. (2010) qui a étudié le modèle de Fine et Gray avec une approche par imputation multiple. D'autre part, les modèles paramétriques pour le CSH et la CIF

n'ont jamais été considérés dans ce contexte, et il est de même pour la modélisation sous MNAR et les analyses de sensibilité.

Dans la littérature, un point de départ pour le développement de méthodes de régression gérant les causes d'évènement manquantes a été un essai clinique randomisé de l'Eastern Cooperative Oncology Group (ECOG). Dans cet essai, 169 femmes âgées avaient été recrutées pour étudier un traitement du cancer du sein de stade II. Après un suivi médiane de 6.7 années, 79 femmes étaient décédées, dont 44 décès par cancer et 17 décès par autres causes. Pour les autres 18 femmes décédées (23%), la cause de décès était manquante, notamment parce que le schéma de l'étude n'était pas adapté à un suivi à long terme (Cummings et al., 1993). Les premières analyses de cet essai, effectuées par Cummings et al. (1985) et Cummings et al. (1986), n'ont pas montré d'effet significatif du traitement sur la survie, mais ont révélé des associations significatives entre la survie et deux facteurs pronostiques: le statut des récepteurs d'estrogène (positive ou négative) et le nombre de ganglions lymphatiques atteints par le cancer (inférieur à 4, ou supérieur ou égale à 4). Étant donné l'importance des causes de décès concurrentes dans cette étude et du haut pourcentage de causes manquantes, ces données on été de grande valeur dans la littérature pour illustrer de nouvelles méthodes pour gérer les causes manquantes (Goetghebeur and Ryan, 1995; Lu and Tsiatis, 2001; Gao, 2006; Nicolaie et al., 2011). Ces auteurs ont notamment étudié les effets des deux facteurs pronostiques mentionnés sur le décès par cancer. Dans cette thèse, nous avons utilisé ces données pour illustrer les méthodes proposées.

Contributions

Une première contribution, répondant au second objectif de la thèse, a été de proposer un cadre général pour la modélisation semi-paramétrique de la CIF sous l'hypothèse MAR. On considère le modèle linéaire généralisé semi-paramétrique suivant:

$$g\{F_j(t|\mathbf{X})\} = \beta_{j0}(t) + \boldsymbol{\beta}_j'\mathbf{X}, \qquad j = 1, 2, \tag{3}$$

où g est une fonction de lien monotone et différentiable et $\beta_{j0}(t)$ est un intercepte dépendant du temps non-spécifié. Le modèle (3) comprend des modèles tels que le modèle de Fine et Gray lorsque g est la fonction cloglog (*complementary log-log*) et le modèle additif lorsque g est la fonction identité. Lorsque toutes les causes d'évènement sont observées, ce modèle peut être ajusté avec la méthode Andersen-Klein (Andersen et al., 2003; Klein and Andersen, 2005). Cette méthode consiste à utiliser des pseudo-valeurs de jackknife de la CIF comme les réponses dans des équations d'estimation généralisées (Liang and Zeger, 1986). Nous avons proposé deux extensions de cette méthode pour gérer les causes d'évènement manquantes. La première extension consiste à utiliser des pseudo-valeurs pondérés par l'inverse de la probabilité que la cause d'évènement soit observée. Nous avons démontré que cette approche amène à des estimateurs consistants et asymptotiquement normaux sous l'hypothèse MAR. De plus, cette approche permet de prendre en compte toute l'information partielle disponible sur les individus avec cause manquante dans les équations d'estimation, impliquant des estimateurs efficaces. La deuxième extension s'agit de l'application de la procédure d'imputation multiple de Bakoyannis et al. (2010) pour imputer les causes manquantes, suivi de l'analyse par la méthode Andersen-Klein des jeux de données complétés. Ces deux méthodes ont été validées et comparées à l'analyse CC dans une étude de simulations, et ensuite appliqués à l'essai ECOG. Nous avons ainsi pu formuler des recommandations sur leur application dans la pratique. Ces travaux ont fait l'objet d'une publication (Moreno-Betancur and Latouche, 2013).

Une deuxième contribution pour la modélisation des risques concurrents avec causes d'évènement manquantes a été d'étudier formellement le concept de l'*ignorabilité* du mécanisme de manquement dans ce contexte. Notamment, nous avons déterminé des conditions suffisantes sur les mécanismes de manquement, de censure et des risques concurrents pour que le mécanisme manquement puisse être ignoré dans une analyse par maximum de vraisemblance. Nous avons ainsi déduit des expressions pour la vraisemblance en fonction du CSH et de la CIF permettant d'ajuster des modèles paramétriques sous ces conditions. Les modèles paramétriques n'avaient pas été considérés dans la

littérature de causes manquantes auparavant. Pour le CSH, les expressions proposées permettent d'ajuster des modèles à risques proportionnels complètement paramétriques:

$$\lambda_j(t|\mathbf{X}) = \lambda_{j0}(\boldsymbol{\alpha}_j, t)\exp(\boldsymbol{\beta}_j'\mathbf{X}), \qquad j = 1, 2$$

où le risque de base λ_{j0} est connu au vecteur de paramètres $\boldsymbol{\alpha}_j$ près. Pour la CIF, le modèle de Jeong and Fine (2007) peut être ajusté:

$$F_j(t|\mathbf{X}) = 1 - \left[1 + \alpha_j \exp(\boldsymbol{\beta}_j'\mathbf{X})\tau_j\{\exp(\rho_j t) - 1\}/\rho_j\right]^{-1/\alpha_j}, \qquad j = 1, 2.$$

La troisième contribution pour cette partie de la thèse a été le développement d'une méthodologie pour effectuer des analyses de sensibilité dans le contexte de causes manquantes. Nous avons repris les idées développées dans le cadre de données longitudinales avec des sorties d'étude. Plus précisément, nous avons considéré une famille de modèles de mélange de profils indexés par un paramètre de sensibilité. Pour modéliser sous MNAR dans ce contexte, on doit considérer la distribution jointe de (T, D, M) pour les individus non-censurés, où M est l'indicateur de manquement, c'est-à-dire, $M = 1$ si la cause est manquante et $M = 0$ sinon. Dénotant par U l'indicateur de censure, avec $U = 1$ si l'individu est censuré et $U = 0$ sinon, la factorisation par mélange de profils s'écrit

$$\begin{aligned} f(t, d, m \mid U = 0) &= f(t, d \mid M = m, U = 0) \times f(m \mid U = 0) \\ &= f(t, m \mid U = 0) \times f(d \mid T = t, M = m, U = 0). \end{aligned} \tag{4}$$

Ici, le modèle d'extrapolation est représenté par le dernier facteur dans (4), $f(d \mid T = t, M = m, U = 0)$. Nous avons considéré des modèles de mélange de profils où le modèle d'extrapolation a la forme suivant:

$$\text{logit}\{\Pi(\mathbf{X}, T, M)\} = \mathbf{h}(\mathbf{X}, T)'\boldsymbol{\gamma} + \kappa M, \tag{5}$$

où $\Pi(\mathbf{X}, T, M) := P(D = 1|\mathbf{X}, T, M, U = 0)$ et $\mathbf{h}(\mathbf{X}, T)$ est un vecteur incluant $\mathbf{X}$, T, et possiblement des interactions et des polynômes de T de plus grand ordre. La première partie du prédicteur linéaire, $\mathbf{h}(\mathbf{X}, T)'\boldsymbol{\gamma}$, détermine complètement la distribution des causes d'évènement pour les individus avec cause observée, pour qui $M_i = 0$. Le paramètre $\boldsymbol{\gamma}$ peut être estimé en ajustant le modèle logistique aux données de ces individus. D'autre part, la distribution des causes d'évènement pour les individus avec cause manquante est identifiée au paramètre κ près, qui est un paramètre de sensibilité. En effet, ce paramètre n'est pas identifiable à partir des données observées.

Le paramètre κ quantifie une différence entre les distributions des causes d'évènements des individus avec cause observée et manquante, ainsi que le départ par rapport à l'hypothèse MAR, celle-ci étant équivalent à la condition $\kappa = 0$. Plus précisément, κ est le logarithme du odds ratio ajusté, comparant les chances d'un évènement par la cause d'intérêt ($D = 1$) entre les individus non-censurés avec cause manquante et ceux avec cause observée. En permettant que κ varie sur un ensemble de valeurs plausibles, on peut évaluer la sensibilité des inférences aux écarts par rapport à MAR. Pour des hypothèses plus détaillées on peut permettre que κ dépend des covariables et du délai de survie, c'est-à-dire $\kappa = \kappa(\mathbf{X}, T)$. De même que pour le données longitudinales, nous proposons une implémentation de cette procédure par imputation multiple. D'abord, les cause manquantes sont imputées en tirant des valeurs du modèle (5) en utilisant l'estimation de $\boldsymbol{\gamma}$ obtenue sous MAR et le $\kappa(\mathbf{X}, T)$ choisi. Ensuite, la méthode de régression souhaitée est utilisée pour analyser chaque jeu de données complété. Ainsi, cette méthode d'analyse de sensibilité est applicable à divers modèles pour risques concurrents, y compris des méthodes paramétriques ou semi-paramétriques pour le CSH et la CIF.

Les premiers résultats d'application de cette méthode à l'essai ECOG ont été encourageants. A la base, cette méthode avait été développée pour une étude épidémiologique sur les différentiels socioéconomiques dans la mortalité par suicide en France. On soupçonnait que la base de causes de décès de cette étude contenait de nombreux suicides codés comme décès à cause manquante suite à un problème de transmission

des informations sur les suicides dans la région de l'étude. Ceci remettait en cause l'hypothèse MAR. Or, avec notre méthode nous avons trouvé que les résultats étaient peu sensibles aux hypothèses de manquement, sans doute grâce au faible pourcentage de causes manquantes dans cette étude (environ 10%).

Le projet sur les différentiels socioéconomiques dans la mortalité par suicide en France nous a amené à étudier en détail les indices utilisés dans la littérature pour mesurer ces différentiels à partir de données de survie et de risques concurrents. Un manuscrit sur cette problématique, externe à celle des données manquantes, est en préparation.

Introduction

Missing data are a common occurrence in medical studies, especially in those involving human beings. For example, when questionnaires are involved, often some individuals will leave some items of the questionnaire unanswered. In medical studies, individuals are often followed over the course of time, e.g. to study the evolution of some variable measured at several points in time or the time to specific events. The prolonged observation of individuals exacerbates the occurrence of missing data because collection of information tends to deteriorate with time. For instance, some people tend to miss scheduled appointments and some drop-out of studies altogether, e.g. for migratory or medical reasons. Such problems are common in both clinical and epidemiological studies, and cannot be entirely prevented even in carefully planned clinical trials with strict protocols.

In medical research, regression modeling is often used to study the effects of covariates (e.g. prognostic factors, exposures) on medical outcomes of interest (e.g. the level of some biological marker, time to death). For such purposes, each medical study recruits individuals according to a predefined set of criteria, and determines in advance the outcomes that will be measured during the study for each individual and the timing of these measurements. The covariate effects of medical interest are those describing the distribution of the entire set of planned outcomes. However, when some outcomes are missing for some of the individuals, the observed outcomes together with the covariates constitute the only available information from which to draw inferences (i.e. point estimates, confidence intervals, hypothesis tests) about these parameters. Apart

from the obvious loss in precision due to the missing information, the validity of any method used to draw inferences about these parameters from the observed data will require that some assumption about the nature of the mechanism leading to missing outcomes, called the *missingness mechanism*, holds.

Rubin (1976) proposed a taxonomy of missingness mechanisms that is key to understanding the statistical issues that arise with missing data. In the context of missing outcomes, the missingness mechanism is said to be MCAR (for *missing completely at random*) if the probability that an outcome is missing is constant. The missingness mechanism is said to be MAR (for *missing at random*) if the probability that an outcome is missing depends on observed data (e.g. covariates, outcomes observed at other time-points) but not on missing outcomes when conditioning on the former. Finally, the missingness mechanism is said to be MNAR (for *missing not at random*) if the probability of missingness depends on missing outcomes even when conditioning on the observed data.

Rubin's taxonomy has crucial implications regarding the modeling requirements to draw valid inferences from the available data. Under an MCAR mechanism, the observed outcomes can be considered to be a random sample of the entire set of planned outcomes. Thus, in many situations, an MCAR mechanism guarantees that valid, albeit inefficient inferences can be obtained with the so-called *complete case* (CC) analysis, which consists in excluding individuals with missing outcomes from the study. Under MAR, missing and observed outcomes have the same conditional distribution given the remaining observed data. This implies that valid inferences can be drawn from the available data without the need to make further assumptions about the distribution of the missing outcomes. Under MNAR, the conditional distributions of the observed and missing outcomes differ, which requires making assumptions about the latter either explicitly or implicitly to address the evident identifiability issues. Thus, regression modeling under MNAR is less straightforward.

Rubin's taxonomy marks an important distinction between different types of methods for handling missing outcomes. However, as Molenberghs et al. (2008) showed in

a quite general setting, it is not possible to assess from the observed data whether the missingness mechanism is MAR or MNAR. In fact, except in special cases of data missing by design, the missingness assumptions underlying any modeling strategy with missing data are unverifiable. Thus, there has been an increasing awareness of the need to perform *sensitivity analyses* to assess the robustness of inferences obtained in a primary analysis to departures from the underlying missingness assumptions. Often, the primary analysis assumes MAR, in which case one possible approach to sensitivity analyses is to assess the discrepancies between the inferences obtained under MAR and those yielded by a family of MNAR models.

The CC analysis and other *ad-hoc* methods used to be common practice, but a growing understanding of the strong and unrealistic assumptions underlying these approaches has stimulated the development, software implementation and use of methods that rely on more relaxed assumptions. This is particularly true in the field of incomplete multivariate data, in which the outcomes are gathered in a vector $(Y_1, \ldots, Y_J)$ for which some components may be missing. Indeed, in this setting, MAR methods are widely available and increasingly used, and several MNAR modeling strategies have also been proposed. On the other hand, although some sensitivity analysis methodology has been developed, this is still an active area of research. Actually, the first aim of this dissertation was to develop a sensitivity analysis approach for continuous longitudinal data with drop-outs. The latter are a special case of incomplete multivariate data in which the outcomes are continuous and ordered in time. Moreover, for each individual, either all the outcomes are observed or all the outcomes are observed up to a certain time-point, at which the individual drops-out and never returns to the study so that all the subsequent outcomes are missing.

The second aim of this thesis was to build on the existing theory for incomplete multivariate data to develop methods for competing risks data with missing causes of failure. The competing risks model is an extension of the standard survival analysis model. In the latter, the time T to the occurrence of one event, often termed a *failure*, is studied. In the competing risks model, there is a distinction between different types

of events or *causes of failure*, such that failure from one cause precludes failure from other causes. Letting D denote the cause of failure, the outcomes of interest in this context are T and D. Usually, T is subject to *right-censoring*, which means that some individuals in the study are observed to be event-free up to a certain time, after which no other information on failure occurrence is available. This particularity of time-to-event data has prompted survival and, in general, event-history analysis to develop as a separate field of statistics, and several specialized methods for regression modeling of competing risks data with right-censoring exist.

Competing risks regression methods generally require that the cause of failure D is observed for all individuals known to have failed, a prerequisite that is not always met in practice. We focused on the setting where some causes of failure are missing. In this case, the outcome vector (T, D) consists of one continuous component subject to right-censoring and a categorical component subject to missingness. A general concept that encompasses both right-censoring and missingness is that of *coarsening* (see Tsiatis, 2006). However, a unified approach to these two problems in this context is not pertinent because, as stated before, there is an entire discipline dedicated to modeling right-censored data, in particular in the competing risks context. Rather, we aimed at accommodating existing regression methods for right-censored competing risks data in order to deal with the missingness part of the problem. Some methods for regression with missing causes of failure under MAR have already been proposed for some of the most common models in competing risks, but other useful models have received little or no attention. Furthermore, to our knowledge, MNAR modeling and sensitivity analysis approaches have never been considered in this setting. Thus, our aim was to address some of these voids in the current missing cause of failure literature.

The manuscript is organized as follows. In Part I, we provide some preliminary background concerning regression modeling of longitudinal data with drop-outs (Chapter 1) and competing risks data with missing causes of failure (Chapter 2). In these chapters, we present the two randomized clinical trials that motivated the methodological developments of this dissertation and that will be used for illustration purposes. Further, we

introduce the notation that will be used throughout the manuscript, discuss regression modeling strategies when there are no missing data and provide an overview of the existing methods to deal with the missing data in each context. In Part II, we present the methodology developed for regression modeling under MAR of longitudinal data with drop-outs (Chapter 3) and competing risks with missing causes of failure (Chapters 4 and 5). In Part III, we present sensitivity analysis methods for longitudinal data with drop-outs (Chapter 6) and competing risks with missing causes (Chapter 7), both of which involve the assessment of the inferences yielded by a family of MNAR models. Hence, in this dissertation, MNAR modeling is considered only in the context of sensitivity analyses. Finally, we present a general discussion of the methods developed and some perspectives on ongoing and future research. In Appendices A and B we provide some supplementary details for the simulation studies of Chapters 3 and 4, respectively. In Appendix C, we provide the current version of a manuscript in preparation that is part of an ongoing project outside the missing data topic.

Part I

Preliminaries

Chapter 1

Background on longitudinal data with drop-outs

1.1 Motivation: The SMI study

Longitudinal data arise when a response or, equivalently, outcome variable of interest is measured for each individual in a study at several time-points, resulting in a sequence of repeated measurements that are ordered in time. The main goal of a longitudinal study is to enable precise assessment of within-individual changes in the outcome variable over time and of the factors that influence those changes. Such within-individual temporal changes cannot in general be studied from cross-sectional data, in which the response variable has been measured at a single time-point for each individual. Hence, longitudinal studies are the key for addressing many questions that arise in medical research.

One example is the case study that motivated this part of the dissertation. It was a randomized clinical trial investigating the benefits of a treatment for sleep-maintenance insomnia (SMI). Patients suffering from this disorder do not have trouble falling asleep when they go to bed, but then wake up during the night and experience difficulties falling back asleep. The trial enrolled 962 patients, randomized to either treatment or

placebo at the beginning of the study, who were asked to complete the Sleep Morning Questionnaire daily during the duration of the trial. This questionnaire includes both quantitative and qualitative interrogations about the quality of sleep during the night. The six main scores of interest in this trial were: the wake-time after sleep onset (WASO); the number of nocturnal awakenings (NAW); the sleep refreshing quality (SLREF), with values ranging from one (excellent) to four (poor); the feeling of sleepiness (FEELC), with values ranging from zero (very sleepy) to nine (not sleepy at all); the total sleep time (TST); and the sleep onset latency (SOL), which is the time it takes to transit from the state of full wakefulness to sleep. Note that a decrease in the WASO, NAW, SLREF or SOL scores or an increase in the FEELC or TST scores would indicate an improvement in the patient's quality of sleep. Thus, from this longitudinal design it is possible to study and compare the temporal evolution of sleep quality in treated and untreated patients.

In the data available from this study, the six scores were recorded as means of the daily measurements for up to six periods, a so-called *baseline* period before the beginning of the study, plus five periods of different lengths after randomization (visits 1 to 5). The length of the baseline period could vary but never exceeded two weeks. Each of the first three periods lasted two weeks while the fourth and fifth periods each lasted three weeks. The actual response variables analyzed were these means, so we were only concerned with continuous outcomes in this study. The actual number of daily measurements contributing to each period mean varied among subjects, visits and scores (see Table 1.1). These counts served as precision weights in all of the analyses of these data to account for the differences in precision among these means.

The main goal of this study was to perform inference about the *treatment effect* on each score, which was defined as the difference in the expected change from baseline at visit 5 of the scores in the treatment and control groups. The weighted mean and standard deviation of each score at baseline and visit 5 calculated for each group from the available data are given in Table 1.2. In principle, estimation of the treatment effect in the SMI study could be based on a regression model. A defining feature of

Table 1.1: Mean number of daily measurements available for each individual per visit per score for each group (standard deviation in brackets).

	Baseline	Visit 1	Visit 2	Visit 3	Visit 4	Visit 5
Control group						
WASO	6.8 (1.2)	12.4 (2.4)	12.0 (2.7)	11.9 (2.9)	17.2 (4.8)	16.7 (4.8)
NAW	6.8 (1.2)	12.5 (2.3)	12.1 (2.7)	12.0 (2.8)	17.3 (4.7)	16.8 (4.7)
SLREF	6.8 (1.2)	12.6 (2.3)	12.3 (2.5)	12.2 (2.7)	17.6 (4.5)	17.1 (4.6)
FEELC	6.8 (1.2)	12.6 (2.3)	12.3 (2.5)	12.2 (2.7)	17.6 (4.5)	17.1 (4.6)
TST	6.8 (1.2)	12.5 (2.3)	12.1 (2.6)	12.0 (2.9)	17.3 (4.9)	16.9 (4.9)
SOL	6.8 (1.1)	12.5 (2.3)	12.1 (2.7)	12.0 (2.9)	17.4 (4.8)	16.9 (4.7)
Treatment group						
WASO	6.8 (1.3)	12.2 (2.7)	11.9 (3.0)	11.8 (2.9)	16.9 (5.0)	16.7 (4.7)
NAW	6.8 (1.3)	12.2 (2.6)	11.9 (2.8)	11.8 (2.9)	17.0 (4.8)	16.7 (4.7)
SLREF	6.8 (1.3)	12.4 (2.6)	12.2 (2.7)	12.0 (2.7)	17.5 (4.5)	17.1 (4.3)
FEELC	6.8 (1.3)	12.4 (2.6)	12.2 (2.7)	12.0 (2.7)	17.5 (4.5)	17.1 (4.3)
TST	6.8 (1.3)	12.3 (2.6)	12.0 (2.9)	11.8 (2.9)	17.1 (4.9)	16.7 (4.7)
SOL	6.8 (1.3)	12.3 (2.6)	12.1 (2.9)	11.9 (2.7)	17.2 (4.8)	16.9 (4.6)

longitudinal data is that the responses collected from one individual are usually not independent from each other. In fact, longitudinal data are a special case of clustered data, in which the measurements within each cluster (i.e. each individual) have a temporal ordering. Thus, the assumption of independence that underlies standard regression methods is violated, a fact that has led to the development of methods that account for the dependence structure. Two important examples are *mixed effects models* (Laird and Ware, 1982) and *marginal models* (Liang and Zeger, 1986). In Section 1.2, we provide a brief overview of these approaches in the context of continuous response variables which are the main focus in this manuscript.

When all the planned outcomes are available for each individual in the study, these methods for longitudinal data provide valid inferences about regression coefficients, i.e. point estimates, confidence intervals (CIs) and hypothesis tests. However, a very common occurrence in longitudinal studies in medical research, especially when the subjects under study are human beings, is that some outcomes are missing for some individuals. Some sources of missing outcomes are individuals leaving items in a questionnaire

Table 1.2: Weighted mean (weighted standard deviation) of each score at baseline and visit 5 calculated from the available data.

	Baseline	Visit 5
Control group		
WASO	108.3 (44.9)	61.7 (46.5)
NAW	2.8 (1.3)	1.9 (1.2)
SLREF	3.2 (0.5)	2.5 (0.6)
FEELC	4.5 (1.7)	5.7 (1.8)
TST	346.6 (48.5)	386.4 (56.3)
SOL	18.7 (8.9)	18.6 (11.7)
Treatment group		
WASO	101.9 (38.0)	47.4 (40.0)
NAW	2.8 (1.5)	1.6 (1.3)
SLREF	3.2 (0.5)	2.5 (0.6)
FEELC	4.4 (1.7)	5.9 (1.7)
TST	343.1 (45.7)	396.5 (57.1)
SOL	19.0 (8.2)	17.6 (11.2)

unanswered, not showing-up for a scheduled measurement, or dropping-out of the study altogether. Actually, the problem of drop-out was prominent in the SMI study and will be the main focus of this part of the dissertation. More precisely, we will focus on the setting where, for each individual, the outcomes are either completely observed, or completely observed up to a certain time-point, at which the individual drops-out and never returns to the study so that all the subsequent outcomes are missing. In clinical trials, common reasons for drop-out are side-effects, lack of efficacy and protocol violation (Molenberghs and Kenward, 2007). In the SMI study, around 22% of the individuals had dropped-out before the study end. The percentages of missing outcomes for each period, group and score are shown in Table 1.3. They were consistently smaller for the treatment group than the control group, but the difference was always small and not significant.

With drop-outs, the validity of inferences obtained with the aforementioned regression methods for longitudinal data will depend on additional assumptions about the drop-out mechanism holding. Alternative approaches, relying on different assump-

Table 1.3: Percentage of missing outcomes per visit per score for each group.

	Baseline	Visit 1	Visit 2	Visit 3	Visit 4	Visit 5
Control group						
WASO	0	0.6	7.2	14.5	19.4	24.1
NAW	0	0.6	7.2	14.5	19.1	24.1
SLREF	0	0.6	7.2	14.5	19.1	23.8
FEELC	0	0.6	7.2	14.5	19.1	23.8
TST	0	0.6	8.1	15.4	19.7	24.6
SOL	0	0.6	7.5	14.8	19.7	24.6
Treatment group						
WASO	0	0.3	6.2	13.0	16.7	21.4
NAW	0	0.3	6.2	12.5	16.5	20.7
SLREF	0	0.3	6.0	12.0	16.5	20.4
FEELC	0	0.3	6.0	12.0	16.5	20.4
TST	0	0.3	6.2	12.2	17.0	21.2
SOL	0	0.3	6.0	12.6	16.7	20.9

tions or providing advantages in terms of ease of implementation, have been developed specifically for modeling longitudinal data with drop-outs. A key to understanding the assumptions underlying different approaches is the classification of missingness mechanisms proposed by Rubin (1976) (cf. Introduction). In Section 1.3, we provide the formal definitions of *missing at random* (MAR) and *missing not at random* (MNAR) drop-out mechanisms. A very thorough review of the methods available for modeling longitudinal data with drop-outs is given by Molenberghs and Kenward (2007), and summarized in Section 1.3 in the context of continuous responses.

Fortunately, principled MAR-based regression methods for longitudinal data with drop-outs have become readily available in common statistical software, and are increasingly being used by practitioners instead of other more questionable ad-hoc approaches that were frequently used before. Several MNAR approaches have also been proposed. However, it is never possible to assess from the observed data whether the missingness mechanism is MAR or MNAR (Molenberghs et al., 2008). Thus, it is essential to assess the potential impact on inferences of departures from the assumptions underlying any analysis by means of a sensitivity analysis. Although some sensitivity analysis method-

ology has been developed, no standard method exists nor should be prescribed as this is still an active area of research (Carroll et al., 2004).

The aim of the "longitudinal data" part of this dissertation was thus to propose a flexible approach for performing sensitivity analyses when dealing with continuous longitudinal data with drop-outs. The family of MNAR models on which the proposed approach relies contains an MAR model as a special case. Hence, the implementation procedure developed provides in particular an alternative method to analyze longitudinal data with drop-outs under the MAR assumption. In the SMI study, our approach to perform sensitivity analyses provided insight about the robustness of the inferences drawn under MAR for the WASO score, which was the primary endpoint, and for the other five scores which defined secondary endpoints. The conclusions regarding the effect of treatment on some of the scores were shown to be reliable, even when considering an alternative definition of the treatment effect based on the expected rate of change of the score. Meanwhile, for other scores the conclusions were found to be fragile and strongly dependent on missingness assumptions.

1.2 Regression for continuous longitudinal outcomes

1.2.1 Notation and general considerations

Although it is not necessary for the methods proposed in Parts II and III of this dissertation, for simplicity of notation and exposition, and because the data studied in this manuscript fulfill these conditions, we consider a longitudinal study in which a fixed number of measurements, say J, of a continuous response variable are intended to be collected from each of the n individuals in the study at a set of J common time-points, $t_1, \ldots, t_J$. However, the measurement occasions, or *visits*, are not assumed to be evenly distributed throughout the duration of the study. The first visit is often referred to as *baseline*.

Let $\mathbf{Y}_i = (Y_{i1}, \ldots, Y_{iJ})$ denote the vector of patient i's planned responses, for $i =$

$1, \ldots, n$. We let $\boldsymbol{\mu}_i = E(\mathbf{Y}_i) := (\mu_{i1}, \ldots, \mu_{iJ})'$ and $\mathbf{V}_i = \text{var}(\mathbf{Y}_i) := \{\text{cov}(Y_{ij'}, Y_{ij''})\}_{j'j''}$ denote the marginal expectation and variance-covariance matrix of $\mathbf{Y}_i$, respectively, where $\mu_{ij} = E(Y_{ij})$.

We focus on the setting where the main interest is in performing inferences about the effects of covariates on the marginal expectations μ_{ij} of the responses. Here, we allow for fully-observed time-fixed covariates measured at baseline, and we will often want to include time-trends in our models. Thus, for each individual i and visit j, we let $\mathbf{X}_{ij}$ denote a p-vector of baseline covariates and polynomial functions of t_j. We denote by $\mathbf{X}_i$ the $J \times p$ matrix whose rows are $\mathbf{X}'_{ij}$ $(j = 1, \ldots, J)$.

The response vectors of different individuals are assumed to be independent, and identically distributed (i.i.d) given the covariates. On the other hand, as mentioned earlier, an important characteristic of longitudinal data is that the set of responses of one individual are usually not independent. With continuous responses, a positive correlation may be expected: individuals with high past responses are more likely to have high future responses, and individuals with low past responses are more likely to have low future responses. An in-depth discussion of the possible sources of such correlations is given by Fitzmaurice et al. (2004, Section 2.5). Thus, the usual independence assumption of standard regression techniques is violated. This aspect of longitudinal data should not be seen as a weakness but as a strength because this correlation is what enables precise estimation of parameters describing within-individual temporal changes from longitudinal studies. Actually, if the correlation structure of longitudinal data is ignored when performing regression, inefficient regression coefficient estimates and biased precision estimates are obtained, the latter resulting in misleading inferences (see example in Section 1.5 of Diggle et al., 2002). Thus, the correlation structure needs to be accounted for when performing regression. This is achieved by allowing non-null values for the off-diagonal elements of $\mathbf{V}_i$, that is, for $\text{cov}(Y_{ij'}, Y_{ij''})$ with $j' \neq j''$.

In the following sections, we present brief overviews of two of the major approaches for regression modeling with continuous responses that account for the inherent correlation of longitudinal data: *linear mixed models* and *marginal models*. Detailed accounts

of these methods are provided by Diggle et al. (2002), Fitzmaurice et al. (2004) and Verbeke and Molenberghs (2000). A third approach that is not discussed here is that of *transition models*, in which the conditional expectation of each response given past responses is modeled (see for example Chapter 10 of Diggle et al., 2002).

1.2.2 Linear mixed models

The linear mixed model (LMM) (Laird and Ware, 1982) is an extension of the normal linear model in which a vector of subject-specific *random effects*, $\mathbf{b}_i$, is introduced to represent the heterogeneity in the regression coefficients across individuals due to unobserved factors. Here, we consider the LMM in which the responses of each individual are assumed to be independent given the fixed and random effects, called the *conditional independence model* by Verbeke and Molenberghs (2000). More precisely, the model makes the following distributional assumptions:

$$Y_{ij} = \mathbf{X}'_{ij}\boldsymbol{\beta} + \mathbf{Z}'_{ij}\mathbf{b}_i + \varepsilon_{ij}, \qquad \varepsilon_{ij} \sim N(0, \sigma^2), \qquad \mathbf{b}_i \sim N(\mathbf{0}, \mathbf{G}), \tag{1.1}$$

for $j = 1, \ldots, J$ and $i = 1, \ldots, n$, where $\boldsymbol{\beta}$ is the fixed effects p-vector common to the population; the q-vectors $\mathbf{Z}'_{ij}$ $(j = 1, \ldots, J)$ are the rows of the $J \times q$ random effects design matrix $\mathbf{Z}_i$, whose columns are a subset of the columns of $\mathbf{X}_i$ (hence $q \leq p$); the zero-mean Gaussian residual errors ε_{ij} are i.i.d. with variance σ^2; and the zero-mean Gaussian subject-specific random effects q-vectors $\mathbf{b}_i$ are i.i.d. with variance $\mathbf{G} = \mathbf{G}(\boldsymbol{\alpha})$, where $\boldsymbol{\alpha}$ is a vector of unknown parameters. Furthermore, the random effects vectors are assumed to be independent of the residual errors and of the covariates.

The introduction of random effects in the normal linear model induces a correlation structure among the responses of an individual, which arises from them sharing these subject-specific effects. Actually, the LMM implies the following marginal distribution for the response vectors:

$$\mathbf{Y}_i \sim N(\mathbf{X}_i\boldsymbol{\beta}, \mathbf{Z}_i\mathbf{G}\mathbf{Z}'_i + \sigma^2\mathbf{I}), \tag{1.2}$$

where $\mathbf{I}$ is the $J \times J$ identity matrix. The implied form for $\mathbf{V}_i$ is thus not diagonal in general, reflecting a correlation among the responses of each individual.

Expression (1.2) also shows that coefficients in the LMM have a marginal, or *population-average*, interpretation because the marginal expectations of the responses satisfy:

$$\mu_{ij} = \mathbf{X}'_{ij}\boldsymbol{\beta}. \tag{1.3}$$

This is a consequence of the linear structure of (1.1), which implies that averaging covariate effects across individuals results in the effect of the covariate on the population average. The latter is no longer true when the linear structure does not hold.

LMMs may be fitted by maximum likelihood or the so-called restricted maximum likelihood. The likelihood function is obtained by integrating the conditional density of the vector of responses given the random effects over the distribution of the random effects. Hypothesis tests and CI building for the fixed effects $\boldsymbol{\beta}$ are commonly based on t or F distributions whose degrees of freedom generally have to be estimated from the data. Several estimation methods exist, but Verbeke and Molenberghs (2000, Chapter 6), who give an overview of this and other aspects of inference with LMMs, argue that in the longitudinal data setting with large samples all of these methods will lead to very similar p-values.

1.2.3 Marginal models

Marginal models are direct regression models for the marginal expectations μ_{ij} of the responses. Thus, in these models the covariate effects automatically have a population-average interpretation. When dealing with continuous longitudinal outcomes, it is standard to assume a linear relation between these expectations and the covariates, like in (1.3). Here, we consider a more general structure that will be useful when we use marginal models in Chapter 4 in the context of competing risks regression. More pre-

cisely, we assume the following mean structure:

$$g(\mu_{ij}) = \mathbf{X}'_{ij}\boldsymbol{\beta}, \tag{1.4}$$

where g is a differentiable monotone link function.

In contrast with LMMs, in marginal models the variance-covariance matrix of $\mathbf{Y}_i$, $\mathbf{V}_i$, is modeled separately. The marginal variances of the responses (i.e. the diagonal terms) are assumed to have the form $\text{var}(Y_{ij}) = \phi\upsilon(\mu_{ij})$, where υ is a known function and ϕ is a scale parameter that could further be allowed to depend on the measurement occasion, i.e. $\phi = \phi_j$ (Fitzmaurice et al., 2004, Chapter 11). The off-diagonal terms, which embody the dependence between the different responses of an individual, are determined by the choice of the so-called *working correlation matrix*, denoted by $\mathbf{R}_i(\boldsymbol{\alpha})$, where $\boldsymbol{\alpha}$ is a vector of unknown parameters. Some possibilities for choosing $\mathbf{R}_i(\boldsymbol{\alpha})$ are: to assume independence (i.e. the identity matrix); to assume that all correlations are equal (the so called *exchangeable* matrix); or to assume that the magnitude of the correlation between two measurements is smaller the further they are apart in time (e.g. the *first order autoregressive* matrix). The covariance matrix of $\mathbf{Y}_i$ can be recovered as $\mathbf{V}_i = \mathbf{A}_i^{1/2}\mathbf{R}_i(\boldsymbol{\alpha})\mathbf{A}_i^{1/2}$ where $\mathbf{A}_i^{1/2}$ is the diagonal matrix with elements $\sqrt{\text{var}(Y_{ij})}$, $j = 1, \ldots, J$, on the diagonal. Thus, $\mathbf{V}_i$ depends on $\boldsymbol{\beta}$, ϕ and $\boldsymbol{\alpha}$.

Under the assumption that $\mathbf{Y}_i$ is multivariate normal, and when g is the identity function and $\upsilon(\mu_{ij}) = 1$, this model corresponds to a common extension of the normal linear model to the longitudinal setting, for which maximum likelihood and restricted maximum likelihood approaches are available (Diggle et al., 2002). But since we are interested in the broader class of models described above, and particularly in the use of a non-standard link function, we require another fitting strategy. A very useful approach that is widely available in current software, and that is also applicable with other types of outcomes (e.g. binary), is that of generalized estimating equations (GEE) (Liang and Zeger, 1986).

GEE is particularly useful for situations where the main interest is in estimation

of the regression parameters in (1.4), as is our case. The principle of GEE is that estimation of $\boldsymbol{\beta}$ may be performed by treating $\mathbf{V}_i$ as a nuisance. Estimation is based on the following estimating equation:

$$U(\boldsymbol{\beta}) = \sum_{i=1}^{n} \left\{ \frac{\partial}{\partial \boldsymbol{\beta}} g^{-1}(\mathbf{X}_i\boldsymbol{\beta}) \right\}' \mathbf{V}_i^{-1} \{\mathbf{Y}_i - g^{-1}(\mathbf{X}_i\boldsymbol{\beta})\} = \sum_{i=1}^{n} U_i(\boldsymbol{\beta}) = 0, \tag{1.5}$$

where $g^{-1}(\mathbf{X}_i\boldsymbol{\beta})$ is short for the J-vector with elements $g^{-1}(\mathbf{X}'_{ij}\boldsymbol{\beta})$ $(j = 1, \ldots, J)$. To bypass the fact that this equation depends on the nuisance parameters ϕ and $\boldsymbol{\alpha}$ through $\mathbf{V}_i$, consistent estimates $\hat{\phi}$ and $\hat{\boldsymbol{\alpha}}$, possibly depending on $\boldsymbol{\beta}$, are plugged in the equation. For example, Liang and Zeger (1986) use product-moment estimates of the nuisance parameters. An iterative algorithm is then performed to obtain final estimates of $\boldsymbol{\beta}$, ϕ and $\boldsymbol{\alpha}$.

Liang and Zeger (1986) showed that, under the assumption that the mean model (1.4) is correctly specified, the estimator $\hat{\boldsymbol{\beta}}$ obtained from the procedure described is consistent even if the covariance structure of $\mathbf{Y}_i$ is misspecified, and asymptotically as efficient as if the true values of the nuisance parameters were known. Furthermore, in many cases it achieves an efficiency close to that of a maximum likelihood estimator based on further distributional assumptions. Finally, $\hat{\boldsymbol{\beta}}$ is asymptotically normal with mean $\boldsymbol{\beta}$ and a variance that can be consistently estimated using the following *sandwich estimator*:

$$\hat{\text{var}}(\hat{\boldsymbol{\beta}}) = I(\hat{\boldsymbol{\beta}})^{-1} \hat{\text{var}}\{U(\boldsymbol{\beta})\} I(\hat{\boldsymbol{\beta}})^{-1}, \tag{1.6}$$

where $I(\boldsymbol{\beta}) = \sum_{i=1}^{n} \{\frac{\partial}{\partial \boldsymbol{\beta}} g^{-1}(\mathbf{X}_i\boldsymbol{\beta})\}' \mathbf{V}_i^{-1} \{\frac{\partial}{\partial \boldsymbol{\beta}} g^{-1}(\mathbf{X}_i\boldsymbol{\beta})\}$, $\hat{\text{var}}\{U(\boldsymbol{\beta})\} = \sum_{i=1}^{n} U_i(\hat{\boldsymbol{\beta}}) U_i(\hat{\boldsymbol{\beta}})'$, and $\mathbf{V}_i$ is estimated by plugging in the final estimates of $\boldsymbol{\beta}$, ϕ and $\boldsymbol{\alpha}$. The sandwich estimator is robust to misspecification of $\mathbf{V}_i$.

1.3 Modeling longitudinal data with drop-outs

Suppose that some individuals drop-out from the study. Then the response vector of individual i ($i = 1, \ldots, n$) is partitioned such that $\mathbf{Y}_i = (\mathbf{Y}_i^{\mathcal{O}}, \mathbf{Y}_i^{\mathcal{M}})$, where $\mathbf{Y}_i^{\mathcal{O}} = (Y_{i1}, \ldots, Y_{i(U_i-1)})$ and $\mathbf{Y}_i^{\mathcal{M}} = (Y_{iU_i}, \ldots, Y_{iJ})$ are the observed and missing parts of $\mathbf{Y}_i$, respectively, and U_i denotes the occasion of the first missing outcome, with $U_i \leq J$ for individuals who dropped-out and the convention that $U_i = J + 1$ for those who completed the study. Henceforth U_i will be referred to as the *drop-out indicator*. Denote by $\mathbf{W}_i$ the fully-observed design matrix of all the covariates that influence the drop-out mechanism, some of which may already be in $\mathbf{X}_i$. Covariates that are in $\mathbf{W}_i$ but not in $\mathbf{X}_i$ are sometimes referred to as *auxiliary covariates*.

Consider the conditional probability mass function of the drop-out indicator given the outcomes and the covariates influencing drop-out, which we denote by

$$f(u|\mathbf{y}^{\mathcal{O}}, \mathbf{y}^{\mathcal{M}}, \mathbf{w}). \tag{1.7}$$

The missing data taxonomy of Rubin (1976) and Little and Rubin (1987) in the setting of longitudinal data with drop-outs can be expressed as follows (Little, 1995):

(i) The drop-out mechanism is said to be MCAR (for *missing completely at random*) if (1.7) does not depend on observed or unobserved responses nor on covariates, that is, $f(u|\mathbf{y}^{\mathcal{O}}, \mathbf{y}^{\mathcal{M}}, \mathbf{w}) = f(u)$.

(ii) The drop-out mechanism is said to be *covariate-dependent* if (1.7) does not depend on observed or unobserved responses, that is, $f(u|\mathbf{y}^{\mathcal{O}}, \mathbf{y}^{\mathcal{M}}, \mathbf{w}) = f(u|\mathbf{w})$.

(iii) The drop-out mechanism is said to be MAR (for *missing at random*) if (1.7) does not depend on $\mathbf{y}^{\mathcal{M}}$, that is, $f(u|\mathbf{y}^{\mathcal{O}}, \mathbf{y}^{\mathcal{M}}, W) = f(u|\mathbf{y}^{\mathcal{O}}, \mathbf{w})$.

(iv) The drop-out mechanism is said to be MNAR (for *missing not at random*) otherwise.

With drop-outs, additional assumptions about the type of drop-out mechanism at play are necessary to identify the distribution of the response vector $\mathbf{Y}$ conditional on covariates. In the following sections, we discuss some existing approaches for regression modeling of longitudinal data with drop-outs and the assumptions underlying these methods.

1.3.1 Ad-hoc approaches

A *complete case* (CC) analysis consists in excluding all the individuals who dropped-out, i.e. with $U_i \leq J$, from statistical inference. This approach can guarantee unbiased effect estimates only under the covariate-dependent drop-out assumption, provided that all the covariates that influence both the outcome and drop-out mechanisms are included in $\mathbf{X_i}$. Furthermore, the exclusion of drop-outs results in the loss of the partial information available from these individuals, i.e. their observed responses and covariates. Thus, a CC analysis is inefficient.

Another common ad-hoc method is the so-called *last observation carried forward* approach. This method consists in replacing all the missing outcomes of each individual who dropped out by his last measured outcome, i.e. by $Y_{i(U_i-1)}$ in our notation. This approach requires a very strong assumption to warrant unbiased coefficient estimates: that outcomes remain constant after drop-out. Molenberghs and Kenward (2007, Section 4.3) show with a simple example that, when this assumption is violated, the size and direction of the resulting bias is not foreseeable as it depends on the true unknown regression coefficients. In addition, this approach underestimates the variability of regression coefficient estimates because the uncertainty concerning the missing outcomes is ignored, and may thus result in misleading inferences.

1.3.2 MAR approaches

Following Rubin (1976), define two parameter vectors $\boldsymbol{\theta_1}$ and $\boldsymbol{\theta_2}$ to be *distinct* if the parameter space of the full vector $(\boldsymbol{\theta_1}', \boldsymbol{\theta_2}')'$ factorizes into the product of the individual

parameter spaces. Rubin (1976) showed that direct likelihood inference based on all the available data yields unbiased and fully-efficient regression coefficient estimates if (i) the drop-out mechanism is MAR and (ii) the parameter vectors of the outcome and drop-out mechanisms are distinct (often called a *separability condition*), provided that all the covariates that simultaneously influence the outcome and drop-out processes are included in the model for the outcomes. When conditions (i) and (ii) hold, the drop-out mechanism is said to be *ignorable*. Thus, under this ignorability assumption, the LMM presented in Section 1.2.2 fitted to the available data provides valid estimates because it is a likelihood-based approach. As mentioned in Section 1.2.3, likelihood-based approaches also exist for fitting marginal models for continuous responses under the assumption that $\mathbf{Y}$ has a multivariate normal distribution. Thus, if all the available data is used, those approaches will also provide unbiased and fully efficient estimates under the ignorability assumption. An important remark concerning the direct likelihood approach is that, in some situations, the maximization of the likelihood function may challenging computationally. Dempster et al. (1977) proposed the *expectation-maximization* algorithm which is valid under the ignorability assumption and may be used in such situations. Also, when using direct likelihood, some caution is needed to obtain precision estimates that are consistent under MAR (Kenward and Molenberghs, 1998).

The GEE approach to fitting marginal models is a non-likelihood frequentist method and guarantees unbiased estimates only under the assumption of a covariate-dependent drop-out mechanism (Liang and Zeger, 1986). Thus, an extension of GEE based on *inverse probability weighting* (IPW) ideas, and that provides unbiased estimates under an MAR drop-out mechanism, was proposed by Robins et al. (1995). Known as *weighted generalized estimating equations* (WGEE), the method consists in weighting the contribution of each observed outcome to the estimating equations by the inverse of the probability of that outcome being observed.

An alternative approach for fitting both LMMs and marginal models under MAR is the multiple imputation (MI) approach of Rubin (1987). Details of this simulation-

based approach are given in a later chapter (see Section 3.1), but briefly it consists in multiply imputing the missing data several times by drawing values from the Bayesian posterior predictive distribution of the missing data given the observed data. This requires building a so-called *imputation model*. The desired *analysis model* is then fitted to each of the completed datasets yielded by this procedure using a method for complete data, and the resulting inferences are then combined into a single final inference using some arithmetically simple formulas.

Several strategies are available for imputing missing outcomes in the longitudinal data setting under MAR. With continuous response variables and missing outcomes due to drop-outs, a simple and commonly used approach is *sequential regression-based imputation* (Rubin, 1987, Section 5.4). This method consists in imputing the missing outcomes at each time-point in chronological order, each time using a univariate imputation model that includes the outcomes at previous visits as predictors. Another widely available approach, applicable in more general settings, assumes that the data follow a multivariate normal distribution and uses Markov chain Monte Carlo methods (Schafer, 1997). In Chapter 3, an alternative approach is proposed, in which missing values are drawn directly from a LMM. For the marginal model, if GEE is the estimation strategy of choice, MI provides a valuable alternative to the WGEE approach under MAR (see for example Chapter 11 of Molenberghs and Kenward, 2007). In contrast, if direct likelihood is considered for either model, MI with a correctly specified imputation model will yield estimates that approximate maximum likelihood estimates, but the latter will be more efficient (Schafer, 1999). Consequently, MI provides no advantage with respect to direct likelihood in the present setting. However, MI is a valuable tool for performing sensitivity analyses such as those presented in Chapter 6. The implementation of this methodology was actually the main motivation for the MI procedure of Chapter 3.

1.3.3 MNAR approaches

Several authors have proposed MNAR models for longitudinal data with drop-outs. Outside of MAR, the joint density of the vector of responses and the drop-out indicator, $f(\mathbf{y}, u)$, must be considered. Little and Rubin (1987, Chapter 11) introduced two classes of MNAR models corresponding to two possible factorizations of the joint density. The first class of models are termed *selection models*, and, omitting covariates and parameters, correspond to the following factorization:

$$f(\mathbf{y}, u) = f(\mathbf{y}^{\mathcal{O}}, \mathbf{y}^{\mathcal{M}}) \times f(u|\mathbf{y}^{\mathcal{O}}, \mathbf{y}^{\mathcal{M}}). \tag{1.8}$$

Thus, in selection models, a model is posited for the marginal distribution of the response vector and another model is posited for the conditional distribution of the drop-out indicator given the observed and missing responses. Fully-parametric selection models may be fitted by maximum likelihood, where the likelihood is obtained by integrating the joint density over the vector of missing outcomes $\mathbf{y}^{\mathcal{M}}$. One of the first examples of such models for continuous responses was the Diggle-Kenward model (Diggle and Kenward, 1994), in which a multivariate normal linear model for the response vector was combined with a logistic model for the drop-out probability at each time-point, which included the current and last measured outcomes as regressors.

The second class of models are known as *pattern-mixture models* (PMMs), and correspond to the following factorization:

$$f(\mathbf{y}, u) = f(\mathbf{y}^{\mathcal{O}}, \mathbf{y}^{\mathcal{M}}|u) \times f(u). \tag{1.9}$$

Thus, in PMMs, a model is posited for the conditional distribution of the response vector given the drop-out indicator and another model is posited for the marginal distribution of the drop-out indicator. Note that the first factor of (1.9) further factorizes as

$$f(\mathbf{y}^{\mathcal{O}}, \mathbf{y}^{\mathcal{M}}|u) = f(\mathbf{y}^{\mathcal{M}}|\mathbf{y}^{\mathcal{O}}, u) \times f(\mathbf{y}^{\mathcal{O}}|u).$$

The first factor in the right-hand side of this equality, $f(\mathbf{y}^{\mathcal{M}}|\mathbf{y}^{\mathcal{O}}, u)$, is under-identified. Hence, when fitting a PMM, additional assumptions must be made to identify this factor. Little (1993) and Little (1994) proposed using so-called *identifying restrictions*, an approach in which unidentified parameters are set equal to functions of the identified parameters. This approach was further explored by other authors (Little, 1995; Little and Wang, 1996; Molenberghs et al., 1998; Thijs et al., 2002; Kenward et al., 2003). Other strategies to fit PMMs include extrapolation of time-trends by fitting a model within each group determined by the drop-out indicator, and treating the drop-out indicator as a covariate (Hedeker and Gibbons, 1997; Verbeke and Molenberghs, 2000; Michiels et al., 2002; Demirtas and Schafer, 2003). The under-identified nature of PMMs is exploited in the sensitivity analysis approach proposed in Chapter 6. Another important aspect of these models is that averaging over the drop-out distribution is necessary to obtain estimates of parameters describing the marginal distribution of the responses. A convenient method is to average implicitly using MI and this is the approach that we use in Chapter 6 (Demirtas and Schafer, 2003).

A third class of MNAR models are the so-called *shared-parameter models* (Wu and Carroll, 1988; Wu and Bailey, 1988, 1989; Little, 1995). A common example is the model that assumes that the measurement and drop-out processes are independent given a vector of subject-specific random effects $\mathbf{b}$, so that the joint density is written as:

$$f(\mathbf{y}, u, \mathbf{b}) = f(\mathbf{y}^{\mathcal{O}}, \mathbf{y}^{\mathcal{M}}|\mathbf{b}) \times f(u|\mathbf{b}) \times f(\mathbf{b}). \tag{1.10}$$

1.3.4 Sensitivity analyses

In the general context of incomplete multivariate data, Molenberghs et al. (2008) showed that for every MNAR model fitted to the observed data, an MAR model producing exactly the same fit to the observed data can be constructed. Consequently, it is not possible to verify from the data itself whether the drop-out mechanism is MAR or MNAR. This means that every model for longitudinal data with drop-outs is subject to

unverifiable assumptions. In MAR models, the main unverifiable assumption is that the drop-out mechanism does not depend on the missing responses when conditioning on the covariates and observed responses. In MNAR models, the underlying unverifiable assumptions are stronger because explicit distributional and structural assumptions are made about either (i) the conditional distribution of drop-out given the observed and missing responses or (ii) the conditional distribution of the missing responses given the observed responses and the drop-out indicator. Hence, these models are more sensitive to model misspecification (Little and Rubin, 1987). In particular, formal tests of MAR versus MNAR such as those constructed by Diggle and Kenward (1994) should be interpreted with caution as they rely on the unverifiable correct specification of the model. Such problems were raised by several authors in the selection model framework in the discussion to Diggle and Kenward (1994).

