Comparison of several analysis methods for recurrent event data for different estimands

Master Thesis presented to the Department of Economics at the

Georg-August-University Göttingen

with a working time of 20 weeks

In partial fulfillment of the requirements for the degree

Master of Science (M. Sc.)

Maria Elisabeth Stark

from: Mering (21556770)

Supervised by:
PD Dr. rer. nat. Norbert Benda
Prof. Dr. sc. hum. Tim Friede

submitted: May 4, 2018

Abstract

Introduction: An estimand describes the treatment effect that should be measured in the context of a clinical trial (Mehrotra et al., 2016, p. 457). In this thesis, the evaluation of the hypothetical estimand and the treatment-policy estimand of a simulation study is performed. The hypothetical estimand captures the treatment effect which could be measured, if no treatment discontinuation occurred (Permutt, 2016, p. 2867). The treatment-policy estimand would evaluate the effect of the combination of the new drug and a rescue medication, if it was needed (Akacha et al., 2017b, p. 10).

Methods: The Log-linear Poisson model, the Log-linear negative-binomial model, the Log rank test, the Cox proportional hazards model, the Likelihood ratio test, the Gamma frailty model, the Shared gamma frailty model and the χ^2-test are used for the evaluation of both estimands. Hereby, the influence of the overdispersion, the mechanism and proportion of the treatment discontinuation and the proportion of the study dropout are examined.

Results: If overdispersion occurs, the type I error rate will increase and the power will shrink. The bias of the methods isn't affected a lot. Treatment discontinuation will have the strongest impact on the results, if the percentage of patients in the treatment group is larger than the one in the placebo group. Hereby, the type I error rate and the bias increase the most, the power shrinks. Study dropout enlarges the bias, the power is diminished. The type I error rate of the methods evaluating the time to the first event grows the most with very high percentages of study dropout. Under treatment discontinuation mechanism missing not at random, the type I error rate and the bias increase and the power shrinks compared to missing completely at random. This is especially valid for the count data models and the Shared gamma frailty models as well as for the χ^2-test.

Conclusions: To yield a valid evaluation of a clinical trial in reality, the proportion of patients with treatment discontinuation in both trial arms should be equal. For the evaluation of these scenarios without overdispersion, the Shared gamma frailty model or the Log-linear Poisson model are proposed. In scenarios with overdispersion, the Cox proportional hazards model or the Log rank test are optimal for the evaluation of the hypothetical estimand. For the treatment-policy estimand, the Shared gamma frailty model is recommended as it respects the whole period of time including the time under rescue medication by providing the largest power and smallest bias. It doesn't reveal an acceptable type I error rate in some scenarios. But their 95%-confidence intervals of the calculated type I error rates include an acceptable type I error rate.

Table of Contents

List of Figures

List of Tables

List of Abbreviations

A	active group (treatment group)
AHR	average hazard ratio
CHMP	Committee for Medicinal Products for Human Use
Cox model	Cox proportional hazards model
df	degree of freedom
DREAM	Dose Ranging Efficacy And safety with Mepolizumab in severe asthma
GF model	Gamma frailty model
GLM NB	Log-linear negative-binomial model
GLM Poisson	Log-linear Poisson model
HR	hazard ratio
LR test	Likelihood ratio test
MAR	missing at random
MCAR	missing completely at random
MLE	maximum likelihood estimator
MNAR	missing not at random
no tr. effect	without treatment effect
OR	odds ratio
PL	placebo group
SGF model	Shared gamma frailty model
tr.-policy	treatment-policy estimand
with tr. effect	with treatment effect

List of Symbols

a	occurrence of at least one asthma exacerbation
$\hat{\alpha}$	calculated type I error rate
α_A	first parameter of Gamma distribution in treatment group
α_{PL}	first parameter of Gamma distribution in placebo group
α_1	first parameter of Gamma distribution before treatment discontinuation
α_{1_A}	first parameter of Gamma distribution in treatment group before treatment discontinuation
$\alpha_{1_{PL}}$	first parameter of Gamma distribution in placebo group before treatment discontinuation
α_2	first parameter of Gamma distribution after treatment discontinuation
α_{2_A}	first parameter of Gamma distribution in treatment group after treatment discontinuation
$\alpha_{2_{PL}}$	first parameter of Gamma distribution in placebo group after treatment discontinuation
β_A	second parameter of Gamma distribution in treatment group
β_{PL}	second parameter of Gamma distribution in placebo group
β_0	the constant in count data model
β_1	the slope of the covariate coding for the trial arm
δ_{A_i}	censoring indicator of patient i in treatment group
δ_i	censoring indicator of patient i
δ_{PL_i}	censoring indicator of patient i in placebo group
Δt	period of time which is of interest
e	Euler's number