As a result of these issues, there is increasing awareness of the need to perform *sensitivity analyses* (e.g. Carroll et al., 2004; Carpenter and Kenward, 2007; National Research Council, 2010; Burzykowski et al., 2010; Little et al., 2012). Such analyses aim at assessing the sensitivity of inferences obtained in a primary analysis to departures from the underlying unverifiable assumptions. Molenberghs and Kenward (2007) recommend that MNAR models be considered only as part of a sensitivity analysis due to their added dependance on unverifiable assumptions, and that the primary analysis be MAR-based.

Several approaches have been proposed for performing sensitivity analyses. However, each specific scientific context may require different considerations when setting up such an analysis, and no standard method exists nor should be prescribed as this is still an active area of research (Carroll et al., 2004). One possibility to assess the sensitivity of a primary analysis to modeling assumptions is to consider a family of MNAR models, each with different structural and/or distributional assumptions. Thus, some authors consider a set of different types of models, e.g. one or several selection, pattern-mixture and shared-parameter models, and assess the discrepancies among the inferences obtained (e.g. Little and Yau, 1996; Kenward, 1998; Kenward and Molenberghs, 1999;

Michiels et al., 2002).

The basis of the methodology developed in Chapter 6, and more generally in Part III of this dissertation, is a more structured version of the latter principle that has been advocated by several authors (Little, 1994; Rotnitzky et al., 1998; Scharfstein et al., 1999; Daniels and Hogan, 2000; Molenberghs et al., 2001a). It consists in positing a family of MNAR models indexed by a scalar or vector parameter that is varied across a set of plausible values determined by subject-matter specialists. Fitting each of these models yields a range of estimates for the parameter of interest and corresponding CIs and significance tests. Comparing the results obtained across the different values of the indexing parameter provides insight about the sensitivity of inferences. Other examples of such analyses are given by Minini and Chavance (2004a,b) and Carpenter et al. (2007) in the selection model framework and by Ratitch et al. (2013) in the PMM framework. The methods presented in Part III rely on the PMM framework, which is viewed by some authors as the most suitable for assessment of sensitivity (Daniels and Hogan, 2000; Daniels and Wang, 2009; Hogan, 2009). Recently, some authors have proposed a formal framework to summarize the results yielded by such sensitivity analysis approaches into a single inference that does not rely on unverifiable missingness assumptions (Molenberghs et al., 2001a; Vansteelandt et al., 2006). Related ideas have been explored by other authors in the Bayesian framework (e.g. Scharfstein et al., 2003).

Another view of sensitivity analyses in the literature is based on the global and local influence ideas of Cook (1977, 1986). Briefly, these approaches consider influence diagnostics from a single model to detect subjects that have a large impact on the conclusions of the analysis (Thijs et al., 2000; Verbeke et al., 2001; Molenberghs et al., 2001b; Jansen et al., 2006).

Chapter 2

Background on competing risks with missing causes of failure

2.1 Motivation: The ECOG clinical trial

Survival analysis concerns the study of time-to-event data, that is, the time to the occurrence of one event, which is often termed a *failure*. The competing risks model arises when there is a distinction between different types of events, or *causes of failure*, such that failure from one cause precludes failure from other causes. In this manuscript, the multi-state model formulation of the competing risks problem is adopted, as depicted in Figure 2.1. At the beginning of the study all individuals are at the "event-free" state 0, which is transient, and when they fail from cause j, $j = 1, \ldots, J$, they move to the absorbing state j. An alternative formulation is based on latent failure times, but this approach has often been criticized because it leads to several interpretation and identifiability issues (Prentice et al., 1978; Andersen et al., 2002).

The competing risks model is suitable for studying phenomena observed in many fields, in particular in clinical research and epidemiology. In these areas, it is often of interest to study the time to competing events such as death from a given cause, relapse

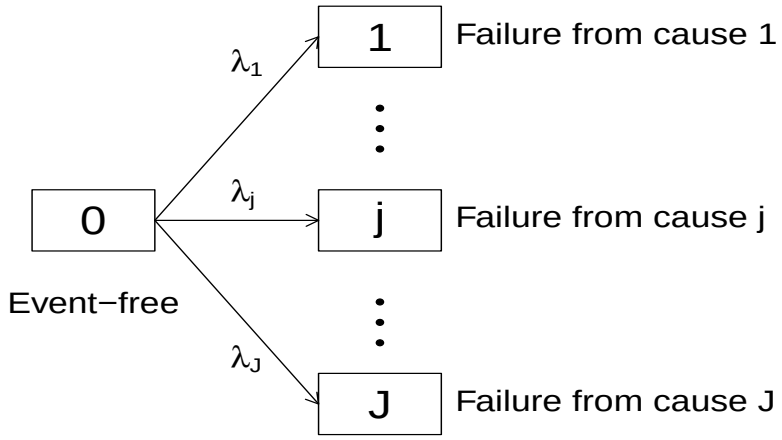

Figure 2.1: Multi-state model representation of the competing risks model.

or response to treatment. A competing risks regression analysis allows disentangling the effect of a risk factor or an intervention on the occurrence of these different events. For example, in a breast cancer clinical trial enrolling elderly women, it may be desirable to distinguish between cancer-related, treatment-related and other-cause mortality, to separately analyze the ability of the treatment to eliminate the cancer and its toxicity when patients are also at risk of dying from other causes.

An extensive literature exists on methods for regression modeling with competing risks data, some of which are now routinely used in clinical and epidemiological studies. A brief overview of the main existing methods is provided in Section 2.2. These methods generally require that the cause of failure is known for all subjects who have failed, a prerequisite that is not always met in practice. For instance, some causes of failure might be missing in studies with a long follow-up because collection of information tends to deteriorate with time due to several factors (e.g. patients move away). Other reasons for missing cause of failure data include forms not being completed by busy practitioners, patients dying without an autopsy, difficulty in assigning a cause of death

to persons with concurrent comorbid illnesses or confusion with definitions when coding causes of death (e.g. underlying cause of death vs. mechanisms of death) (Andersen et al., 1996; Manola and Gray, 2011).

A starting point for methodological developments in the literature for regression modeling of competing risks data with missing causes of failure was the E1178 clinical trial from the Eastern Cooperative Oncology Group (ECOG) (Goetghebeur and Ryan, 1995). The study enrolled 169 women with stage II breast cancer, aged between 65 and 84 years (median age was 71 years), who were randomized to receive tamoxifen or placebo during 24 months. At the cut-off date of the data available, the median follow-up time was of 6.7 years and 79 women had died (53% censored). Because of their advanced age, these women were at high risk of death from causes unrelated to their cancer. Indeed, among the deceased patients, 44 had died from cancer whereas 17 had died from other causes. For the remaining 18 women the cause of death was unknown (23% of deaths with missing cause). Cummings et al. (1993), who performed a later analysis of this trial with a median follow-up of 10 years, attribute these missing data to the trial design which did not foresee the possible obstacles to long-term data collection (e.g. patients moving away without a forwarding address, change in physicians, etc.), in addition to the usual compliance issues that arise in all trials.

Cummings et al. (1985) and Cummings et al. (1986) reported the first results of this trial (median follow-up of 3.4 and 4.6 years, respectively) and found no significant effect of tamoxifen on survival. On the other hand, two prognostic factors, the estrogen-receptor (ER) status of the primary tumor (positive vs. negative) and the degree of positive axillary lymph node involvement (<4 nodes vs. ≥ 4 nodes), were found to be significantly associated with survival. Table 2.1 shows the total number of women and of deceased women in each combined category of ER status and nodal involvement, and the observed causes of death in the data available.

Given the importance of competing causes of death in this study and the high percentage of missing causes, these data have been very valuable for exemplifying the practical value of novel methods for competing risks regression with missing causes of

Table 2.1: Total number of women and deceased women in each combined category of ER-status and nodal involvement, and observed causes of death.

ER status	Nodes	Total	Dead	Causes of death		
				Cancer	Other	Missing
Negative	< 4	1	0	0	0	0
	≥ 4	5	5	5	0	0
Positive	< 4	89	33	18	6	9
	≥ 4	74	41	21	11	9

failure in the literature (Goetghebeur and Ryan, 1995; Lu and Tsiatis, 2001; Gao, 2006; Nicolaie et al., 2011). Particular focus has been given to analyzing the effects of the two mentioned prognostic factors on cancer-related death. In Section 2.3, we provide the formal definitions of *missing at random* (MAR) and *missing not at random* (MNAR) missingness mechanisms in this context and an overview of the existing methods for handling missing causes of failure.

The aim of the "competing risks" part of the dissertation was to address some voids in the current missing cause of failure literature. First, we wished to propose a general framework for fitting regression models for the probability or *risk* of observing a specific event by a given time in the missing cause setting under the MAR assumption. Most of the existing methods focus on models for the *rate* of occurrence of events, but the risk is another essential quantity in understanding the competing risks model. Second, we wished to study the construction of the likelihood under the MAR assumption. In addition to making the fitting of parametric models for several useful functionals straightforward, this construction is interesting in its own right as it provides us with a better understanding of each individual's contribution to estimation with missing causes. Finally, we aimed at proposing an approach for assessing the robustness of inferences to departures from the MAR assumption. To our knowledge, neither MNAR modeling nor sensitivity analysis methodology have ever been in this setting. The work presented here is a first step in that direction. The ECOG trial data are used to illustrate the proposed methods throughout the manuscript.

2.2 Competing risks regression: An overview

2.2.1 Notation and basic functionals

We consider the competing risks setting where an individual may fail from two causes. We can focus on this simplified setting without loss of generality because, with more than two possible causes, failure from each cause can be analyzed separately by lumping all other causes together in an "other cause" category. In this dissertation we focus on regression modeling, thus we assume that there is interest in studying the influence on the competing risks mechanism of a p-vector $\mathbf{X}$ of fully-observed time-fixed covariates measured at baseline (i.e. at time 0). When stated explicitly, we may also consider time-dependent covariates of the *external* type, that is, such that the occurrence of a failure in the present may depend on the history of the covariate but not on its future path (see the formal definition in Kalbfleisch and Prentice, 2002, Chapter 6).

The response variables that are the target of inference when studying the competing risks model with two causes of failure are the *failure time* T and the *cause of failure* D, with $D = 1$ for failures from the cause of interest and $D = 2$ for failures from other causes. The failure time is often subject to *censoring* or *truncation*, which are the two particularities of time-to-event data that have prompted survival and event-history analysis to develop as a separate field of statistics. In this document only the phenomenon of *right-censored* data is considered, which means that some individuals in the study are observed to be event-free up to a certain time C, called the *censoring time*, after which no other information on failure occurrence is available. Thus, for these censored individuals the failure time and the cause of failure are not observed. Defining the observed time to failure or censoring as $\tilde{T} = \min\{T, C\}$ and the censoring indicator $U = I(C < T)$, the observed data are $(\tilde{T}_i, U_i, \mathbf{X}_i)$ for censored individuals and $(\tilde{T}_i, U_i, D_i, \mathbf{X}_i)$ for uncensored individuals. We assume that the data from different individuals are i.i.d. given the covariates.

It is important to emphasize that censoring is considered to be a nuisance phenomenon that prevents observation of the otherwise observable response vector of in-

terest, (T, D). Common examples are censoring due to the ending of a clinical study or patients moving away. Thus, the event of being censored is not a clinically relevant competing event, while all the clinically relevant competing events in the study are represented by either of the two states of the posited model. This condition ensures that it is clinically interesting and realistic to consider the situation without censoring, i.e. to make inferences about the joint distribution of (T, D) (Andersen and Keiding, 2012).

In these circumstances, it is often convenient to make the so-called *independent censoring* assumption. This assumption basically implies that the pairs (T_i, D_i) observed for uncensored patients are in some sense representative of the pairs that would have been observed for all individuals if there had been no censoring, the main consequence being that targeted distribution, i.e. the joint distribution of (T, D) given $\mathbf{X}$, is identifiable from the observed data. The *random censoring* assumption, i.e. the assumption of statistical independence between (T, D) and C given $\mathbf{X}$, would ensure this identifiability. However, the independent censoring assumption is weaker and suffices for application of the martingale and counting processes theory underlying most of the main results in survival analysis and competing risks (Andersen et al., 1993). This assumption can be interpreted as the condition that $P(t \leq T < t + h, D = j | T \geq t, C \geq t, \mathbf{X}) = P(t \leq T < t + h, D = j | T \geq t, \mathbf{X})$ for an infinitesimal h and $j = 1, 2$ (Fleming and Harrington, 2005). Thus, an individual's instantaneous probability of experiencing a failure from cause j, for $j = 1, 2$, given that the individual is still event-free at time t and covariates, is not disturbed by the additional information that the individual is also still uncensored. The formal definition of independent censoring in terms of counting processes can be found in Andersen et al. (1993, Chapter 3).

Given the stated assumptions, it is possible to study how $\mathbf{X}$ influences the competing risks mechanism by positing regression models for two identifiable functionals for each cause. The first functional is the cause-specific hazard rate (CSH), defined at each time-point t as the rate at which the specific event occurs among patients still alive just

before time t:

$$\lambda_j(t) := \lim_{h \to 0} \frac{1}{h} P(t \leq T < t + h, D = j | T \geq t), \qquad j = 1, 2.$$

The CSHs are the transition rates in the multi-state model depicted in Figure 2.1, and can be interpreted as quantifying the instantaneous forces that drive individuals out of the "event-free" state (Beyersmann et al., 2009). Thus, regression modeling of the CSHs is suitable for addressing etiologic questions (Koller et al., 2012).

The second functional is the cumulative incidence function (CIF), defined as the probability of observing the specific event before time t:

$$F_j(t) := P(T \leq t, D = j), \qquad j = 1, 2.$$

In the model of Figure 2.1, the CIF for cause j represents the probability of having transitioned to state j by time t given that the individual was in state 0 at time 0, and is interpreted as the actual risk of failure from cause j in the time-frame $[0, t]$. Regression models for the CIFs are useful when the focus is on prognosis. Note that, when viewed as a function of t, the CIF is not a true probability distribution function because $F_j(\infty) = P(D = j) < 1$. Thus the CIF is said to be a *subdistribution* function.

One key feature of the competing risks model is that both the CSH and the CIF, which represent measures of the rate and the risk of an event respectively, need to be studied in order to fully understand the competing risks mechanism (Andersen et al., 2012; Latouche et al., 2013). This contrasts with the standard survival model, that is, when studying all-cause failure. To see this, note that in the latter setting the probability distribution function of the failure times $F(t) := P(T \leq t) = F_1(t) + F_2(t)$, or equivalently the survival function $S(t) := P(T > t) = 1 - F(t)$, is the corresponding measure of the risk, and the overall hazard rate $\lambda(t) := \lim_{h \to 0} h^{-1} P(t \leq T < t + h | T \geq t) = \lambda_1(t) + \lambda_2(t)$ is the measure of the rate. The identity $S(t) = \exp\{-\int_0^t \lambda(u) du\}$ shows that there is a one-to-one relation between these two quantities. Thus, regression

analysis of either the overall hazard rate or the survival function will lead to the same qualitative conclusions. In contrast, in the competing risks model, the one-to-one correspondence between the rate and the risk breaks down, with the CIF being a complex non-linear function of the CSHs of all the competing causes, and vice-versa:

$$F_j(t) = \int_0^t \lambda_j(u) S(u^-) du, \qquad \lambda_j(t) = \frac{f_j(t)}{1 - F_1(t) - F_2(t)}, \qquad j = 1, 2.$$

Here, $f_j(t) := \frac{d}{dt} F_j(t)$, $j = 1, 2$, are improper density functions because they do not integrate to 1. Therefore, an increase in the CSH of one cause will not necessarily reflect an increase in the corresponding CIF; this will depend on how the other causes' CSHs behave (Beyersmann et al., 2007). As a result, methods for direct regression modeling of each of these functionals have been developed.

2.2.2 Regression models for the CSH

Among the existing methods for regression modeling of the CSH, two classes of models stand out because of their practical value in terms of the interpretability of regression coefficients, availability of software and acceptance by the scientific community. The first is the class of *proportional hazards* models, which are widely used in clinical and epidemiological studies in the standard survival analysis setting, and also for modeling the CSHs in the competing risks setting. A very popular proportional hazards model is the semi-parametric *Cox model* (Cox, 1972) which for cause j has the form

$$\lambda_j(t|\mathbf{X}) = \lambda_{j0}(t) \exp(\boldsymbol{\beta}_j' \mathbf{X}), \tag{2.1}$$

where $\boldsymbol{\beta}_j$ is a p-vector of regression coefficients and the baseline CSH $\lambda_{j0}(t)$ is left unspecified. Thus, covariates are assumed to have a multiplicative effect on the hazard. The "proportional hazards" property refers to the fact that in these models the ratio of the CSHs of any two individuals (called the *hazard ratio*) is constant in time. In fact, the popularity of proportional hazards models may be in part explained by the fact

that regression coefficients represent (log) hazard ratios and thus have a direct "relative rate" interpretation. For instance, if $\mathbf{X}$ represents a single continuous covariate, say X, then the exponential of its effect, say $\exp(\beta_j)$, represents the relative increase in the CSH for cause j for an increase in one unit of the covariate:

$$\exp(\beta_j) = \frac{\lambda_j(t|X = x + 1)}{\lambda_j(t|X = x)}.$$

The regression coefficients of the Cox model may be estimated by maximizing a partial likelihood. The large sample properties of the resulting estimators were established by Andersen and Gill (1982) and Andersen et al. (1985). The partial likelihood actually factorizes into two components, one for each cause. Actually, if $\boldsymbol{\beta}_1$ and $\boldsymbol{\beta}_2$ are assumed to be distinct (cf. Section 1.3.2), the model for each cause may be fitted using standard software for Cox regression by censoring the individuals who failed from a competing cause (Prentice et al., 1978). Similar remarks apply when considering the full likelihood function in the presence of competing risks (cf. Chapter 5).

Fully-parametric versions of the Cox model may also be considered, which have the form

$$\lambda_j(t|\mathbf{X}) = \lambda_{j0}(\boldsymbol{\alpha}_j, t) \exp(\boldsymbol{\beta}_j' \mathbf{X}), \tag{2.2}$$

where the baseline CSHs λ_{j0} are known up to the parameter vectors $\boldsymbol{\alpha}_j$. Maximum likelihood estimation for this type of models has been studied by Kalbfleisch and Prentice (2002, Chapters 3 and 5), Borgan (1984) and Andersen et al. (1985). Some possible parametrizations for the baseline CSHs are: the piecewise constant model $\lambda_{j0}(\boldsymbol{\alpha}_j, t) = \sum_1^K \alpha_{jk} I(t \in [t_{k-1}, t_k])$; the Cox-exponential model $\lambda_{j0}(\boldsymbol{\alpha}_j, t) = \alpha_{j1}$; the Cox-Weibull model $\lambda_{j0}(\boldsymbol{\alpha}_j, t) = \alpha_{j1}\alpha_{j2}t^{\alpha_{j2}-1}$; and the Cox-Gompertz model $\lambda_{j0}(\boldsymbol{\alpha}_j, t) = \alpha_{j1} \exp(\alpha_{j2}t)$ (Bender et al., 2005).

Additive hazards models constitute a second important class of models that provide valuable insight into the competing risks mechanism. A general form of additive hazards model is given by the nonparametric *Aalen model*: $\lambda_j(t|\mathbf{X}) = \lambda_{j0}(t) + \boldsymbol{\beta}_j(t)'\mathbf{X}$, where both $\lambda_{j0}(t)$ and $\boldsymbol{\beta}_j(t)$ are unspecified functions of time (Aalen, 1980). This model

enables the study of time-varying covariate effects. A semi-parametric version of the Aalen model, that can be seen as the additive counterpart of the Cox model, is the model with constant covariate effects i.e. $\boldsymbol{\beta}_j(t) = \boldsymbol{\beta}_j$ (McKeague and Sasieni, 1994; Lin and Ying, 1994):

$$\lambda_j(t|\mathbf{X}) = \lambda_{j0}(t) + \boldsymbol{\beta}_j'\mathbf{X}. \tag{2.3}$$

Model (2.3) is interesting because its regression coefficients have an "excess rate" interpretation, thus giving an *absolute* rather than a *relative* measure of covariate effects, which may be of particular interest in certain contexts. A thorough review of estimation with additive hazards models is given by Martinussen and Scheike (2006).

Klein (2006) advocates the use of such models with competing risks because, unlike the multiplicative structure of proportional hazards models, the additive structure can hold simultaneously for each CSH and for the all-cause hazard. Furthermore, in this case the vector of covariate effects on the all-cause hazard, say $\boldsymbol{\beta}$, is properly partitioned into the sum of the effects on the CSHs, i.e. $\boldsymbol{\beta} = \boldsymbol{\beta}_1 + \boldsymbol{\beta}_2$.

2.2.3 Regression models for the CIF

The most common regression models for the CIF are semi-parametric generalized linear models of the form

$$g\{F_j(t|\mathbf{X})\} = \beta_{j0}(t) + \boldsymbol{\beta}_j'\mathbf{X}, \tag{2.4}$$

where g is a monotone differentiable link function and $\beta_{j0}(t)$ is an unspecified time-dependent intercept. Model (2.4) encompasses models such as the *Fine and Gray model* if g is the complementary log-log (cloglog) function (Fine and Gray, 1999), the *additive model* if g is the identity function (Klein, 2006) and the *absolute risk model* if g is the log function (Gerds et al., 2012). In the latter two models, interpretation of the regression coefficients is straightforward: in the additive model, they have an "excess risk" interpretation and in the absolute risk model they have a "relative risk" interpretation. In the Fine and Gray model, interpretation of the regression coefficients is more subtle.

In fact, the coefficients in this model are (log) ratios of *subdistribution hazard rates*, which are defined as:

$$\lambda_j^*(t) := \lim_{h \to 0} \frac{1}{h} P(t \leq T < t + h, D = j | T \geq t \cup \{T < t, D \neq j\}), \qquad j = 1, 2.$$

The subdistribution hazard for cause j represents the rate at which failures from cause j are occurring among individuals who have not yet failed or who have failed from another cause. Thus, these quantities have no clinically relevant interpretation in the competing risks model (Andersen and Keiding, 2012). However, their importance arises from the fact that, unlike the CSHs, they do have a one-to-one relation with the CIF since

$$F_j(t) = 1 - \exp\left(-\int_0^t \lambda_j^*(u) du\right).$$

Hence, an increase (decrease) in the subdistribution hazard reflects an increase (decrease) in the corresponding CIF.

To estimate the regression coefficients of model (2.4), so-called *inverse probability of censoring weighting* (IPCW) techniques have been proposed (Fine and Gray, 1999; Fine, 2001; Scheike et al., 2008). An alternative approach was proposed by Klein and Andersen (2005) which is henceforth referred to as the *Andersen-Klein pseudo-value approach*. A detailed account of this approach is given in Chapter 4 as it was the basis for the developments presented there.

A fully-parametric model for the CIF has also been proposed, which is a generalized odds rate model with a Gompertz distribution for the logarithm of the baseline cumulative subdistribution hazard (Jeong and Fine, 2007):

$$F_j(t|\mathbf{X}) = 1 - \left[1 + \alpha_j \exp(\boldsymbol{\beta}_j' \mathbf{X}) \tau_j \{\exp(\rho_j t) - 1\}/\rho_j\right]^{-1/\alpha_j}. \tag{2.5}$$

Interesting submodels include the proportional odds model ($\alpha_j = 1$) and the proportional subdistribution hazards model ($\alpha_j \to 0$). This model is fitted using maximum likelihood.

2.3 Regression with missing causes of failure

Consider the setting where the cause of failure is missing for some of the uncensored individuals. Thus, for uncensored individuals a missingness indicator M is observed, with $M = 1$ if the cause of failure is missing and $M = 0$ otherwise. Denote by $\mathbf{W}$ the r-vector of fully-observed covariates that influence the missingness mechanism, some of which may already be in $\mathbf{X}$. Then the observed data are now $(\tilde{T}_i, U_i, \mathbf{X}_i, \mathbf{W}_i)$ for censored individuals, $(\tilde{T}_i, U_i, M_i, D_i, \mathbf{X}_i, \mathbf{W}_i)$ for uncensored individuals with observed cause and $(\tilde{T}_i, U_i, M_i, \mathbf{X}_i, \mathbf{W}_i)$ for uncensored individuals with missing cause.

Let $\pi := P(M = 1|T, D, \mathbf{W}, U = 0)$ be the conditional probability that the cause of failure is missing among uncensored individuals ($U = 0$), upon whom the missingness mechanism acts. With this notation, the missing data taxonomy of Rubin (1976) and Little and Rubin (1987) in the missing cause setting can be expressed as follows:

(i) The mechanism driving missingness is said to be MCAR (for *missing completely at random*) if π is constant.

(ii) The mechanism driving missingness is said to be MAR (for *missing at random*) if π does not depend on the cause of failure, that is $\pi = P(M = 1|T, \mathbf{W}, U = 0)$.

(iii) The mechanism driving missingness is said to be MNAR (for *missing not at random*) otherwise.

With missing causes of failure, the CSHs and CIFs are no longer identifiable without further assumptions about which type of mechanism is driving missingness. Thus, the validity of any strategy for regression modeling in this setting will require that some assumption about the missingness mechanism holds. In the following sections, we discuss the currently available methods for regression modeling with missing causes of failure and their underlying assumptions.

2.3.1 Ad-hoc approaches

A *complete case* (CC) analysis in the present context consists in excluding the uncensored individuals with missing cause of failure from statistical inference. Note that the term "complete cases" is less appropriate in this context because we do not observe a cause of failure for censored individuals, yet they are not excluded in a CC analysis. The inefficiency of a CC analysis in this setting is due to the discarding of the partial information available from the individuals with missing cause of failure, i.e. about T and $\mathbf{X}$. Furthermore, unlike with longitudinal data, the CC analysis in the missing cause setting is not necessarily unbiased when missingness depends only on covariates, which is why we did not consider the additional distinction between MAR and covariate-dependent missingness in the classification above. Actually, the CC analysis does not guarantee unbiased estimates even under an MCAR missingness mechanism.

To explain this, consider the case where the effect of a binary covariate on the CSH or the CIF is to be estimated. Broadly, such estimation requires comparing, not means, but event frequencies between the two groups determined by the covariate. A CC analysis implies a reduction in the numerators and denominators of these frequencies, which are the numbers of events and the risk sets in the two groups, respectively. Under MCAR, the expected percentage reduction in the numerator is the same for each group, equal to $100\pi\%$. However, the percentage reduction in the denominator is not necessarily the same in both groups, particularly due to censoring. Hence, the ratio of these probabilities is generally modified even under MCAR, as is the difference between these probabilities. These modifications will imply biased covariate effect estimates. When missingness depends on the covariate larger biases may be expected. A possibly more intuitive explanation of the bias in the CC analysis is given by Andersen et al. (1996), who note that the CC analysis leads to excluding higher risk individuals because only failed individuals, and no censored individuals, are removed. Hence, the estimated covariate effects actually measure the association of covariates with failure among lower-risk individuals, which is not the main target of interest.

Another approach is to consider "failure with missing cause" as an additional competing event, i.e. to augment the state space of the underlying multi-state model to include 'missing cause' as a type of failure. This approach will be henceforth referred to as an *extra state* (ES) analysis. As mentioned in the beginning of Section 2.2.1, when modeling the CSH or CIF for the cause of interest ($D = 1$) it is only necessary to know whether failures were due to the cause of interest or not. Thus, to guarantee unbiased estimates, the ES analysis requires the strong assumption that $\pi = 0$ if $D = 1$. When this assumption is violated, the ES approach leads to potentially misclassifying failures from the cause of interest as failures from other causes, resulting in an underestimation of the marginal CSH and CIF of the cause of interest. Moreover, the misclassification rate may depend on the covariates under study because missingness probabilities may depend on covariates, as noted by Bakoyannis et al. (2010). This so-called *differential misclassification* may result in either upward or downward biases in regression coefficient estimates. The ES approach is also inefficient because treating all missing causes as cause 2 failures, means that potential cause 1 failures that would help increase precision are oversighted. Finally, the uncertainty concerning the missing causes is completely disregarded, so the precision estimates yielded by this approach are misleading.

A related approach, that we do not explore further but that leads to similar problems as the ES analysis, is to consider "failure with missing cause" as a failure from the cause of interest. In this case, $\pi = 0$ if $D \neq 1$ is the required assumption to guarantee unbiased estimates. Andersen et al. (1996) discuss the pitfalls of this approach in detail.

2.3.2 MAR regression methods for the CSH

Several authors have addressed the problem of fitting semi-parametric CSH models in the missing cause of failure setting under the MAR assumption. Goetghebeur and Ryan (1995) and Andersen et al. (1996) studied a partial likelihood approach for fitting a Cox model for each cause by assuming that the ratio between the baseline CSHs for causes

1 and 2 is constant. Nicolaie et al. (2011) studied a modification of this approach that relaxes the constant baseline hazard ratio assumption. Lu and Tsiatis (2001) explored a multiple imputation (MI) approach for fitting a Cox model for one or both causes. Gao and Tsiatis (2005) and Gao (2006) proposed so-called *augmented inverse probability weighted* estimators for fitting linear transformation models, a large class of models that includes the Cox model as a particular case. Lu and Liang (2008) proposed augmented inverse probability weighted estimators for the semi-parametric additive hazards model (2.3).

On the other hand, parameter estimation for the fully-parametric model (2.2) in the missing cause setting has received no attention despite the potential usefulness of this model in many applications. The direct likelihood approach is explored in Chapter 5.

2.3.3 MAR regression methods for the CIF

Recently, some authors have considered estimation of the CIF based on estimates of other related quantities obtained through regression modeling under MAR. For instance, Lee et al. (2011) use estimates of each of the CSHs obtained via the MI approach of Lu and Tsiatis (2001), and Nicolaie et al. (2011) consider the so-called *vertical modeling* approach to competing risks, the details of which are given in Chapter 5. However, to our knowledge, direct regression modeling of the CIF in the missing cause setting has been addressed only by Bakoyannis et al. (2010) using MI, with the analysis model being the Fine and Gray model fitted with the IPCW approach as in Fine and Gray (1999). No other missing data approaches, such as inverse probability weighting (IPW), have yet been proposed. Moreover, the performance of the MI procedure of Bakoyannis et al. (2010) when applied to flexible modeling strategies for the CIF such as the Andersen-Klein pseudo-value approach or when dealing with continuous covariates has not yet been explored.

One of the main contributions of this dissertation is the proposal of a general framework for semi-parametric regression modeling of the CIF under MAR, encompassing

key models such as the Fine and Gray and additive models (Klein, 2006). Two extensions of the Andersen-Klein pseudo-value approach are considered. The first extension is a novel approach grounded on the inverse probability weighting paradigm for dealing with missing data. The second extension is the MI procedure of Bakoyannis et al. (2010) coupled with the Andersen-Klein approach. These developments are presented in Chapter 4.

As an additional contribution, we derive the expression of the likelihood that would enable the fitting of fully-parametric models for the CIF such as (2.5) under the assumption of an ignorable missingness mechanism (Chapter 5).

2.3.4 MNAR and sensitivity analyses

Under MNAR mechanisms, it is necessary to model the joint distribution of (T, D, M) for uncensored individuals, upon whom the missingness mechanism acts. For this purpose, the different classes of models discussed in Section 1.3.3 for longitudinal data with drop-outs could be transposed to the missing cause of failure setting. However, as with longitudinal data, in the missing cause setting it is not possible to assess from the observed data whether the missingness mechanism is MAR or MNAR. Thus, any modeling strategy relies on unverifiable assumptions and sensitivity analyses should be routinely performed to assess the robustness of inferences to departures from these assumptions. To our knowledge, neither MNAR modeling nor sensitivity analysis methodology have ever been considered in the missing cause of failure setting. The pattern-mixture modeling framework in this context is considered in Chapter 7, and an approach for performing sensitivity analyses is proposed.

Part II

MAR modeling

Chapter 3

An MI procedure for continuous longitudinal data with drop-outs

Several multiple imputation (MI) procedures have been proposed in the literature for dealing with continuous longitudinal data with drop-outs (cf. Section 1.3.2). In this chapter we present an alternative procedure in which missing outcomes are drawn directly from a linear mixed model (LMM) like (1.1). The main purpose of developing this tool was to enable the implementation the sensitivity analysis approach of Chapter 6. In Section 3.1 we provide a general overview of MI since this approach plays a major role in all of the subsequent chapters of this dissertation. In Section 3.2 we describe the MI procedure proposed. In Section 3.3 we present a simulation study performed to validate this approach. Section 3.4 contains some concluding remarks. The application of the proposed method to the analysis of the SMI study is deferred until Chapter 6, where we conduct a sensitivity analysis of the inferences obtained.

3.1 Background on MI

MI is a general approach to deal with missing data proposed by Rubin (1987). An MAR missingness mechanism is assumed in many applications of MI, and we will focus on that setting in this chapter. Essentially, MI consists in the production of several,

say $m > 1$, plausibly completed datasets accounting for all the levels of uncertainty concerning the missing values. The procedure can be summarized in four main steps:

Step 1 A prediction model, called the *imputation model*, is built for the conditional distribution of the missing data given the observed data;

Step 2 The missing data are imputed m times by drawing values from the imputation model to obtain m completed datasets;

Step 3 The *analysis model* (i.e. the main model of interest) is then fitted to each completed dataset using the method that would have been chosen had the data been complete;

Step 4 The results obtained from each completed dataset are combined into a single inference by means of some arithmetically simple formulas, henceforth referred to as *Rubin's formulas*.

Step 1 requires positing an appropriate model for the type of missing data, e.g. a normal linear model for independent continuous data, a logistic model for independent binary data, etc. Furthermore, it is recommended that the imputation model includes as predictors all the variables known to influence missingness and all the variables and associations present in the analysis model (Schafer, 1999). Under the MAR assumption, the parameters of the imputation model may be estimated by fitting the model to the available data. To illustrate this, consider the case of longitudinal data with drop-outs. Following the notation introduced in Chapter 1, an imputation model must be built for $f(\mathbf{y}^{\mathcal{M}}|\mathbf{y}^{\mathcal{O}}, \mathbf{x}, \mathbf{w}, u)$. The MAR assumption, formally defined in this context in Section 1.3, is equivalent to the condition that $f(\mathbf{y}^{\mathcal{M}}|\mathbf{y}^{\mathcal{O}}, \mathbf{x}, \mathbf{w}, u)$ does not depend on u. Indeed, MAR implies:

$$f(\mathbf{y}^{\mathcal{M}}|\mathbf{y}^{\mathcal{O}}, \mathbf{x}, \mathbf{w}, u) = \frac{f(u|\mathbf{y}^{\mathcal{O}}, \mathbf{y}^{\mathcal{M}}, \mathbf{x}, \mathbf{w}) f(\mathbf{y}^{\mathcal{M}}|\mathbf{y}^{\mathcal{O}}, \mathbf{x}, \mathbf{w})}{f(u|\mathbf{y}^{\mathcal{O}}, \mathbf{x}, \mathbf{w})} = f(\mathbf{y}^{\mathcal{M}}|\mathbf{y}^{\mathcal{O}}, \mathbf{x}, \mathbf{w}).$$

The proof of the other implication of the equivalence is analogous. It can further be shown that this expression is equivalent to the condition that, given $s \in \{2, \ldots, J\}$ and

$s' \leq s$,

$$f(y_{\cdot s}|y_{\cdot 1}, \ldots, y_{\cdot(s-1)}, \mathbf{x}, \mathbf{w}, u = s') = f(y_{\cdot s}|y_{\cdot 1}, \ldots, y_{\cdot(s-1)}, \mathbf{x}, \mathbf{w}, u > s)$$

(see formal proof in Molenberghs et al., 1998). That is, at a given time-point s, the conditional distribution of the missing outcomes given the past outcomes and covariates is the same as the conditional distribution of the outcomes available at that time-point. Hence, the former can be estimated from the latter under MAR.

Step 2 requires performing m independent random draws from the posterior predictive distribution of the missing data under the assumed imputation model and the specified prior distributions for its parameters. To see how this may be done in practice, the concept of a *proper* imputation procedure is essential. Barnard et al. (1998) briefly define an imputation procedure to be proper if appropriate variability is incorporated across the m imputations, regarding the uncertainty both in the imputation model parameters and in the missing outcomes given the assumed imputation model (see full formal definition in Rubin, 1987, 1996). Thus, as described by Barnard et al. (1998), the following two-stage procedure is often useful: first, a vector of imputation model parameters is drawn from its posterior distribution; second, the missing value is drawn from the model implied by the drawn parameters. With large samples, it is possible to approximate the posterior distribution of the imputation model's parameters by their maximum-likelihood-estimated asymptotic distribution (Rubin, 1987). Of course, the specific imputation procedure used in practice will depend on the nature of the imputation model. In this dissertation we study procedures for drawing imputations from an LMM (this chapter) and from a logistic model (Chapter 4).

Step 3 is performed using standard software for complete data according to the analysis model, and yields m estimates $\hat{\boldsymbol{\theta}}^{(1)}, \ldots, \hat{\boldsymbol{\theta}}^{(m)}$ of the parameter of interest $\boldsymbol{\theta}$, as well as m variance estimates $\hat{\text{var}}(\hat{\boldsymbol{\theta}}^{(1)}), \ldots, \hat{\text{var}}(\hat{\boldsymbol{\theta}}^{(m)})$.

Step 4 consists in combining these estimates according to Rubin's formulas (Rubin, 1987). The MI estimator of $\boldsymbol{\theta}$ is given by:

$$\hat{\boldsymbol{\theta}}^* = \frac{1}{m}\sum_{l=1}^{m}\hat{\boldsymbol{\theta}}^{(l)}. \tag{3.1}$$

The variance of this estimator is the sum of a within-imputation component ($\mathbf{W}$) and a between-imputation component ($\mathbf{B}$). It is estimated by

$$\begin{aligned}\hat{\text{var}}(\hat{\boldsymbol{\theta}}^*) &= \hat{\mathbf{W}} + \left(1+m^{-1}\right)\hat{\mathbf{B}} \\ &= \frac{1}{m}\sum_{l=1}^{m}\hat{\text{var}}(\hat{\boldsymbol{\theta}}^{(l)}) + \left(1+m^{-1}\right)\frac{\sum_{l=1}^{m}(\hat{\boldsymbol{\theta}}^{(l)}-\hat{\boldsymbol{\theta}}^*)(\hat{\boldsymbol{\theta}}^{(l)}-\hat{\boldsymbol{\theta}}^*)'}{m-1}.\end{aligned} \tag{3.2}$$

To test the hypothesis $H_0(\boldsymbol{\theta} = \boldsymbol{\theta}_0)$ for scalar $\boldsymbol{\theta}$, a t distribution can be used as reference since $(\hat{\boldsymbol{\theta}}^* - \boldsymbol{\theta}_0)/\sqrt{\hat{\text{var}}(\hat{\boldsymbol{\theta}}^*)} \sim t_v$ under H_0, where $v = (m-1)[1+\hat{\mathbf{W}}/\{(1+m^{-1})\hat{\mathbf{B}}\}]^2$ (Rubin and Schenker, 1991). Alternatively, Li et al. (1991) proposed using a scaled statistic with an F reference distribution for performing hypothesis tests for multivariate $\boldsymbol{\theta}$ (see also Meng and Rubin, 1992). This procedure is implemented in the R *mice* package (function *pool.compare* with the default method) and it is what we use to perform hypothesis tests in this chapter and in Chapter 6, but for scalar $\boldsymbol{\theta}$, like in our case, this procedure and the above t-test should lead to very similar results.

A $100(1-\alpha)\%$ confidence interval (CI) for scalar $\boldsymbol{\theta}$ has endpoints $\hat{\boldsymbol{\theta}}^* \pm t_v^{\alpha/2}\sqrt{\hat{\text{var}}(\hat{\boldsymbol{\theta}}^*)}$, where $t_v^{\alpha/2}$ is the upper $(\alpha/2)^{th}$ percentile of a t distribution with v degrees of freedom.

Rubin's formulas require some conditions to be met to yield valid inferences. First, the imputation procedure must be proper as defined above. Second, the imputation model must be correctly specified. Finally, the imputation and analysis models should be compatible, or *congenial*, which means that the imputation model and the analysis model can be derived from the same overarching model (Meng, 1994). If these conditions are met, the MI estimator (3.1) is consistent, asymptotically normal and its asymptotic

variance is correctly estimated by (3.2) (Tsiatis, 2006, Chapter 14). A remarkable feature of MI is that often a small number imputations, say $m = 5$ or $m = 10$, are enough to achieve a very good efficiency compared to the estimator based on an infinite number of imputations.

Under MAR and when the aforementioned conditions are fulfilled, MI provides little gain in the context of longitudinal data with drop-outs compared to the direct likelihood approach, with MI being generally less efficient (cf. Section 1.3.2). The real value of MI in this setting is its applicability even outside these conditions (see Section 9.7 of Molenberghs and Kenward, 2007). In particular, allowing for uncongenial MNAR imputations makes MI a valuable tool for performing sensitivity analyses. The MAR-based MI procedure studied in the following sections is an important ingredient of the developments presented in Chapter 6, where we explore the use of MI in the context of sensitivity analyses.

3.2 Imputation procedure

In the following sections we omit the auxiliary covariate matrix $\mathbf{W}$, and assume that the matrix $\mathbf{X}$ already includes all the covariates influencing the drop-out probability. With longitudinal data, an appropriate imputation model for missing values due to drop-out is an LMM of the form (1.1). As shown in Section 3.1, under the MAR assumption the parameters of such an imputation model can be estimated by fitting the model to the available data. Let $\hat{\boldsymbol{\beta}}$ and $\hat{\text{var}}(\hat{\boldsymbol{\beta}})$ be the estimates obtained for the fixed effects vector and the estimator's variance–covariance matrix, respectively, and let $\hat{\mathbf{b}}_i$ and $\hat{\text{var}}(\hat{\mathbf{b}}_i)$ be the predictor of the i^{th} random effects vector and its estimated variance–covariance matrix, respectively, for $i = 1, \ldots, n$. Also, let $\hat{\sigma}^2$ be the estimated residual variance.

The proposed procedure for the l^{th} imputation of the missing outcomes, where $l \in \{1, \ldots, m\}$, is as follows:

(a) Draw $\boldsymbol{\beta}^{(l)} \sim N(\hat{\boldsymbol{\beta}}, \hat{\text{var}}(\hat{\boldsymbol{\beta}}))$ and $\mathbf{b}_i^{(l)} \sim N(\hat{\mathbf{b}}_i, \hat{\text{var}}(\hat{\mathbf{b}}_i))$ for $i = 1, \ldots, n$.

(b) For each missing outcome Y_{ij}, calculate the linear predictor:

$$\eta_{ij}^{(l)} = \mathbf{X}'_{ij}\boldsymbol{\beta}^{(l)} + \mathbf{Z}'_{ij}\mathbf{b}_i^{(l)}.$$

(c) Draw $\sigma^{2(l)} \sim d\hat{\sigma}^2/\chi^2_d$ where d, the residual degrees of freedom, is estimated as described below. Draw an error $\varepsilon_{ij}^{(l)} \sim N(0, \sigma^{2(l)})$.

(d) Impute each missing outcome Y_{ij} as $\eta_{ij}^{(l)} + \varepsilon_{ij}^{(l)}$.

In LMMs, the residual degrees of freedom d must be estimated, and for this purpose we follow the approach suggested by Bates (2006) (see also Spiegelhalter et al., 2002). It consists in estimating the effective number of parameters as the trace of the *hat matrix* (see formula 22 of Bates, 2010). The residual degrees of freedom d is then estimated by subtracting this quantity from the number of observations used to fit the model.

We implemented this procedure as an imputation method to be passed on to the function *mice* of the `R` *mice* package, which is a generic software for MI (van Buuren and Groothuis-Oudshoorn, 2011). The `R` functions required for this can be found in the Supplementary Material of the corresponding published manuscript (Moreno-Betancur and Chavance, 2013).

3.3 Simulation study

We performed a simulation study to validate the proposed imputation procedure. The main aim of the study was to examine the statistical properties of the MI inferences yielded by the procedure. A secondary aim was to compare the performance of MI to that of the complete case (CC) analysis, which consists in excluding drop-outs from the analysis (cf. Section 1.3.1).

3.3.1 Study design

We mimicked the design of the SMI study (cf. Section 1.1). More precisely, we considered the scenario of a clinical trial performed on n patients, with $n/2$ in the treatment group and $n/2$ in the control group, for whom outcomes had to be measured at six equally spaced visits including baseline.

We generated Gaussian outcomes using an LMM with the expected population trajectory intercepting the origin. The time-slopes were assumed to have a null expectation for individuals in the control group and a non-negative expectation β for those in the treatment group. The random part of the model included subject-specific random intercepts and time-slopes inducing correlations among the outcomes of each individual. Outcomes were thus generated as

$$Y_{ij} = j\beta X_i + b_{0i} + jb_{1i} + \varepsilon_{ij}, \tag{3.3}$$

for individual $i \in \{1, \ldots, n\}$ and visit $j \in \{0, \ldots, 5\}$, where X_i was the group indicator, with $X_i = 1$ if subject i was in the treatment group and $X_i = 0$ otherwise, $\mathbf{b}_i = (b_{0i}, b_{1i})' \sim N(\mathbf{0}, \mathbf{I}_2)$, $i \in \{1, \ldots, n\}$, were the i.i.d. vectors of subject-specific random effects, assumed to be independent of X_i and of the i.i.d. errors ε_{ij}, which were assumed to follow a standard normal distribution. The values considered in the simulation scenarios for β, the mean time-slope for the treatment group (which was also the difference between the expected time-slopes for the two groups), were 0 and 0.2.

The primary endpoint had to be measured at the last scheduled visit and the parameter of interest was the treatment effect, which was defined as the difference between the expected outcomes for the treated and control populations at this visit. The analysis model we considered was thus given by

$$Y_{i5} = \theta_0 + \theta X_i + \epsilon_i, \tag{3.4}$$

where the $\epsilon_i's$ were i.i.d. zero-mean Gaussian errors, and the parameter of interest was

$\theta = 5\beta$, which took as value either 0 or 1.

3.3.2 Simulation of drop-outs

Drop-outs were simulated according to different MAR and MNAR mechanisms, assuming that the baseline outcome was observed for all individuals. The first type of MAR drop-out mechanism considered was in particular a covariate-dependent drop-out mechanism, in which drop-out probability depended on the individual's group, either assigning the treatment group a 0.1 drop-out probability and the control group a 0.4 drop-out probability or vice-versa. As briefly mentioned in Section 1.3.1, in these scenarios the CC analysis does not lead to biased estimates. The latter is due to the fact that the drop-out probability does not depend on the outcomes, so for each group the individuals remaining in the study at the last visit are a simple representative sub-sample of the entire group. However, we can expect a loss of precision.

For the second type of MAR drop-out mechanism, we considered scenarios in which drop-out probability depended on the last observed outcome, with subjects with higher values having a greater probability of dropping out than those with lower values or vice-versa. For MAR data, we also considered a third mechanism, including scenarios in which the drop-out probability depended on the last observed outcome and the individual's group. More precisely, two scenarios were considered: first, subjects in the treatment group had a 0.1 marginal probability of dropping out whereas those in the control group had a 0.4 probability, and within each group, the probability of dropping out was lower for individuals with higher values than for those with lower values; second, subjects in the treatment group had a 0.4 marginal drop-out probability whereas those in the control group had a 0.1 probability, and within each group, the probability of dropping out was higher for individuals with higher values than for those with lower values. For an MNAR drop-out mechanism, we considered scenarios in which subjects with higher values for the first missing outcome had either a higher or a lower probability of dropping out at each visit.

In the second and third types of MAR scenarios and in the MNAR scenario a selection bias can be expected when performing a CC analysis under $H_1(\theta \neq 0)$. Note that there is no bias in the CC analysis under H_0 ($\theta = 0$) if the drop-out mechanism is the same for both groups. The latter holds true for all mechanisms considered, except for the third MAR-type mechanism, for which a bias might be observed when performing a CC analysis under H_0.

For each mechanism, the probability of dropping out was the same at each visit, excluding baseline. For the scenarios in which drop-out probability depended only on outcome values, we assigned to each individual i and each time j, a probability p_{ij} of dropping out that depended on the considered outcome $Y_{ij'}$ (where $j' = j$ or $j' = j - 1$ depending on the scenario) through a logistic model:

$$\text{logit}(p_{ij}) = \lambda_0 + \lambda_1 Y_{ij'}.$$

Here, λ_0 and λ_1 were chosen so that they yielded a marginal probability of dropping out p=0.1 or 0.4. For the scenarios in which the drop-out probability depended on outcome values and group, a separate logistic model like the one above was used to simulate drop-outs within each group, both with equal λ_1 but with different λ_0 so as to yield different marginal drop-out probabilities within each group.

3.3.3 Analysis of the generated data sets

For each drop-out mechanism, 1000 datasets of size $n = 1000$ were generated. Four analyses were performed on each dataset to estimate θ and its variance, obtain CIs and test H_0. First, the complete data analysis was conducted before simulating drop-outs to serve as a reference. Second, after simulating drop-outs, a CC analysis was performed. The third and fourth analyses were MI analyses using the proposed imputation procedure, with $m = 5$ and $m = 20$ imputations, respectively. The imputation model in both cases was an LMM including fixed effects for the treatment group X_i, the measurement time t_j ($= j$) and their interaction $X_i t_j$, and random intercepts and slopes.

The quantities computed to summarize the results of each analysis across the 1000 simulated datasets in each scenario were: the mean of the 1000 estimates $\hat{\theta}^{(s)}$ noted as $\hat{\theta}$; the mean of the 1000 estimated standard deviations $\hat{\sigma}_{\theta}^{(s)}$ noted as $\hat{\sigma}_{\theta}$; the observed standard deviation of the estimates $\hat{\theta}^{(s)}$ noted as $\mathrm{SD}\,(\hat{\theta}^{(s)})$; the empirical coverage probability (CP) of the true value θ by the 95% confidence interval (CI) for θ; the percentage of simulations in which H_0 was rejected (i.e. the type I error rate in the case of a null treatment effect and the power otherwise); and the mean squared error (MSE) of the estimates across the simulations.

3.3.4 Results

Results for the scenarios in which MAR data were generated with drop-out probability depending solely on the individual's group are shown in Table 3.1 for the case in which subjects from the control group had a higher drop-out probability. Results for the opposite case were similar (see Table A.1 in Appendix A). As expected, no bias was introduced by the CC analysis. The MI analyses also led to unbiased estimates. In all analyses, the standard deviation of the estimator was correctly estimated and the CP was satisfactory. However, for both values of θ, the CC analysis led to a loss in precision in comparison to the complete data analysis. The MI analyses partially recovered this loss. This implied a gain in power under H_1 and an MSE about 20% smaller with MI compared to the CC analysis.

The results for the MAR scenario in which drop-out probability was inversely related to the last observed outcome are shown in Table 3.2. Under H_1 a small bias was introduced by the CC analysis which, as expected, increased with the marginal drop-out probability p. In contrast, the MI estimates were unbiased. The CC analyses yielded correct standard deviation estimates and, in the MI analyses, the standard deviation estimates were satisfactory when the marginal drop-out probability was 0.1 and slightly overestimated when this probability was 0.4. Globally, both analyses provided correct CPs, even though they were sometimes slightly higher than expected under H_1, when

Table 3.1: Estimates of treatment effect: MAR scenario in which drop-out probability depended on the individual's group, with drop-out probabilities of 0.1 and 0.4 for the treatment and control groups, respectively. Results of 1000 simulations.

θ	Analysis	$\hat{\theta}$	$\hat{\sigma}_\theta$	SD $(\hat{\theta}^{(s)})$	CP	% reject H_0	MSE
0	Complete data	-0.018	0.328	0.326	95.4	4.5	0.107
	Complete cases	-0.014	0.387	0.374	96.2	3.8	0.140
	5 imputations	-0.015	0.365	0.348	95.6	4.2	0.121
	20 imputations	-0.018	0.362	0.343	95.4	4.6	0.118
1	Complete data	1.012	0.329	0.337	94.6	86.2	0.114
	Complete cases	1.021	0.387	0.397	93.4	73.5	0.158
	5 imputations	1.014	0.368	0.363	95.7	76.4	0.132
	20 imputations	1.013	0.362	0.358	95.5	80.2	0.128

the marginal drop-out probability was 0.4. MI displayed a higher precision with respect to the CC analyses, resulting in improved power under H_1 and a slightly smaller MSE. Results for the case in which drop-out probability was positively associated with the last observed outcome were similar (see Table A.2 in Appendix A).

Results for the MAR scenario in which drop-out probability was inversely related to the last observed outcome and the control group had the highest percentage of drop-outs are shown in Table 3.3. In this context, the CC analysis introduced considerable bias under H_0 and H_1, whereas MI estimates were again unbiased. The CC analyses gave correct standard deviation estimates and the MI standard deviation estimates were satisfactory under H_0 and slightly overestimated under H_1. The CC analyses yielded very poor CPs under H_0 and H_1. On the other hand, MI generated correct CPs. MI improved precision with respect to the CC analyses, thereby resulting in significantly improved power under H_1 and much smaller MSE. Results for the scenario in which drop-out probability was positively related to the last observed outcome and the treatment group had the highest percentage of drop-outs were similar (see Table A.3 in Appendix A).

Table 3.4 shows results for the MNAR scenario according to which the probability

Table 3.2: Estimates of treatment effect: MAR scenario in which drop-out probability was inversely related to the last observed outcome. Results of 1000 simulations.

θ	p	Analysis	$\hat{\theta}$	$\hat{\sigma}_\theta$	SD $(\hat{\theta}^{(s)})$	CP	% reject H_0	MSE
0	0.1	Complete data	0.000	0.329	0.339	94.1	5.9	0.115
		Complete cases	−0.001	0.334	0.353	93.6	6.4	0.124
		5 imputations	−0.003	0.337	0.346	94.5	5.5	0.120
		20 imputations	−0.002	0.336	0.345	94.6	5.4	0.119
	0.4	Complete data	−0.008	0.329	0.323	96.1	3.9	0.104
		Complete cases	−0.015	0.381	0.363	96.2	3.8	0.132
		5 imputations	−0.006	0.390	0.361	95.7	3.7	0.131
		20 imputations	−0.007	0.384	0.347	96.2	3.8	0.120
1	0.1	Complete data	0.989	0.329	0.313	95.4	87.8	0.098
		Complete cases	0.915	0.334	0.311	95.0	80.0	0.104
		5 imputations	0.991	0.338	0.319	95.4	85.9	0.102
		20 imputations	0.991	0.336	0.318	95.5	86.4	0.101
	0.4	Complete data	1.010	0.329	0.317	95.7	88.1	0.101
		Complete cases	0.804	0.381	0.363	93.4	56.4	0.170
		5 imputations	1.005	0.391	0.350	96.9	73.1	0.122
		20 imputations	1.004	0.385	0.338	97.2	76.9	0.114

Table 3.3: Estimates of treatment effect: MAR scenario in which drop-out probability depended on the individual's group and his last observed outcome, with marginal drop-out probabilities of 0.1 and 0.4 for the treatment and control groups, respectively, and drop-out probability inversely related to the last observed outcome within each group. Results of 1000 simulations.

θ	Analysis	$\hat{\theta}$	$\hat{\sigma}_\theta$	SD $(\hat{\theta}^{(s)})$	CP	% reject H_0	MSE
0	Complete data	−0.022	0.329	0.341	93.7	6.2	0.116
	Complete cases	−1.182	0.363	0.359	9.2	90.8	1.527
	5 imputations	−0.022	0.364	0.365	94.4	5.0	0.133
	20 imputations	−0.021	0.361	0.360	94.6	5.4	0.130
1	Complete data	1.005	0.328	0.324	96.3	86.4	0.105
	Complete cases	−0.170	0.363	0.353	9.3	7.2	1.494
	5 imputations	1.012	0.366	0.343	96.2	79.2	0.117
	20 imputations	1.007	0.361	0.341	96.7	80.3	0.116

Table 3.4: Estimates of treatment effect: MNAR scenario in which drop-out probability was inversely related to the first missing outcome. Results of 1000 simulations.

θ	p	Analysis	$\hat{\theta}$	$\hat{\sigma}_\theta$	SD $(\hat{\theta}^{(s)})$	CP	% reject H_0	MSE
0	0.1	Complete data	−0.012	0.329	0.326	94.5	5.4	0.106
		Complete cases	−0.013	0.326	0.323	95.3	4.7	0.105
		5 imputations	−0.013	0.332	0.331	94.4	5.6	0.109
		20 imputations	−0.012	0.331	0.329	94.5	5.5	0.109
	0.4	Complete data	−0.006	0.329	0.336	94.5	5.4	0.113
		Complete cases	−0.011	0.360	0.365	94.9	5.0	0.133
		5 imputations	−0.002	0.374	0.366	95.2	4.4	0.133
		20 imputations	−0.006	0.368	0.356	95.6	4.4	0.126
1	0.1	Complete data	1.016	0.329	0.330	94.8	86.9	0.109
		Complete cases	0.898	0.327	0.327	93.7	78.6	0.117
		5 imputations	0.999	0.333	0.328	95.1	85.6	0.108
		20 imputations	0.999	0.332	0.326	94.8	86.1	0.106
	0.4	Complete data	1.006	0.329	0.350	94.0	84.9	0.122
		Complete cases	0.728	0.361	0.368	87.2	52.4	0.209
		5 imputations	0.946	0.374	0.365	95.6	69.1	0.136
		20 imputations	0.950	0.369	0.357	95.5	73.7	0.130

of dropping out was inversely related to the first missing outcome. Results for the opposite scenario were similar (see Table A.4 in Appendix A). Under H_1, a bias was introduced by the CC analysis that, as expected, increased with the marginal drop-out probability p. MI analyses partially corrected this bias.

3.4 Discussion

The results of the simulation study showed that the proposed imputation procedure yields unbiased coefficient estimates for MAR drop-out mechanisms and approximately unbiased variance estimates. Furthermore, the procedure displayed satisfactory CPs and controlled type I error rates under H_0. In addition to correcting the bias introduced by the CC estimator, the MI estimator led to a gain in precision resulting in lower MSEs

and improved power under H_1.

In the design of the simulation study, we underline that, in the considered scenarios, a strong variability of the random effects was assumed with respect to the variability of the residual errors. Indeed, we had $\sigma_{b_{0i}} = \sigma_{b_{1i}} = \sigma_{\varepsilon_{ij}} = 1$. This assumption led to biases of varying degrees in the CC analyses, depending on the drop-out mechanism. In some situations with real data we may expect the variability of subject-specific random effects to be weaker. Additional simulations were run considering a lower variability for the random effects ($\sigma_{b_{0i}} = \sigma_{b_{1i}} = 0.25$). As expected, the biases induced in this setting by the CC analyses were smaller (data not shown).

Concerning our imputation procedure, a sensitive issue was how to deal with the residual degrees of freedom d in step (c). This number was a parameter estimated from the data. However, it was not drawn in each imputation from an estimated asymptotic distribution and thus did not vary from imputation to imputation. Overall, no underestimation of the variance nor decrease in the empirical CP were observed in the results. However, this could be an issue with small sample studies. Another potential problem with small samples concerns the computational issues that may arise from the maximum-likelihood-based approximation of the posterior distribution of the parameters (Demirtas and Schafer, 2003).

When the MNAR scenario results were described, we indicated that the MI analysis achieved partial bias correction. That correction can be attributed to the drop-out probability in those simulations, which, although conditionally independent of the observed outcomes given the first missing outcome, was marginally correlated to the observed outcomes because of the within-subject correlations. Thus, the available outcomes taken into account in the MI analyses provided partial information about the missing outcomes, which in turn yielded the partial bias correction observed. This bias correction was enabled by our simulation study design and might not be observed in other situations.

Chapter 4

MAR regression modeling of the CIF using pseudo-values

In Chapter 2 we indicated that direct regression modeling of the cumulative incidence function (CIF) in the missing cause setting has been addressed only by Bakoyannis et al. (2010) using multiple imputation (MI), with the analysis model being the Fine and Gray model fitted by an inverse probability of censoring weighting (IPCW) approach as in Fine and Gray (1999). No other missing data approaches, such as inverse probability weighting (IPW), have yet been proposed. Moreover, there has been no assessment of the performance of the MI procedure of Bakoyannis et al. (2010) when applied to flexible modeling strategies for the CIF such as the Andersen-Klein pseudo-value approach, nor when dealing with continuous covariates. In this chapter, we propose a general framework for semi-parametric regression modeling of the CIF with missing causes of failure under the MAR assumption. More precisely, we consider two extensions of the Andersen-Klein pseudo-value approach, the details of which are given in Section 4.1.1. The first extension, presented in Section 4.1.2, is a novel approach grounded on the IPW paradigm for dealing with missing data. The second extension, described in Section 4.1.3, is the MI procedure of Bakoyannis et al. (2010) coupled with the Andersen-Klein approach. We evaluated the small-sample performances of these approaches and

compared them to the naive complete case (CC) analysis by means of an extensive simulation study, which is presented in Section 4.2. In our simulations, we considered both binary and continuous covariates. In Section 4.3, we illustrate the practical value and ease of implementation of the proposed approaches by analyzing the data from the ECOG clinical trial (cf. Section 2.1). Some elements for discussion and concluding remarks are presented in Section 4.4.

4.1 Methodology

In this chapter we omit the auxiliary covariate vector $\mathbf{W}$, assuming that all the covariates influencing the missingness process are already included in $\mathbf{X}$. Thus, the observed data are $(\tilde{T}_i, U_i, \mathbf{X}_i)$ for censored individuals, $(\tilde{T}_i, U_i, M_i, D_i, \mathbf{X}_i)$ for uncensored individuals with observed cause of failure and $(\tilde{T}_i, U_i, M_i, \mathbf{X}_i)$ for uncensored individuals with missing cause, where $\mathbf{X}_i = (X_{i1}, \ldots, X_{ip})$. The data of different individuals are assumed to be i.i.d given the covariates.

Suppose that we are interested in determining the effect of $\mathbf{X}$ on the CIF of the cause of interest ($D = 1$). We model the CIF conditional on $\mathbf{X}$, given by $F_1(t|\mathbf{X}) = P(T \leq t, D = 1|\mathbf{X})$, by means of a generalized linear model of the type described in Section 2.2.3. More precisely, we assume that

$$g\{F_1(t|\mathbf{X}_i)\} = \beta_0(t) + \sum_{h=1}^{p} \beta_h X_{ih}, \qquad t > 0, \quad i = 1, \ldots, n, \tag{4.1}$$

where g is a monotone differentiable link function, β_h is the effect of X_{ih}, the h^{th} component of $\mathbf{X}_i$, and $\beta_0(t)$ is a time-dependent intercept. Model (4.1) encompasses models such as the Fine and Gray model if g is the complementary log-log (cloglog) function and the additive model if g is the identity function (cf. Section 2.2.3). Note that the additive model is not defined near 0, because the CIF is 0 at $t = 0$. Thus, relationship (4.1) can be expected to hold only for $t \geq t_0 > 0$, where t_0 is some chosen time-point (Klein, 2006).

4.1.1 The Andersen-Klein approach

When the cause of failure is known for all patients who have failed, Klein and Andersen (2005) proposed using the pseudo-value approach to fit model (4.1) (Andersen et al., 2003). Let $t_1 < \cdots < t_K$ be the observed failure times irrespective of the cause. In the absence of censoring and with fully observed causes of failure, the indicator variables $I(T_i \leq t_k, D_i = 1)$, for $k = 1, \ldots, K$, constitute a set of fully-observed longitudinal binary outcomes for each patient $i = 1, \ldots, n$. Since $F_1(t|\mathbf{X}) = E\{I(T \leq t, D = 1)|\mathbf{X}\}$, model (4.1) can be fitted using readily available regression techniques for repeated binary data such as the GEE approach of Liang and Zeger (1986) (cf. Section 1.2.3). In contrast, when there is censoring, some of these indicators will be missing and the aforementioned methods are not suitable. The Andersen-Klein approach consists in using pseudo-values from a jackknife statistic constructed from the CIF instead of these incomplete outcomes. These pseudo-values are set as the response variable when performing regression for both censored and uncensored individuals, and actually lead to consistent estimates of the coefficients of model (4.1) as is explained further on.

To explain the construction of the pseudo-values, let Y_k denote the number of subjects at risk at time t_k, d_{1k} the number of type 1 events at time t_k and d_k the total number of events at time t_k. The Aalen-Johansen estimator of the CIF of cause 1 at time t, is given by

$$\hat{\theta}(t) = \sum_{t_k \leq t} \frac{d_{1k}}{Y_k} \prod_{t_l < t_k} \frac{Y_l - d_l}{Y_l}. \tag{4.2}$$

This estimator is approximately unbiased under the usual independent censoring assumption (Andersen et al., 1993, § IV.4), and particularly when the censoring time is stochastically independent of the failure time and the cause of failure. The pseudo-value, or pseudo-observation, of individual i at time t is is denoted by $\hat{\theta}_i(t)$ and defined as the weighted difference between the whole sample estimator of the CIF, $\hat{\theta}(t)$, and the leave-one out estimator, $\hat{\theta}^{-i}(t)$, obtained by excluding individual i from the sample:

$$\hat{\theta}_i(t) = n\hat{\theta}(t) - (n-1)\hat{\theta}^{-i}(t), \qquad t > 0, \quad i = 1, \ldots, n.$$

The pseudo-observation of an individual represents his contribution to the estimate of the CIF at time t on the sample of size n. In the absence of censoring, $\hat{\theta}_i(t)$ reduces to the indicator variable $I(T_i \leq t, D_i = 1)$; with right-censoring, $\hat{\theta}_i(t)$ is a good approximation of this indicator (see Andersen and Perme, 2010, Figures 6 and 7). Note that in the latter case, pseudo-observations are continuous variables. Moreover, the pseudo-observations at a fixed time-point t exhibit the following two properties:

(P1) the $\hat{\theta}_i(t)$'s are approximately i.i.d., and

(P2) the $\hat{\theta}_i(t)$'s are conditionally unbiased given the covariates, that is,

$$E\{\hat{\theta}_i(t)|\mathbf{X}_i\} = F_1(t|\mathbf{X}_i) + o_p(1).$$

These two properties were established by Graw et al. (2009, Lemma 2), and hold under the following conditions:

(i) the censoring time C is independent of T, D and X, and

(ii) $t < t^*$ where t^* is such that the survival function of the censoring time, $G(c) := P(C > c)$, satisfies $G(t^*) > \upsilon$ for a fixed $\upsilon > 0$.

Properties (P1) and (P2) make pseudo-observations suitable to use as alternative outcomes for regression purposes when there is censoring. Indeed, suppose that pseudo-observations are set as the outcome variables for both censored and uncensored individuals. Then the GEE approach of Liang and Zeger (1986), described in Section 1.2.3, can be used to fit model (4.1), with properties (P1) and (P2) guaranteeing the consistency and asymptotic normality of the estimates obtained in this way (Graw et al., 2009, Theorem 2).

To fit the model using GEE, pseudo-observations must be calculated at a grid of time-points $\tau_1 < \cdots < \tau_S$ so that the outcome of each individual is multivariate, given by $\hat{\boldsymbol{\theta}}_{\boldsymbol{i}} = \{\hat{\theta}_i(\tau_1), \ldots, \hat{\theta}_i(\tau_S)\}$. Although a single time-point would be enough to identify the coefficients of model (4.1), it is recommended to use several time-points

because more efficient estimates are obtained and the function $\beta_0(t)$ is identified as well (Graw et al., 2009). The ideal choice for the grid would be the set of observed failure times $t_1, \ldots, t_K$ because pseudo-observations only change at these times. However, Andersen and Klein suggest using between 5 to 10 time-points, which are taken to be equidistant on the event-time scale (Klein and Andersen, 2005; Andersen and Klein, 2007). Since they showed empirically that this choice suffices to obtain good estimates of regression coefficients and that little is gained by using more time-points, we follow their recommendation.

Additionally, the use of a grid with few time-points makes it reasonable to directly estimate the time-specific intercepts $\beta_0(\tau_s)$ parametrically, as was suggested originally (Klein and Andersen, 2005; Klein et al., 2008; Andersen and Perme, 2010). Indeed, if the grid includes a large number of time-points, too many parameters would have to be estimated and alternative approaches to model the function $\beta_0(t)$, such as smoothing techniques, would be more suitable (Andersen and Perme, 2010; Andersen and Klein, 2007). To estimate the time-specific intercepts parametrically it suffices to include indicator variables in $\mathbf{X}_i$, resulting in augmented covariate vectors $\mathbf{X}_i^{(s)} = \{I(\tau_q = \tau_s) : q = 1, \ldots, S; \mathbf{X}_i\}$ for $s \in \{1, \ldots, S\}$. External time-dependent covariates, as defined in Section 2.2.1, can also be included by incorporating their values at each time-point of the grid in the vectors $\mathbf{X}_i^{(s)}$.

The representation of model (4.1) using pseudo-observations and the augmented covariate vectors is

$$g[E\{\hat{\theta}_i(\tau_s)|\mathbf{X}_i^{(s)}\}] = \boldsymbol{\beta}'\mathbf{X}_i^{(s)}, \qquad s = 1, \ldots, S \quad i = 1, \ldots, n,$$

where $\boldsymbol{\beta} = \{\beta_0(\tau_1), \ldots, \beta_0(\tau_S), \beta_1, \ldots, \beta_p\}$ is the parameter vector. Following the theory of GEE outlined in Section 1.2.3, this marginal model may be fitted by solving the following generalized estimating equation:

$$U(\boldsymbol{\beta}) = \sum_{i=1}^{n} \left\{ \frac{\partial}{\partial \boldsymbol{\beta}} g^{-1}(\boldsymbol{\beta}'\mathbf{X}_i^{(\bullet)}) \right\}' \mathbf{V}_i^{-1} \{\hat{\boldsymbol{\theta}}_{\boldsymbol{i}} - g^{-1}(\boldsymbol{\beta}'\mathbf{X}_i^{(\bullet)})\} = \sum_{i=1}^{n} U_i(\boldsymbol{\beta}) = 0, \qquad (4.3)$$

where $g^{-1}(\boldsymbol{\beta}'\mathbf{X}_i^{(\bullet)})$ is short for the S-vector with elements $g^{-1}(\boldsymbol{\beta}'\mathbf{X}_i^{(s)})$ and $\mathbf{V}_i$ is the variance-covariance matrix of $\hat{\boldsymbol{\theta}}_{\boldsymbol{i}}$, which must be modeled and/or estimated separately as explained in Section 1.2.3.

A first possibility for modeling $\mathbf{V}_i$ is to assume independence between the elements of $\hat{\boldsymbol{\theta}}_i$, i.e. take the identity matrix to be the working correlation matrix. Another possibility arises from the fact that the $\hat{\theta}_i(\tau_s)$'s are binary in the absence of censoring. The covariance between two elements of $\hat{\boldsymbol{\theta}}_i$ in that context suggests modeling $\mathbf{V}_i$ as the 'exact' matrix, in which

$$\text{cov}\{\hat{\theta}_i(\tau_{s_1}), \hat{\theta}_i(\tau_{s_2})\} = F_1(\tau_{s_1}|\mathbf{X}_i)\{1 - F_1(\tau_{s_2}|\mathbf{X}_i)\}, \qquad \text{for } \tau_{s_1} \leq \tau_{s_2}.$$

In this case $\mathbf{V}_i$ depends on $\boldsymbol{\beta}$. A third possibility is to use the usual product-moment correlation matrix of the $\hat{\theta}_i(\tau_s)$'s as a plug-in estimate for $\mathbf{V}_i$. Klein and Andersen (2005) performed a simulation study comparing each of these possibilities and their results showed that their method is robust to the path chosen. Thus, they suggest using the identity matrix as the working correlation matrix. We follow this recommendation in our simulation study and in the application of our methods to the ECOG clinical trial.

The consistency and asymptotic normality of the solution $\hat{\boldsymbol{\beta}}$ of equation (4.3) have been established (Graw et al., 2009, Theorem 2) and, as mentioned above, follow from properties (P1) and (P2) of the pseudo-observations. In particular, the asymptotic unbiasedness of equation (4.3) follows from the conditional unbiasedness of the pseudo-values given the covariates, property (P2), under the assumption that model (4.1) is correct. In the case of the additive model, if the model is assumed to hold for $t \geq t_0 > 0$, then the asymptotic unbiasedness of (4.3) will hold only if $\tau_s \geq t_0$ for all s. Thus, estimates will still be consistent if the pseudo-values are calculated at a grid of time-points starting after t_0, i.e. if $\tau_1 \geq t_0$.

The variance of $\hat{\boldsymbol{\beta}}$ can be consistently estimated by using the sandwich estimator

$$\hat{\text{var}}(\hat{\boldsymbol{\beta}}) = I(\hat{\boldsymbol{\beta}})^{-1}\hat{\text{var}}\{U(\boldsymbol{\beta})\}I(\hat{\boldsymbol{\beta}})^{-1}, \tag{4.4}$$

where

$$I(\boldsymbol{\beta}) = \sum_{i=1}^{n}\{\frac{\partial}{\partial\boldsymbol{\beta}}g^{-1}(\boldsymbol{\beta}'\mathbf{X}_i^{(\bullet)})\}'\mathbf{V}_i^{-1}\{\frac{\partial}{\partial\boldsymbol{\beta}}g^{-1}(\boldsymbol{\beta}'\mathbf{X}_i^{(\bullet)})\} \quad \text{and} \quad \hat{\text{var}}\{U(\boldsymbol{\beta})\} = \sum_{i=1}^{n}U_i(\hat{\boldsymbol{\beta}})U_i(\hat{\boldsymbol{\beta}})'.$$

Alternatively, a jackknife variance estimator can be used (Yan and Fine, 2004). The approximate jackknife (AJ) variance estimator is the most recommended in this setting and is also the least burdensome computationally (Klein et al., 2008). Otherwise, a non-parametric bootstrap procedure has been suggested, where the approximately independent pseudo-observations $\hat{\boldsymbol{\theta}}_{\boldsymbol{i}}$ would be resampled (Andersen et al., 2003).

4.1.2 Inverse probability weighted pseudo-values

When the cause of failure is missing for some individuals, assumptions about the missingness mechanism are necessary to identify $F_1(t|\mathbf{X})$. We consider the following MAR-type assumption about the mechanism driving missingness:

$$P(M=0|\mathbf{X}, T\leq t, U=0, D) = P(M=0|\mathbf{X}, T\leq t, U=0) =: \pi_t(\mathbf{X}), \quad t\geq 0. \tag{4.5}$$

That is, at each time t, the probability that the cause of failure is observed among individuals who have already failed is independent of the cause of failure when conditioning on covariates. Under this assumption, the following relation holds for all $t \geq 0$:

$$F_1(t|\mathbf{X}) = \frac{\tilde{F}_1(t|\mathbf{X})}{\pi_t(\mathbf{X})}, \tag{4.6}$$

where $\tilde{F}_1(t|\mathbf{X}) = P(T \leq t, D = 1, M = 0|\mathbf{X})$. Therefore, under assumption (4.5), $F_1(t|\mathbf{X})$ becomes identifiable as the quotient of two identifiable quantities, $\tilde{F}_1(t|\mathbf{X})$ and

$\pi_t(\mathbf{X})$. Indeed, $\tilde{F}_1$ is the CIF of the cause of interest when all failures with missing cause are regarded as due to a cause other than the cause of interest, i.e. $\tilde{F}_1(t|\mathbf{X}) = P(T \leq t, \tilde{D} = 1|\mathbf{X})$ where $\tilde{D}$ is defined for uncensored individuals as $\tilde{D} = 2$ if $M = 1$, and $\tilde{D} = D$ if $M = 0$. Hence, $\tilde{F}_1$ is identifiable. In fact, this function could be modeled from a dataset where the latter recoding procedure has been performed, i.e. where $\tilde{D}$ has been computed for uncensored individuals. Such an approach would correspond to the extra state (ES) analysis, one of the aforementioned ad-hoc techniques to deal with missing causes (cf. Section 2.3.1). Furthermore, the quantities $\pi_t(\mathbf{X})$ are also identifiable and can be estimated from the original data because M is fully observed among individuals who have failed.

Motivated by relation (4.6), we propose the following estimation procedure to fit model (4.1) under assumption (4.5):

1. Determine a grid of time-points $\tau_1 < \ldots < \tau_S$ equidistant on the event-time scale.

2. For each time-point τ_s, obtain a plug-in estimate of $\pi_{\tau_s}(\mathbf{X})$ as follows:

   (i) If $\mathbf{X}$ contains only a few discrete finite-ranged covariates, then for each value $\mathbf{x}$ of $\mathbf{X}$, $\pi_{\tau_s}(\mathbf{x})$ can be estimated by the proportion of failures with known cause among failures occurring before τ_s for individuals with $\mathbf{X}_i = \mathbf{x}$:

   $$\hat{\pi}_{\tau_s}(\mathbf{x}) = \frac{\sum_{i=1}^{n}(1 - M_i) \times I(U_i = 0 \wedge \tilde{T}_i \leq \tau_s \wedge \mathbf{X}_i = \mathbf{x})}{\sum_{i=1}^{n} I(U_i = 0 \wedge \tilde{T}_i \leq \tau_s \wedge \mathbf{X}_i = \mathbf{x})}. \tag{4.7}$$

   The grid of time-points can be modified if required, to ensure that $\hat{\pi}_{\tau_1}(\mathbf{x}) < \infty$ (denominator of (4.7) different from 0) and $\hat{\pi}_{\tau_1}(\mathbf{x}) > 0$ for every value $\mathbf{x}$ of $\mathbf{X}$.

   (ii) If $\mathbf{X}$ contains continuous or many covariates, $\pi_{\tau_s}(\mathbf{X})$ can be modeled using the data of uncensored patients who failed before τ_s by means of a parametric model including the components of $\mathbf{X}$ as predictors. For example, a logistic model can be used:

   $$\text{logit}\{\pi_{\tau_s}(\mathbf{X})\} = \boldsymbol{\delta}_s' \mathbf{r}_s(\mathbf{X}), \tag{4.8}$$

where $\mathbf{r}_s(\mathbf{X})$ is a vector including the components of $\mathbf{X}$ and possibly interaction terms.

3. Code the failures with a missing cause as due to a cause other than the cause of interest (i.e. set $D_i = 2$ if $U_i = 0$ and $M_i = 1$) to obtain a new dataset without missing causes of failure among uncensored patients.

4. Compute the pseudo-observations $\tilde{\boldsymbol{\theta}}_i = \{\tilde{\theta}_i(\tau_1), \ldots, \tilde{\theta}_i(\tau_S)\}$, $i \in \{1, \ldots, n\}$, based on the Aalen-Johansen estimator of $\tilde{F}_1$, the CIF of cause 1 in the new dataset, as in Section 4.1.1.

5. Compute the *inverse probability weighted pseudo-values* (IPWpv) as follows:

$$\hat{\theta}_i(\tau_s) = \frac{\tilde{\theta}_i(\tau_s)}{\hat{\pi}_{\tau_s}(\mathbf{X}_i)}, \quad s = 1, \ldots, S, \quad i = 1, \ldots, n.$$

6. Solve estimating equation (4.3), where the $\hat{\boldsymbol{\theta}}_{\boldsymbol{i}}$'s are now the IPWpv's from Step 5, to obtain an estimate $\hat{\boldsymbol{\beta}}$ of the regression coefficients.

The theoretical validity of this procedure follows from the following lemma.

Lemma 4.1.1 *Fix $t > 0$. Define the IPWpv's at time t as $\hat{\theta}_i(t) = \tilde{\theta}_i(t)/\hat{\pi}_t(\mathbf{X}_i)$, for $i = 1, \ldots, n$, where the $\tilde{\theta}_i(t)$'s are the pseudo-observations obtained from the Aalen-Johansen estimator of $\tilde{F}_1$, and $\hat{\pi}_t(\mathbf{X}) \xrightarrow{p} \pi_t(\mathbf{X}) > 0$. Assume that (4.5) and conditions (i) and (ii) of Section 4.1.1 hold. Then properties (P1) and (P2), where the $\hat{\theta}_i(t)$'s are now IPWpv's, still hold.*

Proof To prove (P1) it suffices to note that the ordinary (unweighted) pseudo-observations $\tilde{\theta}_i(t)$, $i \in \{1, \ldots, n\}$, are approximately i.i.d. variables and the $\mathbf{X}_i$'s are i.i.d. The IPWpv's, being functions of these quantities, are approximately i.i.d. as well.

The proof of (P2) follows from the conditional unbiasedness given the covariates of the ordinary (unweighted) pseudo-observations $\tilde{\theta}_i(t)$, $i\{1, \ldots, n\}$, relation (4.6), Slutsky's Theorem and the properties of convergence in probability (see van der Vaart, 2000), noting that the weights $1/\pi_t(\mathbf{X}_i)$ are bounded:

$$\begin{aligned} E\{\hat{\theta}_i(t)|\mathbf{X}_i\} &= E\{\tilde{\theta}_i(t)/\hat{\pi}_t(\mathbf{X}_i)|\mathbf{X}_i\} \\ &= 1/\hat{\pi}_t(\mathbf{X}_i) \times E\{\tilde{\theta}_i(t)|\mathbf{X}_i\} \\ &= \{1/\pi_t(\mathbf{X}_i) + o_p(1)\} \times \{\tilde{F}_1(t|\mathbf{X}_i) + o_p(1)\} \\ &= \{1/\pi_t(\mathbf{X}_i)\} \times \tilde{F}_1(t|\mathbf{X}_i) + o_p(1) \\ &= F_1(t|\mathbf{X}_i) + o_p(1). \end{aligned}$$

∎

The key consequence of Lemma 4.1.1 is that the estimates obtained from the procedure above are consistent and asymptotically normal. In fact, once (P1) and (P2) have been established for IPWpv, the proof of these asymptotic properties is the same as that of Graw et al. (2009, Theorem 2) (see also Graw et al., 2008, Theorem 3 and Appendix B).

The variance estimators available for the Andersen-Klein approach are not suitable to estimate the variance of $\hat{\boldsymbol{\beta}}$ obtained via IPWpv. Indeed, the sandwich estimator (4.4), where the $\hat{\boldsymbol{\theta}}_{\boldsymbol{i}}$'s are now IPWpv, would regard the weights $1/\pi_t(\mathbf{X})$ as known even though they are estimated, thus neglecting their variability. This is also the case for the bootstrap procedure suggested by Andersen et al. (2003) because, when resampling the IPWpv, the weights $1/\pi_t(\mathbf{X})$ remain fixed. Alternatively, a bootstrap procedure in which individuals are resampled from the original data seems appropriate because it results in updated weights at each resample. Hence, the variability in the estimated weights is reflected in the resampling scheme and the variance can be correctly estimated.

More precisely, this bootstrap procedure consists in sampling individuals from the original data with replacement to obtain a new sample of size n. This is repeated R times to obtain R datasets to which model (4.1) is then fitted by following the IPWpv procedure described above. Hence, R estimates of the regression coefficients, $\hat{\boldsymbol{\beta}}^{(1)}, \ldots, \hat{\boldsymbol{\beta}}^{(R)}$, are obtained. The bootstrap variance estimator is obtained by the sample variance of the R estimates produced, i.e. $\hat{\text{var}}(\hat{\boldsymbol{\beta}}) = \frac{1}{R-1}\sum_{r=1}^{R}(\hat{\boldsymbol{\beta}}^{(r)} - \bar{\boldsymbol{\beta}})(\hat{\boldsymbol{\beta}}^{(r)} - \bar{\boldsymbol{\beta}})'$ where

$\bar{\boldsymbol{\beta}}$ is their mean. R should be at least 100, but preferably larger; modern computers can easily perform several hundred or thousand replications. In our simulation study, we assess the impact of accounting for the variability in the estimated weights at different levels of missingness by comparing the bootstrap estimator to the (readily available) sandwich and AJ estimators based on IPWpv.

4.1.3 Multiple imputation

Bakoyannis et al. (2010) proposed an MI approach to fit the Fine and Gray model by IPCW when there are missing causes of failure. In principle, the same MI method can be used to extend the Andersen-Klein approach to the missing cause setting. We explore this approach in the simulation study in the next section and compare it to the IPWpv approach. In this section we briefly describe the MI approach of Bakoyannis et al. (2010).

Recall from Section 3.1 that the first step in MI consists in building the imputation model. In the missing cause setting, a model for $\Pi(\mathbf{X}, T) := P(D = 1|M = 1, U = 0, \mathbf{X}, T)$ must be built. MAR is equivalent to the assumption that this probability does not depend on the missingness indicator M. Indeed, assuming MAR, we have:

$$\begin{aligned}\Pi(\mathbf{X}, T) &= \frac{P(M = 1|D = 1, U = 0, \mathbf{X}, T) \times P(D = 1|U = 0, \mathbf{X}, T)}{P(M = 1|U = 0, \mathbf{X}, T)} \\ &= P(D = 1|U = 0, \mathbf{X}, T).\end{aligned}$$

The proof of the other implication of the equivalence is analogous. Hence, under MAR, the probability of failure from the cause of interest is the same for individuals with an observed and a missing cause, i.e. $\Pi(\mathbf{X}, T) = P(D = 1|M = 0, U = 0, \mathbf{X}, T)$. This means that a model for $\Pi(\mathbf{X}, T)$ can be constructed from the individuals with an observed cause. To this end, one can use a logistic regression model of the form

$$\text{logit}\{\Pi(\mathbf{X}, T)\} = \mathbf{h}(\mathbf{X}, T)'\boldsymbol{\gamma}, \tag{4.9}$$

where $\mathbf{h}(\mathbf{X}, T)$ is a vector including $\mathbf{X}$, T, and possibly interaction terms and higher order polynomials (Lu and Tsiatis, 2001; Bakoyannis et al., 2010). Let $\hat{\boldsymbol{\gamma}}$ and $\hat{\text{var}}(\hat{\boldsymbol{\gamma}})$ be the estimates of the parameter vector of the imputation model and its variance-covariance matrix, respectively, obtained by fitting this model to the individuals with an observed cause.

The second step consists in imputing the missing causes $m > 1$ times by drawing values from the imputation model to obtain m completed datasets. The procedure for the l^{th} imputation of the missing causes, $l \in \{1, \ldots, m\}$, consists of the following steps:

(a) Draw a vector $\boldsymbol{\gamma}^{(l)}$ from the normal distribution with mean $\hat{\boldsymbol{\gamma}}$ and variance $\hat{\text{var}}(\hat{\boldsymbol{\gamma}})$.

(b) For each patient i who has failed ($U_i = 0$) and with missing cause of failure ($M_i = 1$), calculate the linear predictor:

$$\eta_i^{(l)} = \mathbf{h}(\mathbf{X}_i, T_i)'\boldsymbol{\gamma}^{(l)}.$$

(c) Calculate the probability of failure from the cause of interest by applying the inverse logit transformation, $\Pi^{(l)}(\mathbf{X}_i, T_i) = e^{\eta_i^{(l)}}/(1 + e^{\eta_i^{(l)}})$, and impute the missing cause by drawing from a Bernoulli distribution with probability of success $\Pi^{(l)}(\mathbf{X}_i, T_i)$, i.e. set $D_i^{(l)} = 1$ in the event of success and $D_i^{(l)} = 2$ otherwise.

This procedure is proper as defined in Section 3.1 (Rubin, 1987).

The third step requires fitting the analysis model (i.e. the model of initial interest) to each of the m completed datasets by applying an appropriate complete data method. In our case, we fit model (4.1) to each dataset by applying the Andersen-Klein approach and obtain m estimates of $\boldsymbol{\beta}$, $\hat{\boldsymbol{\beta}}^{(1)}, \ldots, \hat{\boldsymbol{\beta}}^{(m)}$, and m sandwich variance estimates, $\hat{\text{var}}(\hat{\boldsymbol{\beta}})^{(1)}, \ldots, \hat{\text{var}}(\hat{\boldsymbol{\beta}})^{(m)}$. In the fourth and last step these estimates are combined using the formulas of Rubin (1987) presented in Section 3.1, yielding the MI coefficient and variance estimates: $\hat{\boldsymbol{\beta}} = \frac{1}{m}\sum_{l=1}^{m} \hat{\boldsymbol{\beta}}^{(l)}$ and $\hat{\text{var}}(\hat{\boldsymbol{\beta}}) = \hat{\mathbf{W}} + (1 + m^{-1})\,\hat{\mathbf{B}}$, respectively. Here, $\hat{\mathbf{W}}$ is the arithmetic mean of the variance estimates across imputations and $\hat{\mathbf{B}}$ is

the sample variance of the m coefficient estimates. Hypothesis tests and CIs may be constructed as described in Section 3.1.

In practice it is very likely that the imputation model is misspecified, leading to biased estimates. However, as shown empirically by Bakoyannis et al. (2010), inclusion of interactions and higher order terms in the imputation model will reduce the bias in coefficient estimates. To correct the bias in variance estimates, Bakoyannis et al. (2010) suggest using a bootstrap estimator for the variances of the imputation model's parameters. However, if the imputation and analysis models are uncongenial, some bias in the variance estimates may persist.

4.2 Simulation study

To evaluate the small-sample performance of the estimators under consideration, we first focused on estimating the effect of a binary covariate on the CIF of the cause of interest, F_1. The purpose of this study (Sections 4.2.1-4.2.3) was twofold: (i) to study and compare the properties of regression coefficient estimates obtained via a CC analysis (cf. Section 2.3.1) and the IPWpv and MI approaches in terms of bias and relative efficiency, and (ii) to evaluate the variance estimators available for each approach in terms of bias with respect to the observed (Monte Carlo) variance of the estimates, and particularly to compare three possible variance estimators for the IPWpv approach. In a second part, presented in Section 4.2.4, we considered a continuous covariate and studied the performance of each approach in terms of bias when estimating its effect on F_1.

4.2.1 Data generation

Let X be a binary covariate. To generate data, we fixed the baseline prevalence of the cause of interest, $p = F_1(\infty|X = 0) = P(D = 1|X = 0)$, at $p = 0.5$ and considered two submodels of model (4.1) for the cause of interest. First, we considered the Fine

and Gray model, which corresponds to the cloglog link function, by generating data according to the following specification (Bajorunaite, 2003):

$$F_1(t|X) = 1-\{1-p(1-e^{-t})\}^{\exp(\beta^{FG}X)}, \quad F_2(t|X) = (1-p)^{\exp(\beta^{FG}X)}(1-e^{-t\exp(\beta^{FG}X)}),$$

with the effect of X on F_1, fixed at $\beta^{FG} = 0.5$. With this model, the prevalence of the cause of interest in the group defined by $X = 1$ is $F_1(\infty|X = 1) = 1-(1-p)^{\exp(\beta^{FG})} \approx 0.68$. Second, we considered the additive model, which corresponds to the identity link, by generating data according to the following specification:

$$F_1(t|X) = p(1 - e^{-t}) + \beta^{AD}X, \quad F_2(t|X) = (1 - p - \beta^{AD}X)(1 - e^{-t}),$$

with the effect of X on F_1 fixed at $\beta^{AD} = 0.15$. Here, the prevalence of the cause of interest among those with $X = 1$ is $F_1(\infty|X = 1) = p + \beta^{AD} = 0.65$.

We generated datasets of size $n = 200$ and 400 and performed 10000 replications. In each dataset, the binary covariate X was balanced. Failure time and cause of failure for each patient were generated from each of the models above by first drawing the cause of failure from a Bernoulli distribution with success probability $P(D = 1|X)$ as given by the model, and then drawing a failure time according to the conditional probabilities $P(T \leq t|X, D = k) = F_k(t|X)/P(D = k|X)$, $k \in \{1, 2\}$, using the inverse transformation method. For the additive model, the latter is not straightforward because the additive relationship can hold only for $t \geq t_0 > 0$ where t_0 is some chosen time-point. Details of the procedure used are given in Section B.1.1 of Appendix B.

Censoring was superimposed to reach either 25% or 50% censoring. Two types of censoring were considered: uniform censoring on the interval $[a, b]$, where a was given by the first quartile of the event times and b was determined empirically to obtain the targeted percentage of censoring; and administrative censoring, where the time of censoring was given by the third quartile or the median of the event times so that all events occurring after that time were censored.

We assigned uncensored individuals a missing cause of failure with a probability determined by a logistic model of the form

$$\text{logit}\big\{P(M=1|T=t,X=x,U=0)\big\}=\alpha_0+\alpha_1 t+\alpha_2 x. \tag{4.10}$$

Different triplets $(\alpha_0,\alpha_1,\alpha_2)$ led to different types of MAR missingness mechanisms. The mechanisms we considered and the labels that will be used to refer to them henceforth are shown in Table 4.1. Parameter α_0 was determined empirically for each scenario to obtain the targeted global percentage of missing causes of failure among uncensored individuals (10%, 20%, 30% or 40%).

Table 4.1: Mechanisms for generating missing causes of failure using logistic model (4.10), which includes the follow-up time T and the covariate X as predictors.

Label	Type	α_1	α_2	Description
MCAR	MCAR	0	0	Constant missingness probability
MARX+	MAR	0	2	Greater missingness probability if $X=1$
MARX-	MAR	0	-2	Smaller missingness probability if $X=1$
MART+	MAR	2	0	Greater missingness probability if longer follow-up
MART-	MAR	-2	0	Smaller missingness probability if longer follow-up
MARXT+	MAR	1	-3	Greater missingness probability if longer follow-up or $X=0$
MARXT-	MAR	-1	3	Smaller missingness probability if longer follow-up or $X=0$

α_1 and α_2 are the effects of T and X, respectively.

4.2.2 Analysis of the generated datasets

For each generated dataset, we performed an analysis of the complete censored data (CCD), that is, before simulating missing causes of failure, for reference. Regression estimates were thus obtained via the Andersen-Klein approach (Section 4.1.1). After simulating missing causes, regression coefficient estimates were obtained (i) from a CC analysis via the Andersen-Klein approach, (ii) from an analysis based on the entire incomplete dataset via IPWpv using the inverse of estimator (4.7) to estimate the

weights (Section 4.1.2), and (iii) from an analysis based on the entire incomplete dataset via MI with $m = 10$ (Section 4.1.3) using a logistic imputation model including X, T and their interaction. Note that this implied a misspecified imputation model (see Bakoyannis et al., 2010). However, for the sake of computing time, we did not use a bootstrap procedure as suggested by Bakoyannis et al. (2010) to correct the potential bias in the variance estimates due to model misspecification. In fact, our results showed that the use of bootstrap was not necessary in our simulation setting (see below).

Following the recommendations of Klein and Andersen (2005), in all cases the working correlation matrix was the identity matrix and the grid of time-points at which pseudo-values were calculated was the set of deciles of the event times of uncensored individuals, excluding the first two (i.e. quantiles 0.3 to 0.9 in steps of 0.1) for a total of 7 time-points. The latter followed from results of a preliminary simulation study where we examined the bias of the estimated CIF at each time-point and of regression coefficients in a CCD analysis.

Variance estimates in the CCD and CC analyses were obtained from the sandwich estimator (4.4). For each completed dataset in the MI analyses, the variance was also estimated using (4.4) and the variance of the MI estimator was obtained from the usual multiple imputation variance estimator, combining between- and within-imputation variances. For the IPWpv approach, several variance estimators were considered for comparison: the bootstrap estimator described in Section 4.1.2 with $R = 100$ and the sandwich and AJ estimators based on IPWpv.

The simulation study was performed in `R`. (Unweighted) pseudo-value calculation was carried out using the function *jackknife.competing.risks* of the `R` *prodlim* package (Gerds, 2011) and estimating equations were solved using the function *geese* of the `R` *geepack* package (Yan, 2002; Halekoh et al., 2006; Yan and Fine, 2004). The sandwich variance estimator was directly implemented for the Fine and Gray and additive models, using the *bdsmatrix* function from `R` *bdsmatrix* package to reduce memory use (Therneau, 2011). The AJ variance estimator was obtained directly from the function *geese*. The bootstrapping scheme was implemented by using the `R` *boot* package (Canty

and Ripley, 2011; Davison and Hinkley, 1997), with the 'multicore' option to accelerate calculations. The `R` code for data generation and analysis was provided in the Supplementary Material of the corresponding published manuscript (Moreno-Betancur and Latouche, 2013). We encountered some problems when analyzing data with high percentages of censoring or missing causes of failure. Details are given in Section B.1.2 of Appendix B.

4.2.3 Simulation results

Figures 4.1 and 4.2 show simulation results for the Fine and Gray and additive models, respectively, for the scenario with $n = 200$, 50% uniform censoring and selected missingness mechanisms. Similar figures showing the results for other missingness mechanisms, $n = 400$, 25% censoring and administrative censoring are provided in Section B.2 of Appendix B.

Each figure presents the evolution of three measures for each estimator as the percentage of missing causes increased. Target values are represented by dotted lines. The first column shows the mean relative bias (MRB) of coefficient estimates with respect to the real value of the parameter across simulations (target value was 0). The relative bias for each replication was calculated as $(\beta^{FG} - \hat{\beta}^{FG})/\beta^{FG}$ and $(\beta^{AD} - \hat{\beta}^{AD})/\beta^{AD}$ for the Fine and Gray and additive models, respectively. The second column shows the square root of the mean squared error (MSE) of estimates across simulations. The CCD analysis sets a lower bound in terms of MSE because estimates are less precise with missing data. Hence, the target was set at that level for this measure. The third column shows the estimated coverage probability (CP) of the 95% confidence interval (CI) with end-points $\hat{\beta}^{FG} \pm 1.96 \times \sqrt{\hat{\text{var}}(\hat{\beta}^{FG})}$ and $\hat{\beta}^{AD} \pm 1.96 \times \sqrt{\hat{\text{var}}(\hat{\beta}^{AD})}$ for the Fine and Gray and additive models, respectively. This probability is estimated as the percentage of times the CI contains the real value of the parameter (β^{FG} or β^{AD}) across simulations (the target was the nominal level of the CI, 95%). For the IPWpv analysis, only the CI built using the bootstrap variance estimator was analyzed, as the

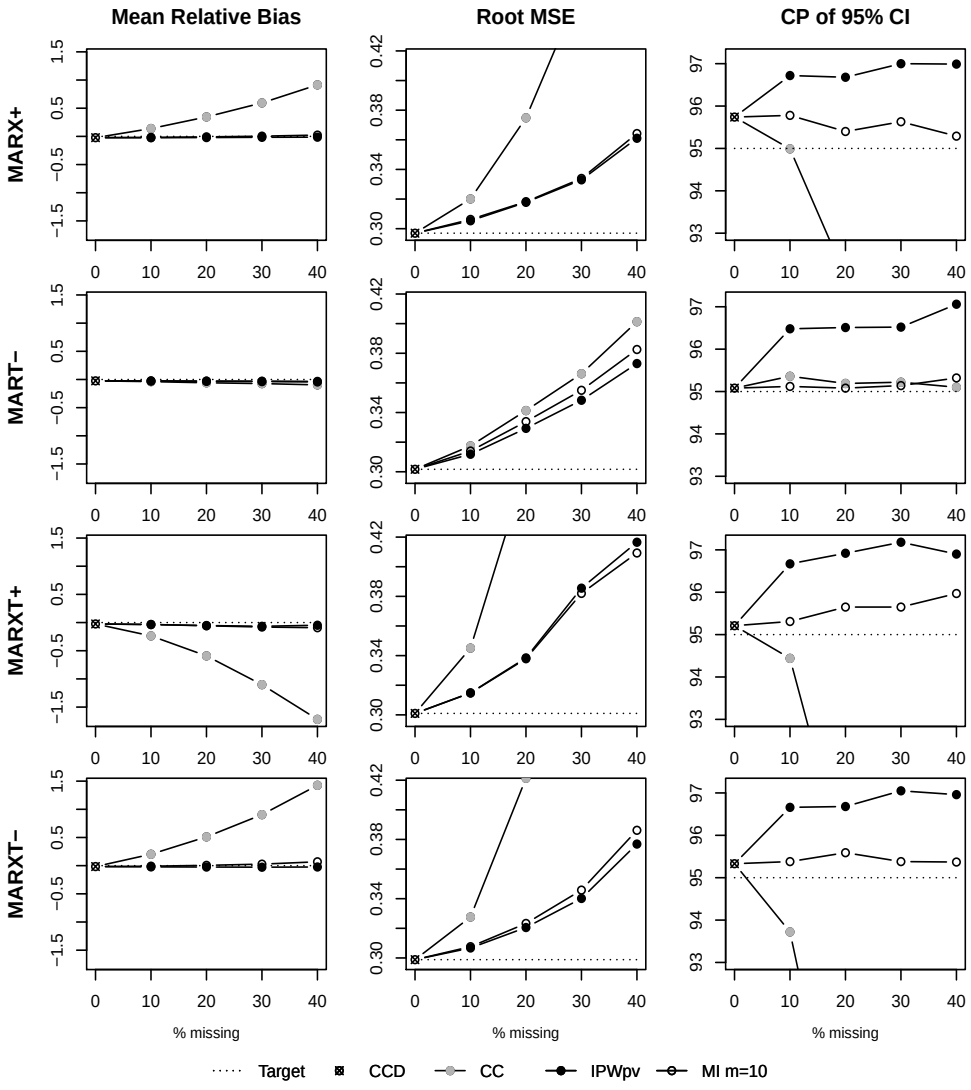

Figure 4.1: Simulation results for coefficient estimation in the Fine and Gray model with a binary covariate, $n = 200$ and 50% uniform censoring for selected missingness mechanisms. Estimates obtained via a complete case analysis (CC), the proposed IPWpv approach (IPWpv) and multiple imputation with $m = 10$ imputations (MI m=10) are compared. For each mechanism and each analysis, the mean relative bias, the square root of the mean squared error (MSE) and the estimated coverage probability (CP) of the 95% confidence interval (CI) are plotted against the percentage of missing causes. With no missing causes, all analyses coincide with the complete censored data analysis (CCD) plotted at 0%. Results are based on 10000 replications.

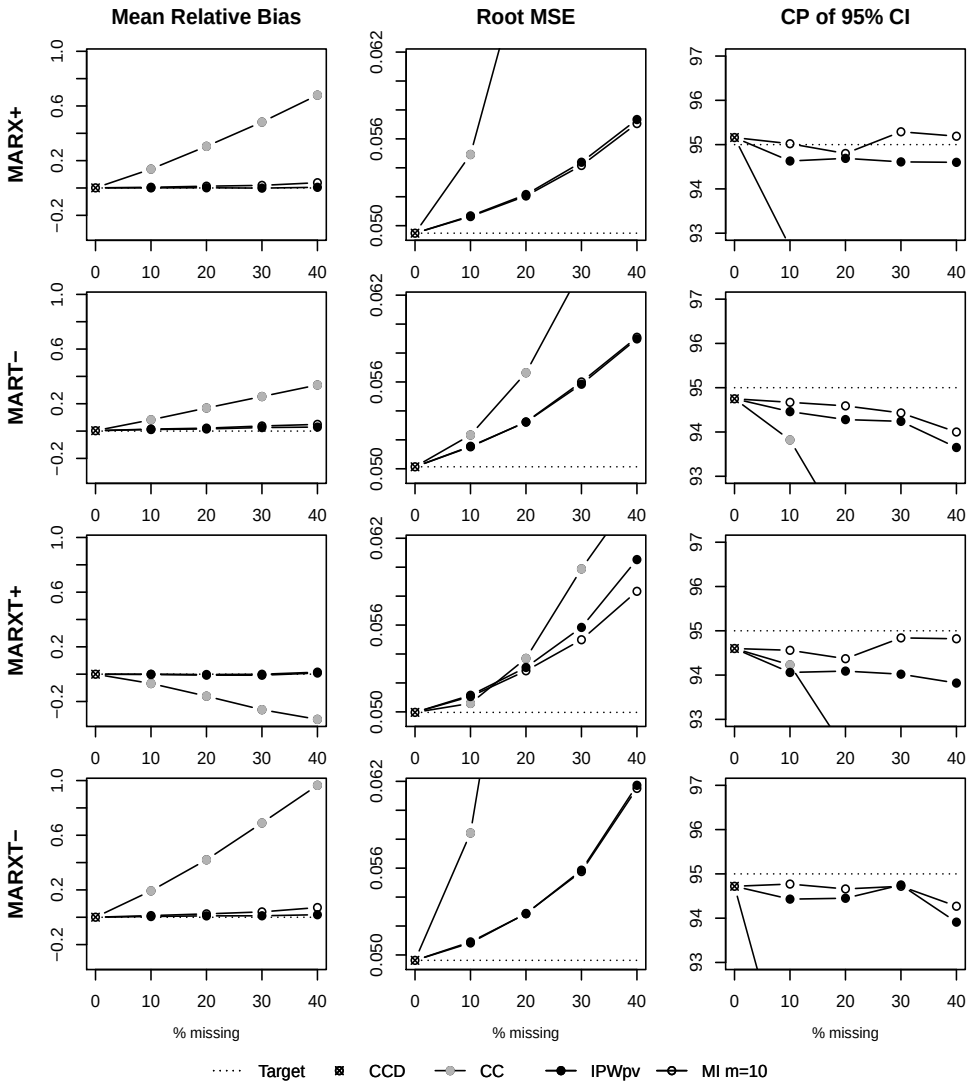

Figure 4.2: Simulation results for coefficient estimation in the additive model with a binary covariate, $n = 200$ and 50% uniform censoring for selected missingness mechanisms. Estimates obtained via a complete case analysis (CC), the proposed IPWpv approach (IPWpv) and multiple imputation with $m = 10$ imputations (MI m=10) are compared. For each mechanism and each analysis, the mean relative bias, the square root of the mean squared error (MSE) and the estimated coverage probability (CP) of the 95% confidence interval (CI) are plotted against the percentage of missing causes. With no missing causes, all analyses coincide with the complete censored data analysis (CCD) plotted at 0%. Results are based on 10000 replications.

other two possible variance estimators were biased (see below).