$\mathrm{E}(...)$	expected value
E_{aj}	expected counts of all possible combinations of a and j
E_j	expected frequency in j
$\mathrm{Exp}(\lambda_E)$	Exponential distribution
η_i	linear predictor of the count data model
$f(...)$	probability density function
$f_{\mathrm{A+PL}}(...)$	common probability density function of treatment and placebo group
$f_{\mathrm{NB}}(...)$	probability density function of t_{TD} if $y_1 \sim \mathrm{PoGa}(\alpha_1, \beta, \vartheta)$
$f_{\mathrm{P}}(...)$	probability density function of t_{TD} if $y_1 \sim \mathrm{Po}(\lambda_{P_1})$
$f_Y(...)$	probability density function of the total number of events until the end of the study
$f_1(...)$	probability density function before treatment discontinuation
$f_2(...)$	probability density function after treatment discontinuation
$F(...)$	cumulative distribution function
$g(...)$	parametric part of the Cox proportional hazards model
γ_A	proportion of patients in treatment group
γ_{PL}	proportion of patients in placebo group
$\mathrm{Ga}(\alpha, \beta)$	Gamma distribution
$\Gamma(...)$	Gamma function
$h(...)$	hazard function
$h_{\mathrm{A}}(...)$	hazard function of treatment group
$h_{\mathrm{PL}}(...)$	hazard function of placebo group
$h_{\mathrm{total}}(...)$	total hazard function

h_0	baseline hazard
$H(...)$	cumulative hazard function
H_0	null hypothesis
$H_{0_{hyp}}$	null hypothesis under the hypothetical estimand
$H_{0_{GF_{hyp}}}$	null hypothesis under the hypothetical estimand of the Gamma frailty model
$H_{0_{GF_{tp}}}$	null hypothesis under the treatment-policy estimand of the Gamma frailty model
$H_{0_{LR_{hyp}}}$	null hypothesis under the hypothetical estimand of the Likelihood ratio test
$H_{0_{LR_{tp}}}$	null hypothesis under the treatment-policy estimand of the Likelihood ratio test
$H_{0_{NB_{hyp}}}$	null hypothesis under the hypothetical estimand of the Log-linear negative-binomial model
$H_{0_{NB_{tp}}}$	null hypothesis under the treatment-policy estimand of the Log-linear negative-binomial model
$H_{0_{Pois_{hyp}}}$	null hypothesis under the hypothetical estimand of the Log-linear Poisson model
$H_{0_{Pois_{tp}}}$	null hypothesis under the treatment-policy estimand of the Log-linear Poisson model
$H_{0_{SGF_{hyp}}}$	null hypothesis under the hypothetical estimand of the Shared gamma frailty model
$H_{0_{SGF_{tp}}}$	null hypothesis under the treatment-policy estimand of the Shared gamma frailty model
$H_{0_{tp}}$	null hypothesis under the treatment-policy estimand
i	patient
$I_B(...)$	indicator function
j	trial arm

$l(...)$	log-likelihood function
$L(...)$	likelihood function
λ	parameter of Exponential distribution, for simplicity without index
$\widehat{\lambda}$	maximum likelihood estimator
$\widehat{\lambda_A}$	maximum likelihood estimator of the treatment group
$\widehat{\lambda_C}$	maximum likelihood estimator of the combined sample
$\widehat{\lambda_{PL}}$	maximum likelihood estimator of the placebo group
λ_A	parameter of Exponential distribution in treatment group, for simplicity without index E
λ_{E_A}	parameter of Exponential distribution in treatment group
$\lambda_{E_{PL}}$	parameter of Exponential distribution in placebo group
λ_{E_1}	parameter of Exponential distribution before treatment discontinuation
$\lambda_{E_{1_A}}$	parameter of Exponential distribution before treatment discontinuation in treatment group
$\lambda_{E_{1_{PL}}}$	parameter of Exponential distribution before treatment discontinuation in placebo group
λ_{E_2}	parameter of Exponential distribution after treatment discontinuation
$\lambda_{E_{2_A}}$	parameter of Exponential distribution after treatment discontinuation in treatment group
$\lambda_{E_{2_{PL}}}$	parameter of Exponential distribution after treatment discontinuation in placebo group
λ_{P_A}	parameter of Poisson distribution in treatment group
λ_{P_i}	outcome of count data model
$\lambda_{P_{PL}}$	parameter of Poisson distribution in placebo group