In the following sections we describe and analyze the results of the simulation study, taking into account all scenarios and considering each criterion of interest according to the main objectives of the study.

4.2.3.1 Bias correction

To examine this aspect we considered the magnitude of the MRB of the estimates. For the additive model (Figure 4.2 and Figures B.8-B.14), the CCD analyses always led to unbiased estimates ($|\text{MRB}| \leq 0.8\%$, all scenarios considered). For the Fine and Gray model (Figure 4.1 and Figures B.1-B.7), CCD estimates were approximately unbiased ($|\text{MRB}| \leq 3.2\%$). The small upward bias in the latter case is explained by the fact that equation (4.3) is only asymptotically unbiased. Actually, at $n = 400$ the MRBs of the Fine and Gray model estimates decreased by around 1% (Figures B.2-B.3). Hence, estimates obtained via the Andersen-Klein approach are more sensitive to sample size for this model.

With missing data, the CC estimator was biased in most cases, either upward or downward, as expected ($|\text{MRB}| \geq 5\%$). It was approximately unbiased ($|\text{MRB}| \leq 5\%$) only in a few scenarios with a low percentage of missing causes (10 to 20%), particularly in MCAR and MART+ scenarios for both models and in the MART- scenario for the Fine and Gray model. The approximate unbiasedness of the CC estimator in the latter scenarios was enabled by our simulation set-up and may not be observed in other situations. As expected, increased percentage of missingness (30-40%), smaller sample size ($n = 200$; compare Figures B.2-B.3 and B.9-B.10 with all other figures) or increased censoring percentage (50%; compare Figures B.6-B.7 and B.13-B.14 with all other figures) generally led to a larger bias in the CC estimator, with the magnitude of its MRB reaching 179% in some cases. The CC analysis was biased even in the MCAR scenario, as expected, with an MRB of magnitude between 10% and 20% with 30-40% missing causes (cf. Section 2.3.1).

In contrast, the IPWpv estimates were always approximately unbiased for both models (|MRB|≤5%, except in a few rare cases where MRB remained under 7%). The same was true for MI estimates (|MRB|≤5%, except in a few cases where the MRB remained under 11%). For the Fine and Gray model, the performance of both approaches was similar, but for the additive model the MRBs of the MI estimates were generally slightly larger. The latter is probably due to the misspecification of the imputation model.

4.2.3.2 Relative efficiency

The root MSE of the CC estimator was generally dominated by its bias and was the largest of the three estimators. For the Fine and Gray model (Figure 4.1 and Figures B.1-B.7), it was only in the MART- scenario with 25% censoring and $n = 200$ that the root MSE of the CC estimator was slightly smaller than that of the MI estimator, but it still remained higher than that of the IPWpv estimator (Figure B.6). For the additive model (Figure 4.2 and Figures B.8-B.14), it was in a very few scenarios in which the CC estimator was approximately unbiased or only moderately biased that its root MSE was slightly smaller than for the other two estimators (see Figures B.8, B.10 and B.12).

Conversely, since the IPWpv and MI estimators were always unbiased or approximately unbiased, their root MSEs were dominated by their variances. Thus, when studying their root relative efficiency (RRE), defined as

$$\text{RRE}=\{(\text{MSE of IPWpv})/(\text{MSE of MI})\}^{\frac{1}{2}},$$

we gained insight into their relative precision. In the case of the Fine and Gray model (Figure 4.3), the IPWpv estimator was in general more precise, and thus had a smaller root MSE than the MI estimator (RRE between 0.96 and 1). Only in the MART+ and MARXT+ scenarios, especially with higher percentages of missing causes, did the MI estimator perform better (RRE between 1 and 1.02). On the other hand, for the additive model (Figure 4.4), the differences in precision between both estimators were smaller, with the MI estimator being generally slightly more precise (RRE between 0.98 and

1.02, except for the MARXT+ scenario with RRE up to 1.04). In general, the observed differences in standard deviations were quite small, of around 4% maximum. Thus, the two estimators exhibited comparable precision. However, the results do indicate that the relative performance of these two estimators depends on the missingness mechanism and the model, particularly with increasing percentages of missing causes. For each model, the biggest differences in precision between the two estimators were observed for the missingness mechanisms in which missingness probability depended on failure time, with symmetrically opposite results depending on whether missingness was more common among earlier failures (mechanisms MART- and MARXT-) or later failures (mechanisms MART+ and MARXT+).

To understand the latter pattern, we explored the factors known to affect the performance of these approaches, mainly the specification of the missingness and imputation models. In the IPWpv approach, the weights were obtained via the non-parametric estimator, i.e. the inverse of the observed frequencies, which are the maximum likelihood estimators of the observation probabilities. Thus, there could not be any issues related to misspecification of this model such as unstable weights. The latter was confirmed in an analysis of the distribution of the weights estimated in these scenarios.

On the other hand, the imputation model used was parametric and misspecified for both models. Thus the simulation set-up was unfavorable for MI compared to IPWpv. In the case of the Fine and Gray model, the linear predictor of the 'true' imputation model includes much more complex effects and interaction terms for the covariate and failure time compared to our simple logistic imputation model (see formula (8) of Bakoyannis et al. (2010), whose simulation model was the same as ours, except for the auxiliary covariate and the effect parameter value). On the other hand, following Beyersmann et al. (2009), for the additive model the 'true' imputation model can be deduced to be

$$P(D=1|T=t,X=x,U=0)=\frac{\lambda_1(t|X=x)}{\lambda_1(t|X=x)+\lambda_1(t|X=x)}=\frac{p}{1-\beta^{AD}x}.$$

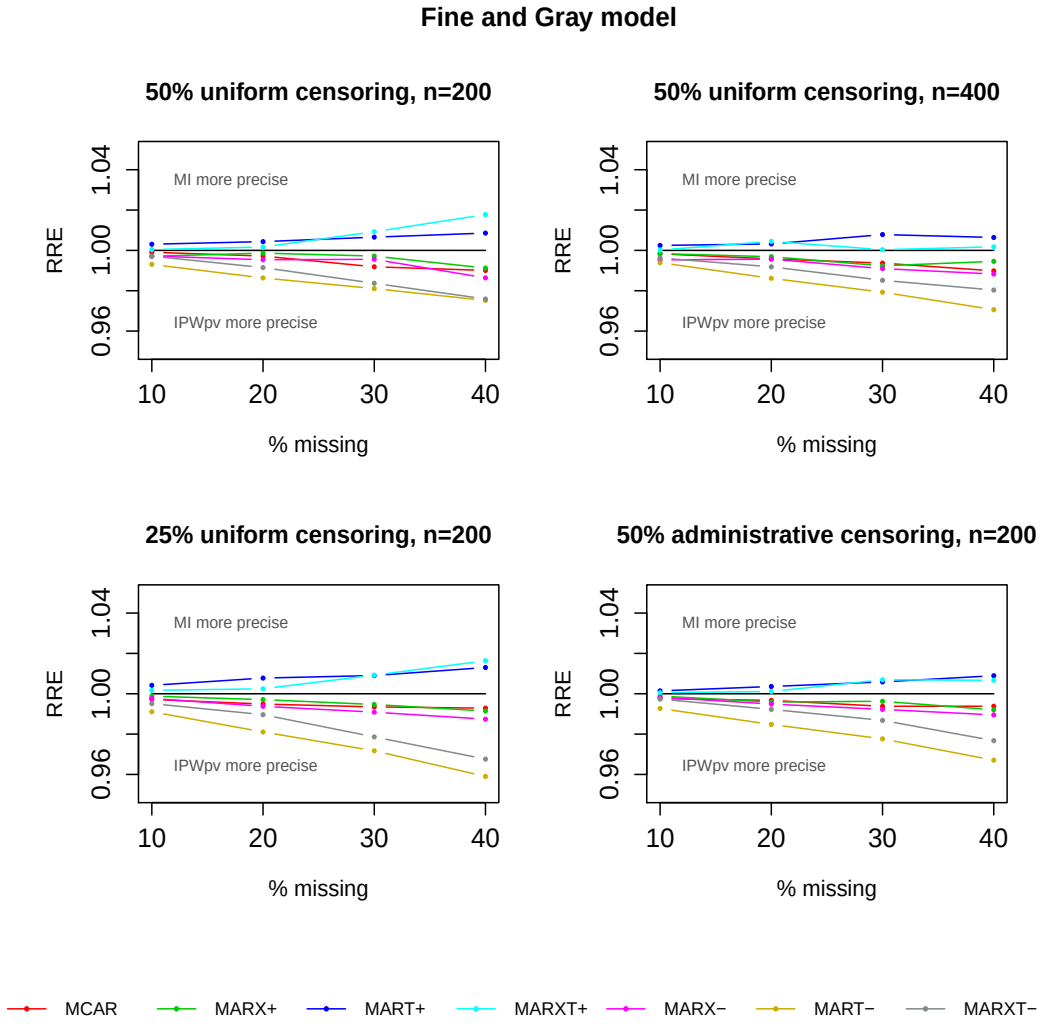

Figure 4.3: Simulation results for the root relative efficiency (RRE) of the IPWpv and MI (m=10) estimators in the Fine and Gray model with a binary covariate, for the four different combinations of censoring percentage, censoring type and sample size considered. For each of these scenarios, the RRE obtained with each of the missingness mechanisms studied is plotted against the percentage of missing causes. An RRE above 1 indicates that MI was more precise; an RRE below 1 means IPWpv was more precise. To facilitate comparison, a solid black line was drawn at RRE=1. Results are based on 10000 replications.

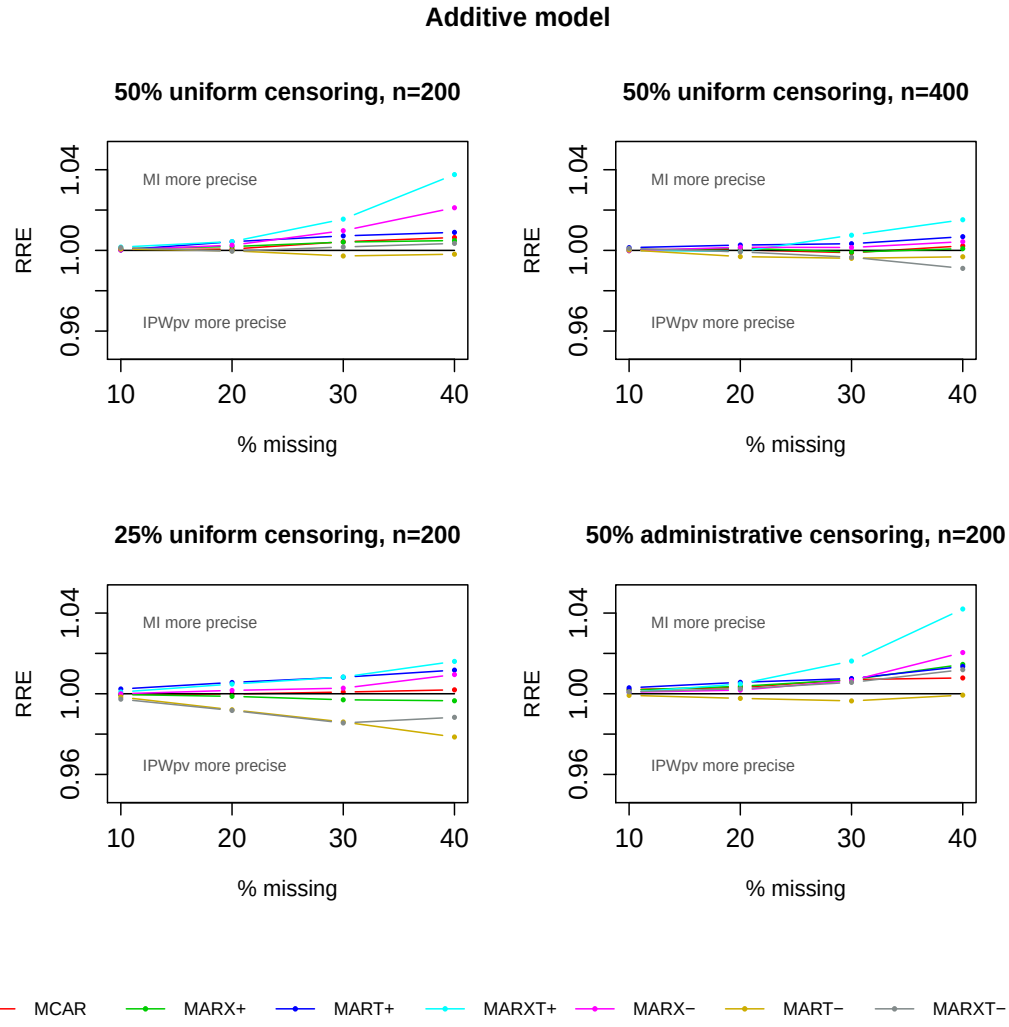

Figure 4.4: Simulation results for the root relative efficiency (RRE) of the IPWpv and MI (m=10) estimators in the additive model with a binary covariate, for the four different combinations of censoring percentage, censoring type and sample size considered. For each of these scenarios, the RRE obtained with each of the missingness mechanisms studied is plotted against the percentage of missing causes. An RRE above 1 indicates that MI was more precise; an RRE below 1 means IPWpv was more precise. To facilitate comparison, a solid black line was drawn at RRE=1. Results are based on 10000 replications.

Thus, the 'true' model was not a logistic model and did not include failure time as a predictor, but still this variable was included in the model used for imputation. Therefore, the impact of the misspecification of the imputation model on efficiency can be expected to differ for both models, as was observed in the results, because the underlying 'true' model was different. Moreover, the impact of misspecification on efficiency can also be expected to depend on the distribution of the failure times among the failures with missing cause, because the effect of failure time was largely misspecified in both cases. This distribution is affected by the missingness mechanism, especially when missingness probabilities depend on failure time, as was the case in the MART-, MARXT-, MART+ and MARXT+ scenarios. In conclusion, the small fluctuations in the relative efficiency of the two approaches across the different scenarios can be attributed to a variation in the performance of MI; the use of a misspecified imputation model had a varying impact on the efficiency of this estimator, depending on both the missingness mechanism and the model.

Of course, a misspecified imputation model has a higher impact with an increased percentage of missing data, but also higher amounts of missing data may require more imputations to achieve the best possible efficiency with MI (see Table 4.1 of Rubin, 1987). We thus performed some additional simulations for some scenarios with increased numbers of imputations (results not shown). As expected, this led to an improvement of the RREs in favor of MI, but in some cases IPWpv remained more precise, even with 100 imputations.

4.2.3.3 Variance estimation

We evaluated the variance estimators used in each analysis by calculating the MRB of the standard deviation estimate with respect to the observed (Monte Carlo) standard deviation. Next we summarize our findings and omit the detailed results, except for IPWpw for which partial results are shown.

In the CCD analyses, the sandwich variance estimator was used and it was always unbiased for both models ($|\text{MRB}| \leq 3\%$). In the CC analyses, the sandwich estimator

was approximately unbiased for both models (|MRB|≤5% except in a few rare cases where it remained under 10%). In the MI approach, the variance estimator was based on the sandwich estimates from the completed datasets and was obtained from the usual MI variance estimator by combining between- and within-imputation variances. This estimator was always approximately unbiased (|MRB|≤4% for both models, except in one rare case where it remained under 6%). Its overall performance seemed comparable to that displayed by a bootstrap estimator for MI in the simulation results presented by Bakoyannis et al. (2010). The advantage of the usual MI variance estimator is that it is straightforward to compute and does not have the limitations of the bootstrap with small sample sizes.

In the IPWpv analyses, the mean relative biases of three variance estimators were compared: the sandwich, AJ and bootstrap estimators. Some results of this comparison for each model and selected mechanisms with $n = 200$ and 50% uniform censoring are shown in Figure 4.5. Results were similar for other scenarios (not shown). With 10% missing causes, all three estimators were approximately unbiased for both models (|MRB|≤10%). The AJ and sandwich estimators were always very close, and as the percentage of missing causes increased, they became biased, as expected. The bias was upward, resulting in negative mean relative biases down to -28% and -50% for the Fine and Gray and additive models, respectively. This overestimation of the variance is likely due to a phenomenon already documented by Robins et al. (1995), who showed that the true asymptotic variance of an estimator obtained by inverse probability weighting when the weights are known is at least as large as when the weights are estimated. This implies that the variance will tend to be overestimated if the uncertainty in the weights is ignored.

Conversely, the bootstrap variance estimator was approximately unbiased for the Fine and Gray model (|MRB|≤10% for $n = 200$ and ≤4% for $n = 400$) and unbiased for the additive model (|MRB|≤2%), all levels of missingness considered. For the Fine and Gray model, however, the bootstrap variance estimator was difficult to obtain in small samples with high percentages of censoring and missing causes, with some of

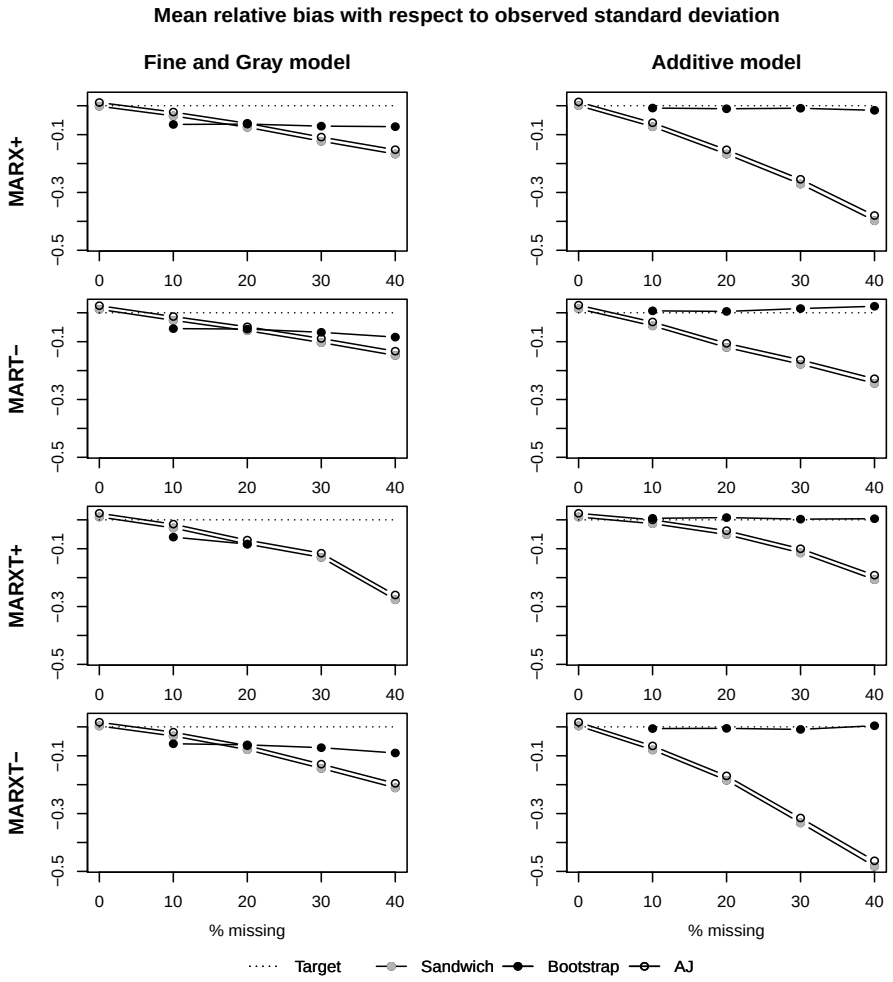

Figure 4.5: Simulation results for variance estimation in the IPWpv analysis for the Fine and Gray and additive models with a binary covariate, $n = 200$ and 50% uniform censoring for selected missingness mechanisms. We considered the sandwich estimator, approximate jackknife (AJ) estimator and the bootstrap estimator with $R = 100$. For each model, mechanism and estimator, the mean relative bias of the standard deviation estimate is plotted against the percentage of missing causes. The bootstrap estimator was not implemented for 0% missing causes, as the sandwich and AJ estimators are already known to be approximately unbiased in that case. For MARXT+ at 30% and 40% missing causes, the mean relative biases of the bootstrap estimates are not plotted as there were several replications where these estimates could not be obtained owing to the small number of observed events. Results are based on 10000 replications.

the missingness mechanisms leading to erroneously divergent variance estimates in a few or several replications. It seems that some bootstrap samples, possibly with very few type 1 events and leading to divergent coefficient estimates, were not identified as such and consequently discarded by our simulation algorithm. When there were two or fewer divergent estimates, they were excluded from the mean relative bias calculation in Figure 4.5. For MART+, MARX- (both not shown) and MARXT+ with 30 and/or 40% missing causes, there were several divergent estimates so the mean relative biases were very large (not plotted in Figure 4.5). The latter shows a limitation of the bootstrap estimator for small samples and high percentages of censoring and missing causes, indicating that care must be taken when implementing the bootstrap in such cases. Nevertheless, in scenarios with 25% uniform censoring, $n = 400$ or a low percentage of missing causes, this problem was very rarely encountered. Therefore, regardless of this drawback, bootstrap estimates could be obtained easily in almost all scenarios and were approximately unbiased, unlike the other two estimators.

4.2.3.4 Coverage probability of the 95% confidence interval

The CPs in the CCD analyses were near the nominal value for both models (for the Fine and Gray model see Figure 4.1 and Figures B.1-B.7; for the additive model see Figure 4.2 and Figures B.8-B.14). As expected, the CC analysis led to very poor CPs in the scenarios where the estimator was biased (the CP curves went down to 2% in some cases - not visible in the graphs). The CP was acceptable only in some scenarios where the CC estimator was approximately unbiased.

As shown in the graphs, a high or low CP in the CCD analysis generally led to high or low CPs for IPWpv and MI analyses, implying that the CP performance for these approaches is highly dependent on the performance of the Andersen-Klein approach with complete data. In the case of the Fine and Gray model, the MI analysis led to CPs very close to those of the CCD analysis and thus to the nominal value. The CPs of the IPWpv analyses were also acceptable with large samples (Figures B.2-B.3 and B.9-B.10), but conservative with small samples (all other figures), reflecting the

moderate overestimation of the variance with the bootstrap method in this case and for this model (Figure 4.5). Furthermore, these coverage probabilities deteriorated as the percentage of missing causes increased, reflecting the limitation of the bootstrap estimator with high percentages of missing data. In the case of the additive model, the IPWpv and MI analyses displayed similar CPs, close to those of the CCD analysis and thus to nominal level, with the IPWpv CPs being generally slightly lower than for MI analyses.

Finally, the type of censoring (uniform or administrative) did not seem to have a notable effect on the performance of any of the approaches considered (compare Figures B.4-B.5 and B.11-B.12 with all other figures).

4.2.4 Continuous covariate

Following a similar set-up as described above, we performed another set of simulations with a continuous covariate $X \sim N(0, 1)$. Again, we focused on estimating the effect of X on the CIF of the cause of interest, F_1. The purpose of this second part was to assess the performance of the different estimators in terms of bias in this setting, particularly for the IPWpv approach with the weights estimated via (4.8). Figure 4.6 shows the MRB of each estimator for each model and for selected missingness mechanisms and several sample sizes (n=200, 400 and 1000), in a scenario with 50% uniform censoring and 40% missing causes. The results for other missingness mechanisms are provided in Appendix B (Figure B.15). The results show that the CC analysis leads to biased estimates in most scenarios, while the IPWpv and MI estimators are both approximately unbiased in all scenarios.

4.3 Application to the ECOG clinical trial

In this section, we revisit the analysis of the ECOG clinical trial (cf. Section 2.1) to illustrate the practical value of the proposed methodology. Assuming MAR, several

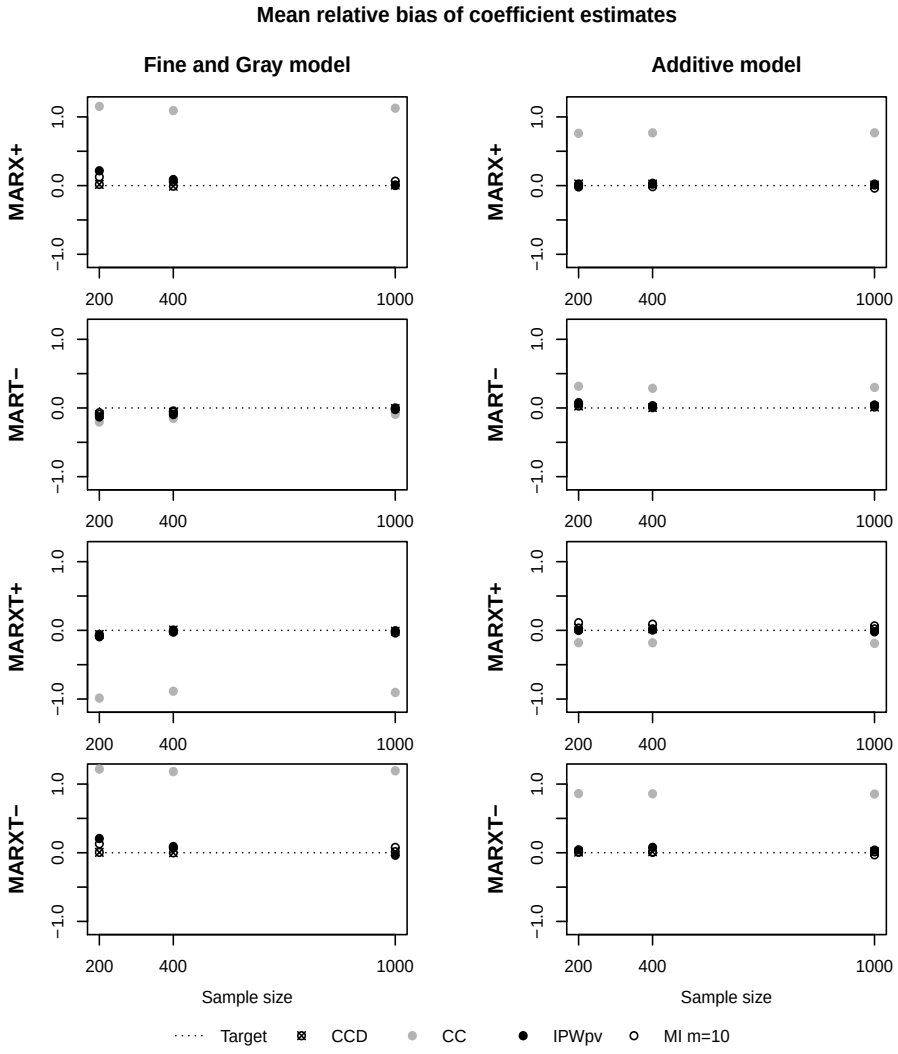

Figure 4.6: Simulation results for coefficient estimation in the Fine and Gray and additive models with a continuous covariate, for selected missingness mechanisms in a scenario with 50% uniform censoring and 40% missing causes. Estimates obtained via a complete case analysis (CC), the proposed IPWpv approach (IPWpv) and multiple imputation with $m = 10$ imputations (MI m=10) are compared. For each mechanism and each analysis, the mean relative bias is plotted against the sample size ($n = 200, 400, 1000$). With no missing causes, all analyses coincide with the complete censored data analysis (CCD), also included in the plots. Results are based on 1000 replications.

authors have studied the effects of the estrogen-receptor (ER) status of the primary tumor (positive vs. negative) and the degree of positive axillary lymph node involvement ($<$4 nodes vs. $\geq$ 4 nodes) on the cause-specific hazard rate of death from cancer (Goetghebeur and Ryan, 1995; Lu and Tsiatis, 2001; Gao, 2006). Furthermore, Nicolaie et al. (2011) performed a vertical modeling analysis of the data. Here we focus on analyzing the effect of these prognostic factors on the CIF of death from cancer.

Figure 4.7 shows non-parametric CIF estimates, obtained via the Aalen-Johansen estimator (4.2), by cause of death, ER status and number of positive nodes, when treating "missing cause" as an additional competing event (i.e. like in an ES analysis). For the combination 'ER-negative and 1-3 nodes' all curves are zero because there were no deaths in this group. From this figure it seems that being ER-negative has a considerable impact on the incidence of cancer death. On the other hand, the impact of the number of positive nodes is not so clear, especially because the deaths with missing cause (right panel) actually belong in either of the other two sets of curves (left and center panels). Thus, in these plots the actual effects of the prognostic factors on cancer death are obscured by the missing data, much like it would be expected in an ES regression analysis (cf. Section 2.3.1). We therefore conducted a regression analysis of the CIF of death from cancer using the proposed methods to estimate these effects under the MAR assumption. We also considered regression models for the CIF of death from other non-cancer causes, but no covariate had a significant effect on it so this analysis is not presented.

We modeled the CIF of death from cancer according to model (4.1) by considering the additive model (identity link) and the Fine and Gray model (cloglog link). In both cases, the model was multivariable, including the indicator variables "$\geq$ 4 nodes" and "ER status", the latter being 1 for patients with an ER-negative primary and 0 for those with an ER-positive primary. Estimates of the regression coefficients were obtained by using the proposed IPWpv and MI (with $m = 10$) approaches, and were compared to those obtained with a CC analysis and an ES analysis (i.e. regarding deaths with a missing cause as due to other non-cancer causes). In all cases, the working correlation

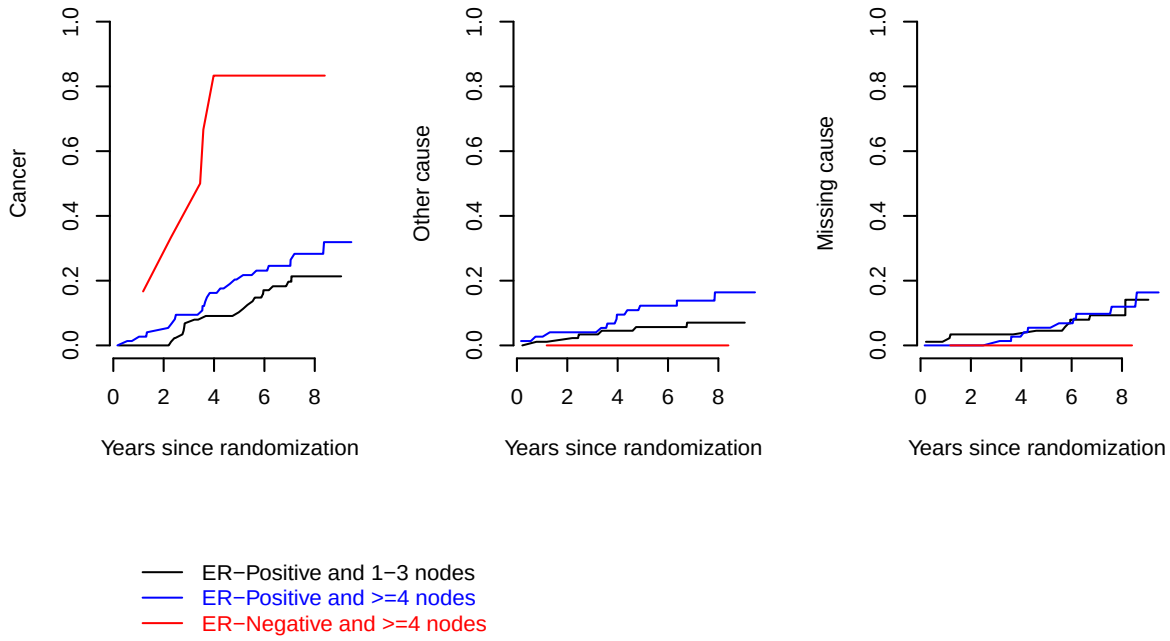

Figure 4.7: Non-parametric CIF estimates by cause of death, ER status and number of positive nodes, for the ECOG clinical trial. For the combination 'ER-negative and 1-3 nodes' all curves are zero because there were no deaths in this group.

matrix was the identity matrix and the grid of time-points used to calculate the pseudo-values in each case was determined by the deciles of the event times excluding the first two, as in the simulation study. The additive model was assumed to be valid for $t \geq t_0 = \tau_1 = 3.0$ years (maximum follow-up time: 9.5 years).

In applying the IPWpv approach, we used the inverse of (4.7) to estimate the weights because only two binary covariates were involved. We came across a difficulty when estimating the weights corresponding to the pseudo-values of the only individual presenting an ER-negative primary and less than four nodes. There were no deaths in this category so these quantities were unidentifiable from the data, whatever the grid chosen. We thus evaluated the sensitivity of the results to the values of these weights by performing two analyses where we assigned, to all the pseudo-values of this individual, either the smallest or the largest of the weights estimated among the pseudo-values of

all the other individuals. These weights were 1 and 1.35, respectively, in the original data, and were recalculated for each bootstrap resample when estimating the variance. Both analyses gave virtually the same results so we present only those for the first analysis.

When applying the MI approach, the difficulty was to build an imputation model including the "ER status" indicator because there were no "other cause" deaths among ER-negative patients. Therefore, the imputation model parameter estimates diverged. This problem had already been documented by Lu and Tsiatis (2001) who also analyzed these data. Since the cause of death was known for all ER-negative patients, they tackled this inconvenience by fitting an imputation model without the "ER status" variable to the individuals with an observed cause and an ER-positive primary. We used the same strategy. Thus, the imputation model included as predictors the number of positive nodes, the time of death and their interaction.

In the IPWpv analysis, the estimator's variance was obtained via the bootstrap variance estimator described in Section 4.1.2 with $R = 1000$ for the additive model and $R = 10000$ for the Fine and Gray model, respectively. In each case, R was chosen to achieve an accuracy to two significant figures in variance estimates. Around 10% of bootstrap samples were discarded and not replaced in each analysis, as in the simulation study (i.e. samples with no deaths or no observed deaths before τ_1 for a category of the covariate vector, or with less than two cancer deaths). In the MI analysis, the variance was estimated by combining sandwich variance estimates obtained from each imputed dataset. In the CC and ES analyses, the estimators' variances were obtained from the sandwich estimator.

The results of these analyses are shown in Table 4.2. In this table, z gives the ratio between the parameter estimate and its standard error (SE). Since the CC, ES and IPWpv estimators are asymptotically normal, we report the p-values of standard two-tailed significance z-tests for these analyses. For MI, two-tailed significance t-tests were performed, with the degrees of freedom of the reference distribution calculated as described in Section 3.1. With the Fine and Gray model, all analyses estimated a highly

significant effect of having an ER-negative primary on the CIF of death from cancer at a 5% level. For the additive model, this effect translated in all analyses to a highly significant increase of around 0.5 in the probability of death from cancer for patients with an ER-negative primary. The latter exemplifies the straightforward 'excess risk' interpretation of covariate effects in the additive model. Conversely, the IPWpv and MI approaches disagreed with the CC analysis on the significance of the effect of having four or more positive nodes at a 5% significance level when using the Fine and Gray model. The CC estimate was borderline significant while the IPWpv and MI estimates were non-significant. Assuming MAR, this difference was due to an overestimation of the impact of having four or more positive nodes in the CC analysis ($\hat{\beta} = 0.79$ for CC against $\hat{\beta} = 0.52$ and $\hat{\beta} = 0.53$ for IPWpv and MI, respectively). The ES analysis also resulted in a moderate overestimation of this effect ($\hat{\beta} = 0.72$), significant at least at a 10% level. The IPWpv and MI approaches corrected these biases. In the additive case, all analyses led to non-significant estimates at a 5% level.

4.4 Discussion

In the present chapter, we provided a general framework for modeling the CIF with missing causes of failure. We proposed an alternative to the methodology of Bakoyannis et al. (2010) and also examined the application of their work to a generic approach for modeling the CIF. Since a large class of models for the CIF can be fit with the Andersen-Klein approach due to the choice of link function, the two extensions considered provide a flexible way to improve goodness of fit in this setting. Simulation results showed that these approaches correct the bias of CC analysis estimates under relaxed assumptions about the missingness mechanism, both for binary and continuous covariates. Asymptotic properties for the novel IPWpv estimator followed readily from results found in the literature. Variance estimators were suggested and evaluated. The ECOG clinical trial highlighted the practical issues that may arise when implementing each approach if there is a low frequency for one category of a covariate. In the ECOG

Table 4.2: Multivariable Fine and Gray and additive models for the CIF of death from cancer including the ER status and the indicator of the presence of four or more positive nodes as covariates

Model	Covariate	Analysis	$\hat{\beta}$	SE	z	p-value
Fine and Gray	ER status	CC	1.85	0.4328	4.286	<0.0001
		ES	1.96	0.4456	4.390	<0.0001
		IPWpv	1.77	0.4859	3.651	0.0003
		MI m=10	1.79	0.4776	3.756	0.0002
	>=4 nodes	CC	0.79	0.3964	1.990	0.0466
		ES	0.72	0.3897	1.842	0.0654
		IPWpv	0.52	0.3905	1.339	0.1807
		MI m=10	0.53	0.3419	1.543	0.1230
Additive	ER status	CC	0.51	0.1377	3.690	0.0002
		ES	0.54	0.1407	3.853	0.0001
		IPWpv	0.51	0.1411	3.621	0.0003
		MI m=10	0.51	0.1410	3.616	0.0003
	>=4 nodes	CC	0.09	0.0539	1.699	0.0893
		ES	0.08	0.0508	1.552	0.1207
		IPWpv	0.07	0.0561	1.291	0.1968
		MI m=10	0.08	0.0546	1.484	0.1378

$\hat{\beta}$ is the estimate of the model-specific covariate effect, obtained via a complete case analysis (CC), an extra state analysis (ES) and the proposed IPWpv and multiple imputation (MI m=10) approaches.

data, there was a small number of patients with an ER-negative primary. Thus, when applying IPWpv, it was impossible to estimate some of the weights, and when applying MI, it was impossible to fit an imputation model including at least all covariates in the analysis model, as is recommended. We found ways to circumvent these problems while still being confident about the results obtained.

The two approaches considered are related to two paradigms of dealing with missing data, IPW and MI, of which there are several comparisons in the literature (see for example Carpenter et al., 2006; Seaman and White, 2013). When choosing between the two approaches, there is a trade-off between modeling missingness probability and the incomplete outcome. In our context, both require modeling a binary variable (M or D), so the complexity of the task is comparable. Also, both approaches can handle auxiliary covariates suspected to influence the missingness mechanism that must thus

be incorporated in the analyses to make MAR-like assumptions plausible. In the IPWpv approach, such covariates can be included as predictors in the weight estimation process, either by including them in the covariate vector of estimator (4.7) or, most likely, in a parametric model like (4.8). In the MI approach, such variables can be included in the imputation model. Actually, in the ECOG data, two additional binary covariates were available, the tumor size ($\leq$ 3 cm vs. $>$ 3 cm) and the treatment group (placebo vs. tamoxifen). However, these variables were not significant predictors of missingness nor of the cause of death, and when they were included in the IPWpv and MI analyses as described, the results were not affected. We thus decided to exclude these covariates from the final analyses: in the implementation of IPWpv, this strategy enabled the use of the preferred non-parametric estimator (4.7) because the number of covariates was low (see below); for MI, it avoided the loss of efficiency that could arise from the inclusion of non-significant predictors in the imputation model (Schafer, 2003).

In the literature, the principal argument against IPW estimators is their lack of efficiency, at least compared to MI (Carpenter et al., 2006; Seaman and White, 2013). Standard IPW approaches are inefficient partly because only the data of the individuals with complete data are included in the estimating equations. In our approach, the IPWpv's of individuals with missing cause, which rely on the pseudo-values of the modified CIF $\tilde{F}_1$, are included in the regression equations, and their data are also accounted for when calculating the IPWpv's of all other individuals. Thus, contrasting with standard IPW methods, our approach incorporates the partial information available from the incomplete cases about the competing risks process (vital status, failure time and covariates) in the estimating equations through the pseudo-values of $\tilde{F}_1$. This may explain why, in our simulation study, the efficiency of IPWpv was comparable to that of MI. Also, it has been noted that the performance of IPW estimators when using a parametric model depends on this model being correctly specified. Otherwise, estimates may be biased and the problem of unstable weights may arise. This is a major reason why the non-parametric estimator (4.7) is preferable when dealing with a few discrete finite-ranged covariates, like in our simulations and in the ECOG example, as

it avoids the problems of misspecification. Actually, this estimator, equal to the (conditional) relative frequency of observed causes, corresponds to the maximum likelihood estimator of the corresponding probability, and therefore it is optimal in terms of efficiency, at least asymptotically. To deal with continuous covariates while keeping the benefits of the non-parametric approach, non-parametric smoothing techniques could be considered (Song et al., 2010). Otherwise, when choosing a parametric approach, it would be desirable to apply weight stabilization techniques or to consider a so-called *doubly-robust* extension.

The main drawback of MI is that both a correctly specified imputation model and congeniality are required to warrant unbiased estimates, particularly for the variance. The latter conditions can be easily violated in practice: misspecified models are the rule rather than the exception, and uncongeniality arises easily in this context because specifying the cause of failure distribution already partially determines the CIF (cf. Chapter 7 for more discussion about this point). Another situation where uncongeniality may easily arise is when the person imputing the data is different from the person who analyzes the imputed datasets, as is sometimes the case (Meng, 1994). Thus, building an appropriate imputation model is generally not a straightforward task. Nevertheless, the missing cause of failure literature seems to suggest that MI is quite robust to misspecification of the imputation model, if the latter includes all important predictors and interaction terms (Lu and Tsiatis, 2001; Bakoyannis et al., 2010). This was confirmed by our simulation study, in which a misspecified but rich imputation model led to approximately unbiased coefficient estimates. However, the use of a misspecified model also led to a varying efficiency of MI relative to IPWpv, with small fluctuations depending on the scenario being investigated. Of course, this empirical finding should be interpreted carefully because our simulation set-up was unfavorable for MI compared to IPWpv: for the former a misspecified parametric model was used to impute while for the latter a non-parametric model was used for the weights. Surprisingly, variance estimates obtained from Rubin's formula were also approximately unbiased in our simulations even though we did not use the bootstrap to correct for bias as advocated by

Bakoyannis et al. (2010).

The ES analysis was not considered for comparison in the simulation study, but the ECOG application exemplified the pitfalls of this ad-hoc approach which were already mentioned in Section 2.3.1. As seen in Table 4.2, for both models the effect of having an ER-negative primary was overestimated by the ES analysis. Since all patients with a missing cause were ER-positive, coding their deaths as due to other non-cancer causes led to an underestimation of the CIF of cancer death in this group while the CIF of cancer in the other group was not affected, leading to the observed biases. On the other hand, the effect of having four or more positive nodes was only moderately overestimated in the Fine and Gray model, and correctly estimated in the additive model. In this case, there was an equal number of failures with missing cause in each category of the covariate. Therefore, when coding these failures as due to non-cancer causes, the cancer CIFs in both groups were underestimated to approximately the same extent, resulting only in small or no biases. Finally, the high precision of the estimates obtained with the ES approach is misleading because the uncertainty concerning the missing causes is completely disregarded. These results confirm that, although at first this approach may seem more sound than a CC analysis, it can lead to similarly spurious results and is highly inadvisable.

Regardless of the limitations of IPWpv and MI, both approaches provide a considerable gain compared to ad-hoc methods in terms of bias correction and precision under relaxed assumptions about the missingness mechanism, and should therefore be considered in a primary analysis. When dealing with a few discrete covariates, we recommend using the IPWpv approach. In these settings, IPWpv is easily implemented, without the need to build a complicated model for the weights nor the need to perform imputations. Indeed, the non-parametric estimator (4.7) for the probabilities can be used and thus the issues arising from misspecification such as bias, unstable weights and loss of efficiency are avoided. With more complex covariate structures, the use of estimator (4.7) becomes unfeasible. Therefore, MI should be used instead; although we did not explore the issue here, in the literature IPW approaches have been found to

be more sensitive to misspecification of the missingness probability model, particularly due to the problem of unstable weights (Seaman and White, 2013), while MI seems to be robust to imputation model misspecification provided the latter is rich enough. A further explanation for this difference was mentioned by Molenberghs and Kenward (2007, Chapter 11), who noted that in IPW approaches all subjects are assigned weights so misspecification of the weight model will affect all of them. In MI, misspecification of the imputation model will affect subjects with missing data but not those with complete data.

Approaches for handling missing data are meant to allow data analysts to achieve the results that they would have obtained with the method that they would have chosen if there were no missing data. Although there are other approaches to model the CIF when all causes of failure are observed (cf. Section 2.2.3), here we focused on extending the Andersen-Klein pseudo-value approach to the missing cause setting. Thus, before using the missing data methods proposed in this work, the analyst has to determine whether he or she would use the pseudo-value approach if there were no missing causes. In the following, we attempt to clarify the advantages and limitations associated with this approach.

The Andersen-Klein pseudo-value approach presents simultaneously several valuable advantages compared to other available regression approaches for the CIF. First, the approach can be easily implemented using readily available software (see Andersen and Perme, 2010; Gerds, 2011). Second, it allows to fit a large class of models for the CIF due to the possibility to choose a link function. Concerning this choice, here we closely examined two possibilities, the Fine and Gray (cloglog link) and additive (identity link) models. An alternative is the recently proposed 'absolute risk model' which corresponds to a logarithmic link function (Gerds et al., 2012). To evaluate the choice of link, goodness of fit tests or diagnostic plots could be performed (Klein, 2006; Fine and Gray, 1999; Klein and Andersen, 2005). In fact, in addition to allowing flexible modeling through the choice of a link function, a third advantage of the Andersen-Klein approach is that it provides the user with an outcome variable which may be used

for graphical goodness of fit assessment (see for example Andersen and Perme, 2010). However, all of these goodness-of-fit assessment techniques require fully-observed causes of failure, so an adaptation to the missing cause setting should be further explored. In any case, inclusion of time-by-covariate interactions is likely to improve fit.

The Andersen-Klein pseudo-value approach has some limitations, particularly regarding the underlying assumptions about the censoring mechanism. Indeed, as mentioned in Section 4.1.1, a condition required to obtain consistent estimates with this approach is that the censoring times be stochastically independent of the failure time, the cause of failure and the covariates. This is a stronger assumption than the usual assumption of 'independent censoring' (cf. Section 2.2.1). The independence of the censoring times from the failure times and causes of failure can of course be unfeasible in some situations. However, the rigorous study of whether this condition may be relaxed has not yet been addressed in the pseudo-value literature and would probably need further consideration. In any case, the extent of the dependence between these variables cannot be assessed from the observed data so such assumptions are unverifiable (Tsiatis, 1975). Thus, as when dealing with missing data, sensitivity analyses should be performed to assess the robustness of inferences to these assumptions. More work is needed on how such an analysis might be performed in our setting, but this is outside the scope of this manuscript.

On the other hand, Binder et al. (2012) explored the performance of the Andersen-Klein pseudo-value approach when the assumption of independence between the censoring times and the covariates is violated, a situation which may be common in practice. Briefly, they found that the estimates obtained using pseudo-values are no longer unbiased under covariate-dependent censoring, but the induced bias was very small in all of the scenarios they studied ($|\text{MRB}| \leq 9\%$ in their simulation study; $|\text{MRB}| \leq 24\%$ in a sensitivity analysis for their real data analysis). The authors propose to correct this bias by calculating the pseudo-values using a modified Aalen-Johansen estimator, which is weighted by the inverse of the probability of censoring. To assess the need for this extended method, the assumption of covariate-independent censoring can be eval-

uated by fitting a Cox model for the censoring times to determine which variables are significant predictors of censoring. A Cox model can also be used to obtain estimates of the censoring probabilities if the extended approach is to be used. If these alternative pseudo-values satisfy properties (P1) and (P2) - something that remains to be shown formally - then the IPWpv and MI approaches would still be valid and could be applied in the same way as described here.

Chapter 5

Direct likelihood for competing risks

Several authors have addressed the problem of fitting semi-parametric regression models for the cause-specific hazard (CSH) in the missing cause of failure setting under the MAR assumption (cf. Section 2.3.2). On the other hand, parameter estimation for fully-parametric models of the form (2.2) with missing causes has received no attention despite the potential usefulness of these models in many applications. In this chapter, we propose a direct likelihood approach for fitting these and other parametric competing risks regression models when the missingness mechanism is assumed to be MAR. More precisely, we show how the concept of *ignorability* mentioned in Section 1.3.2 applies in this setting, when, in addition to MAR, random censoring and a parameter separability condition are assumed (Section 5.1). Using this result, we derived expressions for the likelihood in terms of several interesting functionals in competing risks, making the fitting of parametric models for these quantities straightforward (Section 5.2). The chapter ends with some concluding remarks (Section 5.3).

5.1 Ignorability

In this section, we use arguments similar to those of Little and Rubin (1987, Section 5.3) to demonstrate the ignorability of the missingness and censoring mechanisms in the competing risks setting with missing causes under the set of assumptions presented

next. We emphasize that all of the arguments in this chapter rely on the assumption that $\mathbf{X}$ and $\mathbf{W}$ include all the covariates that act on the competing risks/censoring and missingness mechanisms, respectively. In particular, all the covariates that influence all three mechanisms simultaneously must be included in $\mathbf{X}$. Consider the following assumptions:

(A1) The missing cause mechanism is MAR.

(A2) The censoring mechanism is random, i.e. (T, D) is independent of C given $\mathbf{X}$.

(A3) The parameter vectors of the competing risks, censoring and missingness mechanisms, denoted by $\boldsymbol{\theta}$, $\boldsymbol{\delta}$ and $\boldsymbol{\psi}$ respectively, are distinct (cf. Section 1.3.2).

The target of inference is the parameter indexing $f(t, d|\mathbf{X})$, i.e. $\boldsymbol{\theta}$. Next we will show that, under assumptions (A1)-(A3), valid direct likelihood inferences about $\boldsymbol{\theta}$ can be performed by ignoring the missingness and censoring mechanisms, provided that all the available data are used in the likelihood construction as detailed below.