λ_{PL}	parameter of Exponential distribution in placebo group, for simplicity without index E
λ_{P_1}	parameter of Poisson distribution before treatment discontinuation
λ_{P_2}	parameter of Poisson distribution after treatment discontinuation
λ_{TD_A}	treatment discontinuation rate in the treatment group
$\lambda_{TD_{PL}}$	treatment discontinuation rate in the placebo group
λ_{1_A}	exacerbation rate in the treatment group before treatment discontinuation
$\lambda_{1_{PL}}$	exacerbation rate in the placebo group before treatment discontinuation
λ_{2_A}	exacerbation rate in the treatment group after treatment discontinuation
$\lambda_{2_{PL}}$	exacerbation rate in the placebo group after treatment discontinuation
$\log(...)$	natural logarithm
$\log(\Delta_{estimated})$	logarithmized estimated treatment effect
$\log(\Delta_{hyp})$	logarithmized hypothetical estimand
$\log(\Delta_{tp})$	logarithmized treatment-policy estimand
$\log(\Delta_{true})$	logarithmized true treatment effect in general
$\log(\Delta_{true_{CD_{hyp}}})$	true logarithmized hypothetical estimand of count data models
$\log(\Delta_{true_{CD_{tp_{MCAR}}}})$	true logarithmized treatment-policy estimand of count data models under MCAR
$\log(\Delta_{true_{CD_{tp_{MNAR}}}})$	true logarithmized treatment-policy estimand of count data models under MNAR

$\log(\Delta_{true_{E_{hyp}}})$	true logarithmized hypothetical estimand of time-to-first-event methods
$\log(\Delta_{true_{E_{tp_{MCAR}}}})$	true logarithmized treatment-policy estimand of time-to-first-event methods under MCAR
$\log(\Delta_{true_{E_{tp_{MNAR}}}})$	true logarithmized treatment-policy estimand of time-to-first-event methods under MNAR
$\log(\Delta_{true_{OR_{hyp}}})$	true logarithmized hypothetical estimand of χ^2-test
$\log(\Delta_{true_{OR_{tp_{MCAR}}}})$	true logarithmized treatment-policy estimand of χ^2-test under MCAR
$\log(\Delta_{true_{OR_{tp_{MNAR}}}})$	true logarithmized treatment-policy estimand of χ^2-test under MNAR
m	total number of trial arms
$M(t)$	process history
n_A	number of patients in treatment group
n_{PL}	number of patients in placebo group
N	total number of patients
N_{conv}	number of converged models
O_{aj}	observed frequency in a and j
O_j	observed frequency in j
ϕ	overdispersion parameter
p	probability of success in each trial
P(...)	probability
p_1	probability of success in each trial before treatment discontinuation
p_2	probability of success in each trial after treatment discontinuation

$\text{Po}(\lambda_P)$	Poisson distribution
$\text{PoGa}(\alpha, \beta, \vartheta)$	Poisson-Gamma distribution
$\%_{0E_A}$	percentage of patients with at least one event in treatment group
$\%_{0E_{PL}}$	percentage of patients with at least one event in placebo group
$\%_{0E_{1_A}}$	percentage of patients with at least one event in treatment group before treatment discontinuation
$\%_{0E_{1_{PL}}}$	percentage of patients with at least one event in placebo group before treatment discontinuation
$\%_{0E_2}$	percentage of patients with at least one event after treatment discontinuation
$\%_{0SD}$	percentage of patients with study dropout
$\%_{0TD_A}$	percentage of patients with treatment discontinuation in treatment group
$\%_{0TD_{PL}}$	percentage of patients with treatment discontinuation in placebo group
r	number of uncensored patients
r_A	number of uncensored patients in treatment group
r_{PL}	number of uncensored patients in placebo group
s	number of the two possibilities of either the occurrence of an event or no occurrence
$s(...)$	score function
$S(...)$	survival function
$S_0(...)$	baseline survival function
t	time
t_E	time until the next exacerbation