Inferences about any component of $\boldsymbol{\varphi} = (\boldsymbol{\theta}, \boldsymbol{\delta}, \boldsymbol{\psi})$ can be made only from the observed data, which consist of $(\tilde{T}_i, U_i, \mathbf{X}_i, \mathbf{W}_i)$ for censored individuals, $(\tilde{T}_i, D_i, U_i, M_i, \mathbf{X}_i, \mathbf{W}_i)$ for uncensored individuals with observed cause and $(\tilde{T}_i, U_i, M_i, \mathbf{X}_i, \mathbf{W}_i)$ for uncensored individuals with missing cause. The data of different individuals are assumed to be i.i.d. given the covariates. Thus, the observed data likelihood is $\mathcal{L}(\boldsymbol{\varphi}) = \prod_{i=1}^{n} L_i(\boldsymbol{\varphi})$, where the contribution to the likelihood of individual i, $L_i(\boldsymbol{\varphi})$, is proportional to conditional joint density of his observed data given the covariates. For censored individuals, for whom $\tilde{T}_i = C_i$ and $U_i = 1$, we have:

$$\begin{aligned} L_i(\boldsymbol{\varphi}) &\propto f(\tilde{t}_i, u_i = 1|\mathbf{x}_i, \boldsymbol{\theta}, \boldsymbol{\delta}) \\ &= f(c_i, u_i = 1|\mathbf{x}_i, \boldsymbol{\theta}, \boldsymbol{\delta}) \\ &= P(T_i > c_i|\mathbf{x}_i, \boldsymbol{\theta}) f(c_i|\mathbf{x}_i, \boldsymbol{\delta}). \end{aligned} \tag{5.1}$$

For uncensored individuals with observed cause of failure, for whom $\tilde{T}_i = T_i$ and $U_i = 0$, we have:

$$\begin{aligned} L_i(\boldsymbol{\varphi}) &\propto f(\tilde{t}_i, d_i, u_i = 0, m_i | \mathbf{x}_i, \mathbf{w}_i, \boldsymbol{\theta}, \boldsymbol{\delta}, \boldsymbol{\psi}) \\ &= f(t_i, d_i, u_i = 0 | \mathbf{x}_i, \boldsymbol{\theta}, \boldsymbol{\delta}) f(m_i | t_i, d_i, u_i = 0, \mathbf{w}_i, \boldsymbol{\psi}) \\ &= f(t_i, d_i, u_i = 0 | \mathbf{x}_i, \boldsymbol{\theta}, \boldsymbol{\delta}) f(m_i | t_i, u_i = 0, \mathbf{w}_i, \boldsymbol{\psi}) \qquad (5.2) \\ &= f(t_i, d_i | \mathbf{x}_i, \boldsymbol{\theta}) P(C_i > t_i | \mathbf{x}_i, \boldsymbol{\delta}) f(m_i | t_i, u_i = 0, \mathbf{w}_i, \boldsymbol{\psi}). \qquad (5.3) \end{aligned}$$

Finally, for uncensored individuals with missing cause of failure, for whom $\tilde{T}_i = T_i$ and $U_i = 0$, the density is obtained by integrating over the missing data, i.e. summing over the set of all possible values of the unobserved D_i:

$$\begin{aligned} L_i(\boldsymbol{\varphi}) &\propto f(\tilde{t}_i, u_i = 0, m_i | \mathbf{x}_i, \mathbf{w}_i, \boldsymbol{\theta}, \boldsymbol{\delta}, \boldsymbol{\psi}) \\ &= \sum_{j=1,2} f(t_i, d_i = j, u_i = 0, m_i | \mathbf{x}_i, \mathbf{w}_i, \boldsymbol{\theta}, \boldsymbol{\delta}, \boldsymbol{\psi}) \\ &= \sum_{j=1,2} \{ f(t_i, d_i = j, u_i = 0 | \mathbf{x}_i, \boldsymbol{\theta}, \boldsymbol{\delta}) f(m_i | t_i, d_i = j, u_i = 0, \mathbf{w}_i, \boldsymbol{\psi}) \} \\ &= \sum_{j=1,2} \{ f(t_i, d_i = j, u_i = 0 | \mathbf{x}_i, \boldsymbol{\theta}, \boldsymbol{\delta}) \} f(m_i | t_i, u_i = 0, \mathbf{w}_i, \boldsymbol{\psi}) \qquad (5.4) \\ &= \sum_{j=1,2} \{ f(t_i, d_i = j | \mathbf{x}_i, \boldsymbol{\theta}) \} P(C_i > t_i | \mathbf{x}_i, \boldsymbol{\delta}) f(m_i | t_i, u_i = 0, \mathbf{w}_i, \boldsymbol{\psi}) \qquad (5.5) \\ &= f(t_i | \mathbf{x}_i, \boldsymbol{\theta}) P(C_i > t_i | \mathbf{x}_i, \boldsymbol{\delta}) f(m_i | t_i, u_i = 0, \mathbf{w}_i, \boldsymbol{\psi}). \qquad (5.6) \end{aligned}$$

Equalities (5.2) and (5.4) follow from (A1). Equalities (5.1), (5.3) and (5.5) follow from (A2). The main consequence of these deductions is that the observed data likelihood factorizes as

$$\mathcal{L}(\boldsymbol{\varphi}) = \mathcal{L}_1(\boldsymbol{\theta}) \mathcal{L}_2(\boldsymbol{\delta}) \mathcal{L}_3(\boldsymbol{\psi}).$$

This fact coupled with the distinctness of the parameters - assumption (A3) - implies that inferences about $\boldsymbol{\theta}$ may be based solely on $\mathcal{L}_1(\boldsymbol{\theta})$. That is, inferences about $\boldsymbol{\theta}$ may

be performed by ignoring the second factor of (5.1) and the second and third factors of (5.3) and (5.6), i.e. by ignoring the censoring and missingness mechanisms.

5.2 Expressions for the likelihood

As a consequence of the arguments in the previous section and the relations between the basic competing risks quantities (cf. Section 2.2), particularly that $f_j(t) = \lambda_j(t)S(t)$ and $f(t) = \lambda(t)S(t)$, $\mathcal{L}_1(\boldsymbol{\theta})$ may be written in terms of the CSHs as follows:

$$\mathcal{L}_1(\boldsymbol{\theta}) = \prod_{i=1}^{n} \left\{ \sum_{j \in \mathcal{J}_i} \lambda_j(\tilde{t}_i|\mathbf{x}_i) \right\}^{1-u_i} \times \exp\left\{ -\int_0^{\tilde{t}_i} \lambda_1(u|\mathbf{x}_i) + \lambda_2(u|\mathbf{x}_i).du \right\}, \tag{5.7}$$

where $\mathcal{J}_i := \{d_i\}$ if $m_i = 0$ and $\mathcal{J}_i := \{1, 2\}$ if $m_i = 1$.

Note that the contribution of individuals with a missing cause is based on $f(t|\mathbf{X}) = \lambda(t|\mathbf{X})S(t|\mathbf{X})$, so they provide information only about the occurrence of failures, all-causes combined, and thus about the distribution of T. With fully-observed causes of failure, the latter expression reduces to the known likelihood with random censoring: the contribution of censored individuals is based on $S(t|\mathbf{X})$, providing information only about survival, and thus about the distribution of T; and the contribution of individuals with an event of type j is based on $f_j(t|\mathbf{X}) = \lambda_j(t|\mathbf{X})S(t|\mathbf{X})$, providing information about the occurrence of cause j failures, and thus about the joint distribution of T and D (Prentice et al., 1978).

The likelihood function $\mathcal{L}_1(\boldsymbol{\theta})$ can be rewritten in terms of the cumulative incidence functions (CIFs) by considering that $\lambda_j(t) = \frac{d}{dt}F_j(t)/S(t)$, $\lambda(t) = \frac{d}{dt}\{F_1(t)+F_2(t)\}/S(t)$ and $S(t) = 1 - F_1(t) - F_2(t)$. Then we have

$$\mathcal{L}_1(\boldsymbol{\theta}) = \prod_{i=1}^{n} \left\{ \sum_{j \in \mathcal{J}_i} \frac{d}{dt} F_j(\tilde{t}_i|\mathbf{x}_i) \right\}^{1-u_i} \times \left\{1 - F_1(\tilde{t}_i|\mathbf{x}_i) - F_2(\tilde{t}_i|\mathbf{x}_i)\right\}^{u_i}. \tag{5.8}$$

With fully-observed causes of failure, this expression reduces to the likelihood function presented by Jeong and Fine (2007).

A third way of rewriting this likelihood is to consider the so-called *vertical modeling* factorization $f_j(t) = f(t)r_j(t)$, where $r_j(t) = P(D = j|T = t)$ is called the *relative hazard* (Nicolaie et al., 2010, 2011). This yields $\lambda_j(t) = \lambda(t)r_j(t)$ and thus

$$\mathcal{L}_1(\boldsymbol{\theta}) = \prod_{i=1}^{n} \lambda(\tilde{t}_i|\mathbf{x}_i)^{1-u_i} \times \exp\left\{-\int_0^{\tilde{t}_i} \lambda(u|\mathbf{x}_i).du\right\} \times \left\{\prod_{j\in\mathcal{J}_i} r_j(\tilde{t}_i|\mathbf{x}_i)\right\}^{(1-u_i)(1-m_i)} . \quad (5.9)$$

The ingredients of this expression, the all-cause and relative hazards, are the main focus of the vertical modeling approach to the study of competing risks. Actually, expression (5.9) reduces to the usual vertical modeling likelihood when there are no missing causes of failure (Nicolaie et al., 2010).

Expressions (5.7) and (5.8) may be used directly for fitting parametric regression models for the CSHs and the CIFs, respectively, yielding unbiased and fully-efficient estimates of covariate effects under assumptions (A1)-(A3). These expressions could also potentially be used as a basis for constructing partial likelihoods for fitting semi-parametric models for these quantities. On the other hand, expression (5.9) is the basis for vertical modeling of competing risks with missing causes of failure as proposed by Nicolaie et al. (2011).

5.3 Discussion

In this chapter, we presented the construction of the likelihood for competing risks with missing causes of failure under MAR, random censoring and a parameter separability condition, concluding that the missingness and censoring mechanisms can be ignored under these assumptions. In addition to making the fitting of parametric models for several interesting functionals straightforward, the rationale underlying this construction is interesting in is own right as it enhances our understanding of each individual's contribution to inferences according to their status (Klein and Moeschberger, 2003, Section 3.5). Although this construction has been briefly outlined elsewhere (Nicolaie et al., 2011), we are not aware of any work presenting the expressions of the likelihood

in terms of the CSHs and the CIFs that we provided in Section 5.2. Since the latter are useful mainly for fitting parametric models, a possible explanation for this void in the literature is the preference for semi-parametric models in the biomedical context.

The contents of this chapter are mainly theoretical. The practical implementation and evaluation of the proposed approach through simulation experiments are the object of ongoing research. Since we are dealing with parametric models, one key aspect to explore will be the impact of model misspecification on inferences. The future application of this approach to the ECOG clinical trial will enable a comparison with the approaches of Goetghebeur and Ryan (1995) and Lu and Tsiatis (2001) in a real dataset.

An important difference between (5.7) and the CSH-based likelihood function with fully-observed causes of failure is that (5.7) does not factorize into two factors, one depending solely on the the CSH of cause 1 and the other solely on the CSH of cause 2. In the setting without missing causes, this factorization is what, in part, justifies that a CSH analysis for one cause be performed by censoring failures from other causes (cf. Section 2.2.2). With missing causes, the individuals with unknown cause of failure provide information about the all-cause hazard, which contains information about both CSHs, and this is what prevents such a factorization. On the other hand, the CIF-based likelihood (5.8) does not factorize into two factors, each expressed in terms of the CIF of one cause, even in the setting with fully-observed causes of failure.

Vertical modeling consists in modeling the all-cause and relative hazards instead of the CSHs or CIFs, providing an alternative angle from which the competing risks mechanism can be understood. Note that the likelihood (5.9) factorizes into two factors, each expressed in terms of solely one of the vertical modeling functionals. Hence, Nicolaie et al. (2010, 2011) argue that the vertical modeling approach is easy to implement because a model for each quantity can be fitted separately by maximizing the corresponding factor using available software. However, even in the scenario with fully-observed causes of failure, the latter approach requires a further separability condition between the parameters of the all-cause hazard $\lambda(t|\mathbf{X})$ and the relative hazards

$r_j(t|\mathbf{X}) = \lambda_j(t|\mathbf{X})/\lambda(t|\mathbf{X})$, $j = 1, 2$, which may be an odd requirement considering the intimate relation between these functionals. In the relative hazard of a given cause, say cause 1, the parameters of $\lambda_1(t|\mathbf{X})$ appear in both the numerator and denominator, so one may consider that they 'cancel' out. On the other hand, the parameters of $\lambda_2(t|\mathbf{X})$ appear only in the denominator of $r_1(t|\mathbf{X})$ and will thus generally be shared with $\lambda(t|\mathbf{X}) = \lambda_1(t|\mathbf{X}) + \lambda_2(t|\mathbf{X})$. For an explicit example, see the data generation models in the simulation study of Lu and Tsiatis (2001), which are easily seen to violate this separability condition. When the separability condition is violated, the estimates obtained using the approach of Nicolaie et al. (2011) are no longer fully-efficient. The loss of efficiency stems from an increase in the number of parameters to estimate (see Shih's discussion to Diggle and Kenward, 1994; Rubin, 1976; Altham, 1984; Shih, 1992). Hence, if the separability condition does not hold, the entire likelihood (5.9) must be considered simultaneously to obtain fully-efficient estimates. Note that this remark applies even with fully-observed causes of failure.

Another approach that could be used to fit parametric competing risks models under MAR is multiple imputation (MI) (cf. Section 3.1). Actually, the same procedure of Bakoyannis et al. (2010) described in Section 4.1.3 could be used, the difference being that the analysis model would be the parametric model of interest. The latter would be fitted to each completed dataset by maximizing the likelihood derived with fully-observed causes of failure. With correctly-specified imputation and analysis models, the MI estimator will approximate the direct likelihood estimator while being less efficient (Schafer, 1999). However, MI may be valuable when there is no desire to include some covariates that influence both the competing risks and missingness mechanism in the competing risks model. Indeed, such covariates may be included in the imputation model and then excluded from the models fitted to each completed dataset. The main interest of MI is, however, that once the data are imputed, any model for any quantity may be fitted, including semi-parametric models as shown for the CIF in Chapter 4. MI is also a valuable tool for performing sensitivity analyses as will be shown in Chapter 7.

Part III

Sensitivity analyses

Chapter 6

Sensitivity analyses for continuous longitudinal data with drop-outs

When modeling longitudinal data with drop-outs, unbiased and fully-efficient regression coefficient estimates can be obtained by a direct likelihood approach under the assumption of an ignorable drop-out mechanism (cf. Section 1.3.2). However, the plausibility of the underlying MAR assumption cannot be assessed from the observed data (cf. Section 1.3.4). Thus, sensitivity analyses should be routinely performed to assess the robustness of inferences to departures from this assumption. However, no standard method exists nor should be prescribed as this is still an active area of research (Carroll et al., 2004). In this chapter, we propose an approach to perform such analyses in the setting where the available data are described by means of a linear mixed model (LMM) in a primary analysis assuming MAR. We consider a family of MNAR pattern-mixture models (PMMs) indexed by a so-called *sensitivity parameter* as the basis to explore the sensitivity of inferences made about a parameter of interest. To specify these models, the analyst must make explicit assumptions about the aspects of the missing data distribution that may diverge from the observed data distribution and affect the parameter under investigation. Thus, our approach targets a fundamental question in a sensitivity analysis: How are MAR-based inferences on the parameter of interest affected if the

missing and the observed data come from different distributions?

In Section 6.1 we provide some background regarding sensitivity analyses based on sensitivity parameters. The proposed methodology is described in Section 6.2. The performance of the approach was explored in a simulation study, the findings of which are presented in Section 6.3. In Section 6.4 we present the analysis of the SMI study (cf. Section 1.1), which actually motivated the proposed methodology. This case study illustrated the practical value of our approach and underlined the need for sensitivity analyses when modeling longitudinal data with drop-outs. Some elements for discussion and concluding remarks are given in Section 6.5.

6.1 Background on sensitivity parameters

As mentioned in Section 1.3.4, one possible road to sensitivity analyses is to consider a family of MNAR models indexed by a scalar or vector parameter that is varied across a set of plausible values. To further explain this idea in the context of PMMs, we consider the setting of longitudinal data with drop-outs and closely follow the ideas of Daniels and Wang (2009) and Hogan (2009). Following the notation introduced in Chapter 1, let $\boldsymbol{\varphi}$ denote the parameter vector of the joint density of the outcomes and the drop-out indicator, $f(\mathbf{y}^{\mathcal{O}}, \mathbf{y}^{\mathcal{M}}, u)$. Usually, the target parameter of inference $\boldsymbol{\theta}$ is a function h of this parameter: $\boldsymbol{\theta} = h(\boldsymbol{\varphi})$. Omitting covariates, this joint density can be written as follows:

$$f(\mathbf{y}^{\mathcal{O}}, \mathbf{y}^{\mathcal{M}}, u|\boldsymbol{\varphi}) = f(\mathbf{y}^{\mathcal{O}}, u|\boldsymbol{\varphi}) \times f(\mathbf{y}^{\mathcal{M}}|\mathbf{y}^{\mathcal{O}}, u, \boldsymbol{\varphi}).$$

The first factor, $f(\mathbf{y}^{\mathcal{O}}, u|\boldsymbol{\varphi})$, represents the model for the observed data; the second factor, $f(\mathbf{y}^{\mathcal{M}}|\mathbf{y}^{\mathcal{O}}, u, \boldsymbol{\varphi})$, represents the model for the conditional distribution of the missing data given the observed data, henceforth called the *extrapolation model*. While the former is identifiable from the observed data, the latter is not because $\mathbf{y}^{\mathcal{M}}$ is not observed. Hence, additional assumptions are required to identify the extrapolation model, and such assumptions are not verifiable from the observed data because they relate strictly

to the missing data. Furthermore, inferences about $\boldsymbol{\varphi}$ depend on these unverifiable assumptions concerning the extrapolation model. The goal of a sensitivity analysis in the present context is to assess the sensitivity of inferences about the parameter of interest $\boldsymbol{\theta}$ to these particular assumptions.

PMMs lend themselves well to such analyses because, as mentioned in Section 1.3.3, the fitting of PMMs requires making explicit assumptions about the extrapolation model. Moreover, it is possible to parametrize these models in such a way that $\boldsymbol{\varphi} = (\boldsymbol{\phi}, \boldsymbol{\kappa})$ where $\boldsymbol{\kappa}$ does not appear in the observed data model, so that this model is indexed solely by $\boldsymbol{\phi}$ and the extrapolation model is indexed by $(\boldsymbol{\phi}, \boldsymbol{\kappa})$:

$$f(\mathbf{y}^{\mathcal{O}}, \mathbf{y}^{\mathcal{M}}, u|\boldsymbol{\phi}, \boldsymbol{\kappa}) = f(\mathbf{y}^{\mathcal{O}}, u|\boldsymbol{\phi}) \times f(\mathbf{y}^{\mathcal{M}}|\mathbf{y}^{\mathcal{O}}, u, \boldsymbol{\phi}, \boldsymbol{\kappa}). \tag{6.1}$$

The implications of this parametrization are the following. First, given a fixed value of $\boldsymbol{\phi}$, any value of $\boldsymbol{\kappa}$ yields the same fit to the observed data, i.e. the observed data likelihood, $\mathcal{L}(\boldsymbol{\phi}, \boldsymbol{\kappa}|\mathbf{y}^{\mathcal{O}}, u)$, regarded as a function of $\boldsymbol{\kappa}$, is constant. Hence, $\boldsymbol{\kappa}$ is not identifiable. Second, $\mathcal{L}(\boldsymbol{\phi}, \boldsymbol{\kappa}|\mathbf{y}^{\mathcal{O}}, u)$ regarded as a function of $\boldsymbol{\phi}$ is non-constant, so this parameter is identifiable. Finally, the parameter of interest, $\boldsymbol{\theta} = h(\boldsymbol{\phi}, \boldsymbol{\kappa})$, (generally) depends on the unidentifiable parameter $\boldsymbol{\kappa}$. To summarize, different values of $\boldsymbol{\kappa}$ yield the same fit to the observed data, but also imply different extrapolation models and (generally) different values for the parameter of interest. A parameter with such properties is called a *sensitivity parameter* because it "embodies" the source of differences in inferences observed under different (unverifiable) assumptions about the extrapolation model.

Sensitivity analyses can therefore be performed by varying $\boldsymbol{\kappa}$ over a range of values and assessing the differences between the resulting inferences and those obtained in a primary analysis. Often, it is possible to find sensitivity parameters that have an intuitive interpretation, so that subject-matter experts can be consulted about plausible ranges of values for these parameters. In this chapter and in Chapter 7, where sensitivity analysis approaches are proposed for longitudinal data with drop-outs and

competing risks with missing causes of failure, respectively, we consider primary analyses that assume MAR. When studying MAR multiple imputation (MI) approaches in these two settings (Chapters 3 and 4), it was shown that MAR is equivalent to the assumption that the observed and missing data have the same distribution. We chose parametrizations such that $\boldsymbol{\kappa} = \mathbf{0}$ was equivalent to the MAR assumption, and actually corresponded to the primary analysis model. Thus, $\boldsymbol{\kappa} \neq \mathbf{0}$ implied a departure from the MAR assumption, and the value of $\boldsymbol{\kappa}$ actually quantified an interpretable difference between the distributions of the observed and missing data. In such settings, sensitivity parameters are often termed *informativity parameters* as they quantify the extent to which the drop-out mechanism is *informative*, which is another name for MNAR.

While PMMs can be easily parametrized in terms of sensitivity parameters, selection and shared-parameter models, at least fully-parametric ones, generally cannot be formulated in terms of such parameters (Daniels and Wang, 2009; Hogan, 2009). The latter is why some authors view the PMM framework as the most suitable for assessment of sensitivity (Daniels and Hogan, 2000; Daniels and Wang, 2009; Hogan, 2009).

6.2 Methodology

6.2.1 A family of PMMs for longitudinal data

In this chapter we omit the auxiliary covariate matrix $\mathbf{W}$, and assume that the matrix $\mathbf{X}$ already includes all the covariates influencing the drop-out mechanism. The family of PMMs considered relies on the assumption that the outcomes arise from a mixture of two distributions: the observed data distribution and the missing data distribution. Thus, we distinguish only between missing and observed outcomes by considering the missingness indicators $R_{ij} := I(U_i \leq j)$. More precisely, we consider LMMs of the form

$$Y_{ij} = \mathbf{X}'_{ij}\boldsymbol{\beta} + \mathbf{Z}'_{ij}\mathbf{b}_i + \kappa R_{ij} + \varepsilon_{ij}, \qquad \varepsilon_{ij} \sim N(0, \sigma^2), \qquad \mathbf{b}_i \sim N(\mathbf{0}, \mathbf{G}), \tag{6.2}$$

where, apart from the term κR_{ij}, the notation and assumptions are the same as in Section 1.2.2. The first part of the linear predictor, $\mathbf{X}'_{ij}\boldsymbol{\beta} + \mathbf{Z}'_{ij}\mathbf{b}_i$, and the variance parameters, σ^2 and $\mathbf{G}$, completely determine the distribution of the observed outcomes, for which $R_{ij} = 0$. The parameters involved are identifiable from the observed data, and may be estimated by fitting the LMM implied by (6.2) for the observed data to these data by maximum likelihood. On the other hand, the distribution of the missing outcomes is identified up to parameter κ, which is a sensitivity parameter. Indeed, this parameter does not appear in the observed data model, and only appears in the extrapolation model. Hence, this parameter is not identifiable from the observed data. Thus, the model in (6.2) corresponds to a parametrization of the type represented in (6.1).

In the family of PMMs represented by (6.2), it is assumed that the distributions of the missing and observed outcomes are the same up to a shift in the expected value, which is quantified by parameter κ. The MAR assumption is equivalent to the assumption that the missing and observed outcomes have the same distribution, and hence is equivalent to assuming that $\kappa = 0$. Therefore, this family of PMMs can be thought of as being 'centered' at MAR, with κ quantifying the degree of departure from this assumption. Thus, κ may also be called an informativity parameter. Finally, if the primary analysis assumes MAR and that the data are described by an LMM like (6.2) without the term κR_{ij}, then κ also quantifies the departure from the primary analysis model.

The interest of this family of models is that κ has an intuitive interpretation as a shift in an expected value and can be allowed to depend on the covariates, i.e. $\kappa = \kappa(\mathbf{X}_{ij})$. The function $\kappa(\mathbf{X}_{ij})$ can be specified in order to closely reflect the key characteristics of the missing data distribution that may differ from the observed data distribution and affect the parameter of interest. For instance, suppose that the parameter of interest is the effect of a fixed binary (0/1)-coded covariate measured at baseline on the expected change from baseline of the outcomes at visit J. Suppose that this covariate is represented in the l^{th} component of $\mathbf{X}_{i1}$, X_{i1l} $(= X_{i2l} = \cdots = X_{iJl})$. Then an

appropriate choice would be $\kappa(\mathbf{X}_{ij}) = k_{X_{i1l}}$ where $k_{X_{i1l}}$ is a group-specific positive or negative constant. With this choice, the distribution of the missing data is obtained by shifting the covariate effect of the observed data distribution (or the intercept, if $k_0 = k_1$), while holding the other parameters fixed. Another possibility is to take $\kappa(\mathbf{X}_{ij}) = k_{X_{i1l}} t_j$, with t_j the time of the j^{th} measurement, in particular if the main parameter of interest is the effect of the covariate on the expected rate of change of the outcome (i.e. the coefficient of the time-by-covariate interaction). With this choice, the distribution of the missing outcomes has a higher or lower time-by-covariate interaction (or time-slope, if $k_0 = k_1$) than the distribution of the observed data. Whatever the choice, varying k_0 and k_1 over a range of plausible values will define a family of MNAR models specifically designed to provide insight into the sensitivity of inferences made about the parameter of interest to departures from MAR. Of note, k_0 and k_1 are themselves informativity parameters.

Multidimensional parametrization of $\kappa(\mathbf{X}_{ij})$ could also be envisioned. For instance, in the example above one could consider a combination of the two possibilities described, defining $\kappa(\mathbf{X}_{ij}) = k^{(1)}_{X_{i1l}} + k^{(2)}_{X_{i1l}} t_j$. Thus, the space of informativity parameters is now two-dimensional, meaning that two informativity parameters, $k^{(1)}_{X_{i1l}}$ and $k^{(2)}_{X_{i1l}}$, must now be varied for each group across a range of plausible values. With this choice, the group-specific intercepts and slopes of the missing data distribution are simultaneously controlled, differing from those of the observed data distribution for non-null values of the informativity parameters.

Similar ideas have been explored in the univariate data setting (White et al., 2007; Resseguier et al., 2011) and evoked in the longitudinal data setting (Carpenter and Kenward, 2007; National Research Council, 2010), with some closely related approaches having been proposed (Daniels and Hogan, 2000; Ratitch et al., 2013).

6.2.2 Practical implementation

PMMs assume that the data arise from a mixture of several distributions, generally one for each drop-out occasion. Thus, as briefly mentioned in Section 1.3.3, averaging over the distribution of the drop-out indicator is necessary to obtain estimates of parameters describing the marginal distribution of the responses, that is, the distribution of $\mathbf{Y}$ not conditioned on the drop-out indicator, which are usually the parameters of interest. A practical, unified approach to achieve this whatever the choice of $\kappa(\mathbf{X}_{ij})$ is to use MI: the imputation model is given by the assumed PMM and the analysis model describes the marginal distribution of the responses. Using MI as a means for averaging, the procedure proposed to perform sensitivity analyses based on the family of PMMs described, consists in repeating the following steps for several choices of $\kappa(\mathbf{X}_{ij})$:

Step 1 Fit the LMM (6.2) without the term κR_{ij} to the available data using maximum likelihood.

Step 2 Impute the missing outcomes m times by drawing values from the LMM (6.2), using the parameter estimates obtained in Step 1 and the chosen $\kappa = \kappa(\mathbf{X}_{ij})$. This yields m completed datasets.

Step 3 Estimate the marginal scalar or vector parameter of interest $\boldsymbol{\theta}$ and its variance from each completed dataset by using a complete data method.

Step 4 Using Rubin's formulas, combine the m parameter and variance estimates yielded by Step 3 to obtain a final estimate of $\boldsymbol{\theta}$ and its variance, construct CIs and perform hypothesis tests.

Note that Step 1 corresponds to the primary analysis and does not have to be repeated for each choice of $\kappa(\mathbf{X}_{ij})$. By repeating Steps 2-4 for several choices of $\kappa(\mathbf{X}_{ij})$ and comparing the results obtained, the robustness of inferences about $\boldsymbol{\theta}$ to departures from the MAR assumption can be assessed. Graphical inspection of the results obtained by using so-called *sensitivity plots* can be useful, as will be illustrated in the analysis of the SMI study.

In principle, Steps 1 and 3 are straightforward using available software, e.g. use package *lme4* in R (Bates et al., 2012) or PROC MIXED in SAS (SAS Institute Inc., 2003) for Step 1. Step 4 is performed by applying the formulas presented in Section 3.1. Finally, Step 2 can be performed by using a modified version of the MI procedure described Section 3.2. The modification consists in replacing step (b) by:

(b') For each missing outcome Y_{ij}, calculate the linear predictor implied by (6.2) and the chosen $\kappa(\mathbf{X}_{ij})$:

$$\eta_{ij}^{(l)} = \mathbf{X}'_{ij}\boldsymbol{\beta}^{(l)} + \mathbf{Z}'_{ij}\mathbf{b}_i^{(l)} + \kappa(\mathbf{X}_{ij})R_{ij}.$$

Note that the estimates $\hat{\boldsymbol{\beta}}$, $\hat{\text{var}}(\hat{\boldsymbol{\beta}})$, $\hat{\sigma}^2$, $\hat{\mathbf{b}}_i$ and $\hat{\text{var}}(\hat{\mathbf{b}}_i)$ $(i = 1, \ldots, n)$ required by the imputation procedure are those obtained in Step 1.

Of course, the modified imputation procedure coincides with the original imputation procedure of Section 3.2 when $\kappa(\mathbf{X}_{ij}) = 0$. Recall that the latter was implemented as a method to be passed on to the function *mice* of the R *mice* package (van Buuren and Groothuis-Oudshoorn, 2011). Thus, to implement the modified procedure it suffices to modify the imputations yielded by *mice* with the original procedure by adding $\kappa(\mathbf{X}_{ij})$. Example R code, showing how to obtain the desired imputations following (6.2) with the functions for the original procedure, can be found in the Supplementary Material of the corresponding published manuscript (Moreno-Betancur and Chavance, 2013).

Recall from Section 3.1 that, in principle, the imputation and analysis models must be congenial to obtain valid inferences with MI. When imputing with a PMM like (6.2) and analyzing with a model for the marginal distribution of the outcomes, congeniality is violated because the former includes R as a predictor while the latter does not. Nevertheless, as pointed out by Meng (1994) and Molenberghs and Kenward (2007), this is not catastrophic; allowing for this violation makes MI a valuable tool in several situations, particularly when performing sensitivity analyses. Some previous examples of MI in the latter setting can be found in the literature (Little and Yau, 1996; Minini and Chavance, 2004a). Furthermore, uncongenial procedures in which the imputation

model is more general than the analysis model yield valid parameter estimates (Schafer, 1999). On the other hand, several authors have shown that when the imputation model is misspecified or when the imputation and analysis models are uncongenial, Rubin's variance estimator (3.2) may be biased (Meng, 1994; Robins and Wang, 2000). The behavior of the MI inferences yielded by the proposed procedure will be assessed in the simulation study presented next.

6.3 Simulation study

We performed a simulation study to explore the performance of the proposed approach. The study design was exactly the same as in the simulation study of Chapter 3. We considered the realistic situation where the family of models defined by (6.2) does not include the true model for the outcomes given the covariates and R that is induced by the data generation process. For this purpose, we simulated MNAR data under a selection model, that is, by generating data according to the design described in Section 3.3.1 and then assigning each individual a drop-out probability at each visit that depended on the current outcome. Then, we obtained inferences on θ under several scenarios by following our procedure, that is, by assuming a PMM from family (6.2) for several choices of $\kappa = \kappa(\mathbf{X}_{ij})$. In performing this analysis, we first identified a scenario in which the assumed PMM was a good approximation of the true selection model used to generate the data in the sense that it led to an unbiased estimate of θ. This allowed us to assess the inferences obtained with our procedure when the assumed model was close to the true model regarding the parameter of interest θ. Then, taking the latter scenario as reference, we assessed the behavior of the inferences obtained with increasing departures from the true model. Particular emphasis was given to evaluating the performance of the MI variance estimator in this setting in which the imputation model was misspecified and the congeniality condition was violated.

6.3.1 Simulation of drop-outs

We considered two types of MNAR drop-out mechanisms, assuming that the baseline outcome was observed for all individuals. In the first type of mechanism, drop-out probability depended on the first missing outcome, with subjects with higher values having a smaller probability of dropping out than those with lower values (henceforth called the MNAR1 mechanism) or vice-versa (MNAR2 mechanism). In the second type of mechanism, the drop-out probability depended on the first missing outcome and the individual's group. More precisely, two scenarios were considered: first, subjects in the treatment group had a 0.1 marginal probability of dropping out whereas those in the control group had a 0.4 probability, and within each group, the probability of dropping out was lower for individuals with higher values than for those with lower values (MNARG1 mechanism); second, subjects in the treatment group had a 0.4 marginal drop-out probability whereas those in the control group had a 0.1 probability, and within each group, the probability of dropping out was higher for individuals with higher values than for those with lower values (MNARG2 mechanism).

For each mechanism, the probability of dropping out was the same at each visit, except baseline. For the MNAR1 and MNAR2 scenarios, in which drop-out probability depended only on outcome values, we assigned each individual i and time j, a probability p_{ij} of dropping out which depended on the current outcome Y_{ij} through a logistic model:

$$\text{logit}(p_{ij}) = \lambda_0 + \lambda_1 Y_{ij}.$$

Here λ_0 and λ_1 were chosen so that they yielded a marginal probability of dropping out of p=0.4. For the MNARG1 and MNARG2 scenarios, in which the drop-out probability depended on outcome values and group, a separate logistic model, like the one above, was used to simulate drop-outs for each group, generating different marginal drop-out probabilities (0.1 or 0.4) within each group.

6.3.2 Analysis of the generated data sets

For each drop-out mechanism and value of θ, 1000 data sets of size $n = 200$ were generated. To analyze each data set we followed the procedure outlined in Section 6.2.2. The model fitted to the available data in Step 1 was an LMM including fixed effects for the group indicator X_i, the measurement time t_j $(= j)$ and their interaction $X_i t_j$, and random intercepts and slopes. For Step 2, we defined $\kappa(X_i, t_j) = k_{X_i}$. Thus, the distribution of the missing data at visit 5 had a shift of magnitude k_{X_i} with respect to the observed data distribution. Several analyses were performed, each corresponding to different values of the informativity parameters k_0 and k_1 which were chosen as described in the next paragraphs. For each pair of parameters (k_0, k_1) considered, $m = 10$ imputations of the missing data were performed. In Step 3, parameter θ was estimated by fitting model (3.4) to each completed data set. In Step 4, final parameter and variance estimates, CIs and significance tests were obtained.

In each scenario, the measures computed to summarize the results of each analysis across the 1000 datasets were the same as those computed in the simulation study of Chapter 3 (see Section 3.3.3). Additional measures included in the tables to facilitate comparisons were: the mean bias (MB) of the parameter estimates, denoted by MB$(\hat{\theta})$; and the mean relative bias (MRB) of the standard deviation estimator relative to the observed standard deviation, denoted by MRB$(\hat{\sigma}_\theta)$, and calculated as $100\times\{\hat{\sigma}_\theta - \text{SD}\,(\hat{\theta}^{(s)})\}/\text{SD}\,(\hat{\theta}^{(s)})$.

For each drop-out mechanism, when choosing the values of the informativity parameters to be considered, we first identified some values $\hat{k}_0$ and $\hat{k}_1$ for which the implied PMM was a good approximation of the true underlying MNAR selection model, at least for the purpose of obtaining an approximately unbiased estimate of θ. This scenario allowed us to assess the inferences obtained with our procedure when the imputation model was close to the true model with regards to the parameter of interest. To determine $\hat{k}_0$ and $\hat{k}_1$, we inspected the biases in the estimates of θ obtained with the procedure described across several values for k_0 and k_1. The values retained for $\hat{k}_0$ and

$\hat{k}_1$ by inspection were one of many other possibilities that also led to approximately unbiased estimates of θ, and for which the associated PMMs may thus also be considered to be good approximations of the true MNAR model with regards to θ. However, in an abuse of language, we will henceforth refer to the PMM associated with the retained values for $\hat{k}_0$ and $\hat{k}_1$ as the 'Best MNAR' model. We emphasize that the latter designation refers only to MNAR PMMs of the form (6.2) and is meant in an approximate sense. In particular, note that the 'Best MNAR' model is a good approximation of the true model at visit 5 but not necessarily at other visits because estimation of θ requires only the outcomes at visit 5.

In practice, the analyst cannot reproduce the approach used to find $\hat{k}_0$ and $\hat{k}_1$ because the real value of θ, necessary for bias assessment, is what he is trying to estimate in the first place. In fact, the main aim of the sensitivity analysis set-up is to try to encompass the 'Best MNAR' model in the scenarios considered in order to obtain a set of correct inferences among the several sets of results yielded by the procedure. The analyst may achieve this by (i) making reasonable and well-thought assumptions about the missing data distribution, which he expresses through the choice of $\kappa(\mathbf{X}_{ij})$ and (ii) performing several analyses corresponding to different scenarios, e.g. by varying the values of the informativity parameters. We tried to mimic this aspect of the approach in the simulation study by choosing values of k_0 and k_1 that covered several scenarios relative to the 'best' values, $\hat{k}_0$ and $\hat{k}_1$. This enabled us to assess the behavior of the inferences obtained with increasing departures from the true model. In particular, we considered the MAR scenario, which corresponded to $k_0 = k_1 = 0$.

6.3.3 Results

Results for the each of the drop-out mechanisms are shown in Tables 6.1-6.4. The upper and lower panels of each table show the results under $H_0(\theta = 0)$ and $H_1(\theta \neq 0)$, respectively. In describing and analyzing the results, we will first focus on lines 1–4 of each panel, which show results for analyses with increasing departures from the MAR

scenario, expressed as multiples of $\hat{k}_0$ and $\hat{k}_1$, with line 3 corresponding to the 'Best MNAR' model. Analyses for other choices of k_0 and k_1 and the CC analysis are shown in the next lines, and commented on in the last paragraph of this section.

When performing a sensitivity analysis, we seek to observe the variation in parameter estimates across different scenarios. The latter allows us to determine a range of values in which the true value may lie. Thus, in our simulations, we do not expect all estimates of θ to be unbiased. In fact, only the 'Best MNAR' model is expected to result in an approximately unbiased estimate of θ because it was defined by this criterion, but in practice there is no way of identifying this model so this is irrelevant. In the other analyses, there may or may not be biases depending on several factors, including the type of drop-out mechanism, and this is what we actually seek to observe and assess. Under H_0, there were no biases in the MNAR1 and MNAR2 scenarios as expected because the drop-out mechanism was the same in both groups, and the small differences observed between the estimates obtained in the two extreme analyses ('MAR' and 'Other 2') are non-significant. On the other hand, under H_0 in the MNARG1 and MNARG2 scenarios and under H_1 in all scenarios (except two analyses in the MNAR1 scenario) there were biases when departing from the 'Best MNAR' analysis, which itself led to unbiased estimates as expected. However, in the MNAR1 and MNAR2 scenarios the observed biases were much smaller than in the MNARG1 and MNARG2 scenarios: in the extreme analyses the magnitude of the mean bias was about 0.1 for the former versus 0.4 for the latter. Thus, as expected, the range of variation of the expected value of the estimator was wider when considering data that were MNARG1 and MNARG2 compared with MNAR1 and MNAR2 data.

The MRB of the mean standard deviation estimates was always positive (except in one case where it was very close to zero) which implies that the MI variance estimator tended to yield conservative estimates, that is, to overestimate the variance. This is probably due to misspecification of the imputation model and uncongeniality of the imputation and analysis models. The bias in the standard deviation estimates was always moderate, with $|\text{MRB}| \leq 10\%$ all scenarios considered. However, this overestimation

did result in higher coverage probabilities than expected.

For the 'Best MNAR' analysis, and for all other analyses where the coefficient estimator was unbiased, the overestimation in the variance estimates led to conservative but acceptable CPs. Furthermore, in these analyses we observed controlled type I error rates under $H_0(\theta = 0)$, with the tests having a size equal to or less than the chosen target (5%).

Line 5 of each panel in each table (the 'Other 3' analysis), shows the results obtained when setting the informativity parameters in both groups to be equal to the mean of the 'best' parameters. Since in every scenario it resulted that $\hat{k}_0$ and $\hat{k}_1$ were very close, if not equal, this scenario gave results that were similar to those of the 'Best MNAR' analysis. When $\hat{k}_0 = \hat{k}_1$, there were still some differences in the results between this analysis and the 'Best MNAR' analysis because of the random nature of the imputation procedure. Lines 6–7 of each panel (the 'Other 4' and 'Other 5' analyses) correspond to other analyses with more extreme departures from the 'Best MNAR' analysis than those considered above. They led to biased coefficient estimates in all cases, with biases generally of larger magnitude than those observed in the previous analyses. The CC analysis (line 8 in each panel) led to biased coefficient estimates in all scenarios except, as expected, in the MNAR1 and MNAR2 scenarios under H_0 in which the drop-out mechanisms of both groups were equal. It is important to note that the biases of the CC analysis in the MNARG1 and MNARG2 scenarios were of much larger magnitude than the biases observed for any of the other analyses.

To summarize, in the scenarios where the assumed PMM was close to the true model with regards to the parameter of interest, the coefficient estimator displayed satisfactory statistical properties in terms of CPs and type I error rates. Furthermore, we were able to observe the variation in the expected value of the coefficient estimator across several scenarios, including MAR and other departures from the true model, with the width of the range of variation depending on the missingness mechanism and other factors. Our approach is therefore suitable for assessing the sensitivity of inferences to assumptions about the missing data.

Table 6.1: Sensitivity analysis of inferences on θ in the MNAR1 scenario, in which drop-out probability was inversely related to the first missing outcome. Results of 1000 simulations.

θ	Analysis	k_0	k_1	$\hat{\theta}$	MB($\hat{\theta}$)	$\hat{\sigma}_\theta$	SD ($\hat{\theta}^{(s)}$)	MRB($\hat{\sigma}_\theta$)	CP	% reject H_0	MSE
0	MAR	0	0	-0.023	-0.023	0.836	0.781	7.1	96.6	4.0	0.610
	Other 1	$\hat{k}_0/2$	$\hat{k}_1/2$	-0.028	-0.028	0.859	0.805	6.7	96.0	4.3	0.648
	Best MNAR	$\hat{k}_0$	$\hat{k}_1$	-0.038	-0.038	0.881	0.829	6.3	95.7	4.5	0.688
	Other 2	$2\hat{k}_0$	$2\hat{k}_1$	-0.065	-0.065	0.929	0.873	6.4	95.5	4.7	0.766
	Other 3	$(\hat{k}_0+\hat{k}_1)/2$	$(\hat{k}_0+\hat{k}_1)/2$	-0.008	-0.008	0.881	0.828	6.5	96.1	4.2	0.684
	Other 4	$\hat{k}_0/2$	$2\hat{k}_1$	-0.985	-0.985	0.895	0.846	5.8	83.3	17.9	1.685
	Other 5	$2\hat{k}_0$	$\hat{k}_1/2$	0.907	0.907	0.892	0.834	6.8	85.9	14.8	1.519
	CC	–	–	-0.027	-0.027	0.817	0.833	-2.0	95.1	4.9	0.694
1	MAR	0	0	0.983	-0.017	0.832	0.811	2.5	95.6	22.4	0.658
	Other 1	$\hat{k}_0/2$	$\hat{k}_1/2$	1.014	0.014	0.854	0.825	3.5	95.6	21.9	0.680
	Best MNAR	$\hat{k}_0$	$\hat{k}_1$	1.034	0.034	0.871	0.863	0.9	95.2	22.2	0.746
	Other 2	$2\hat{k}_0$	$2\hat{k}_1$	1.096	0.096	0.918	0.925	-0.7	94.3	23.2	0.863
	Other 3	$(\hat{k}_0+\hat{k}_1)/2$	$(\hat{k}_0+\hat{k}_1)/2$	1.055	0.055	0.872	0.866	0.8	94.6	22.3	0.751
	Other 4	$\hat{k}_0/2$	$2\hat{k}_1$	0.148	-0.852	0.885	0.870	1.8	86.1	5.7	1.482
	Other 5	$2\hat{k}_0$	$\hat{k}_1/2$	1.946	0.946	0.885	0.875	1.2	82.9	58.9	1.660
	CC	–	–	0.759	-0.241	0.812	0.830	-2.1	93.9	16.2	0.746

For $\theta = 0$, $\hat{k}_0 = -1.4$ and $\hat{k}_1 = -1.5$
For $\theta = 1$, $\hat{k}_0 = -1.5$ and $\hat{k}_1 = -1.5$

Table 6.2: Sensitivity analysis of inferences on θ in the MNAR2 scenario, in which drop-out probability was positively related to the first missing outcome. Results of 1000 simulations.

θ	Analysis	k_0	k_1	$\hat{\theta}$	MB($\hat{\theta}$)	$\hat{\sigma}_\theta$	SD ($\hat{\theta}^{(s)}$)	MRB($\hat{\sigma}_\theta$)	CP	% reject H_0	MSE
0	MAR	0	0	0.005	0.005	0.818	0.797	2.7	96.5	3.6	0.635
	Other 1	$\hat{k}_0/2$	$\hat{k}_1/2$	0.006	0.006	0.836	0.817	2.4	95.8	4.2	0.667
	Best MNAR	$\hat{k}_0$	$\hat{k}_1$	-0.006	-0.006	0.858	0.836	2.6	96.7	3.4	0.699
	Other 2	$2\hat{k}_0$	$2\hat{k}_1$	-0.014	-0.014	0.907	0.897	1.1	95.5	4.9	0.804
	Other 3	$(\hat{k}_0+\hat{k}_1)/2$	$(\hat{k}_0+\hat{k}_1)/2$	0.004	0.004	0.858	0.840	2.1	96.0	4.2	0.705
	Other 4	$\hat{k}_0/2$	$2\hat{k}_1$	0.833	0.833	0.873	0.861	1.4	84.0	16.5	1.436
	Other 5	$2\hat{k}_0$	$\hat{k}_1/2$	-0.837	-0.837	0.872	0.855	2.1	85.8	14.7	1.430
	CC	–	–	0.022	0.022	0.797	0.836	-4.7	93.7	6.3	0.699
1	MAR	0	0	0.931	-0.069	0.830	0.757	9.6	96.3	17.2	0.578
	Other 1	$\hat{k}_0/2$	$\hat{k}_1/2$	0.974	-0.026	0.841	0.780	7.9	95.9	18.6	0.608
	Best MNAR	$\hat{k}_0$	$\hat{k}_1$	1.015	0.015	0.864	0.802	7.8	96.0	19.2	0.642
	Other 2	$2\hat{k}_0$	$2\hat{k}_1$	1.100	0.100	0.910	0.853	6.8	95.9	21.0	0.736
	Other 3	$(\hat{k}_0+\hat{k}_1)/2$	$(\hat{k}_0+\hat{k}_1)/2$	0.999	-0.001	0.863	0.801	7.8	96.6	18.9	0.640
	Other 4	$\hat{k}_0/2$	$2\hat{k}_1$	1.883	0.883	0.885	0.817	8.3	86.1	56.4	1.447
	Other 5	$2\hat{k}_0$	$\hat{k}_1/2$	0.195	-0.805	0.875	0.820	6.7	87.8	4.4	1.321
	CC	–	–	0.644	-0.356	0.806	0.801	0.7	92.5	11.3	0.767

For $\theta = 0$, $\hat{k}_0 = 1.5$ and $\hat{k}_1 = 1.5$
For $\theta = 1$, $\hat{k}_0 = 1.4$ and $\hat{k}_1 = 1.4$

Table 6.3: Sensitivity analysis of inferences on θ in the MNARG1 scenario, in which drop-out probability was inversely related to the first missing outcome within each group, and each group had different marginal drop-out probabilities (0.4 for the control group and 0.1 for the treatment group). Results of 1000 simulations.

θ	Analysis	k_0	k_1	$\hat{\theta}$	MB($\hat{\theta}$)	$\hat{\sigma}_\theta$	SD ($\hat{\theta}^{(s)}$)	MRB($\hat{\sigma}_\theta$)	CP	% reject H_0	MSE
0	MAR	0	0	-0.375	-0.375	0.796	0.767	3.8	92.6	7.4	0.729
	Other 1	$\hat{k}_0/2$	$\hat{k}_1/2$	-0.174	-0.174	0.810	0.785	3.2	94.9	5.4	0.646
	Best MNAR	$\hat{k}_0$	$\hat{k}_1$	0.034	0.034	0.824	0.797	3.4	95.2	5.0	0.637
	Other 2	$2\hat{k}_0$	$2\hat{k}_1$	0.442	0.442	0.864	0.838	3.1	92.8	7.8	0.896
	Other 3	$(\hat{k}_0+\hat{k}_1)/2$	$(\hat{k}_0+\hat{k}_1)/2$	0.011	0.011	0.827	0.802	3.0	95.3	5.0	0.643
	Other 4	$\hat{k}_0/2$	$2\hat{k}_1$	-0.438	-0.438	0.827	0.802	3.1	92.3	7.9	0.835
	Other 5	$2\hat{k}_0$	$\hat{k}_1/2$	0.704	0.704	0.845	0.824	2.6	88.3	12.3	1.174
	CC	–	–	-1.413	-1.413	0.788	0.783	0.6	58.6	41.4	2.610
1	MAR	0	0	0.584	-0.416	0.792	0.771	2.8	92.9	10.8	0.766
	Other 1	$\hat{k}_0/2$	$\hat{k}_1/2$	0.788	-0.212	0.806	0.788	2.3	94.6	16.1	0.665
	Best MNAR	$\hat{k}_0$	$\hat{k}_1$	1.001	0.001	0.823	0.805	2.2	95.6	23.1	0.648
	Other 2	$2\hat{k}_0$	$2\hat{k}_1$	1.424	0.424	0.860	0.854	0.6	92.6	38.9	0.909
	Other 3	$(\hat{k}_0+\hat{k}_1)/2$	$(\hat{k}_0+\hat{k}_1)/2$	1.017	0.017	0.824	0.805	2.3	95.0	22.9	0.648
	Other 4	$\hat{k}_0/2$	$2\hat{k}_1$	0.547	-0.453	0.825	0.801	3.0	92.2	9.3	0.846
	Other 5	$2\hat{k}_0$	$\hat{k}_1/2$	1.667	0.667	0.845	0.839	0.7	89.3	51.8	1.149
	CC	–	–	-0.566	-1.566	0.792	0.771	2.7	48.7	9.8	3.047

For $\theta = 0$, $\hat{k}_0 = -1.4$ and $\hat{k}_1 = -1.4$
For $\theta = 1$, $\hat{k}_0 = -1.4$ and $\hat{k}_1 = -1.5$

Table 6.4: Sensitivity analysis of inferences on θ in the MNARG2 scenario, in which drop-out probability was positively related to the first missing outcome within each group, and each group had different marginal drop-out probabilities (0.1 for the control group and 0.4 for the treatment group). Results of 1000 simulations.

θ	Analysis	k_0	k_1	$\hat{\theta}$	MB($\hat{\theta}$)	$\hat{\sigma}_\theta$	SD ($\hat{\theta}^{(s)}$)	MRB($\hat{\sigma}_\theta$)	CP	% reject H_0	MSE
0	MAR	0	0	-0.395	-0.395	0.789	0.745	5.9	93.9	6.7	0.710
	Other 1	$\hat{k}_0/2$	$\hat{k}_1/2$	-0.194	-0.194	0.806	0.754	7.0	96.1	4.4	0.605
	Best MNAR	$\hat{k}_0$	$\hat{k}_1$	0.008	0.008	0.822	0.778	5.6	96.7	3.4	0.605
	Other 2	$2\hat{k}_0$	$2\hat{k}_1$	0.417	0.417	0.858	0.827	3.8	92.5	7.6	0.857
	Other 3	$(\hat{k}_0+\hat{k}_1)/2$	$(\hat{k}_0+\hat{k}_1)/2$	-0.012	-0.012	0.823	0.777	5.8	96.2	4.2	0.604
	Other 4	$\hat{k}_0/2$	$2\hat{k}_1$	0.676	0.676	0.841	0.800	5.2	90.2	10.0	1.095
	Other 5	$2\hat{k}_0$	$\hat{k}_1/2$	-0.463	-0.463	0.823	0.776	6.0	93.4	7.0	0.816
	CC	–	–	-1.376	-1.376	0.784	0.744	5.4	58.4	41.6	2.446
1	MAR	0	0	0.583	-0.417	0.802	0.760	5.6	92.8	10.5	0.750
	Other 1	$\hat{k}_0/2$	$\hat{k}_1/2$	0.792	-0.208	0.815	0.775	5.2	95.8	15.4	0.643
	Best MNAR	$\hat{k}_0$	$\hat{k}_1$	1.019	0.019	0.834	0.791	5.4	96.8	21.6	0.626
	Other 2	$2\hat{k}_0$	$2\hat{k}_1$	1.440	0.440	0.873	0.838	4.2	93.4	36.3	0.895
	Other 3	$(\hat{k}_0+\hat{k}_1)/2$	$(\hat{k}_0+\hat{k}_1)/2$	1.042	0.042	0.835	0.796	4.8	96.5	22.4	0.635
	Other 4	$\hat{k}_0/2$	$2\hat{k}_1$	1.736	0.736	0.854	0.825	3.5	87.5	54.3	1.221
	Other 5	$2\hat{k}_0$	$\hat{k}_1/2$	0.506	-0.494	0.838	0.797	5.1	92.2	8.9	0.878
	CC	–	–	-0.650	-1.650	0.800	0.770	3.9	46.3	11.6	3.315

For $\theta = 0$, $\hat{k}_0 = 1.4$ and $\hat{k}_1 = 1.5$
For $\theta = 1$, $\hat{k}_0 = 1.5$ and $\hat{k}_1 = 1.4$

6.4 Application to the SMI study

Recall from Section 1.1 that six scores indicative of the quality of sleep were recorded for each of the patients in the SMI study. The primary endpoint was the WASO score and the other five scores were secondary endpoints. For a given score, let Y_{ij} denote the outcome of patient i at the j^{th} visit for $i \in \{1, \ldots, 962\}$ and $j \in \{0, \ldots, 5\}$, and X_i denote the group indicator, with $X_i = 1$ for the treatment group and $X_i = 0$ for the control group. The parameter of interest was the treatment effect θ, defined as the difference between the expected change from baseline of the scores in the treatment and control groups at visit 5, i.e. $\theta = E(Y_{i5} - Y_{i0}|X_i = 1) - E(Y_{i5} - Y_{i0}|X_i = 0) = E(Y_{i5}|X_i = 1) - E(Y_{i5}|X_i = 0)$, the latter equality arising from the randomization at baseline. Thus, with complete data, the treatment effect could be estimated by fitting the linear model

$$Y_{i5} = \theta_0 + \theta X_i + \epsilon_i, \tag{6.3}$$

where the ϵ_i's are i.i.d. zero-mean Gaussian errors. In addition to obtaining a point estimate, a CI for θ may be built and a hypothesis test may be conducted to test whether θ is significantly different from zero. In the following sections we focus on this last task, i.e. on testing whether there is a significant effect of treatment, and on assessing the sensitivity of the outcome of this test to assumptions about the drop-out mechanism. This test is often of high relevance in clinical trials because drugs are marketed only if their observed effect is found to be statistically significant. Thus, in the clinical trial setting, positive findings, i.e. results leading to the conclusion that there is a significant treatment effect, should be the main focus of sensitivity analyses.