t_{A_i}	time until the first event of patient i in treatment group
t_i	time until the first event of patient i
t_{PL_i}	time until the first event of patient i in placebo group
t_{SD}	time until study dropout
t_{SD-TD}	time between treatment discontinuation and study dropout
t_{TD}	time until treatment discontinuation
Var(...)	variance
$w(...)$	weight function
x_i	covariate representing trial group
$\chi^2_{Ind.}$	test statistic of the Chi-squared test of independence
$\chi^2_{Logrank}$	test statistic of the Log rank test
χ^2_{LR}	test statistic of Likelihood ratio test
χ^2_1	Chi-squared distribution with one degree of freedom
y	number of events during the study
y_i	number of exacerbations of the individual patient i
y_1	number of exacerbations before treatment discontinuation
y_2	number of exacerbations after treatment discontinuation
z_i	individual frailty
Z	frailties

1 Introduction

In the context of a clinical trial, the treatment effect of a newly developed drug can be investigated. But before a clinical is permitted, a protocol must be elaborated. In recommendation 1 of the National Research Council's report, which is called 'The Prevention and Treatment of Missing Data in Clinical Trials' (National Research Council et al., 2010, p. 26) requirements of such a clinical trial protocol are named:

> "The trial protocol should explicitly define (a) the objective(s) of the trial; (b) the associated primary outcome or outcomes; (c) how, when and on whom the outcome or outcomes will be measured; and (d) the measures of intervention effects, that is, the causal estimands of primary interest. These measures should be meaningful for all study participants, and estimable with minimal assumptions. Concerning the latter, the protocol should address the potential impact and treatment of missing data."

This recommendation announces several important aspects that should be addressed in a clinical trial. Concerning a confirmatory clinical trial in asthma, the objective of the trial should be the investigation of the treatment effect of a newly developed drug. Hence, information about efficacy and safety of this drug can be provided to a regulatory agency (Committee for Medicinal Products for Human Use (CHMP) et al., 2017, p. 3). The 'Guideline on the clinical investigation of medicinal products for the treatment of asthma' proposes the exacerbation rate as the primary outcome (Committee for Medicinal Products for Human Use (CHMP) et al., 2015). Other possibilities can be the time to the first asthma exacerbation or the occurrence of at least one exacerbation. A study duration of at least six months is required by the Guideline (Committee for Medicinal Products for Human Use (CHMP) et al., 2015, p. 11). The grade of severity of the asthma disease of the recruited patients should be comparable to the grade of severity to which the new product is intended to be applied (Committee for Medicinal Products for Human Use (CHMP) et al., 2015, p. 6). Furthermore, a stratified randomisation of the patients regarding the aspects of smoking history, prior number of exacerbations and use of further asthma drugs is proposed by the relevant Guideline Committee for Medicinal Products for Human Use (CHMP) et al., 2015, p. 6). The 'ICH E9 addendum on estimands and sensitivity analysis in clinical trials to the guideline on statistical principles in clinical trials' refers to the measures of the intervention effect. Consequently it proposes different estimands for the estimation of the treatment effect. Choosing the appropriate estimand, the strategy to address intercurrent events must be predefined (Committee for Medicinal Products for Human Use (CHMP) et al., 2017, p. 5). An intercurrent event in a confirmatory clinical trial can be the change

of medication due to medical reasons. If the hypothetical estimand is measured, the intercurrent events will be ignored (Committee for Medicinal Products for Human Use (CHMP) et al., 2017, p. 7). In this case, the trial is evaluated as if the intercurrent event had not occurred. Otherwise, if the treatment effect is based on data before and after the intercurrent event, the treatment-policy estimand will be an appropriate choice. If a change of medication after treatment discontinuation is performed, the estimand will include the effect of the combination of treatments that is actually applied in practice. An estimation of the hypothetical estimand with minimal assumptions would be possible if a complete follow-up of all subjects in the absence of any intercurrent event was given. Whereas the treatment-policy estimand would be estimable with minimal assumptions if all patients were observed until the end of the study (Committee for Medicinal Products for Human Use (CHMP) et al., 2017, p. 13). The impossibility of a complete follow-up, e.g. because of a study dropout, leads to missing data (Leuchs et al., 2017, p. 13). For the hypothetical estimand data after a treatment discontinuation is not considered and a change of medication will lead to missing data. Whereas, for the treatment-policy estimand, treatment discontinuation and study dropout are not the same. In this thesis, hypothetical and treatment-policy estimands are estimated with different methods in the context of a simulated placebo-controlled confirmatory clinical trial in asthma. In particular, the treatment-policy estimand is simulated and evaluated with and without study dropout. The type I error rate, power and bias of these methods are compared.

The 'Guideline on the clinical investigation of medicinal products for the treatment of asthma' proposes the implementation of several endpoints to picture different aspects of this multidimensional disease (Committee for Medicinal Products for Human Use (CHMP) et al., 2015, p. 12). In this thesis, the same three endpoints are chosen for each estimand: the number of asthma exacerbations, the time to the first exacerbation or the time between all suceeding exacerbations respectively and the occurrence of at least one exacerbation. These endpoints are evaluated with adequate methods. The first one representing the number of events is evaluated with a Log-linear Poisson model (GLM Poisson) and a Log-linear negative-binomial model (GLM NB). The Log rank test, Cox proportional hazards model (Cox Model), Likelihood ratio test (LR test), Gamma frailty model (GF model), and Shared Gamma frailty model (SGF model) are used for the analysis of the second endpoint considering the time to the first event or the time between all succeeding events respectively. The third endpoint which contains the occurrence of at least one exacerbation is assessed with a χ^2-test of independence. These methods are tested with the help of simulations. Hereby, a confirmatory clinical trial in asthma with a study duration of one year is created. Its most important characteristic is the generation of data for patients with either the same grade of severity of the

disease or with different frailties. In this context, it is of interest, whether those methods that are able to take into account different severities of the asthma disease show a better performance than those which can not. Furthermore, treatment discontinuation mechanism can either occur missing completely at random (MCAR) or missing not at random (MNAR). For the hypothetical estimand 48 scenarios are simulated. The treatment policy estimand is evaluated on the basis of 144 scenarios. Hereby, additionally to the treatment discontinuation, study dropout is considered in some scenarios. After the treatment discontinuation, a change of medication from either the new drug or placebo to a rescue medication takes place.

The thesis is structured the following way: first a theoretical background about treatment effects in a confirmatory clinical asthma trial is given. Second, the term 'estimand' is briefly defined, the conditions for the equivalence of null hypotheses of the hypothetical and treatment-policy estimand are explained and the consequences of the choice of an estimand for the design of a clinical trials are pictured. The third section deals with the missing data mechanisms in a confirmatory clinical trial. Fourth, the chosen endpoints and their underlying distributions are presented. Hereby, the true treatment effects for the hypothetical and treatment-policy estimand are derived. In the fifth section, the statistical methods are described. Furthermore, the null hypotheses of the parametric methods are transformed into each other to ensure the comparability of these models. Sixth, the choice of scenarios and parameters as well as the generation process of the data are shown. In the seventh, eighth and ninth section, the results are presented and discussed and a conclusion of this thesis is given.

The goal of this thesis is to compare these methods with respect to their type I error rate, power and bias to give advice concerning the analysis of such a study under different conditions in reality. Hence, the influence of the treatment discontinuation mechanism, the proportion of patients with treatment discontinuation in treatment and placebo arm and the occurrence of missing data are examined.

2 Treatment effects in a confirmatory clinical asthma trial

In the following, the concept of different treatment effects, which are called estimands, is explained. First, a brief definition of the term 'estimand' is given. In this context, the advantages and disadvantages of the hypothetical and treatment-policy estimand, which are compared in this thesis, are discussed. Second, the conditions for equivalence of the null hypotheses of the hypothetical and treatment-policy estimand are presented. Third, the consequences for the design of a clinical trial resulting from the choice of an estimand are explored.