6.4.1 Primary analysis

In a primary analysis of these data, the drop-out mechanism may be assumed to be ignorable so that valid inferences about θ may be obtained through a direct likelihood

analysis based on an LMM with random slope and intercept:

$$Y_{ij} = \beta_0 + \beta_1 X_i + \beta_2 t_j + \beta_3 X_i t_j + b_{0i} + b_{1i} t_j + \varepsilon_{ij}, \tag{6.4}$$

where t_j is the time of visit j and $\mathbf{b}_i = (b_{0i}, b_{1i})'$, $i = 1, \ldots, n$, are the i.i.d. zero-mean Gaussian vectors of subject-specific random effects, which are assumed to be independent of X_i and of the zero-mean i.i.d. Gaussian errors ε_{ij}. In this model, $\theta = \beta_1 + \beta_3 t_5$.

An alternative to the direct likelihood approach is to use the MI approach presented in Chapter 3, with (6.4) as the imputation model and (6.3) as the analysis model. Even though MI may be slightly less efficient than direct likelihood (cf. Section 1.3.2), we preferred to use MI because its two-stage nature provided us with additional flexibility in two aspects. First, after careful residual analyses, some scores had to be transformed to improve the fit of model (6.4) regarding the normality assumption of the errors. The transformed scores were: WASOt = $\sqrt{\text{WASO}}$, NAWt = $\sqrt{\text{NAW}}$ and SOLt = Log(SOL+1). Following Rubin (1987), we first performed imputations from model (6.4) built on the transformed scores, then applied the inverse transformation to the imputed scores and finally fitted model (6.3). This enabled us to directly estimate the effect of treatment on the untransformed scores, which was our main interest. The estimate of θ obtained in this way was normal because the asymptotic conditions implied the normality of the means in each group at visit 5. Hence, the MI approach allowed us to correct for the non-normality of the scores but still obtain an estimate of the actual parameter of interest with nice properties. Second, even though the distribution of the residuals obtained when fitting model (6.4) on the transformed scores better approximated normality, for most of the scores it had heavier tails than expected for a normal distribution. The MI approach allowed us to investigate the sensitivity of the results to this phenomenon. We did so by modifying part (c) of the imputation procedure, drawing errors from a logistic distribution instead of a normal distribution, the former resembling the latter but with heavier tails. All analyses were performed

with the original and modified imputation procedures. The results were similar in both cases, so only the results pertaining to the Gaussian case are presented.

The results obtained with the MI approach with $m = 20$ imputations are presented in Table 6.5. For comparison, we also provide the results obtained when performing a CC analysis, i.e. when fitting model (6.3) to the available outcomes at visit 5, thus excluding any data from patients who dropped-out from the analysis. With the MI approach, the treatment effect was highly significant for the WASO and NAW scores, significant for the SLREF and FEELC scores and non-significant for the TST and SOL scores at a 5% level. The CC analysis led to similar conclusions for all scores except for the TST score, with all the estimates being less precise as expected. For the TST score, the treatment effect was larger and significant at a 5% level according to the CC analysis. The discrepancy between the two analyses is not surprising since the CC analysis requires the stronger assumption of covariate-dependent drop-out to guarantee unbiased estimates. In fact, since the covariate-dependent drop-out assumption implies the MAR assumption, if the former held then both analyses should yield very similar results. Since this is not the case, we can suspect that the CC estimate is biased and that its corresponding significance test is unreliable.

6.4.2 Sensitivity analysis for θ

A sensitivity analysis was performed following the procedure outlined in Section 6.2.2 to evaluate the robustness of the results obtained with the MI approach in the primary analysis regarding the significance of the effect of treatment at a 5% level. In this clinical trial it was particularly interesting to determine whether positive findings still held under MNAR scenarios. Thus, the analysis was performed for all scores except TST and SOL, for which the treatment effect was deemed non-significant by the primary MAR-based analysis.

In Step 1, model (6.4) was fitted to the available data after applying the transformations described in Section 6.4.1. For Step 2, since θ measured the effect of treatment

Table 6.5: Inferences on the treatment effect θ for each score, obtained from the MAR-based MI approach (MAR) and the CC analysis (CC).

	Analysis	$\hat{\theta}$	SE	95% CI	p-value
WASO	MAR	−14.31	3.10	(−20.39, −8.22)	<0.001
	CC	−14.25	3.25	(−20.63, −7.86)	<0.001
NAW	MAR	−0.34	0.08	(−0.51, −0.18)	<0.001
	CC	−0.38	0.10	(−0.57, −0.19)	<0.001
SLREF	MAR	−0.09	0.04	(−0.16, −0.01)	0.03
	CC	−0.09	0.04	(−0.18, 0.00)	0.04
FEELC	MAR	0.24	0.12	(0.01, 0.47)	0.04
	CC	0.26	0.13	(0.01, 0.52)	0.05
TST	MAR	7.39	4.14	(−0.73, 15.51)	0.07
	CC	10.06	4.37	(1.49, 18.63)	0.02
SOL	MAR	−0.90	0.82	(−2.52, 0.71)	0.27
	CC	−1.00	0.88	(−2.72, 0.72)	0.25

on the expected change from baseline at visit 5, we defined $\kappa(X_i, t_j) = k_{X_i}\hat{\varsigma}_5$, where the informativity parameter k_{X_i} of each group was allowed to vary between -0.5 and 0.5 in 0.05 increments, and $\hat{\varsigma}_5$ was the sample standard deviation of the transformed scores at visit 5. The latter was chosen because only the distribution of missing outcomes at visit 5 matters in the estimation of θ. Thus, the distribution of the missing data at visit 5 was shifted by k_{X_i} standard deviations with respect to the observed data distribution. For each pair of parameters (k_0, k_1), $m = 20$ imputations of the missing data were performed. In Step 3, parameter θ was estimated from the untransformed scores by fitting model (6.3), as in the primary analysis.

Results for scores with a highly significant (respectively, borderline significant) treatment effect, WASO and NAW (respectively, SLREF and FEELC), were very similar so we show only the results for the WASO and SLREF scores. Figure 6.1 shows contour plots identifying the regions where a given pair of parameters (k_0, k_1) yielded significant treatment effects ($p < 0.05$) for each score. For the WASO score, we can see that it was only in the small gray region, in which the control group is strongly advantaged and the treatment group strongly disadvantaged (as clinical improvement is associated with

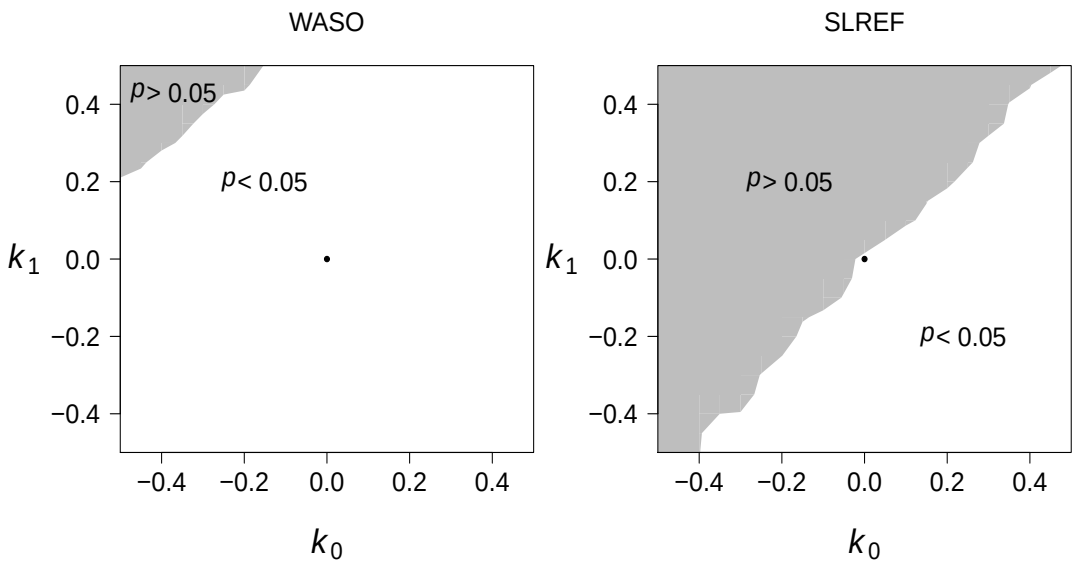

Figure 6.1: Contour plots for the WASO and SLREF scores showing the regions in which the treatment effect θ was significant (white) or non-significant (gray), according to a given pair of informativity parameters (k_0, k_1), corresponding to the control and treatment groups, respectively. The point plotted at (0,0) corresponds to the MAR analysis result.

a lower WASO score), that the treatment effect became non-significant. The contour plot for the SLREF score shows that, even though the treatment effect on this score was significant under MAR, this conclusion is very sensitive to departures from this assumption.

Several interesting scenarios can be visualized on the contour plots in Figure 6.1. A first example is the scenario in which both groups have the same shift, which is represented by the identity line ($k_0 = k_1$) on the contour plot. In this case, the missing data are implicitly advantaged or disadvantaged (depending on the sign of the informativity parameters), so the group with the highest drop-out rate, the control group, is also advantaged or disadvantaged. To further illustrate this situation, the treatment effect estimates for each score are plotted in Figure 6.2 against the common parameter $k = k_0 = k_1$ of both groups. This parameter was allowed to vary from -3 to 3 in increments of 0.5, so that the expectation of the missing data was up to three standard

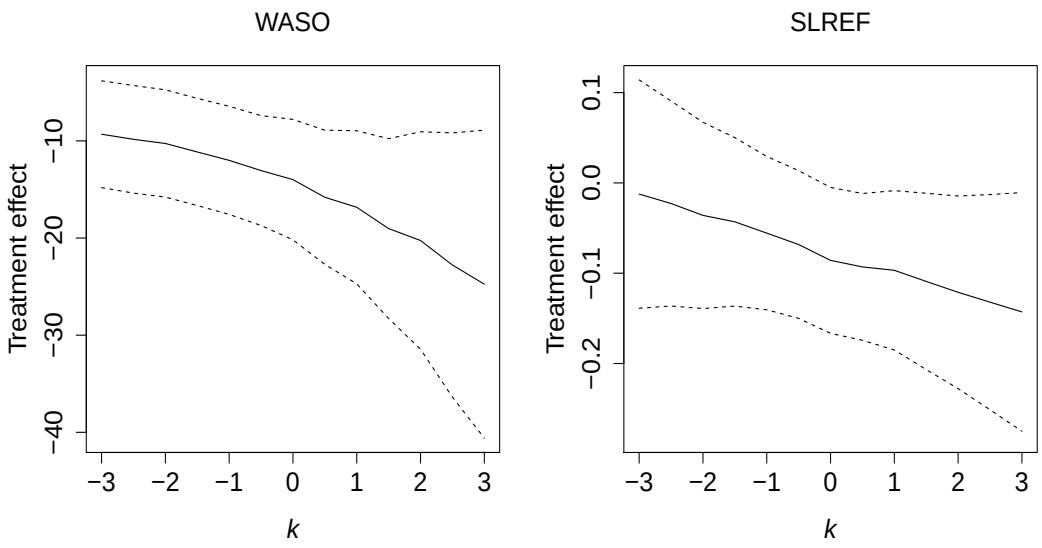

Figure 6.2: Sensitivity analysis of the estimate of the treatment effect θ for the WASO and SLREF scores when $k_0 = k_1 = k$. Dashed lines show the corresponding 95% confidence intervals.

deviations below or above that of the observed data. Note that the value of the curve at $k = 0$ corresponds to the MAR estimate given in Table 6.5. For the WASO score, we can see that even when the missing outcomes are advantaged ($k < 0$) and assumed to have a mean that is three standard deviations away from the mean of the observed outcomes, the treatment effect is still large and significant. Obviously, in the opposite case ($k > 0$), the treatment effect is still significant and even larger. On the other hand, for the SLREF score, we can see that when the missing outcomes are advantaged ($k < 0$), the treatment effect vanishes, reflecting the fact that the drop-out rate was higher in the control group than in the treatment group. Symmetrically, of course, the treatment effect was stronger and more significant when the missing outcomes were disadvantaged ($k > 0$).

Another notable case is when the groups have shifts of equal magnitude and opposite signs (i.e. the line $k_0 = -k_1$ in Figure 1), so that when one group is advantaged, the other is disadvantaged to an equal extent. An interesting scenario in this case is when the control group is advantaged and the treatment group disadvantaged (which

corresponds to $k_1 > 0$ for both scores). For the WASO score, we can see that it is in this extreme case and when $k_1 > 0.25$ approximately that the treatment effect becomes non-significant. For the SLREF score, the treatment effect is null and non-significant for almost every value of k_1 in this scenario.

6.4.3 Sensitivity analysis for β_3

We considered a second definition of the treatment effect, given by the difference in the time-slopes of the two groups when considering the transformed scores, i.e. parameter β_3 of model (6.4). We performed this analysis only for the (transformed) WASO and NAW scores, for which positive findings were found to be robust to missingness assumptions in the previous analyses. For the WASOt score, the first line of Table 6.6 shows the inferences on β_3 obtained by direct likelihood under the assumption of an ignorable drop-out mechanism, i.e. by fitting model (6.4) to the available data using maximum likelihood. The treatment effect defined in this way was significant for this score at a 5% level, with $\hat{\beta}_3 = -0.031 \pm 0.005$, as for the NAWt score (results not shown).

On the basis of the procedure outlined in Section 6.2.2, a sensitivity analysis was performed to evaluate how this conclusion could change under an MNAR drop-out mechanism. As before, in Step 1 model (6.4) was fitted to the available (transformed) scores. For Step 2, we considered a multidimensional parametrization of the term $\kappa(\mathbf{X}_{ij})$, defining $\kappa(X_i, t_j, S_i) = k^{(1)}_{X_i}\hat{\varsigma} + k^{(2)}_{X_i}\hat{\beta}_2(t_j - S_i)$, where $\hat{\varsigma}$ was the sample standard deviation of the observed transformed scores, $\hat{\beta}_2$ was the time-slope estimated in the direct likelihood analysis assuming ignorability, S_i was the time of the last observed outcome for individual i, and parameters $k^{(1)}_{X_i}$ and $k^{(2)}_{X_i}$ were positive or negative, and possibly depended on group. Thus, the group-specific intercepts and time-slopes of the missing data distribution were simultaneously controlled; while the group-specific intercepts were shifted by $k^{(1)}_{X_i}$ standard deviations with respect to the observed data distribution, the group-specific time-slopes were expressed as fractions of the time-slope of the observed data distribution. The purpose of the correction term $-S_i$ will become

Table 6.6: Sensitivity analysis of inferences on the time-slope difference between the two groups, β_3, for the WASOt score. Several values of the informativity parameters $k_{X_i}^{(1)}$ and $k_{X_i}^{(2)}$ are considered. The results obtained under MAR using maximum likelihood are also provided.

Scenario	$k_{X_i}^{(1)}$	$k_{X_i}^{(2)}$	$\hat{\beta}_3$	SE	p-value
MAR	–	–	−0.031	0.005	<0.001
Modified slopes	0	0.25	−0.030	0.006	<0.001
	0	0.5	−0.028	0.006	<0.001
	0	1	−0.026	0.007	<0.001
	0	$(-1)^{X_i} * 0.25$	−0.015	0.006	0.008
	0	$(-1)^{X_i} * 0.5$	0.005	0.006	0.397
Modified intercepts	−0.25	0	−0.029	0.006	<0.001
	−0.5	0	−0.027	0.007	<0.001
	−1	0	−0.021	0.006	0.001
	$(-1)^{1-X_i} * 0.1$	0	−0.018	0.006	0.001
	$(-1)^{1-X_i} * 0.25$	0	0.003	0.006	0.584
Modified slopes and intercepts	−0.25	0.25	−0.029	0.007	<0.001
	−0.5	0.5	−0.025	0.006	<0.001
	−1	1	−0.017	0.007	0.019
	$(-1)^{1-X_i} * 0.1$	0.5	−0.016	0.006	0.004
	$(-1)^{1-X_i} * 0.1$	1	−0.012	0.006	0.040
	−0.1	$(-1)^{X_i} * 0.25$	−0.014	0.006	0.027
	−0.25	$(-1)^{X_i} * 0.25$	−0.010	0.007	0.151
	$(-1)^{1-X_i} * 0.05$	$(-1)^{X_i} * 0.05$	−0.019	0.006	0.001
	$(-1)^{1-X_i} * 0.1$	$(-1)^{X_i} * 0.1$	−0.011	0.006	0.066

clear below, when we analyze in closer detail the scenarios implied by different choices of $k_{X_i}^{(1)}$ and $k_{X_i}^{(2)}$. As before, we performed $m = 20$ imputations. In Step 3, the estimate of β_3 was obtained by fitting model (6.4) on the transformed scores in each completed dataset using maximum likelihood.

In terms of the underlying clinical assumptions, three general scenarios can be distinguished regarding the choice of the informativity parameters $k_{X_i}^{(1)}$ and $k_{X_i}^{(2)}$. To illustrate, these scenarios are depicted for the WAOSt score in Figure 6.3 for a patient in the treatment group. From left to right, the first plot illustrates the scenario with $k_1^{(1)} = 0$ and $k_1^{(2)} \neq 0$. In this case, the patient is assumed to have experienced an increase in the absolute rate of change in time of the score after his last observed measurement. Here, the correction term $-S_i$ shifted the imputation model so that the expected trajectory of the missing data under MNAR intersected the expected trajectory under MAR at the patient's last visit. This means that the modification in the rate of change of the score in time is assumed to have occurred immediately after the patient's last visit. This assumption can be modified by taking a different correction factor, somewhere between S_i and S_{i+1}, or even after S_{i+1} for more intricate assumptions which may however be justifiable in certain settings. The second plot illustrates the scenario with $k_1^{(1)} \neq 0$ and $k_1^{(2)} = 0$. Here, the patient is assumed to have experienced a sudden drop in his score at some point between his last visit and the next visit, with no modification of the rate of change of the score in time after that decrease. The correction term $-S_i$ does not intervene in this case. In the third plot, $k_1^{(1)} \neq 0$ and $k_1^{(2)} \neq 0$, showing the resulting trajectory of a patient experiencing both a sudden drop in the score and a modification in the rate of change of the score in time. Here, the correction term $-S_i$ ensures that, up to the shift $k_1^{(1)}\hat{\varsigma}$, the MNAR and MAR trajectories cross at the patient's last visit. Thus, as before, the modification in the rate of change of the score in time is assumed to have occurred immediately after the patient's last visit. Meanwhile, the sudden drop in the score could have happened at any moment between the patient's last visit and the next visit.

Table 6.6 shows the results of the sensitivity analysis for the WASOt score, in each of

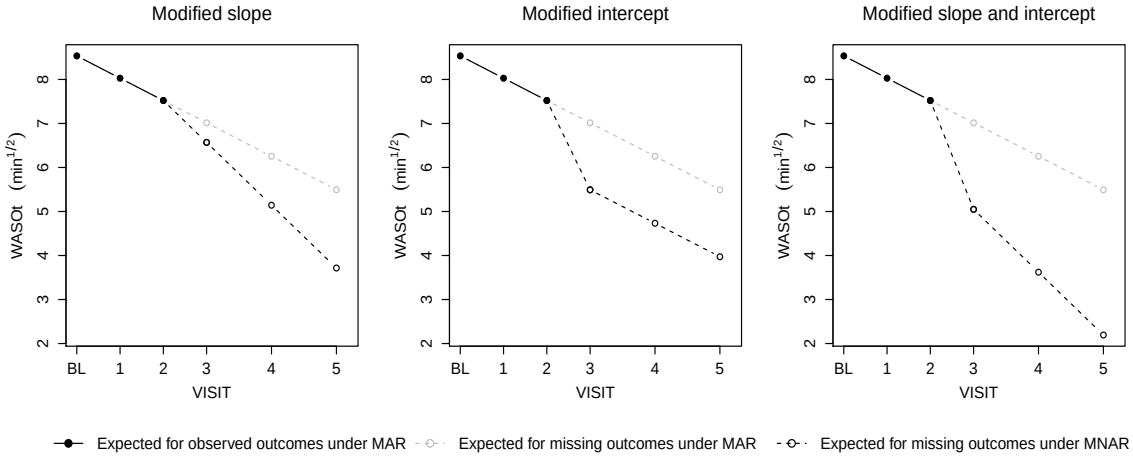

Figure 6.3: Expected trajectories of the WASOt score (measured in minutes$^{\frac{1}{2}}$) for a patient in the treatment group who dropped-out after visit 2. The plot shows the expected trajectory of his observed outcomes under MAR, of his missing outcomes under MAR and of his missing outcomes under the assumed MNAR PMM. In the latter, the slope, intercept or both are modified with respect to the expected trajectory under MAR. The vectors $(k_1^{(1)}, k_1^{(2)})$ of informativity parameters used in the plots were, from left to right, $(0, 1)$, $(-0.5, 0)$ *and* $(-0.5, 1)$. *BL=Baseline.*

the three general scenarios described above and for several values of $k_{X_i}^{(1)}$ and $k_{X_i}^{(2)}$. Lines 2–6 show results for the scenario in which only the slopes were modified, i.e. $k_{X_i}^{(1)} = 0$ and $k_{X_i}^{(2)} \neq 0$. Within this scenario, a first remarkable case was when $k_0^{(2)} = k_1^{(2)} > 0$ (lines 2–4). In this case, the magnitude of the mean time-slope of the missing data distribution was larger than that of the observed data, hence advantaging missing outcomes. We can see that the time-slope difference between the groups, β_3, remained significant even when the time-slope of the missing data distribution was twice that of the observed data distribution ($k_0^{(2)} = k_1^{(2)} = 1$). This finding also held true in the opposite case ($k_0^{(2)} = k_1^{(2)} < 0$), in which the time-slope of the missing data distribution was assumed to be smaller in magnitude than that of the observed data (results not shown). In lines 5–6, we present results for the interesting case in which $k_{X_i}^{(2)}$ depended on group, such that the treatment group was disadvantaged ($k_1^{(2)} < 0$) and the control group was advantaged ($k_0^{(2)} > 0$). When $k_{X_i}^{(2)} = (-1)^{X_i} * 0.25$, the treatment effect remained

significant, while it was no longer significant when $k_{X_i}^{(2)} = (-1)^{X_i} * 0.5$.

Concerning the scenario in which only the intercepts were modified, i.e. $k_{X_i}^{(1)} \neq 0$ and $k_{X_i}^{(2)} = 0$, lines 7–9 show results for the case $k_0^{(1)} = k_1^{(1)} < 0$, in which the missing outcomes were advantaged. The time-slope difference between the groups, β_3, remained significant in this case, even when the intercept of the missing data distribution was shifted by one standard deviation. The same result was found in the opposite case, when $k_0^{(1)} = k_1^{(1)} > 0$ (results not shown). Another interesting case was when $k_{X_i}^{(1)}$ depended on group (lines 10–11), such that the treatment group was disadvantaged ($k_1^{(1)} > 0$) and the control group was advantaged ($k_0^{(1)} < 0$). When $|k_{X_i}^{(1)}| = 0.1$ the treatment effect was still significant, but not when $|k_{X_i}^{(1)}| = 0.25$.

In the scenario in which both the intercepts and the slopes were modified, four distinct cases were considered. In the first case, neither of the informativity parameters depended on group (lines 12–14), with $k_0^{(1)} = k_1^{(1)} < 0$ and $k_0^{(2)} = k_1^{(2)} > 0$ so that the missing outcomes were doubly advantaged. The time-slope difference between the groups, β_3, remained significant in this case for all the values of the parameters considered. The same result was observed when the missing outcomes were doubly disadvantaged, taking $k_0^{(1)} = k_1^{(1)} > 0$ and $k_0^{(2)} = k_1^{(2)} < 0$ (results not shown). In the second case (lines 15–16), the intercept parameter depended on group such that the treatment group was disadvantaged ($k_1^{(1)} > 0$) and the control group was advantaged ($k_0^{(1)} < 0$). On the other hand, the slope parameter $k_{X_i}^{(2)}$ did not depend on group. When $|k_{X_i}^{(1)}| = 0.1$, the treatment effect remained significant for $k_0^{(2)} = k_1^{(2)} = 0.5$ and $k_0^{(2)} = k_1^{(2)} = 1$. In the third case (lines 17–18), $k_{X_i}^{(1)}$ did not depend on group while $k_{X_i}^{(2)}$ did, such that the treatment group was disadvantaged ($k_1^{(2)} < 0$) and the control group was advantaged ($k_0^{(2)} > 0$). When $|k_{X_i}^{(2)}| = 0.25$, the treatment effect remained significant for $k_0^{(1)} = k_1^{(1)} = -0.1$ but not for $k_0^{(1)} = k_1^{(1)} = -0.25$. In the last case, both parameters depended on group, such that both disadvantaged the treatment group and advantaged the control group (lines 19–20). When $|k_{X_i}^{(1)}| = |k_{X_i}^{(2)}| = 0.05$, the treatment effect was still significant, but it was no longer significant at a 5% level when $|k_{X_i}^{(1)}| = |k_{X_i}^{(2)}| = 0.1$.

When some similar analyses were conducted for the NAWt score, the results were

qualitatively similar to those obtained for the WASOt score. Thus, the treatment effect defined in this alternative way was also significant for the WASO and NAW scores under MAR, and remained so under a large range of departures from the MAR assumption.

6.5 Discussion

In this chapter, we proposed a method for performing sensitivity analyses when dealing with longitudinal data with drop-outs. The family of PMMs on which the approach is based is 'centered' at a model that would be a natural choice in a primary analysis assuming ignorable drop-out. Hence, the sensitivity parameter κ has a three-fold interpretation as quantifying the distance between the missing and observed data distributions, the degree of departure from the MAR assumption and the degree of departure from this primary analysis. The major strength of the approach lies in the fact that the data analyst needs to consider, and make explicit assumptions about, the main characteristics of the missing data distribution that may differ from the observed data distribution and affect inferences on the parameter of interest. Moreover, these assumptions can be easily expressed in the analysis through the choice of $\kappa = \kappa(\mathbf{X}_{ij})$. The proposed MI-based implementation procedure makes it easy to perform and compare several analyses over a range of possible choices for $\kappa(\mathbf{X}_{ij})$, determining a range of possible distributions for the missing outcomes. This is crucial in any sensitivity analysis. The simulation study results confirmed that this approach is suitable for assessing the sensitivity of inferences to assumptions about the missing data. In the SMI study, our approach provided insight about the robustness of the conclusions regarding treatment effect drawn from the WASO score, as well as from the other five scores which defined secondary endpoints: Some of the conclusions drawn under the MAR assumption were shown to be reliable, while others were found to be fragile and strongly dependent on missingness assumptions.

The estimates obtained with the proposed approach require careful interpretation because the model in Step 3 is likely to be misspecified. For example, in the SMI study,

the treatment effects in the observed and missing data were assumed to differ in some scenarios when generating imputations, with the result that the model fitted to the completed datasets was misspecified. However, the maximum likelihood estimator in a misspecified model can be interpreted as a natural estimator for the parameter that minimizes the Kullback–Leibler information criterion between the true and the assumed distributions (White, 1982; Kullback and Leibler, 1951). The latter should not be seen as a drawback since the goal of a sensitivity analysis is to investigate the stability of inferences obtained in a primary analysis across several MNAR scenarios, rather than to provide definitive estimates of the parameter of interest (Carroll et al., 2004).

Several authors have shown that when the imputation model is misspecified or when the imputation and analysis models are uncongenial, Rubin's variance estimator (3.2) may be biased (Meng, 1994; Robins and Wang, 2000). This was confirmed by the results of our simulation study, in which an upward bias was observed. Although the magnitude of the MRB of the estimator was always moderate, estimated at less than 10%, it did result in higher CPs than expected. Thus, an alternative variance estimator would be desirable. For instance, as suggested by Bakoyannis et al. (2010) in the competing risks setting, a bootstrap estimator could possibly be used to correct this bias, at least partially.

In the simulation study, the chosen values for the informativity parameters in terms of magnitude and sign enabled the relatively good behavior observed for the estimator when departing from the true model. The largest k_0 and k_1 considered had magnitude $|2\hat{k}_0| \approx |2\hat{k}_1| \approx 3$, that is, a small magnitude relative to the expected standard deviation of the outcomes at visit 5, which was $\sqrt{\text{var}(Y_{i5})} = \sqrt{27} \approx 5.2$ according to the simulation model. Larger deviations would lead of course to a poorer behavior of the estimator because such scenarios would represent larger departures from the true model. Also, for each mechanism, we only considered values of the informativity parameters which had the same sign as those corresponding to the 'Best MNAR' model. If values of the opposite sign were studied, we could expect the estimator to behave more poorly. In practice, the knowledge or intuition of the analyst regarding the plausibility of the

scenarios studied is of great importance to ensure the pertinence of the results obtained from a sensitivity analysis. For example, the missing data distribution would not be expected to deviate from that of the observed data by up to three standard deviations. Thus, the analyst can feel confident about inferences that are stable when considering reasonable departures from MAR, as was the case for the WASO and NAW scores in the SMI study.

When MNAR data were generated in the simulation study of Chapter 3, we indicated that the MI MAR-based analysis achieved partial bias correction because the available outcomes used to build the MI model provided partial information about the missing outcomes. Similar remarks apply to the MI procedure considered in this chapter, and explain the small magnitude of the 'best' values, $\hat{k}_0$ and $\hat{k}_1$, yielded by the MNAR mechanisms considered, as well as the moderate biases observed for our approach when imputing under MAR and other 'wrong' models, especially in the MNAR1 and MNAR2 mechanisms. Considering a stronger effect for the first missing outcome in the logistic model used to simulate drop-out would probably diminish this effect.

In the SMI study, there were two levels of missing data: missing daily scores and missing scores for a whole period (visit). Table 1.1 shows that the mean number of daily measurements available for each score and period was about the same for control and treatment group patients. If the daily selection mechanism was the same for both groups, then no bias would be introduced by this level of missingness. We did not find any reason why the groups' daily selection mechanisms should differ, so missing daily scores were not explicitly modeled in the analyses. Regardless of this additional complexity due to the data collection method, the data of this study were of high quality and thus suitable for performing sensitivity analyses. Indeed, for this purpose it is crucial that the drop-out rate is moderate; with a high rate of missing data, inferences rely more heavily on missingness assumptions and a small departure from the MAR assumption can invalidate the conclusion of an MAR-based primary analysis. Hence, it is essential to minimize the percentage of missing data in any clinical trial, as insisted upon by the PSI missing data expert group (Burzykowski et al., 2010).

From a clinical perspective, the results obtained for the SMI study suggest the efficacy of the treatment under analysis. To see this, we recall the meaning of each score. The NAW and WASO scores are the two quantitative measures that can be considered as the most indicative of the severity of the disorder, because they represent respectively the number of sleep interruptions during the night and the sum of their durations. Results for these two scores were very robust and reflected a marked treatment effect. On the other hand, SLREF and FEELC scores are qualitative scores indicating the quality of sleep and are therefore only indirectly indicative of the disorder's severity. In line with this appreciation, the results for these two scores yielded more uncertain conclusions about the effect of treatment. This was also true for the quantitative TST score, representing the total sleep time, which by itself is not directly indicative of the disorder severity, either. In fact, it is the TST relative to the total wake time (WASO) that would give a direct indication of severity. Finally, treatment effect was non-significant for the SOL score, which measures the sleep onset latency. The latter is not a pertinent score to determine the severity of SMI because affected people do not have trouble falling asleep.

Even though we focused on missing data due to drop-out, the approach can also be used to analyze data with intermittent missingness because LMMs can be fitted to this type of data. In fact, in the SMI study there were also a few intermittently-missing outcomes. Here, these missing outcomes were handled in the same way as missing values after drop-out, i.e. the same imputation model was used to impute them, but other strategies could be considered (Carpenter and Kenward, 2007). Furthermore, although the approach was described for continuous Gaussian outcomes, in principle it could be extended to any type of outcome variable that can be modeled by means of a generalized linear mixed model (e.g. binary and count data). Of course, the implementation and performance of the approach in these cases needs further investigation.

Chapter 7

Sensitivity analyses for competing risks with missing causes of failure

As with longitudinal data with drop-outs, in the competing risks setting with missing causes of failure it is not possible to assess from the observed data whether the missingness mechanism is MAR or MNAR. Thus, sensitivity analyses play an important role in this setting too, for assessment of the robustness of inferences to departures from unverifiable assumptions about the missingness mechanism. However, to our knowledge, neither MNAR modeling nor sensitivity analysis methodology have ever been considered in the missing cause of failure setting. In this chapter, we consider pattern-mixture models (PMMs) in this context, and propose an approach for performing sensitivity analyses following the ideas behind the methodology of Chapter 6. The proposed methodology, described in Section 7.1, is applicable to various competing risks regression models, particularly when modeling the cause-specific hazard (CSH) and the cumulative incidence function (CIF), either by parametric or semi-parametric regression techniques. In Section 7.2, we illustrate the approach by revisiting the analysis of the ECOG clinical trial. Some discussion points are given in Section 7.3.

7.1 Methodology

7.1.1 A family of PMMs for competing risks

MNAR modeling in the missing cause of failure setting requires consideration of the joint density of (T, D, M) for uncensored individuals $(U = 0)$, upon whom the missingness mechanism acts. Omitting covariates, the pattern-mixture modeling factorization of this density is given by

$$f(t, d, m \mid U = 0) = f(t, d \mid M = m, U = 0) \times f(m \mid U = 0) \tag{7.1}$$

$$= f(t, m \mid U = 0) \times f(d \mid T = t, M = m, U = 0). \tag{7.2}$$

Equation (7.1) shows that PMMs in this context imply different competing risks mechanisms for individuals with an observed cause of failure $(M = 0)$ and those with a missing cause $(M = 1)$. Note that the first factor is not identifiable because there is no data on D for individuals with missing cause. Equation (7.2) clarifies the source of this non-identifiability: The first factor represents the model for the data that is completely observed for all individuals, while the second factor represents the model for the incompletely observed data, i.e. the extrapolation model (cf. Section 6.1). The former model is identifiable from the observed data while the latter is not. Hence, additional, yet unverifiable assumptions are required to identify the second factor of (7.2). Furthermore, inferences about parameters indexing $f(t, d)$, which are usually the parameters of interest with competing risks, will generally depend on these assumptions.

Following the ideas of Section 6.1, a sensitivity analysis can be performed by means of a sensitivity parameter, that is, a non-identified parameter indexing the extrapolation model, conditional upon which the parameter of interest is identifiable. For this purpose, we consider the family of PMMs for which the extrapolation model is of the form:

$$\text{logit}\{\Pi(\mathbf{X}, T, M)\} = \mathbf{h}(\mathbf{X}, T)'\boldsymbol{\gamma} + \kappa M, \tag{7.3}$$

where $\Pi(\mathbf{X}, T, M) := P(D = 1|\mathbf{X}, T, M, U = 0)$ and $\mathbf{h}(\mathbf{X}, T)$ is a vector including $\mathbf{X}$, T, and possibly interaction terms and higher order polynomials. The first part of the linear predictor, $\mathbf{h}(\mathbf{X}, T)'\boldsymbol{\gamma}$, completely determines the cause of failure distribution for individuals with observed cause, for whom $M_i = 0$. Parameter $\boldsymbol{\gamma}$ may be estimated by fitting the logistic model to the data of these individuals. On the other hand, the cause of failure distribution for individuals with missing cause is identified up to parameter κ, which is a sensitivity parameter. Indeed, this parameter is not identifiable from the observed data.

In the family of PMMs implied by (7.3), it is assumed that the cause of failure distributions of individuals with missing and observed cause are the same up to a shift in the linear predictor, which is quantified by parameter κ. As shown in Section 4.1.3, MAR is equivalent to the assumption that $\Pi(\mathbf{X}, T, M) = P(D = 1|\mathbf{X}, T, U = 0)$, i.e. under this assumption the cause of failure distributions of individuals with missing and observed cause are identical. Hence, MAR is equivalent to $\kappa = 0$, and the family of PMMs implied by (7.3) can be thought of as being 'centered' at MAR, with κ quantifying the degree of departure from this assumption. Thus, κ may also be called an informativity parameter. By varying κ across a range of values, the sensitivity of inferences obtained under the MAR assumption can be assessed.

Under MAR, the extrapolation model is identifiable from the data of individuals with observed cause, and this is what is done, at least implicitly, in all MAR approaches such as the direct likelihood approach of Chapter 5. A model for the cause of failure distribution that does not depend on M is usually considered explicitly in MAR-based approaches such as the multiple imputation (MI) approach of Chapter 4 (cf. model (4.9)) and the vertical modeling approach of Nicolaie et al. (2011) described in Chapter 5. Note that the extrapolation model concerns what Nicolaie et al. (2011) call the *relative hazard*. If the primary analysis assumes MAR, and that the cause of failure distribution follows a model like (7.3) without the term κM, then κ also quantifies the departure from the primary analysis model. Thus, in that case κ has a threefold interpretation as the sensitivity parameter in Chapter 6.

Parameter κ represents the (adjusted) logarithm of the odds ratio (or in short, the log odds ratio) comparing the odds of dying from the cause of interest ($D = 1$) between deceased patients with missing and observed cause of failure. That is,

$$\exp(\kappa) = \frac{\Pi^{mis}/(1 - \Pi^{mis})}{\Pi^{obs}/(1 - \Pi^{obs})},$$

where Π^{mis} and Π^{obs} are shorthand notations for $\Pi(\mathbf{X}, T, M = 1)$ and $\Pi(\mathbf{X}, T, M = 0)$, respectively. Thus, possible values for κ may be chosen by assessing the plausibility of the implied odds ratio $\exp(\kappa)$ in the context of the study.

Moreover, parameter κ may be allowed to depend on follow-up time and covariates, i.e. $\kappa = \kappa(\mathbf{X}, T)$. In specifying $\kappa(\mathbf{X}, T)$, the analyst should try to capture the differences between the cause of failure distributions of individuals with missing and observed cause that could have an impact on inferences about the parameter of interest. For instance, if X is a binary exposure of interest taking values 0 and 1, then the choice $\kappa(\mathbf{X}, T) = k_X$, where k_X is a group-specific constant, could be used to reflect different (adjusted) odds ratios, comparing cause 1 mortality between individuals with missing and observed cause, within in each category of X. Indeed, in that case

$$\exp(k_0) = \frac{\Pi_0^{mis}/(1 - \Pi_0^{mis})}{\Pi_0^{obs}/(1 - \Pi_0^{obs})}, \qquad \exp(k_1) = \frac{\Pi_1^{mis}/(1 - \Pi_1^{mis})}{\Pi_1^{obs}/(1 - \Pi_1^{obs})},$$

where the subscripts indicate the category of X. Of course, the plausibility of the range of values assigned to k_0 and k_1 needs to be assessed. For this purpose, note that the difference between these two informativity parameters is the (adjusted) log odds ratio comparing the odds ratios of cause 1 mortality according to X between those with missing and observed cause:

$$k_1 - k_0 = \log\left\{\frac{\Pi_1^{mis}/(1 - \Pi_1^{mis})}{\Pi_0^{mis}/(1 - \Pi_0^{mis})}\right\} - \log\left\{\frac{\Pi_1^{obs}/(1 - \Pi_1^{obs})}{\Pi_0^{obs}/(1 - \Pi_0^{obs})}\right\}.$$

The second term of this expression can always be estimated from individuals with observed cause. Meanwhile, the first term is not identifiable and assumptions must be

made about it. In some settings, it is possible to conduct an investigation to recover the causes of failure for a small sample of the individuals with missing cause. If such additional data are available, they can be used to obtain at least a rough estimate of the first term, which can serve as a basis to choose a range of plausible values for k_0 and k_1.

In general, for $\mathbf{X}$ a p-vector of covariates, multidimensional parametrizations such as

$$\kappa(\mathbf{X}, T) = \mathbf{k}'_{\mathbf{X}}\mathbf{X} + k_T T,$$

where $\mathbf{k}_{\mathbf{X}}$ is a p-vector of informativity parameters and k_T is a scalar informativity parameter, may be considered. Varying $\mathbf{k}_{\mathbf{X}}$ and k_T over ranges of plausible values will define a family of MNAR models designed to provide insight into the sensitivity of inferences made about the parameter of interest to departures from MAR. Similar ideas have been evoked elsewhere (White et al., 2007; Resseguier et al., 2011).

7.1.2 Practical implementation

Models of the form (7.3) assume that the causes of failure arise from a mixture of two distributions, one for individuals with observed cause and one for individuals with missing cause. Usually, the parameter of interest is a parameter indexing the CSHs or the CIFs, which are quantities describing the marginal distribution of T and D, $f(t, d|\mathbf{X})$. However, model (7.3) already partially determines this marginal distribution, and thus the CSHs and the CIFs. Thus, in principle, imposing model (7.3) precludes all of the common regression strategies that model the CSH and CIF directly, without modeling $f(d|\mathbf{X}, T, M, U = 0)$ explicitly. Of course, a regression model for the first factor of (7.2) could be posited, and then the resulting $f(t, d|\mathbf{X})$, CSHs and CIFs could be recovered by integrating over the distribution of M. However, the non-linearity of all of the models involved implies that the effects of covariates on the quantities of interest will be hard, if not impossible, to recover and will have complex interpretations.

Fortunately, the two-stage nature of MI provides us with a way to work around

these seeming problems, enabling the application of the proposed approach to assess the sensitivity of inferences obtained with various regression models for competing risks data. The idea is to multiply impute the missing causes according to (7.3) for the chosen $\kappa(\mathbf{X}, T)$, and then the desired model (e.g. a direct regression model for the CSH or the CIF) can be fitted to each completed dataset. The procedure can be summarized in the following steps:

Step 1 Fit the logistic model (7.3) without the term κM to the individuals with observed cause.

Step 2 Multiply impute the missing causes m times by drawing from model (7.3), using the parameter estimates obtained in Step 1 and the chosen $\kappa(\mathbf{X}, T)$. This yields m completed datasets.

Step 3 Fit the competing risks model of choice to each completed dataset to obtain m estimates of the parameter of interest $\boldsymbol{\theta}$ and its variance.

Step 4 Using Rubin's formulas, combine the m parameter and variance estimates yielded by Step 3 to obtain a final estimate of $\boldsymbol{\theta}$ and its variance, construct CIs and perform hypothesis tests.

Note that Step 1 does not have to be repeated for each choice of $\kappa(\mathbf{X}, T)$. By repeating Steps 2-4 for several choices of $\kappa(\mathbf{X}, T)$ and comparing the results obtained, the robustness of inferences about $\boldsymbol{\theta}$ to departures from the MAR assumption can be assessed. Sensitivity plots like those provided for the SMI study in Chapter 6 can be useful.

In principle, Steps 1 and 3 are straightforward using available software, e.g. use the *glm* function in `R` (R Core Team, 2013) or PROC LOGISTIC in SAS (SAS Institute Inc., 2003) for Step 1. Step 4 is performed by applying the formulas presented in Section 3.1. Finally, Step 2 can be performed by using a modified version of the MI procedure of Bakoyannis et al. (2010), described in Section 4.1.3. The modification consists in replacing step (b) by:

(b') For each patient i who has failed ($U_i = 0$) and with missing cause of failure ($M_i = 1$), calculate the linear predictor implied by (7.3) and the chosen $\kappa(\mathbf{X}, T)$:

$$\eta_i^{(l)} = \mathbf{h}(\mathbf{X}_i, T_i)'\boldsymbol{\gamma}^{(l)} + \kappa(\mathbf{X}_i, T_i)M_i.$$

Note that the estimates $\hat{\boldsymbol{\gamma}}$ and $\hat{\text{var}}(\hat{\boldsymbol{\gamma}})$ required by the imputation procedure are those obtained in Step 1. Also, note that the modified imputation procedure coincides with the procedure of Bakoyannis et al. (2010) when $\kappa(\mathbf{X}, T) = 0$.

7.2 Application to the ECOG clinical trial

In this section, we apply the proposed procedure to the analysis of the ECOG clinical trial (cf. Section 2.1). As mentioned in Section 4.3, several authors have studied the effects of the estrogen-receptor (ER) status (positive vs. negative) and the number of positive nodes (<4 nodes vs. $\geq$ 4 nodes) on the CSH of death from cancer under the MAR assumption. For instance, Goetghebeur and Ryan (1995) and Lu and Tsiatis (2001) considered the Cox model (2.1) for the CSH of cancer including as predictors the indicator variables "$\geq$ 4 nodes" and "ER status", the latter being 1 for patients with an ER-negative primary and 0 for those with an ER-positive primary.

The results of Goetghebeur and Ryan (1995) (GR) and Lu and Tsiatis (2001) (LT) are shown in Table 7.1. The table also shows the results obtained when fitting this model by means of a complete case (CC) analysis, an extra state (ES) analysis (cf. Section 2.3.1) and the proposed procedure with $\kappa = 0$. In the latter, the imputation model was a logistic model including the follow-up time, the number of nodes and their interaction as predictors, and was built from the complete cases with ER-positive status just like in Section 4.3. For this procedure, $m = 100$ imputations were performed. In Table 7.1, the p-value column is blank for the GR and LT approaches because these were not provided in the corresponding manuscripts. Note that both our procedure with $\kappa = 0$ and the approach of Lu and Tsiatis (2001) are MAR-based MI approaches.

However, while the former corresponds to the original Bayesian MI of Rubin (1987), with parameters being drawn from their posterior distribution or an approximation of it (cf. Section 3.1), the latter is what Tsiatis (2006) calls *frequentist imputation*; With this approach, the parameters are not drawn from their posterior distribution, so the imputation procedure is not proper and requires deriving a variance estimator that accounts explicitly for the uncertainty in the estimation of the imputation model's parameters.

Table 7.1: Multivariable Cox model for the CSH of death from cancer including the ER status and the indicator of the presence of four or more positive nodes as covariates. Estimates obtained via a complete case analysis (CC), an extra state analysis (ES), the approaches of Goetghebeur and Ryan (1995) (GR) and Lu and Tsiatis (2001) (LT), and the proposed procedure with $\kappa = 0$ ($\kappa = 0$) are shown.

	Analysis	$\hat{\beta}$	SE	p-value
ER status	CC	1.71	0.4866	<0.001
	ES	1.83	0.4857	<0.001
	GR	1.59	0.4822	–
	LT	1.61	0.4794	–
	$\kappa = 0$	1.62	0.4801	<0.001
≥4 nodes	CC	0.66	0.3090	0.032
	ES	0.62	0.3090	0.045
	GR	0.57	0.2803	–
	LT	0.60	0.2618	–
	$\kappa = 0$	0.55	0.2812	0.051

$\hat{\beta}$ is the estimate of the covariate-specific effect

From Table 7.1 we can see that the estimated effect of having an ER-negative primary on the CSH of cancer is strong in all analyses, and deemed significantly different from zero at a 5% level by the tests performed. The effect of having 4 or more four positive nodes displays stronger fluctuations across the different approaches relative to the effect size, and is deemed borderline significant at a 5% level by the tests performed.

Following the proposed procedure, we analyzed the sensitivity of the outcome of the significance test for the effect of having four or more positive nodes, to departures from

the MAR assumption. In Step 1, the same logistic model used under MAR was fitted to the data of deceased patients with observed cause of failure and an ER-positive status. For Step 2, we considered an informativity parameter that depended on the indicator "$\geq$ 4 nodes", that we denote by X, i.e. $\kappa(X) = k_X$, where k_0 and k_1 were allowed to vary between -3 and 3, so that $0.05 \leq \exp(k_X) \leq 20$. For each choice of k_0 and k_1, $m = 100$ imputations of the missing causes were performed. In Step 3, we refitted the Cox model for the CSH of cancer including the indicator variables "$\geq$ 4 nodes" and "ER status" as predictors. The parameter of interest was the effect β_N of the indicator "$\geq$ 4 nodes".

Figure 7.1 shows a contour plot identifying the regions where a given pair of parameters (k_0, k_1) yielded a significant effect of "$\geq$ 4 nodes" ($p \leq 0.05$). This contour plot suggests that the outcome of the significance test is quite sensitive to departures from the MAR assumption.

To further explore the source of this sensitivity, we first considered the interesting scenario in which both groups have the same value of the informativity parameter, i.e. $k_0 = k_1 = k$. The common parameter k represents the adjusted log odds ratio comparing the odds of dying from cancer between deceased women with missing and observed cause. Thus, when $k > 0$ (respectively, $k < 0$), a higher (respectively, lower) proportion of cancer deaths is expected among women with missing cause of death compared to women with observed cause with the same characteristics. The estimates and corresponding confidence intervals (CIs) obtained in this scenario are plotted in Figure 7.2.

As we can see in Figure 7.2, the outcome of the significance test changes when the adjusted odds ratio comparing cancer mortality between women with missing and observed cause is above $\exp(0.2) \approx 1.2$. However, the effect estimate does not change substantially across the scenarios considered even though quite large, and probably unrealistic, values for the informativity parameters were considered. The hazard ratio was always comprised between 1.67 and 1.75, with the small fluctuations observed probably due to the random nature of the imputation procedure. These results seem

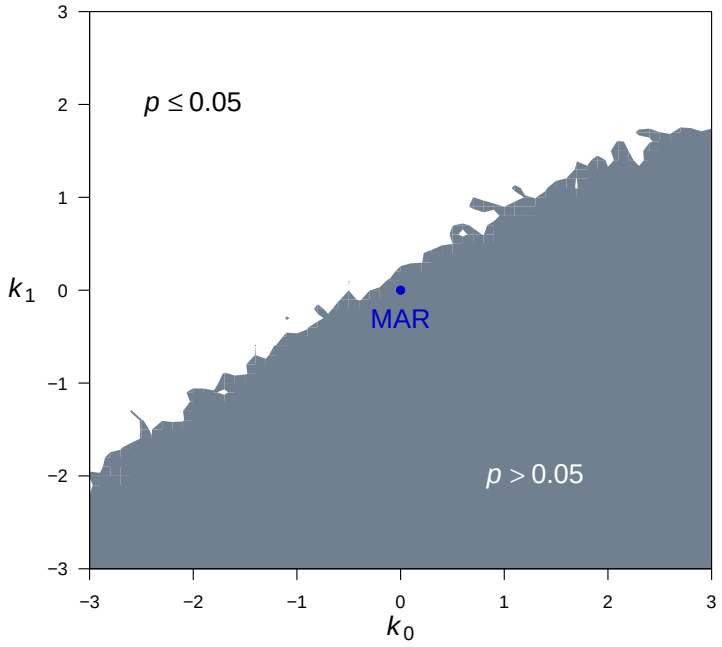

Figure 7.1: Contour plot showing the regions in which β_N, the effect of having 4 or more positive nodes on the (log) CSH of cancer death, was significant (white) or non-significant (gray), according to a given pair of informativity parameters (k_0, k_1), corresponding to the groups with less than 4 nodes and the group with 4 or more nodes, respectively. The point plotted at (0,0) in blue corresponds to the MAR analysis result.

thus to indicate that the effect estimate is not very sensitive to departures from the MAR assumption of the type depicted in Figure 7.2, i.e. with $k_0 = k_1$. To explain this, recall that both groups exhibit the same number of missing causes and approximately the same proportion of cancer deaths (cf. Table 2.1). Thus, the number of cancer deaths among those with missing causes is increased (if $k > 0$) or decreased (if $k < 0$) to approximately the same extent in both groups with respect to the MAR scenario. This means that the CSHs of cancer are also increased or decreased to approximately the same extent in both groups, implying a hazard ratio of approximately the same magnitude as under MAR.

The CI does not change drastically either, with the change from non-significance to significance clearly not being a consequence of a substantial change in the effect

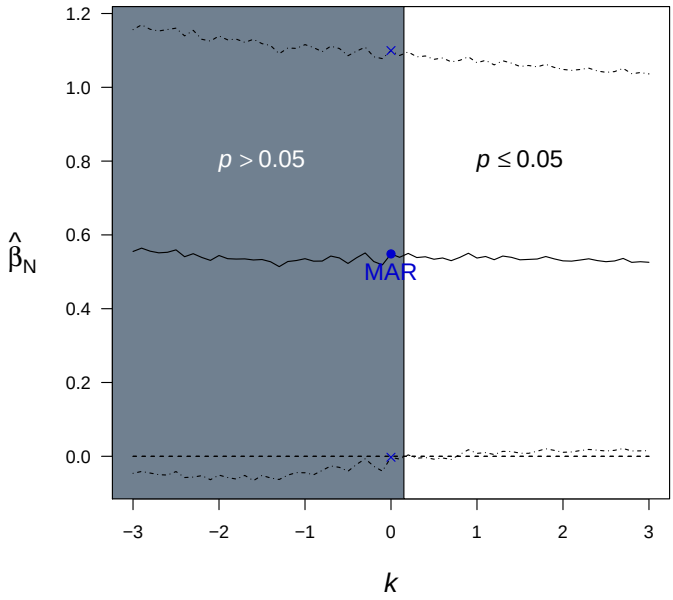

Figure 7.2: Sensitivity analysis of the estimate of β_N, the effect of having 4 or more positive nodes, on the (log) CSH of cancer death when $k_0 = k_1 = k$. Dashed lines show the corresponding 95% confidence intervals, and the white region indicates the values of k for which the effect was significantly different from zero. The MAR analysis result, corresponding to $k = 0$, is indicated in blue. The dotted line at $\hat{\beta}_N = 0$ is included for reference.

estimate, as was observed for the SMI study in Figure 6.2. Rather, this change seems to be a consequence of an improvement in the precision of the estimate resulting from the increased amount of cancer death events in both groups as k increases. Actually, β_N remained borderline significant across all values of k considered, with a p-value fluctuating between 0.04 and 0.08. Thus, the conclusion of this analysis is that the outcome of the test is not substantially sensitive to departures from MAR of the type implied by the condition $k_0 = k_1$.

Next we considered the scenario where $k_0 = -k_1$, so that the adjusted odds ratios comparing cancer mortality between individuals with missing and observed cause in the two groups were inversely proportional. The estimates obtained in this scenario are

plotted in Figure 7.3. In this scenario, the effect estimate does change substantially, with the implied hazard ratio fluctuating between 1.3 and 2.3. Of course, this is due to the induced large differences in the proportions of cancer deaths among the patients with missing cause in the two groups. Thus, the jump from significance to non-significance as k_0 increases in this scenario does arise from a change in the effect and not solely from an increased precision. Actually, the p-value across all values of k_0 considered fluctuated between 0.005 and 0.33. The conclusion here is that the outcome of the test is substantially sensitive to departures from MAR of the type implied by the condition $k_0 = -k_1$.

7.3 Discussion

In this ongoing work, we propose a procedure to assess the sensitivity of inferences to missing data assumptions in the competing risks setting with missing causes of failure. In the family of PMMs proposed, the sensitivity parameter κ quantifies the degree of departure from the MAR assumption and has a useful interpretation as the (adjusted) log odds ratio comparing the odds of dying from the cause of interest between failed patients with missing and observed cause of failure. In the specification of $\kappa = \kappa(\mathbf{X}, T)$, the data analyst needs to make explicit assumptions about the main differences between the cause of failure distributions of individuals with missing and observed cause. The proposed MI-based implementation procedure makes it possible to apply this approach to various competing risks regression models (e.g. parametric or semi-parametric models for the CSH or the CIF). MI also facilitates the performance of several analyses over a range of possible choices for $\kappa(\mathbf{X}, T)$. The first results of application of the procedure with the ECOG clinical trial are encouraging.

Recall from Section 3.1 that, in principle, the imputation and analysis models must be congenial to obtain valid inferences with MI. As mentioned at the beginning of Section 7.1.2, imposing a model like (7.3) will partially determine the CSHs and the CIFs, which are the usual quantities of focus in the analysis model. Moreover, model

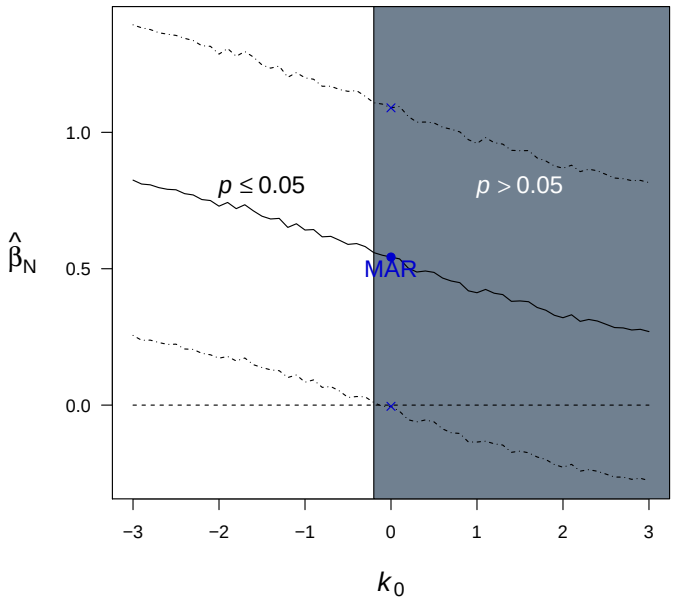

Figure 7.3: Sensitivity analysis of the estimate of β_N, the effect of having 4 or more positive nodes, on the (log) CSH of cancer death when $k_0 = -k_1$. Dashed lines show the corresponding 95% confidence intervals, and the white region indicates the values of k_0 for which the effect was significantly different from zero. The MAR analysis result, corresponding to $k_0 = 0$, is indicated in blue. The dotted line at $\hat{\beta}_N = 0$ is included for reference.

(7.3) assumes a dependence on M that will not likely be included in the analysis model. These two observations imply that the imputation and analysis models will usually be uncongenial in this context. Thus, the next necessary step in this work is to perform a simulation study similar to the one presented in Chapter 6, to assess the statistical properties of the MI inferences yielded by the proposed procedure and evaluate its overall performance. For instance, we can expect Rubin's variance estimator (3.2) to be biased (Meng, 1994; Robins and Wang, 2000), so it would be desirable to propose and evaluate alternative variance estimators.

The developments in this chapter were motivated by an epidemiological study of socioeconomic inequalities in suicide mortality in France. The data in this study corre-

sponded to a permanent, cross-sectionally representative 1% sample of the population called the *permanent demographic sample* (PDS), which was started in 1968 by the French National Institute of Statistics and Economic Studies (Insee) (Couet, 2007). This study will thus be henceforth referred to as the PDS study. For the individuals in the sample, survival data were available from civil registration records, socioeconomic information was available from exhaustive population censuses and cause of death data were retrieved from the national cause of death registry (Menvielle et al., 2007). The cause of death was missing for around 10% of the individuals in the sample. The French national cause of death registry was suspected to contain many suicides coded as deaths with unknown cause because of an area-specific reporting issue. The MAR assumption was therefore implausible and it was desirable to perform sensitivity analyses. Furthermore, another study had been previously performed to retrieve the cause of death for a sub-sample of the individuals with missing cause in the national cause of death registry, which provided us with additional information to choose the range of values for the sensitivity parameters. The analyses performed following the proposed approach evidenced that inferences were broadly insensitive to missingness assumptions, doubtlessly due to the small percentage of missing data, which is why we chose the ECOG data to illustrate the approach instead.

The two scenarios depicted for the ECOG trial in Figures 7.2 and 7.3, emphasize the need for assessing the plausibility of, not only the magnitude, but also the sign of the values considered for the informativity parameters. Here we considered a range of values for the sensitivity parameters such that $0.05 \leq \exp(k_X) \leq 20$, representing very extreme values for an odds ratio. However, in some scenarios such large differences may be plausible, e.g. in the PDS study according to data available from a previous study. The ECOG example also emphasizes the need to direct sensitivity analyses toward the inferential procedure of interest according to the context, as pointed out by Carroll et al. (2004). Very different conclusions about sensitivity may result from focusing on effect estimates and from focusing on hypothesis tests or CIs, which depend on both the effect estimates and their (estimated) precision.

We conclude this chapter by noting that MI is not required to apply the proposed approach to inferences obtained in vertical modeling of competing risks. Indeed, in this approach $f(d|\mathbf{X}, T, U = 0)$, called by Nicolaie et al. (2010) the relative hazard, is modeled explicitly (cf. Chapter 5). Under the MAR assumption, vertical modeling can be applied in almost the same way as when all the causes of failure are observed, the only difference being that the model for the relative hazard is fitted exclusively to the individuals with observed cause (Nicolaie et al., 2011). Our approach can thus be considered to extend the vertical modeling approach with missing causes of failure to the MNAR setting by allowing a dependance of the model for the relative hazard on M.

General discussion and future research

In this dissertation, we have presented and discussed a variety of methods to perform regression with missing outcomes for two outcome data structures typically encountered in medical research: continuous longitudinal data and competing risks data. While in the former context parametric models are often the choice with no missing data, in the latter there is a preference for semi-parametric models. Furthermore, special methodology to deal with right-censoring is required for regression modeling of competing risks. Nevertheless, we were able to apply many of the ideas developed for longitudinal data with drop-outs, and more generally for incomplete multivariate data, to the competing risks setting with missing causes: direct likelihood, MI and IPW under MAR, and pattern-mixture modeling under MNAR and for sensitivity analyses. Hence, a first conclusion of this work is that missing data concepts and modeling ideas are transposable across different settings, transcending contextual particularities.

In some cases, this process results in pleasant surprises. When modeling the CIF under MAR (Chapter 4), the application of IPW ideas coupled with the use of pseudo-values to deal with right-censoring led to a pleasant surprise in that we were able to construct an IPW estimator that uses information more efficiently than usual IPW estimators. Indeed, the use of pseudo-values as alternative outcomes enabled us to include the partial information available from the individuals with missing cause in the estimating equations, while standard IPW estimators use only the complete cases

in these equations. This resulted in an IPW estimator that displayed an efficiency comparable to that of MI in our simulation study, contrary to what is usually expected (Seaman and White, 2013).

While MAR approaches for handling missing data are now widely available in current software and are often used by practitioners, sensitivity analysis methodology is still an active area of research. In the longitudinal data setting, Daniels and Hogan (2000) and Ratitch et al. (2013) proposed sensitivity analysis approaches that are similar in spirit to that proposed in Chapter 6. A remarkable aspect of our method is that it is easily applicable to studies where there is a large number of measurements or where the timing and number of measurements differ across individuals. This flexibility is inherited from linear mixed models, which allow for the inclusion of time-trends and random effects covariance structures. Furthermore, the sensitivity parameters in our approach are easily interpreted as they quantify specific aspects of the expected trajectories of individuals (e.g. intercepts, time-slopes), thus facilitating the formulation of assumptions about the distribution of the missing outcomes.

The sensitivity analysis methods proposed in this manuscript are, in their present state, exploratory tools, albeit conducted rigorously. Indeed, the information resulting from these methods, and from similar approaches found in the literature, is extremely detailed. Hence, an important matter is to determine how this information could be summarized to facilitate its reporting and use. This issue is particularly relevant in the clinical trial setting, in which regulations demand that protocols specify in advance the statistical analyses and decision rules that will be adopted. Although this is still a somewhat open question, a big step towards an answer is given by the work of Molenberghs et al. (2001a) and Vansteelandt et al. (2006). As briefly mentioned earlier, these authors propose a formal framework to summarize the results of such sensitivity analyses into a single inference that does not rely on unverifiable missingness assumptions. They define the concept of a *region of ignorance*, which they relate to the parameter regions yielded by these sensitivity analysis approaches, and use this to extend the notion of a confidence region to what they call a *region of uncertainty*. These regions account

for the uncertainty due to missing data as well as finite sampling. Other authors have explored related ideas in the Bayesian framework (e.g. Scharfstein et al., 2003).

It is important to emphasize that the gain in using the methods discussed will generally depend on the percentage of individuals who drop-out before the study end in longitudinal studies, and on the percentage of missing causes of failure among those who failed in competing risks studies. In the modest experience accumulated throughout these last few years, we noticed that the impact of using these methods on inferences starts to become evident when these percentages are above around 10%. At levels of 10% or less, inferences are relatively insensitive to missingness assumptions and ad-hoc approaches will often yield approximately valid inferences. This observation was supported by our findings in the PDS study mentioned in the discussion of Chapter 7. On the other hand, when these percentages are above around 40%, inferences will often be extremely sensitive to missingness assumptions and small departures from a primary analysis may easily invalidate the conclusions drawn from it. Although missing outcome methods will help shed light on this sensitivity, they cannot remedy the considerable lack of information. Thus, one will rarely be able to draw a final definitive conclusion with great confidence in such studies. In conclusion, it is in studies where these percentages are between around 10 and 40% that missing outcome methods are most valuable and can help avoid incorrect inferences and conclusions. Nevertheless, in every study there should be an effort to minimize the amount of missing data as much as possible, for example by anticipating possible sources of missingness and strategies to avoid these in the planning phase.

An essential condition for the widespread use of any statistical method, in particular missing data approaches, is the availability of software. For the methods of Chapters 3, 4 and 6, published `R` code is available to implement the proposed procedures. The approach of Chapter 7 is relatively easy to implement using any available MI software that includes an imputation procedure for binary data. For the direct likelihood approach of Chapter 5, software implementation is less straightforward and is part of ongoing work. Once finished, we will be able to evaluate the approach through simulation experiments

and apply it to the ECOG clinical trial for comparison with other approaches. The assessment of the approach of Chapter 7 through simulations is also intended future work.

Another ongoing project, which was prompted by the PDS study, concerns the indexes used in the epidemiological and economic literature to measure socioeconomic inequalities with survival and competing risks data. The subject of this project being outside the missing data theme of this dissertation, we decided not to included it as a separate chapter. The manuscript in preparation is targeted at an epidemiology journal and its current version is provided in Appendix C.

A future perspective of the work performed on competing risks with missing causes of failure was prompted by a question of one the reviewers of Moreno-Betancur and Latouche (2013). It concerns the statistical problem that arises when all causes of failure are observed but are subject to misclassification, with known misclassification probabilities (see for example van Rompaye et al., 2012). We found that an extension of the Andersen-Klein pseudo-value approach to that setting may be possible owing to a result similar to that of Lemma 4.1.1 for some other form of modified pseudo-values. Since there is currently no approach for regression modeling of the CIF in the misclassified cause of failure setting, we plan to investigate this extension in the future.

Bibliography

Aalen, O. (1980). A model for nonparametric regression analysis of counting processes. In *Mathematical statistics and probability theory*, pages 1–25. Springer, New York.

Altham, P. (1984). Improving the precision of estimation by fitting a model. *Journal of the Royal Statistical Society. Series B*, 46(1):118–119.

Andersen, J., Goetghebeur, E., and Ryan, L. (1996). Missing cause of death information in the analysis of survival data. *Statistics in Medicine*, 15:2191–2201.

Andersen, P. K., Abildstrom, S. Z., and Rosthoj, S. (2002). Competing risks as a multi-state model. *Statistical Methods in Medical Research*, 11(2):203–215.

Andersen, P. K., Borgan, O., Gill, R. D., and Keiding, N. (1993). *Statistical Models Based on Counting Processes*. Springer, Berlin and New York.

Andersen, P. K., Borgan, O., Hjort, N. L., Arjas, E., Stene, J., and Aalen, O. (1985). Counting process models for life history data: A review. *Scandinavian Journal of Statistics*, 12(2):97–158.

Andersen, P. K., Geskus, R. B., de Witte, T., and Putter, H. (2012). Competing risks in epidemiology: possibilities and pitfalls. *International Journal of Epidemiology*, 41(3):861–870.

Andersen, P. K. and Gill, R. D. (1982). Cox's regression model for counting processes: a large sample study. *The annals of statistics*, 10(4):1100–1120.

Andersen, P. K. and Keiding, N. (2012). Interpretability and importance of functionals in competing risks and multistate models. *Statistics in medicine*, 31(11-12):1074–1088.

Andersen, P. K. and Klein, J. P. (2007). Regression analysis for multistate models based on a pseudo-value approach, with applications to bone marrow transplantation studies. *Scandinavian Journal of Statistics*, 34(1):3–16.

Andersen, P. K., Klein, J. P., and Rosthoj, S. (2003). Generalised linear models for correlated pseudo-observations, with applications to multi-state models. *Biometrika*, 90(1):15–27.

Andersen, P. K. and Perme, M. (2010). Pseudo-observations in survival analysis. *Statistical Methods in Medical Research*, 19(1):71–99.

Bajorunaite, R. (2003). *Comparison of Failure Probabilities in the Presence of Competing Risks*. PhD thesis, Graduate School of Biomedical Sciences of the Medical College of Wisconsin.

Bakoyannis, G., Siannis, F., and Touloumi, G. (2010). Modelling competing risks data with missing cause of failure. *Statistics in Medicine*, 29(30):3172–3185.

Barnard, J., Rubin, D. B., and Schenker, N. (1998). Multiple imputation methods. In Armitage, P. and Colton, T., editors, *Encyclopedia of Biostatistics*, pages 2772–2780. John Wiley & Sons, Ltd., Chichester.

Bates, D. (2006). *Residual df in lmer and simulation results*. R-help Main R Mailing List, Available at https://stat.ethz.ch/pipermail/r-help/2006-July/110133.html (accessed May 10, 2010).

Bates, D. (2010). *Penalized least squares versus generalized least squares representations of linear mixed models*. Research Report, Department of Statistics, University of Wisconsin Madison, Available at http://cran.r-project.org/web/packages/lme4/vignettes/PLSvGLS.pdf (accessed May 10, 2010).

Bates, D., Maechler, M., and Bolker, B. (2012). *lme4: Linear mixed-effects models using S4 classes*. R package version 0.999999-0, Available at Http://cran.r-project.org/package=lme4 (accessed 30 August, 2012).

Bender, R., Augustin, T., and Blettner, M. (2005). Generating survival times to simulate Cox proportional hazards models. *Statistics in medicine*, 24(11):1713–1723.

Beyersmann, J., Dettenkofer, M., Bertz, H., and Schumacher, M. (2007). A competing risks analysis of bloodstream infection after stem-cell transplantation using subdistribution hazards and cause-specific hazards. *Statistics in Medicine*, 26(30):5360–5369.

Beyersmann, J., Latouche, A., Buchholz, A., and Schumacher, M. (2009). Simulating competing risks data in survival analysis. *Statistics in Medicine*, 28(6):956–971.

Binder, N., Gerds, T. A., and Andersen, P. K. (2012). *Pseudo-observations for competing risks with covariate dependent censoring - Research Report 12/6*. Department of Biostatistics, University of Copenhagen, Denmark, Available at https://ifsv.sund.ku.dk/biostat/biostat_annualreport/images/2/2f/Research_Report_12-06.pdf (accessed August 28, 2012).

Borgan, O. (1984). Maximum likelihood estimation in parametric counting process models, with applications to censored failure time data. *Scandinavian Journal of Statistics*, 11(1):1–16.

Burzykowski, T., Carpenter, J., Coens, C., Evans, D., France, L., Kenward, M., Lane, P., Matcham, J., Morgan, D., Phillips, A., Roger, J., Sullivan, B., White, I., and Yu, L.-M. (2010). Missing data: Discussion points from the PSI missing data expert group. *Pharmaceutical Statistics*, 9(4):288–297.

Canty, A. and Ripley, B. D. (2011). *boot: Bootstrap R (S-Plus) Functions*. R package version 1.3-3, Available at http://cran.r-project.org/web/packages/boot/index.html (accessed July 15, 2011).

Carpenter, J. R. and Kenward, M. G. (2007). *Missing data in randomised clinical trials – a practical guide*. Monograph. Medical Statistics Unit, London School of Hygiene and Tropical Medicine, UK, Available at http://www.hta.nhs.uk/nihrmethodology/reports/1589.pdf (accessed 10 August, 2012).

Carpenter, J. R., Kenward, M. G., and Vansteelandt, S. (2006). A comparison of multiple imputation and doubly robust estimation for analyses with missing data. *Journal of the Royal Statistical Society. Series A (Statistics in Society)*, 169(3):571–584.

Carpenter, J. R., Kenward, M. G., and White, I. R. (2007). Sensitivity analysis after multiple imputation under missing at random: a weighting approach. *Statistical Methods in Medical Research*, 16(3):259–275.

Carroll, R., Davidian, M., Dubin, J., Fitzmaurice, G., Kenward, M., Mohlenberghs, G., and Roy, J. D. (2004). Emerging Issues in Longitudinal and Functional Data Analysis - Discussion of two important missing data issues. *Statistica Sinica*, 14:615–629.

Cook, R. D. (1977). Detection of influential observation in linear regression. *Technometrics*, 19(1):15–18.

Cook, R. D. (1986). Assessment of local influence. *Journal of the Royal Statistical Society. Series B (Methodological)*, 48(2):133–169.

Couet, C. (2007). INSEE's Permanent Demographic Sample (EDP). *Courrier des statistiques, English series*, 13:29–38.

Cox, D. (1972). Regression models and life-tables. *Journal of the Royal Statistical Society. Series B (Methodological)*, 34(2):187–220.

Cummings, F. J., Gray, R., Davis, T. E., Tormey, D. C., Harris, J. E., Falkson, G., and Arseneau, J. (1985). Adjuvant tamoxifen treatment of elderly women with stage II breast cancer: A double-blind comparison with placebo. *Annals of Internal Medicine*, 103(3):324–329.

Cummings, F. J., Gray, R., Davis, T. E., Tormey, D. C., Harris, J. E., Falkson, G. G., and Arseneau, J. (1986). Tamoxifen versus placebo: double-blind adjuvant trial in elderly women with stage II breast cancer. *NCI monographs : a publication of the National Cancer Institute*, (1):119–123.

Cummings, F. J., Gray, R., Tormey, D. C., Davis, T. E., Volk, H., Harris, J., Falkson, G., and Bennett, J. M. (1993). Adjuvant tamoxifen versus placebo in elderly women with node-positive breast cancer: long-term follow-up and causes of death. *Journal of Clinical Oncology*, 11(1):29–35.

Daniels, M. J. and Hogan, J. W. (2000). Reparameterizing the pattern mixture model for sensitivity analyses under informative dropout. *Biometrics*, 56(4):1241–1248.

Daniels, M. J. and Wang, C. (2009). Comments on: Missing data methods in longitudinal studies: a review. *TEST: An Official Journal of the Spanish Society of Statistics and Operations Research*, 18(1):51–58.

Davison, A. C. and Hinkley, D. V. (1997). *Bootstrap Methods and Their Applications*. Cambridge University Press, Cambridge.

Demirtas, H. and Schafer, J. L. (2003). On the performance of random-coefficient pattern-mixture models for non-ignorable drop-out. *Statistics in Medicine*, 22(16):2553–2575.

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

Diggle, P., Heagerty, P., Liang, K. Y., and Zeger, S. L. (2002). *Analysis of Longitudinal Data*. Oxford University Press, Oxford, 2nd edition.

Diggle, P. and Kenward, M. G. (1994). Informative drop-out in longitudinal data analysis (with discussion). *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, 43(1):49–93.

Fine, J. P. (2001). Regression modeling of competing crude failure probabilities. *Biostatistics*, 2(1):85–97.

Fine, J. P. and Gray, R. J. (1999). A proportional hazards model for the subdistribution of a competing risk. *Journal of the American Statistical Association*, 94(446):496–509.

Fitzmaurice, G. M., Laird, N. M., and Ware, J. H. (2004). *Applied longitudinal analysis*. John Wiley & Sons, Ltd., Hoboken, N.J.

Fleming, T. R. and Harrington, D. P. (2005). *Counting processes and survival analysis*. John Wiley & Sons, Ltd., Hoboken, N.J.

Gao, G. Z. (2006). *Semiparametric Estimators for the Regression Coefficients in the Linear Transformation Competing Risks Models with Missing Cause of Failure*. PhD thesis, North Carolina State University, Raleigh.

Gao, G. Z. and Tsiatis, A. A. (2005). Semiparametric estimators for the regression coefficients in the linear transformation competing risks model with missing cause of failure. *Biometrika*, 92(4):875–891.

Gerds, T. A. (2011). *prodlim: Product Limit Estimation*. R package version 1.2.9., Available at http://cran.r-project.org/package=prodlim (accessed March 1, 2012).

Gerds, T. A., Scheike, T. H., and Andersen, P. K. (2012). Absolute risk regression for competing risks: interpretation, link functions, and prediction. *Statistics in Medicine*, 31(29):3921–3930.

Goetghebeur, E. and Ryan, L. (1995). Analysis of competing risks survival data when some failure types are missing. *Biometrika*, 82(4):821–833.

Graw, F., Gerds, T. A., and Schumacher, M. (2008). *On pseudo-values for regression analysis in competing risks models - Research report 08/4*. Department of Biostatistics, University of Copenhagen, Denmark, Available at https://ifsv.sund.ku.dk/biostat/biostat_annualreport/images/3/3b/ResearchReport-2008-4.pdf (accessed October 1, 2010).

Graw, F., Gerds, T. A., and Schumacher, M. (2009). On pseudo-values for regression analysis in competing risks models. *Lifetime Data Analysis*, 15:241–255.

Halekoh, U., Hojsgaard, S., and Yan, J. (2006). The R package geepack for generalized estimating equations. *Journal of Statistical Software*, 15(2):1–11.

Hedeker, D. and Gibbons, R. D. (1997). Application of random-effects pattern-mixture models for missing data in longitudinal studies. *Psychological Methods*, 2(1):64–78.

Hogan, J. (2009). Comments on: Missing data methods in longitudinal studies: a review. *TEST: An Official Journal of the Spanish Society of Statistics and Operations Research*, 18(1):59–64.

Jansen, I., Hens, N., Molenberghs, G., Aerts, M., Verbeke, G., and Kenward, M. G. (2006). The nature of sensitivity in monotone missing not at random models. *Computational Statistics & Data Analysis*, 50(3):830–858.

Jeong, J. H. and Fine, J. P. (2007). Parametric regression on cumulative incidence function. *Biostatistics*, 8(2):184–196.

Kalbfleisch, J. D. and Prentice, R. L. (2002). *The statistical analysis of failure time data*. John Wiley & Sons, Ltd., Hoboken, N.J., 2nd edition.

Kenward, M. G. (1998). Selection models for repeated measurements with non-random dropout: an illustration of sensitivity. *Statistics in Medicine*, 17(23):2723–2732.

Kenward, M. G. and Molenberghs, G. (1998). Likelihood based frequentist inference when data are missing at random. *Statistical Science*, 13(3):236–247.

Kenward, M. G. and Molenberghs, G. (1999). Parametric models for incomplete continuous and categorical longitudinal data. *Statistical Methods in Medical Research*, 8:51–83.

Kenward, M. G., Molenberghs, G., and Thijs, H. (2003). Pattern-mixture models with proper time dependence. *Biometrika*, 90(1):53–71.

Klein, J. P. (2006). Modelling competing risks in cancer studies. *Statistics in Medicine*, 25:1015–1034.

Klein, J. P. and Andersen, P. K. (2005). Regression modeling of competing risks data based on pseudovalues of the cumulative incidence function. *Biometrics*, 61:223–229.

Klein, J. P., Gerster, M., Andersen, P. K., Tarima, S., and Perme, M. P. (2008). SAS and R functions to compute pseudo-values for censored data regression. *Computer Methods and Programs in Biomedicine*, 89:289–300.

Klein, J. P. and Moeschberger, M. L. (2003). *Survival Analysis: Techniques for Censored and Truncated Data*. Springer, New York, 2nd edition.

Koller, M. T., Raatz, H., Steyerberg, E. W., and Wolbers, M. (2012). Competing risks and the clinical community: irrelevance or ignorance? *Statistics in Medicine*, 31:1089–1097.

Kullback, S. and Leibler, R. A. (1951). On Information and Sufficiency. *The Annals of Mathematical Statistics*, 22(1):79–86.

Laird, N. M. and Ware, J. H. (1982). Random-effects models for longitudinal data. *Biometrics*, 38(4):963–974.

Latouche, A., Allignol, A., Beyersmann, J., Labopin, M., and Fine, J. P. (2013). A competing risks analysis should report results on all cause-specific hazards and cumulative incidence functions. *Journal of Clinical Epidemiology*, 66(6):648–653.

Lee, M., Cronin, K. A., Gail, M. H., Dignam, J. J., and Feuer, E. J. (2011). Multiple imputation methods for inference on cumulative incidence with missing cause of failure. *Biometrical Journal*, 53(6):974–993.

Li, K. H., Raghunathan, T. E., and Rubin, D. B. (1991). Large-sample significance levels from multiply imputed data using moment-based statistics and an F reference distribution. *Journal of the American Statistical Association*, 86(416):1065–1073.

Liang, K. Y. and Zeger, S. L. (1986). Longitudinal data analysis using generalized linear models. *Biometrika*, 73(1):13–22.

Lin, D. Y. and Ying, Z. L. (1994). Semiparametric analysis of the additive risk model. *Biometrika*, 81(1):61–71.

Little, R. J. (1993). Pattern-Mixture Models for Multivariate Incomplete Data. *Journal of the American Statistical Association*, 88(421):125–134.

Little, R. J. (1994). A class of pattern-mixture models for normal incomplete data. *Biometrika*, 81(3):471–483.

Little, R. J. (1995). Modeling the drop-out mechanism in repeated-measures studies. *Journal of the American Statistical Association*, 90(431):1112–1121.

Little, R. J., Agostino, R. D., Cohen, M. L., Dickersin, K., Emerson, S. S., Farrar, J. T., Frangakis, C., Hogan, J. W., Molenberghs, G., Murphy, S. A., Neaton, J. D., Rotnitzky, A., Scharfstein, D., Shih, W. J., Siegel, J. P., and Stern, H. (2012). The prevention and treatment of missing data in clinical trials. *New England Journal of Medicine*, 367(14):1355–1360.

Little, R. J. and Rubin, D. B. (1987). *Statistical Analysis with Missing Data*. Wiley Series in Probability and Statistics. Wiley, New York, 1st edition.

Little, R. J. and Wang, Y. (1996). Pattern-mixture models for multivariate incomplete data with covariates. *Biometrics*, 52(1):98–111.

Little, R. J. and Yau, L. (1996). Intent-to-treat analysis for longitudinal studies with drop-outs. *Biometrics*, 52(4):1324–1333.

Lu, K. F. and Tsiatis, A. A. (2001). Multiple imputation methods for estimating regression coefficients in the competing risks model with missing cause of failure. *Biometrics*, 57(4):1191–1197.

Lu, W. and Liang, Y. (2008). Analysis of competing risks data with missing cause of failure under additive hazards model. *Statistica Sinica*, 18:219–234.

Manola, J. B. and Gray, R. J. (2011). When bad things happen to good studies. *Journal of Clinical Oncology*, 29(26):3497–3499.

Martinussen, T. and Scheike, T. H. (2006). *Dynamic regression models for survival data*. Springer, New York.

McKeague, I. W. and Sasieni, P. D. (1994). A partly parametric additive risk model. *Biometrika*, 81(3):501–514.

Meng, X. L. (1994). Multiple-imputation inferences with uncongenial sources of input. *Statistical Science*, 9(4):538–558.

Meng, X. L. and Rubin, D. B. (1992). Performing likelihood ratio tests with multiply-imputed data sets. *Biometrika*, 79(1):103–111.

Menvielle, G., Chastang, J., Luce, D., and Leclerc, A. (2007). Temporal trends in socioeconomic inequality in mortality in France between 1968 and 1996. Study of educational differences by cause of death. *Revue d'Épidémiologie et de Santé Publique*, 55:e1–e8.

Michiels, B., Molenberghs, G., Bijnens, L., Vangeneugden, T., and Thijs, H. (2002). Selection models and pattern-mixture models to analyse longitudinal quality of life data subject to drop-out. *Statistics in Medicine*, 21:1023–1041.

Minini, P. and Chavance, M. (2004a). Sensitivity analysis of longitudinal binary data with non-monotone missing values. *Biostatistics*, 5(4):531–544.

Minini, P. and Chavance, M. (2004b). Sensitivity analysis of longitudinal normal data with drop-outs. *Statistics in Medicine*, 23(7):1039–1054.

Molenberghs, G., Beunckens, C., Sotto, C., and Kenward, M. G. (2008). Every missingness not at random model has a missingness at random counterpart with equal fit. *Journal of the Royal Statistical Society. Series B (Statistical Methodology)*, 70(2):371–388.

Molenberghs, G. and Kenward, M. G. (2007). *Missing data in clinical studies*. Wiley, Chichester.

Molenberghs, G., Kenward, M. G., and Goetghebeur, E. (2001a). Sensitivity analysis for incomplete contingency tables: the Slovenian plebiscite case. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, 50(1):15–29.

Molenberghs, G., Michiels, B., Kenward, M. G., and Diggle, P. (1998). Monotone missing data and pattern-mixture models. *Statistica Neerlandica*, 52(2):153–161.

Molenberghs, G., Verbeke, G., Thijs, H., Lesaffre, E., and Kenward, M. G. (2001b). Influence analysis to assess sensitivity of the dropout process. *Computational Statistics & Data Analysis*, 37:93–113.

Moreno-Betancur, M. and Chavance, M. (2013). Sensitivity analysis of incomplete longitudinal data departing from the missing at random assumption: Methodology and application in a clinical trial with drop-outs. *Statistical Methods in Medical Research [Epub ahead of print on May 22, 2013]*.

Moreno-Betancur, M. and Latouche, A. (2013). Regression modeling of the cumulative incidence function with missing causes of failure using pseudo-values. *Statistics in Medicine*, 32(18):3206–3223.

National Research Council (2010). *The prevention and treatment of missing data in clinical trials*. The National Academies Press, Washington, DC.

Nicolaie, M. A., van Houwelingen, H. C., and Putter, H. (2010). Vertical modeling: A pattern mixture approach for competing risks modeling. *Statistics in Medicine*, 29(11):1190–1205.

Nicolaie, M. A., van Houwelingen, H. C., and Putter, H. (2011). Vertical modeling: Analysis of competing risks data with missing causes of failure. *Statistical Methods in Medical Research [Epub ahead of print on December 16, 2011]*.

Prentice, R. L., Kalbfleisch, J. D., Peterson, A. V., Flournoy, N., Farewell, V. T., and Breslow, N. E. (1978). The analysis of failure times in the presence of competing risks. *Biometrics*, 34(4):541–554.

R Core Team (2013). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

Ratitch, B., O'Kelly, M., and Tosiello, R. (2013). Missing data in clinical trials: from clinical assumptions to statistical analysis using pattern mixture models. *Pharmaceutical statistics [Epub ahead of print on January 4, 2013]*.

Resseguier, N., Giorgi, R., and Paoletti, X. (2011). Sensitivity analysis when data are missing not-at-random. *Epidemiology*, 22(2):282–283.

Robins, J. M., Rotnitzky, A., and Zhao, L. P. (1995). Analysis of semiparametric regression models for repeated outcomes in the presence of missing data. *Journal of the American Statistical Association*, 90(429):106–121.

Robins, J. M. and Wang, N. (2000). Inference for imputation estimators. *Biometrika*, 87(1):113–124.

Rotnitzky, A., Robins, J. M., and Scharfstein, D. O. (1998). Semiparametric regression for repeated outcomes with nonignorable nonresponse. *Journal of the American Statistical Association*, 93(444):1321–1339.

Rubin, D. B. (1976). Inference and missing data. *Biometrika*, 63(3):581–592.

Rubin, D. B. (1987). *Multiple Imputation for Nonresponse in Surveys*. John Wiley & Sons, Ltd., New York.

Rubin, D. B. (1996). Multiple imputation after 18+ years. *Journal of the American Statistical Association*, 91(434):473–489.

Rubin, D. B. and Schenker, N. (1991). Multiple imputation in health-care databases: an overview and some applications. *Statistics in Medicine*, 10(4):585–598.

SAS Institute Inc. (2003). *SAS/STAT user's guide, Version 9.1*. SAS Institute Inc., Cary.

Schafer, J. L. (1997). *Analysis of incomplete multivariate data*. Chapman and Hall, London.

Schafer, J. L. (1999). Multiple imputation: a primer. *Statistical methods in medical research*, 8:3–15.

Schafer, J. L. (2003). Multiple imputation in multivariate problems when the imputation and analysis models differ. *Statistica Neerlandica*, 57(1):19–35.

Scharfstein, D. O., Daniels, M. J., and Robins, J. M. (2003). Incorporating prior beliefs about selection bias into the analysis of randomized trials with missing outcomes. *Biostatistics*, 4(4):495–512.

Scharfstein, D. O., Rotnitzky, A., and Robins, J. M. (1999). Adjusting for nonignorable drop-out using semiparametric nonresponse models. *Journal of the American Statistical Association*, 94(448):1096–1120.

Scheike, T. H., Zhang, M. J., and Gerds, T. A. (2008). Predicting cumulative incidence probability by direct binomial regression. *Biometrika*, 95(1):205–220.

Seaman, S. R. and White, I. R. (2013). Review of inverse probability weighting for dealing with missing data. *Statistical Methods in Medical Research*, 22(3):278–295.

Shih, W. J. (1992). On informative and random dropouts in longitudinal studies. *Biometrics*, 48(3):970–972.

Song, X., Sun, L., Mu, X., and Dinse, G. E. (2010). Additive hazards regression with censoring indicators missing at random. *Canadian Journal of Statistics*, 38(3):333–351.

Spiegelhalter, D. J., Best, N. G., Carlin, B. P., and Van Der Linde, A. (2002). Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society. Series B (Methodological)*, 64(4):583–639.

Therneau, T. (2011). *bdsmatrix: Routines for Block Diagonal Symmetric matrices*. R package version 1.3, Available at http://cran.r-project.org/package=bdsmatrix (accessed September 15, 2012).

Thijs, H., Molenberghs, G., Michiels, B., Verbeke, G., and Curran, D. (2002). Strategies to fit pattern-mixture models. *Biostatistics*, 3(2):245–265.

Thijs, H., Molenberghs, G., and Verbeke, G. (2000). The milk protein trial: influence analysis of the dropout process. *Biometrical Journal*, 42(5):617–646.

Tsiatis, A. A. (1975). A nonidentifiability aspect of the problem of competing risks. *Proceedings of the National Academy of Sciences*, 72(1):20–22.

Tsiatis, A. A. (2006). *Semiparametric Theory and Missing Data*. Springer, New York.

van Buuren, S. and Groothuis-Oudshoorn, K. (2011). mice: Multivariate Imputation by Chained Equations in R. *Journal of Statistical Software*, 45(3):1–67.

van der Vaart, A. W. (2000). *Asymptotic Statistics*. Cambridge University Press, Cambridge.

van Rompaye, B., Jaffar, S., and Goetghebeur, E. (2012). Estimation with Cox models: cause-specific survival analysis with misclassified cause of failure. *Epidemiology*, 23(2):194–202.

Vansteelandt, S., Goetghebeur, E., Kenward, M. G., and Molenberghs, G. (2006). Ignorance and uncertainty regions as inferential tools in a sensitivity analysis. *Statistica Sinica*, 16:953–979.

Verbeke, G. and Molenberghs, G. (2000). *Linear Mixed Models for Longitudinal Data*. Springer, New York.

Verbeke, G., Molenberghs, G., Thijs, H., Lesaffre, E., and Kenward, M. G. (2001). Sensitivity analysis for nonrandom dropout: a local influence approach. *Biometrics*, 57(1):7–14.

White, H. (1982). Maximum likelihood estimation of misspecified models. *Econometrica*, 50(1):1–25.

White, I. R., Carpenter, J., Evans, S., and Schroter, S. (2007). Eliciting and using expert opinions about dropout bias in randomized controlled trials. *Clinical Trials*, 4(2):125–139.

Wu, M. C. and Bailey, K. R. (1988). Analysing changes in the presence of informative right censoring caused by death and withdrawal. *Statistics in Medicine*, 7:337–346.

Wu, M. C. and Bailey, K. R. (1989). Estimation and comparison of changes in the presence of informative right censoring: conditional linear model. *Biometrics*, 45(3):939–955.

Wu, M. C. and Carroll, R. J. (1988). Estimation and comparison of changes in the presence of informative right censoring by modeling the censoring process. *Biometrics*, 44(1):175–188.

Yan, J. (2002). geepack: Yet Another Package for Generalized Estimating Equations. *R-News*, 2/3:12–14.

Yan, J. and Fine, J. P. (2004). Estimating equations for association structures. *Statistics in Medicine*, 23:859–874.

Appendix A

Supplementary results for the simulation study of Chapter 3

Table A.1: Estimates of treatment effect: MAR scenario in which drop-out probability depended on the individual's group, with drop-out probabilities of 0.4 and 0.1 for the treatment and control groups, respectively. Results of 1000 simulations.

θ	Analysis	$\hat{\theta}$	$\hat{\sigma}_\theta$	SD $(\hat{\theta}^{(s)})$	CP	% reject H_0	MSE
0	Complete data	-0.003	0.328	0.331	94.3	5.6	0.110
	Complete cases	-0.003	0.387	0.385	94.8	4.9	0.148
	5 imputations	-0.007	0.365	0.350	95.3	4.4	0.123
	20 imputations	-0.007	0.361	0.345	95.4	4.6	0.119
1	Complete data	0.990	0.328	0.331	94.8	84.5	0.109
	Complete cases	0.989	0.387	0.387	94.6	72.5	0.150
	5 imputations	0.997	0.366	0.356	95.4	78.1	0.127
	20 imputations	0.993	0.362	0.353	95.2	78.5	0.124

Table A.2: Estimates of treatment effect: MAR scenario in which drop-out probability was positively associated with the last observed outcome. Results of 1000 simulations.

θ	p	Analysis	$\hat{\theta}$	$\hat{\sigma}_\theta$	SD $(\hat{\theta}^{(s)})$	CP	% reject H_0	MSE
0	0.1	Complete data	−0.006	0.329	0.329	95.5	4.5	0.108
		Complete cases	−0.010	0.334	0.332	95.6	4.4	0.110
		5 imputations	−0.007	0.337	0.338	95.6	4.4	0.114
		20 imputations	−0.008	0.336	0.336	95.5	4.5	0.113
	0.4	Complete data	0.017	0.328	0.336	93.6	6.4	0.113
		Complete cases	0.016	0.380	0.378	95.4	4.5	0.143
		5 imputations	0.016	0.389	0.372	95.4	4.0	0.139
		20 imputations	0.015	0.383	0.365	95.7	4.3	0.133
1	0.1	Complete data	1.018	0.329	0.321	96.5	87.8	0.103
		Complete cases	0.941	0.333	0.328	94.4	81.6	0.111
		5 imputations	1.018	0.337	0.325	96.2	85.9	0.106
		20 imputations	1.018	0.336	0.324	96.2	86.9	0.105
	0.4	Complete data	1.005	0.329	0.337	93.8	86.1	0.114
		Complete cases	0.796	0.380	0.382	91.8	53.6	0.187
		5 imputations	0.999	0.389	0.376	94.8	71.5	0.141
		20 imputations	1.002	0.383	0.370	95.0	75.9	0.137

Table A.3: Estimates of treatment effect: MAR scenario in which drop-out probability depended on the individual's group and his last observed outcome, with marginal drop-out probabilities of 0.4 and 0.1 for the treatment and control groups, respectively, and drop-out probability positively associated with the last observed outcome within each group. Results of 1000 simulations.

θ	Analysis	$\hat{\theta}$	$\hat{\sigma}_\theta$	SD $(\hat{\theta}^{(s)})$	CP	% reject H_0	MSE
0	Complete data	0.006	0.328	0.335	94.0	6.0	0.112
	Complete cases	−1.158	0.363	0.362	10.9	89.0	1.472
	5 imputations	0.005	0.364	0.359	94.9	5.0	0.129
	20 imputations	0.006	0.360	0.352	95.3	4.7	0.124
1	Complete data	0.986	0.328	0.331	94.6	84.9	0.110
	Complete cases	−0.230	0.362	0.357	7.1	9.2	1.641
	5 imputations	0.984	0.362	0.350	94.3	77.1	0.122
	20 imputations	0.985	0.360	0.348	94.4	78.7	0.121

Table A.4: Estimates of treatment effect: MNAR scenario in which drop-out probability was positively related to the first missing outcome. Results of 1000 simulations.

θ	p	Analysis	$\hat{\theta}$	$\hat{\sigma}_\theta$	SD $(\hat{\theta}^{(s)})$	CP	% reject H_0	MSE
0	0.1	Complete data	−0.021	0.328	0.329	94.8	5.2	0.109
		Complete cases	−0.021	0.326	0.325	94.6	5.4	0.106
		5 imputations	−0.019	0.332	0.330	94.9	5.1	0.109
		20 imputations	−0.020	0.331	0.329	95.4	4.6	0.109
	0.4	Complete data	0.013	0.329	0.334	95.0	5.0	0.112
		Complete cases	0.013	0.360	0.362	94.5	5.5	0.131
		5 imputations	0.014	0.374	0.368	95.0	4.4	0.136
		20 imputations	0.013	0.369	0.356	94.9	5.1	0.127
1	0.1	Complete data	1.013	0.329	0.356	92.6	84.5	0.127
		Complete cases	0.892	0.326	0.341	91.9	75.8	0.128
		5 imputations	0.997	0.332	0.358	93.1	82.2	0.128
		20 imputations	0.999	0.331	0.356	93.0	83.3	0.126
	0.4	Complete data	0.996	0.328	0.327	95.1	87.3	0.107
		Complete cases	0.707	0.360	0.373	86.2	51.4	0.225
		5 imputations	0.948	0.369	0.346	95.3	72.9	0.122
		20 imputations	0.948	0.366	0.339	95.9	76.4	0.118

Appendix B

Supplementary information for the simulation study of Chapter 4

In this appendix we provide further technical details (Section B.1) and additional results (Section B.2) for the simulation study presented in Chapter 4, Section 4.2.

B.1 Simulation study details

B.1.1 Generating data from the additive model

The additive model is not defined near 0 because the CIF is 0 at $t = 0$. The model is thus expected to hold only for $t \geq t_0 > 0$ where t_0 is some chosen time. However, to generate data via the inverse transformation method, the entire distribution has to be specified. To understand this, consider the case with a binary covariate and let $u := P(T \leq t|X, D = 1)$, the probability used to generate failure times for individuals failing from the cause of interest. For the chosen additive simulation model, this probability is

$$u = \frac{F_1(t|X)}{P(D = 1|X)} = \frac{p(1 - e^{-t}) + \beta^{AD}X}{p + \beta^{AD}X},$$

leading to $t = \log(p) - \log\{p(1-u) - \beta^{AD}X(u-1)\}$ when applying the inverse transformation. Thus, valid (positive) times are generated only if $X = 0$ or $u > \beta^{AD}/(p+\beta^{AD}) = 0.15/(0.5+0.15) \approx 0.23$. Therefore, our additive simulation model implies that approximately the first quartile of patients with $X = 1$ and failing from cause 1 must have already failed at $t = 0$. Nevertheless, as the model is expected to hold only from some chosen point t_0, it suffices that all of these patients fail at a positive time before t_0 to guarantee an additive structure after t_0 and a proper CIF (i.e. which is 0 at $t = 0$). For our purposes, these patients may be assigned any positive failure time before t_0 because pseudo-values are calculated at a grid starting at $\tau_1 \geq t_0$ and thus will be the same regardless of how these failure times are chosen. In our simulations, we assigned these patients the smallest failure time generated among all other patients. Therefore, these patients, who represented about $P(X = 1) \times P(D = 1|X = 1) \times 0.25 \approx 8\%$ of all patients given the parameters, all failed before the first decile of all failure times (irrespective of X and cause of failure). After superimposing censoring, they all failed before the second decile of the failure times because of the way censoring times were generated. Thus, choosing $t_0 = \tau_1$, the third decile of the distribution in the final dataset, guaranteed that these patients failed before t_0.

B.1.2 Analyzing data with high percentages of censoring and missing causes

When generating data with high percentages of censoring and/or missing causes of failure, some of the missingness configurations or entire datasets obtained had to be discarded and replaced because one of the analyses could not be performed, usually because there were very few events. Briefly, in the binary covariate case, datasets or missing causes were redrawn in the following cases: (a) violation of the condition $\tau_1 \geq t_0$, where τ_1 is the first point of the grid chosen in the CC analysis of the additive model, which occurred at least once for at most 0.3% of the replications in each scenario; (b) impossibility of estimating the weights $1/\pi_{\tau_s}(X)$ in the IPWpv analysis because for a

category of the covariate there were no events or no observed events before τ_1, which occurred, respectively, for at most 0.1% and 2% of the replications in each scenario (except for MARX- and MARXT+ with $n = 200$, 40% missing causes and 50% censoring where the latter occurred for up to 8% of the replications); (c) impossibility to obtain GEE estimates in CC or IPWpv analyses or to fit the imputation model in datasets with less than two observed type 1 or type 2 events for $X = 0$ or $X = 1$, which occurred at least once for at most 1% of the replications in each scenario (except for the MARXT+ mechanism with 40% missing causes and 50% censoring where it occurred at least once for up to 7% of the replications). Similar comments apply to the continuous covariate case.

When analyzing bootstrap samples with high percentages of censoring and missing causes, there were datasets or patterns of missingness where (b) or (c) above also occurred, so the IPWpv analysis could not be performed. In these cases, the bootstrap sample was discarded and not replaced so the bootstrap estimate was based on fewer than 100 bootstrap samples. The mean percentage of bootstrap samples discarded in each scenario exceeded 5% only in a few rare cases.

In practice, if either (a) or (b) occur, the grid can possibly be modified to ensure that $\tau_1 \geq t_0$ or that there is at least one event/one observed event before τ_1 for each value of the covariate, hopefully while keeping the time-points approximately evenly spaced. However, in the simulation study we decided to discard/redraw missing causes so that the grid could be chosen in the same way for each generated dataset, hence avoiding further variability due to differences in grid choice. If an appropriate grid cannot be found in practice or if (c) occurs, a workaround might be found depending on the data structure, as illustrated in our analysis of the ECOG data.

B.2 Additional simulation study results

Figures B.1 and B.8 show simulation results for the Fine and Gray and additive models, respectively, with a binary covariate for the scenario with $n = 200$, 50% uniform

censoring and the missingness mechanisms omitted from the main text of Chapter 4 (Section 4.2.3). Also for the binary covariate case, results for the scenarios with $n = 400$ and 50% uniform censoring are provided in Figures B.2-B.3 and B.9-B.10, results for $n = 200$ and 50% administrative censoring are found in Figures B.4-B.5 and B.11-B.12 and results corresponding to $n = 200$ and 25% uniform censoring are given in Figures B.6-B.7 and B.13-B.14. Finally, Figure B.15 shows the results for coefficient estimation in the Fine and Gray and additive models with a continuous covariate in a scenario with 50% uniform censoring and 40% missing causes for the missingness mechanisms omitted from the main text of of Chapter 4.

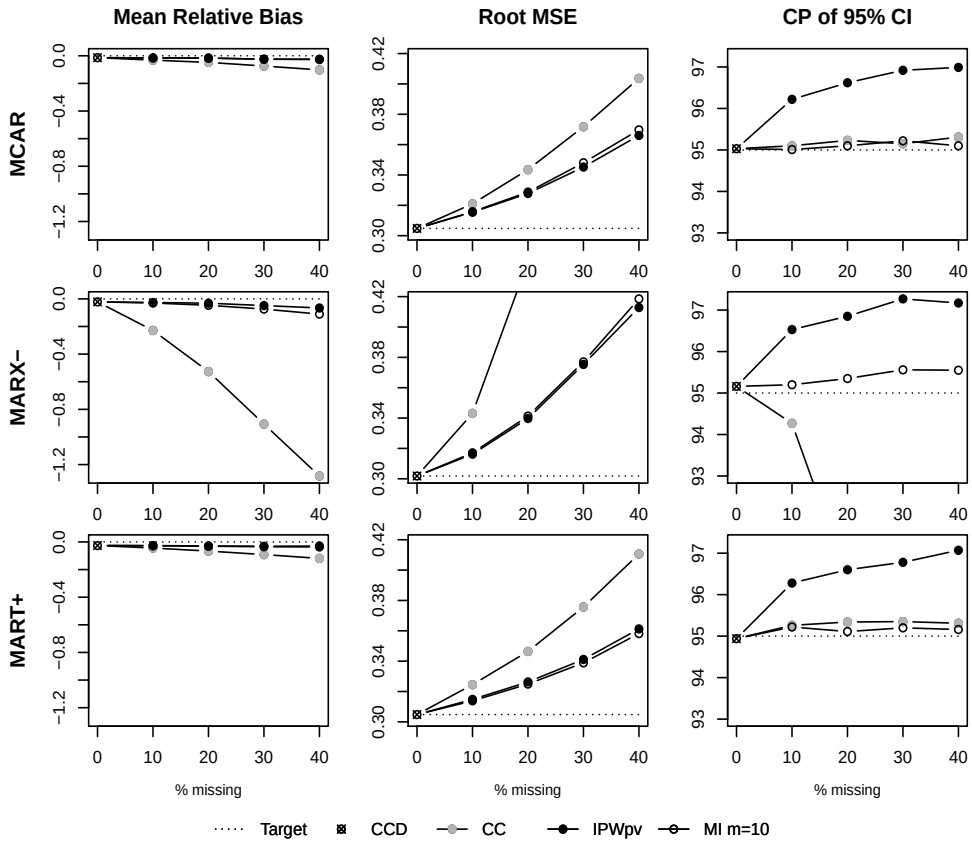

Figure B.1: Simulation results for coefficient estimation in the Fine and Gray model with a binary covariate, $n = 200$ and 50% uniform censoring for other missingness mechanisms. Estimates obtained via a complete case analysis (CC), the proposed IPWpv approach (IPWpv) and multiple imputation with $m = 10$ imputations (MI m=10) are compared. For each mechanism and each analysis, the mean relative bias, the square root of the mean squared error (MSE) and the estimated coverage probability (CP) of the 95% confidence interval (CI) are plotted against the percentage of missing causes. With no missing causes, all analyses coincide with the complete censored data analysis (CCD) plotted at 0%. Results are based on 10000 replications.

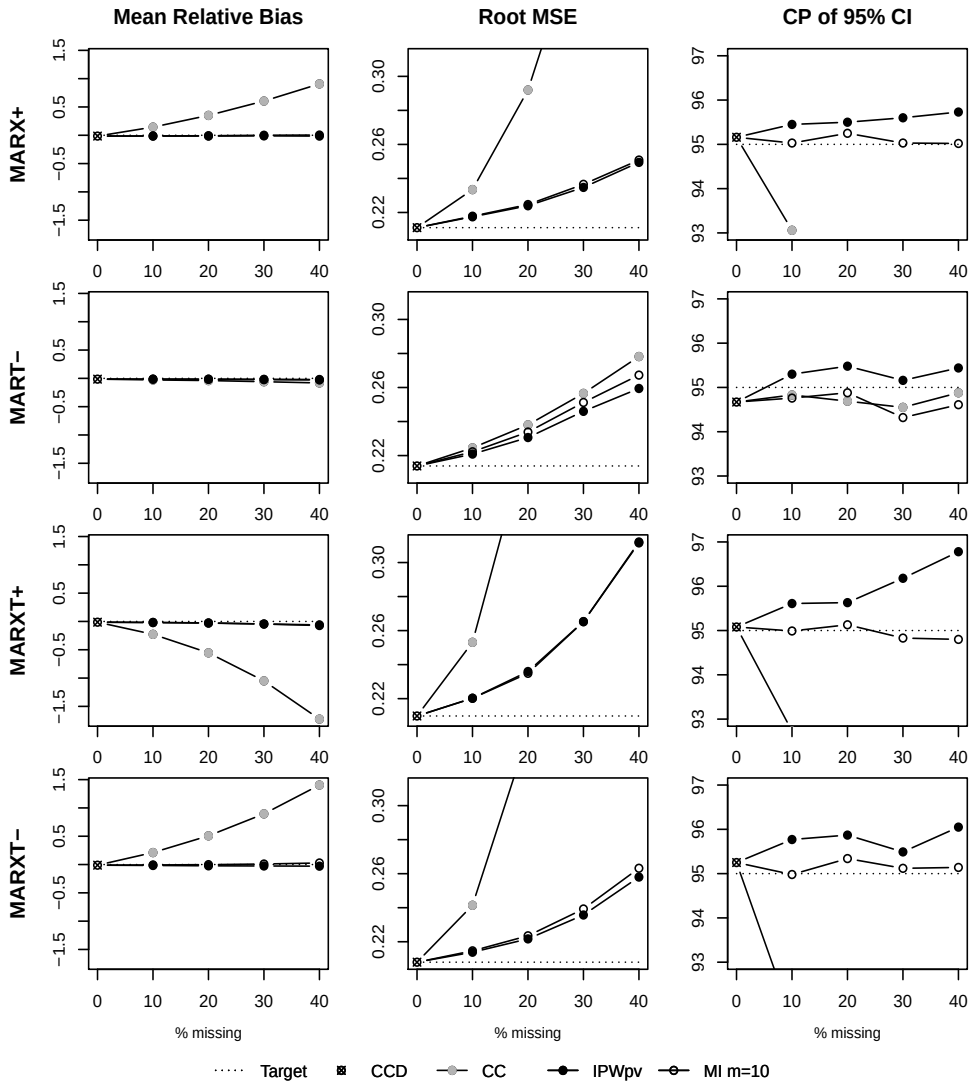

Figure B.2: Simulation results for coefficient estimation in the Fine and Gray model with a binary covariate, $n = 400$ and 50% uniform censoring for selected missingness mechanisms. Estimates obtained via a complete case analysis (CC), the proposed IPWpv approach (IPWpv) and multiple imputation with $m = 10$ imputations (MI m=10) are compared. For each mechanism and each analysis, the mean relative bias, the square root of the mean squared error (MSE) and the estimated coverage probability (CP) of the 95% confidence interval (CI) are plotted against the percentage of missing causes. With no missing causes, all analyses coincide with the complete censored data analysis (CCD) plotted at 0%. Results are based on 10000 replications.

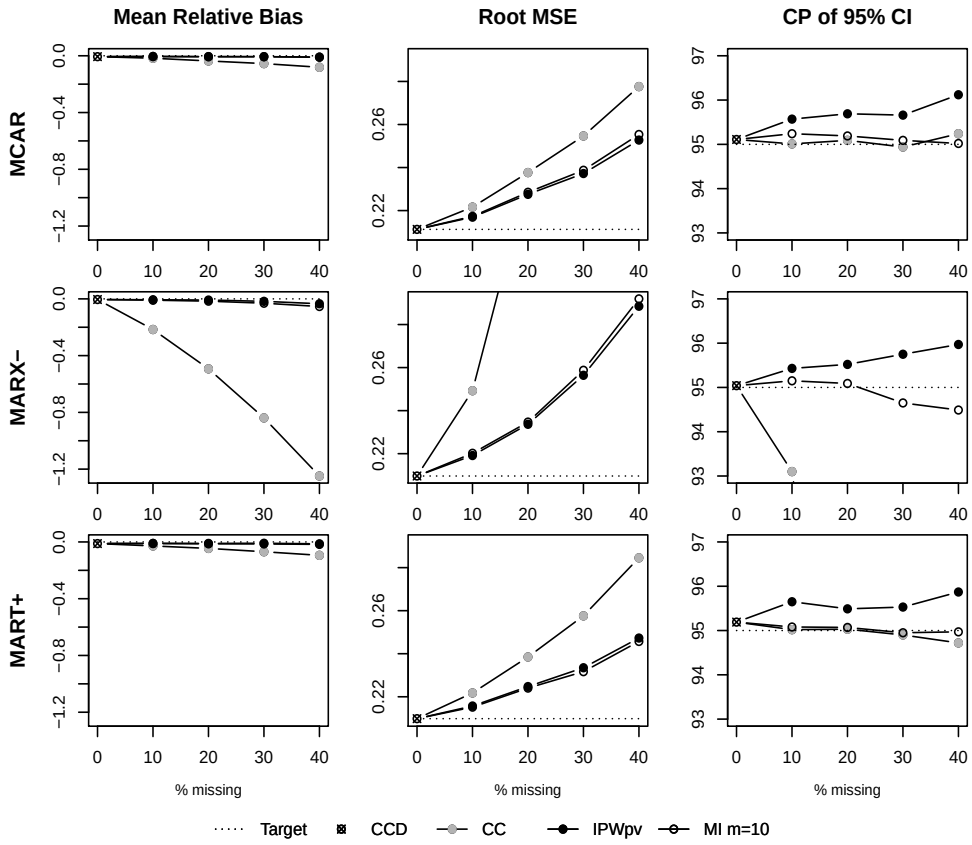

Figure B.3: Simulation results for coefficient estimation in the Fine and Gray model with a binary covariate, $n = 400$ and 50% uniform censoring for other missingness mechanisms. Estimates obtained via a complete case analysis (CC), the proposed IPWpv approach (IPWpv) and multiple imputation with $m = 10$ imputations (MI m=10) are compared. For each mechanism and each analysis, the mean relative bias, the square root of the mean squared error (MSE) and the estimated coverage probability (CP) of the 95% confidence interval (CI) are plotted against the percentage of missing causes. With no missing causes, all analyses coincide with the complete censored data analysis (CCD) plotted at 0%. Results are based on 10000 replications.

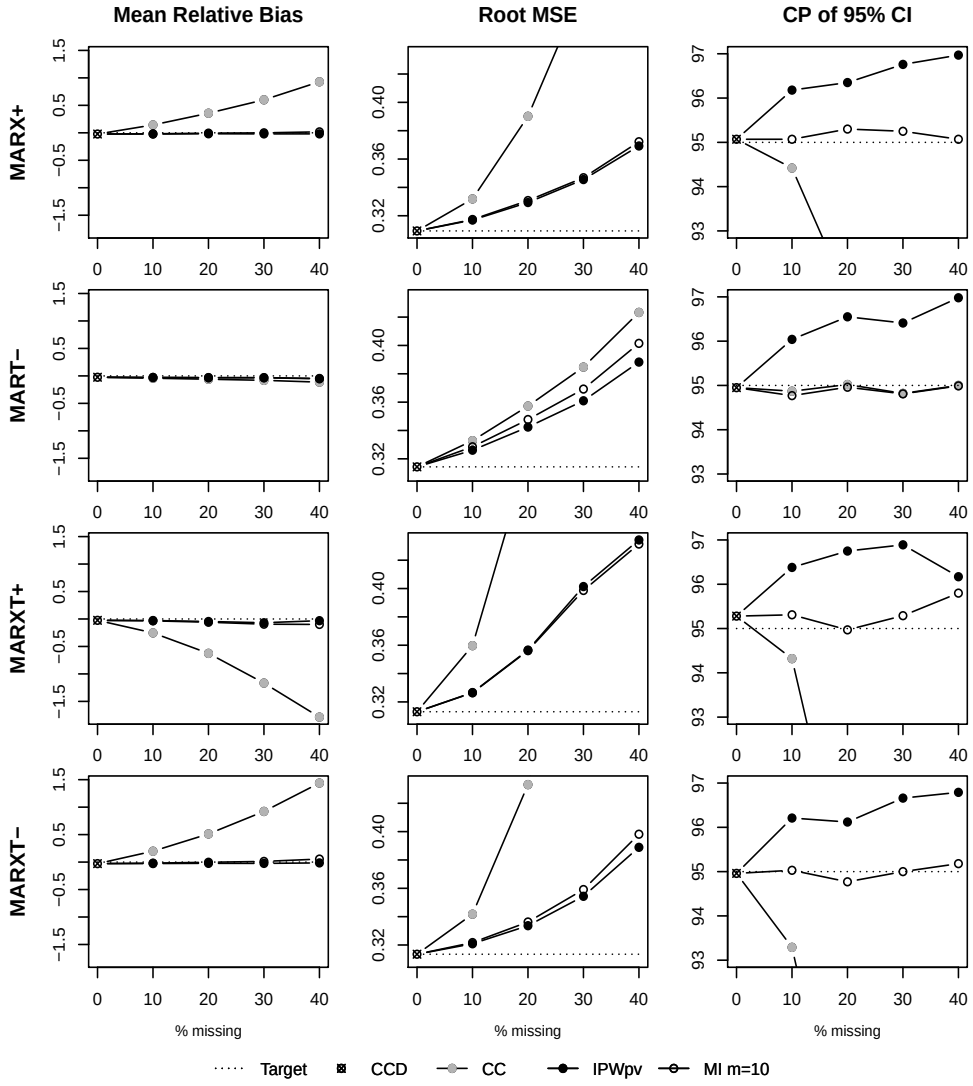

Figure B.4: Simulation results for coefficient estimation in the Fine and Gray model with a binary covariate, $n = 200$ and 50% administrative censoring for selected missingness mechanisms. Estimates obtained via a complete case analysis (CC), the proposed IPWpv approach (IPWpv) and multiple imputation with $m = 10$ imputations (MI m=10) are compared. For each mechanism and each analysis, the mean relative bias, the square root of the mean squared error (MSE) and the estimated coverage probability (CP) of the 95% confidence interval (CI) are plotted against the percentage of missing causes. With no missing causes, all analyses coincide with the complete censored data analysis (CCD) plotted at 0%. Results are based on 10000 replications.

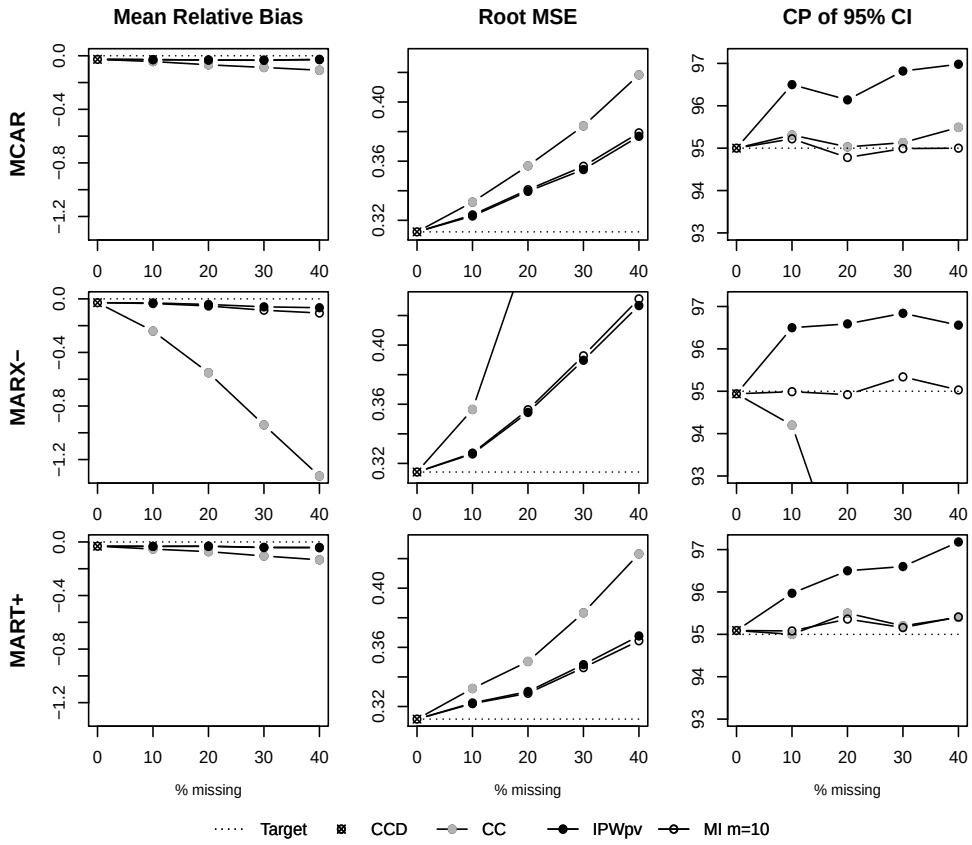

Figure B.5: Simulation results for coefficient estimation in the Fine and Gray model with a binary covariate, $n = 200$ and 50% administrative censoring for other missingness mechanisms. Estimates obtained via a complete case analysis (CC), the proposed IPWpv approach (IPWpv) and multiple imputation with $m = 10$ imputations (MI m=10) are compared. For each mechanism and each analysis, the mean relative bias, the square root of the mean squared error (MSE) and the estimated coverage probability (CP) of the 95% confidence interval (CI) are plotted against the percentage of missing causes. With no missing causes, all analyses coincide with the complete censored data analysis (CCD) plotted at 0%. Results are based on 10000 replications.

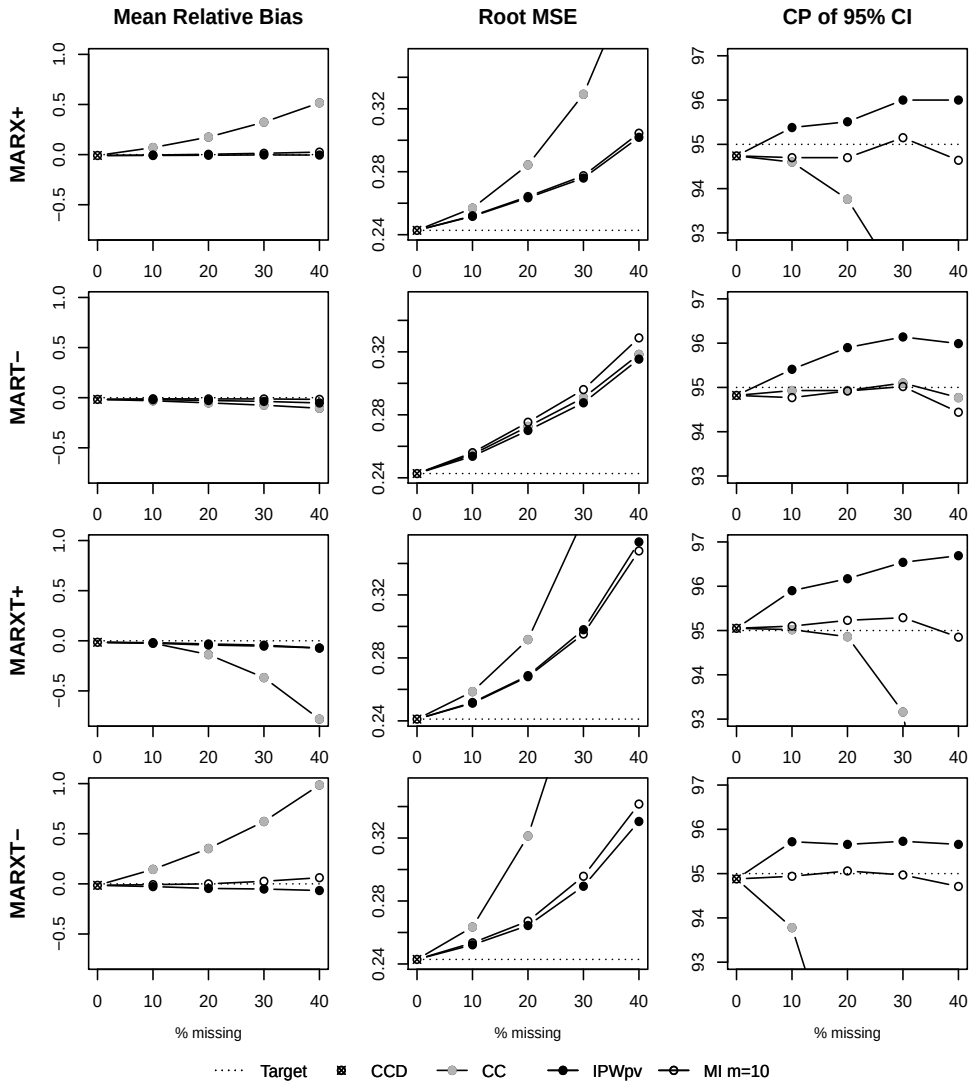

Figure B.6: Simulation results for coefficient estimation in the Fine and Gray model with a binary covariate, $n = 200$ and 25% uniform censoring for selected missingness mechanisms. Estimates obtained via a complete case analysis (CC), the proposed IPWpv approach (IPWpv) and multiple imputation with $m = 10$ imputations (MI m=10) are compared. For each mechanism and each analysis, the mean relative bias, the square root of the mean squared error (MSE) and the estimated coverage probability (CP) of the 95% confidence interval (CI) are plotted against the percentage of missing causes. With no missing causes, all analyses coincide with the complete censored data analysis (CCD) plotted at 0%. Results are based on 10000 replications.

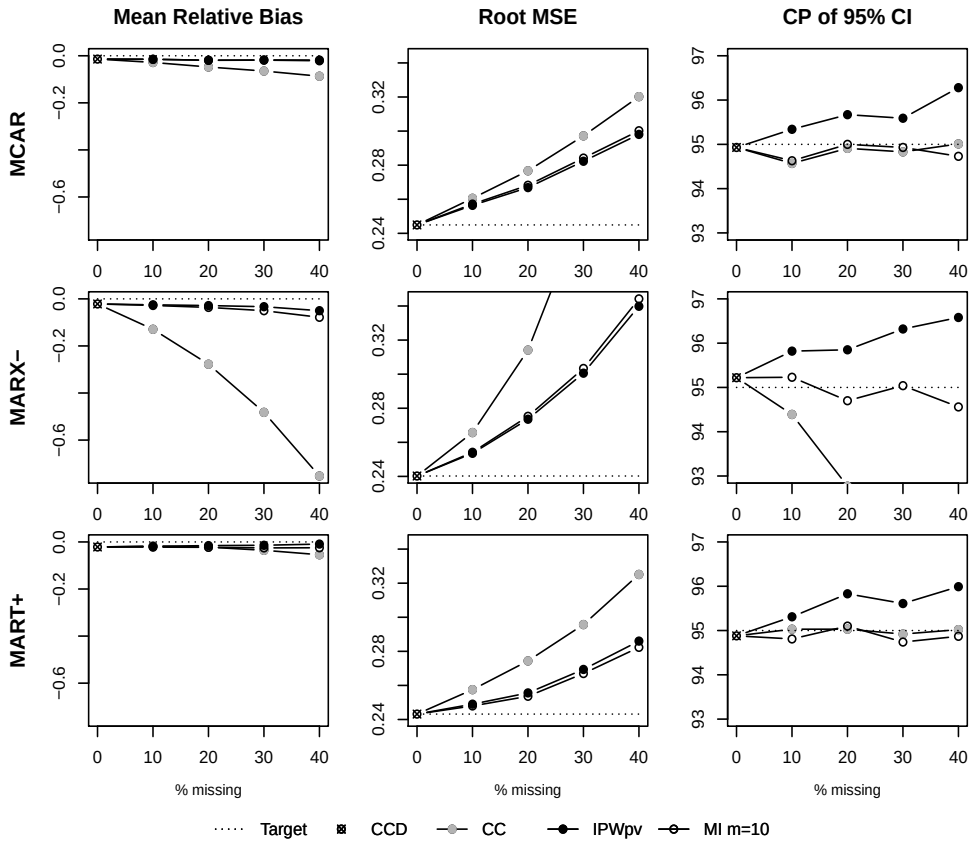

Figure B.7: Simulation results for coefficient estimation in the Fine and Gray model with a binary covariate, $n = 200$ and 25% uniform censoring for other missingness mechanisms. Estimates obtained via a complete case analysis (CC), the proposed IPWpv approach (IPWpv) and multiple imputation with $m = 10$ imputations (MI m=10) are compared. For each mechanism and each analysis, the mean relative bias, the square root of the mean squared error (MSE) and the estimated coverage probability (CP) of the 95% confidence interval (CI) are plotted against the percentage of missing causes. With no missing causes, all analyses coincide with the complete censored data analysis (CCD) plotted at 0%. Results are based on 10000 replications.

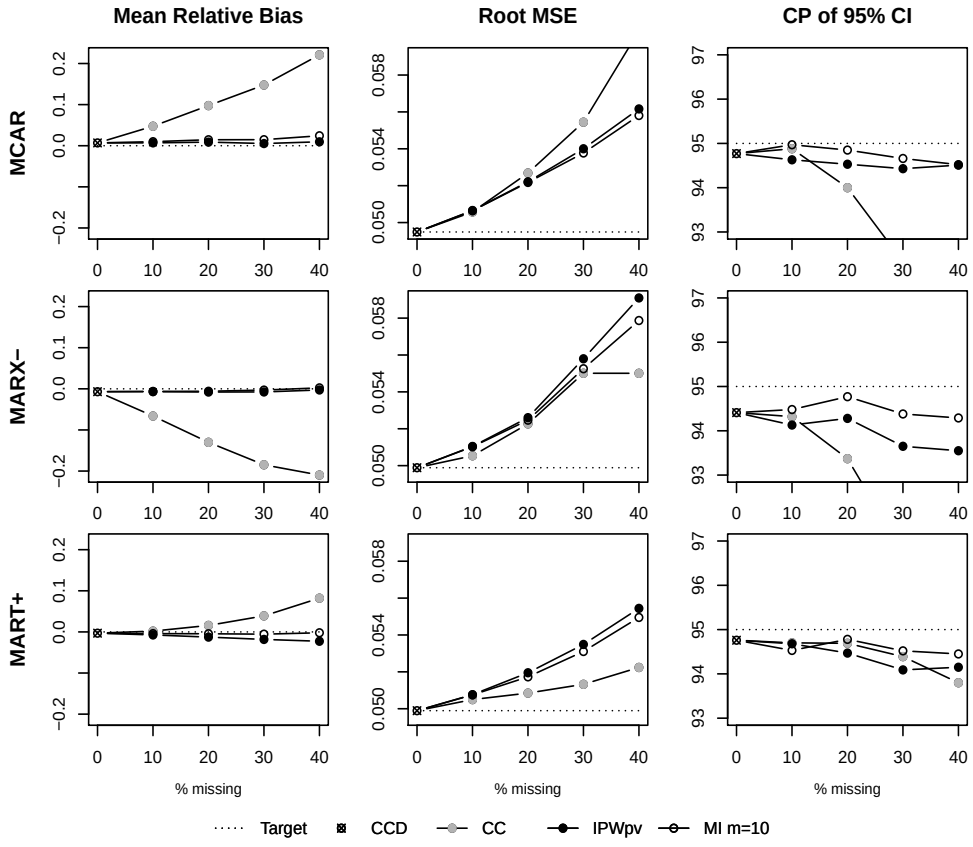

Figure B.8: Simulation results for coefficient estimation in the additive model with a binary covariate, $n = 200$ and 50% uniform censoring for other missingness mechanisms. Estimates obtained via a complete case analysis (CC), the proposed IPWpv approach (IPWpv) and multiple imputation with $m = 10$ imputations (MI m=10) are compared. For each mechanism and each analysis, the mean relative bias, the square root of the mean squared error (MSE) and the estimated coverage probability (CP) of the 95% confidence interval (CI) are plotted against the percentage of missing causes. With no missing causes, all analyses coincide with the complete censored data analysis (CCD) plotted at 0%. Results are based on 10000 replications.

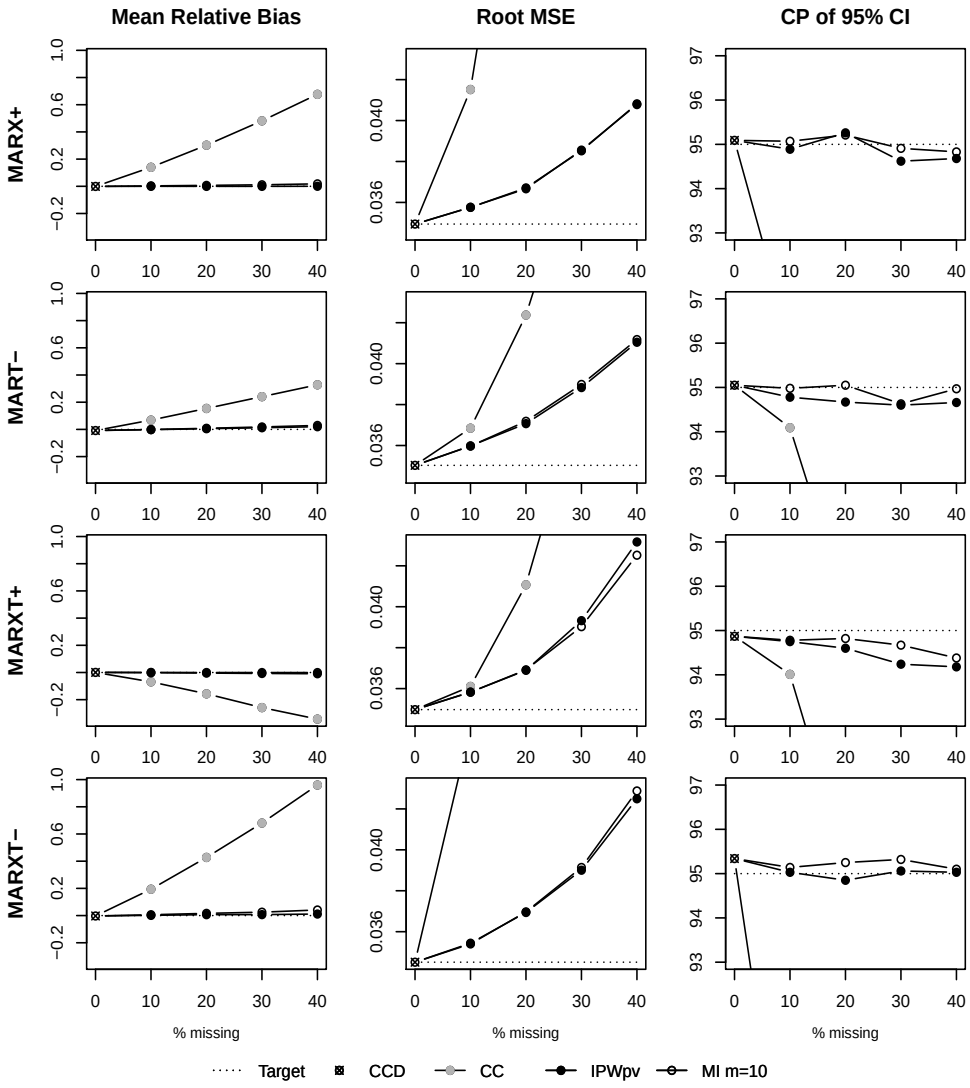

Figure B.9: Simulation results for coefficient estimation in the additive model with a binary covariate, $n = 400$ and 50% uniform censoring for selected missingness mechanisms. Estimates obtained via a complete case analysis (CC), the proposed IPWpv approach (IPWpv) and multiple imputation with $m = 10$ imputations (MI m=10) are compared. For each mechanism and each analysis, the mean relative bias, the square root of the mean squared error (MSE) and the estimated coverage probability (CP) of the 95% confidence interval (CI) are plotted against the percentage of missing causes. With no missing causes, all analyses coincide with the complete censored data analysis (CCD) plotted at 0%. Results are based on 10000 replications.

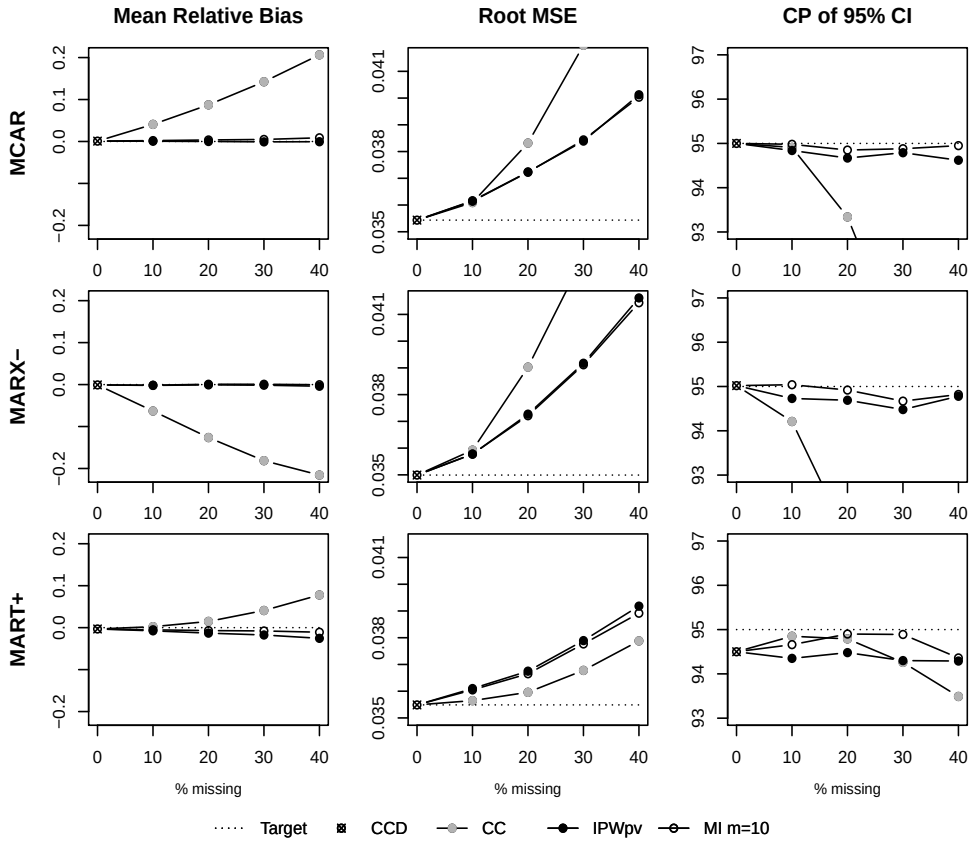

Figure B.10: Simulation results for coefficient estimation in the additive model with a binary covariate, $n = 400$ and 50% uniform censoring for other missingness mechanisms. Estimates obtained via a complete case analysis (CC), the proposed IPWpv approach (IPWpv) and multiple imputation with $m = 10$ imputations (MI m=10) are compared. For each mechanism and each analysis, the mean relative bias, the square root of the mean squared error (MSE) and the estimated coverage probability (CP) of the 95% confidence interval (CI) are plotted against the percentage of missing causes. With no missing causes, all analyses coincide with the complete censored data analysis (CCD) plotted at 0%. Results are based on 10000 replications.

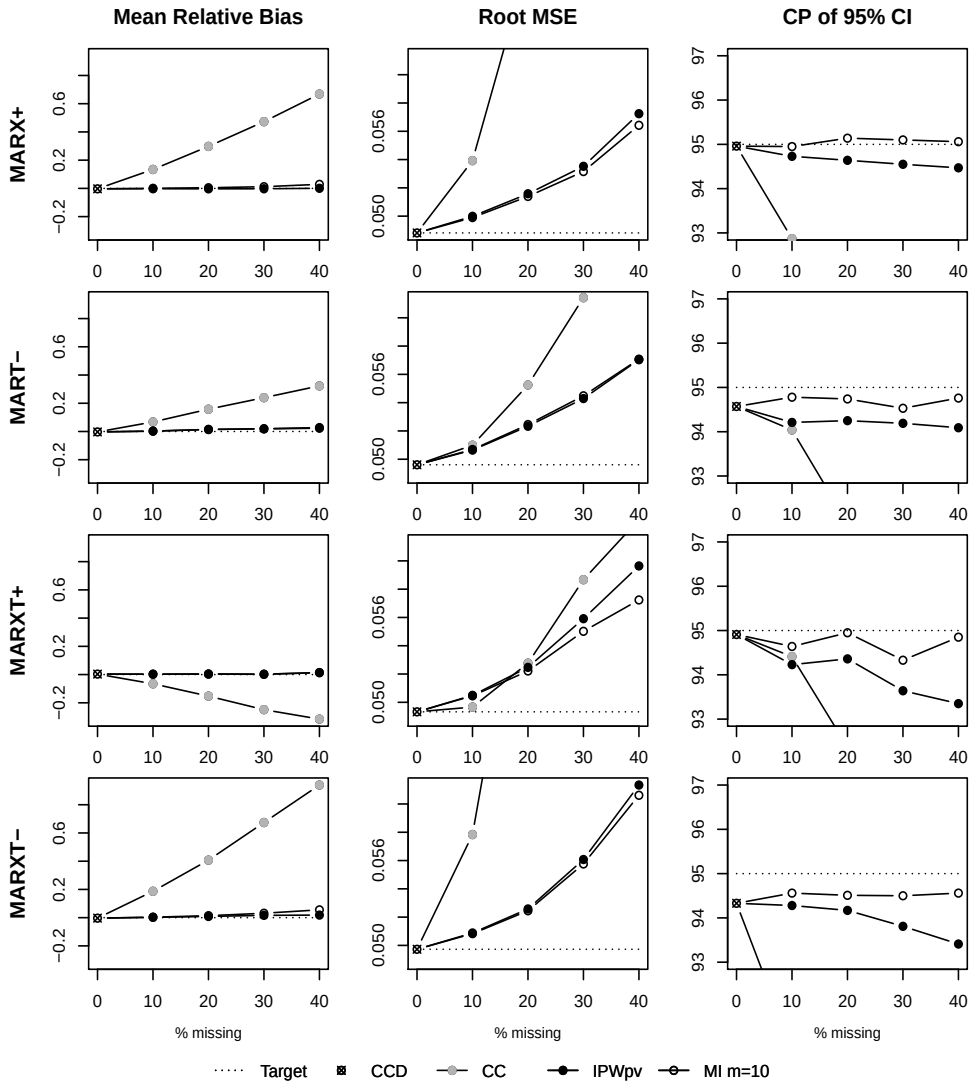

Figure B.11: Simulation results for coefficient estimation in the additive model with a binary covariate, $n = 200$ and 50% administrative censoring for selected missingness mechanisms. Estimates obtained via a complete case analysis (CC), the proposed IPWpv approach (IPWpv) and multiple imputation with $m = 10$ imputations (MI m=10) are compared. For each mechanism and each analysis, the mean relative bias, the square root of the mean squared error (MSE) and the estimated coverage probability (CP) of the 95% confidence interval (CI) are plotted against the percentage of missing causes. With no missing causes, all analyses coincide with the complete censored data analysis (CCD) plotted at 0%. Results are based on 10000 replications.

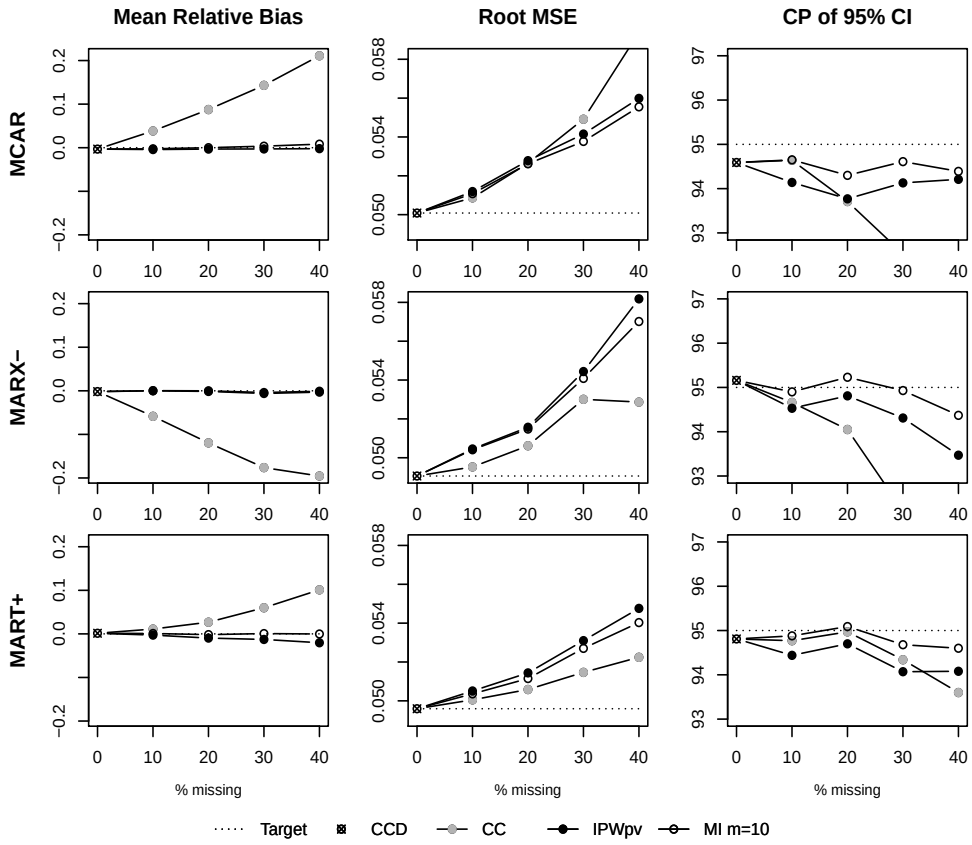

Figure B.12: Simulation results for coefficient estimation in the additive model with a binary covariate, $n = 200$ and 50% administrative censoring for other missingness mechanisms. Estimates obtained via a complete case analysis (CC), the proposed IPWpv approach (IPWpv) and multiple imputation with $m = 10$ imputations (MI m=10) are compared. For each mechanism and each analysis, the mean relative bias, the square root of the mean squared error (MSE) and the estimated coverage probability (CP) of the 95% confidence interval (CI) are plotted against the percentage of missing causes. With no missing causes, all analyses coincide with the complete censored data analysis (CCD) plotted at 0%. Results are based on 10000 replications.

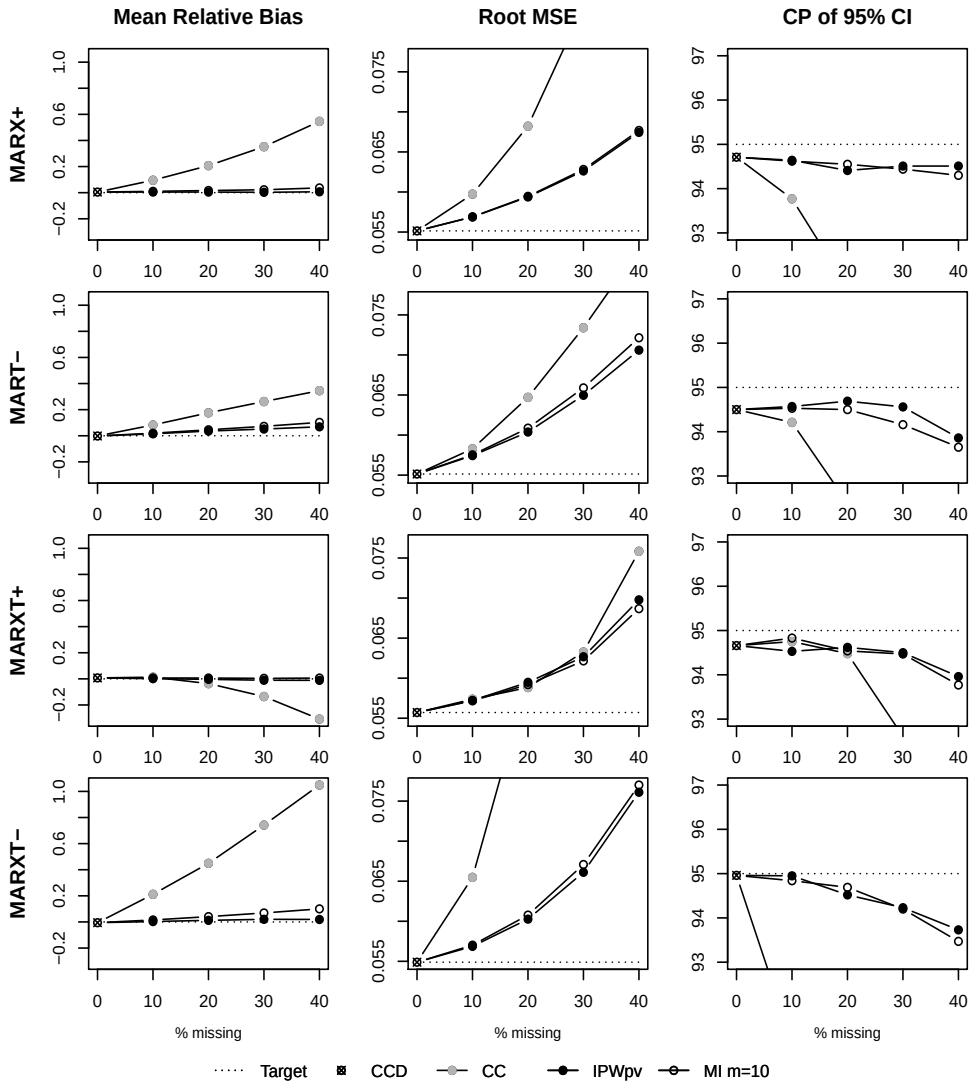

Figure B.13: Simulation results for coefficient estimation in the additive model with a binary covariate, $n = 200$ and 25% uniform censoring for selected missingness mechanisms. Estimates obtained via a complete case analysis (CC), the proposed IPWpv approach (IPWpv) and multiple imputation with $m = 10$ imputations (MI m=10) are compared. For each mechanism and each analysis, the mean relative bias, the square root of the mean squared error (MSE) and the estimated coverage probability (CP) of the 95% confidence interval (CI) are plotted against the percentage of missing causes. With no missing causes, all analyses coincide with the complete censored data analysis (CCD) plotted at 0%. Results are based on 10000 replications.

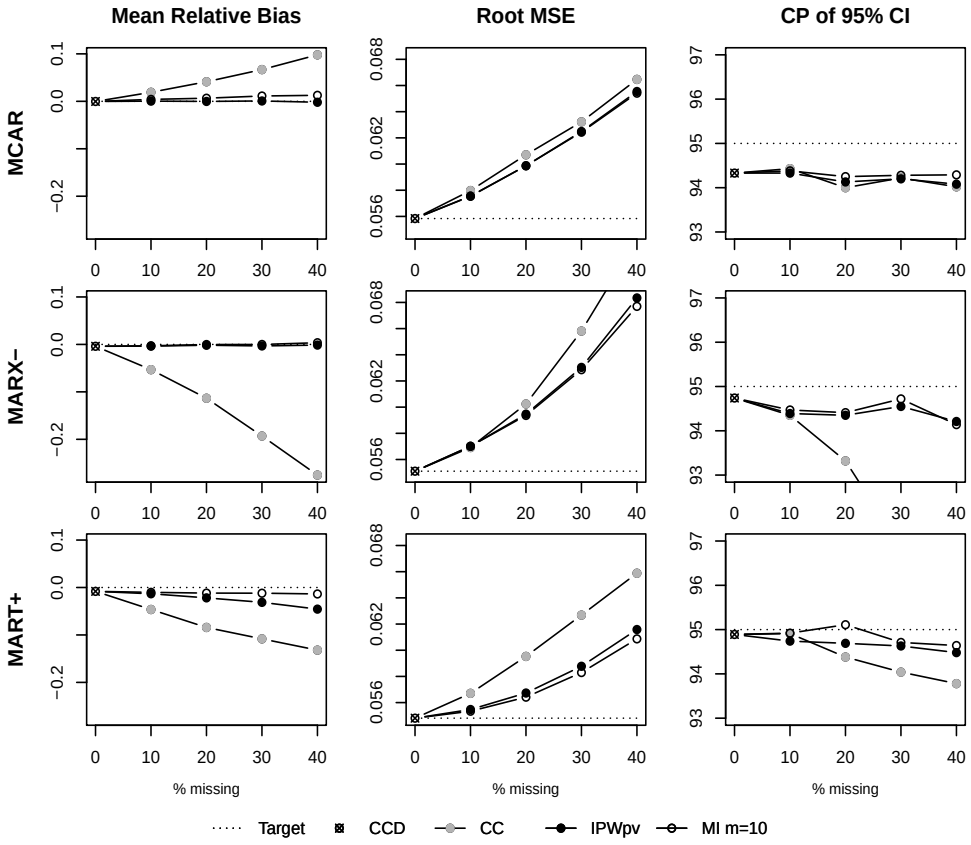

Figure B.14: Simulation results for coefficient estimation in the additive model with a binary covariate, $n = 200$ and 25% uniform censoring for other missingness mechanisms. Estimates obtained via a complete case analysis (CC), the proposed IPWpv approach (IPWpv) and multiple imputation with $m = 10$ imputations (MI m=10) are compared. For each mechanism and each analysis, the mean relative bias, the square root of the mean squared error (MSE) and the estimated coverage probability (CP) of the 95% confidence interval (CI) are plotted against the percentage of missing causes. With no missing causes, all analyses coincide with the complete censored data analysis (CCD) plotted at 0%. Results are based on 10000 replications.

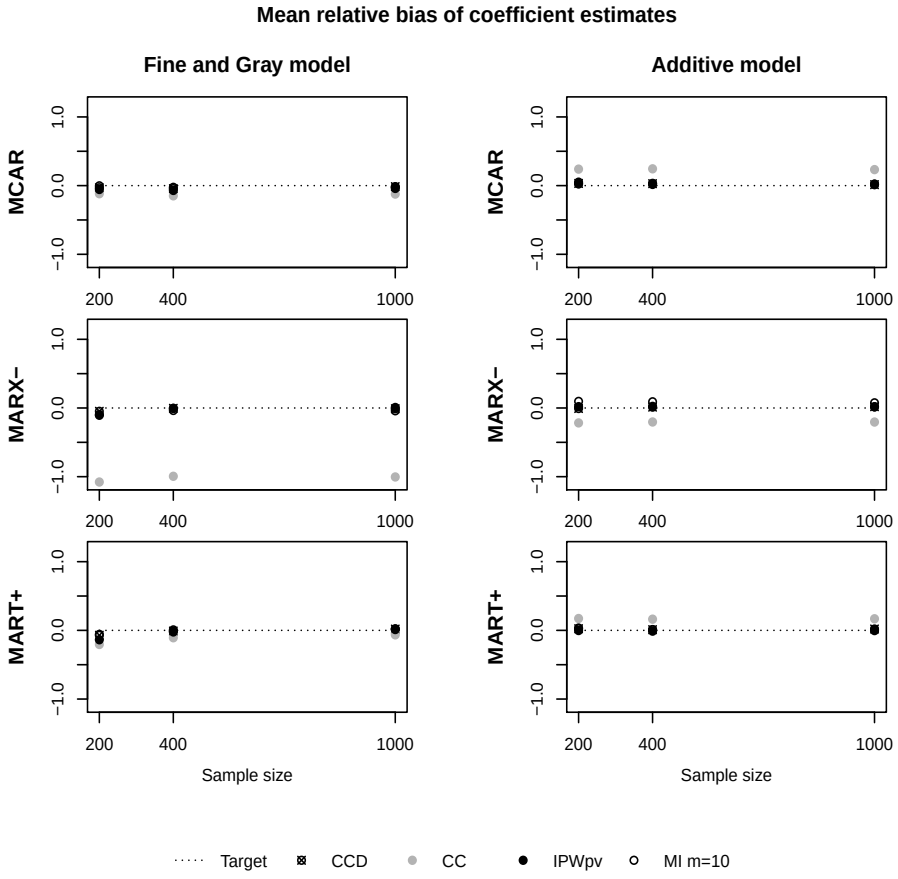

Figure B.15: Simulation results for coefficient estimation in the Fine and Gray and additive models with a continuous covariate, for other missingness mechanisms in a scenario with 50% uniform censoring and 40% missing causes. Estimates obtained via a complete case analysis (CC), the proposed IPWpv approach (IPWpv) and multiple imputation with $m = 10$ *imputations (MI m=10) are compared. For each mechanism and each analysis, the mean relative bias is plotted against the sample size (*$n = 200, 400, 1000$*). With no missing causes, all analyses coincide with the complete censored data analysis (CCD), also included in the plots. Results are based on 1000 replications.*

Appendix C

Manuscript in preparation

Moreno-Betancur, M., Latouche, A., Rey, G. A structured framework for estimation of the relative index and slope index of inequality in studies of socioeconomic gradients in event rates and risks (*in preparation*).