2.1 Definition of an estimand

The term 'estimand' denotes the treatment effect that "is to be estimated" with the help of a clinical trial (Mehrotra et al., 2016, p. 457). This treatment effect is measured in summary statistics which compare the effect of the new drug in relation to the one of a placebo or alternative medication (National Research Council et al., 2010, p. 22). There are mainly five different estimands: The hypothetical estimand, the treatment-policy estimand, the composite estimand, the principle stratum estimand and the while-on-treatment estimand (Committee for Medicinal Products for Human Use (CHMP) et al., 2017, pp. 7–8). In the context of this thesis, only the hypothetical and treatment-policy estimands are relevant as they represent to opposed possibilities to handle the occurrence of intercurrent events. Therefore, only these two estimands and their way to deal with intercurrent events are explained in greater detail in this section. But in general, an estimand is characterized by three components (Akacha et al., 2017a, p. 270; Phillips et al., 2016, p. 7):

1. The population that is included into the study sample
2. The variable denoting the clinical outcome to measure the treatment effect
3. The handling of an intercurrent event

The hypothetical estimand represents the treatment effect which could be measured, if no treatment discontinuation had occurred (Permutt, 2016, p. 2867). The population which is included into the study for the estimation of the hypothetical estimand is chosen through characteristics reflecting the examined disease. Referring to a confirmatory clinical asthma trial which is the example for the simulations in this thesis, the population consists of suffering from asthma. The variable depends on the chosen endpoint, which is in this thesis either the number of events at the end of the study, the time until the first

event or the time between all succeeding events respectively or the occurrence of at least one event. The endpoints are briefly described in section 4. Treatment discontinuation, which represents an intercurrent event, is ignored, as if it had not occurred.

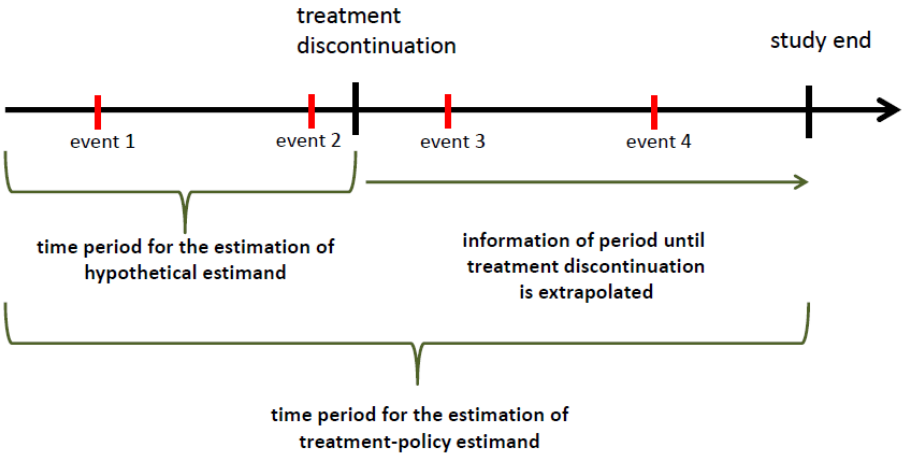

Figure 1: Comparison of hypothetical and treatment-policy estimand

Whereas the treatment-policy estimand evaluates the effect of the combination of treatments that is actually applied in practice (Akacha et al., 2017b, p. 10). The three components characterizing an estimand only differ between the hypothetical and the treatment-policy estimand in the way of handling treatment discontinuation. For the treatment-policy estimand data are analyzed regardless of the usage of rescue medication after treatment discontinuation.

Figure 1 compares the hypothetical and treatment-policy estimand. The graphic shows that for the hypothetical estimand, only the time period until treatment discontinuation is taken into account. For the treatment effect of the endpoint considering the number of events at the end of the study, the number of events until treatment discontinuation is extrapolated until the end of the study. The treatment effect of the second endpoint considering the time until the first event contains only those patients, who suffer their first event before their individual time of treatment discontinuation. All other persons are censored. If the time between all succeeding events are of interest, only the time between those events are evaluated, which occur before treatment discontinuation. The third endpoint again evaluates if during the time until treatment discontinuation, at least one event per patient had occurred. Therefore, the hypothetical estimand is appropriate for the estimation of efficacy whereby no treatment discontinuation occurs. Efficacy denotes the treatment effect under ideal conditions (Hernán et al., 2013, p. 561). The effectiveness of the combination of the new drug and an alternative medication is assessed,

if treatment-policy estimand is used. Effectiveness represents the treatment effect under a realistic scenario (Hernán et al., 2013, p. 561). The treatment-policy estimand evaluates the effect of the new treatment of the endpoint considering the number of events until the end of the study on the basis of all available data. This is also valid for the two further endpoints.

Both estimands have their specific advantages and disadvantages. With the help of the hypothetical estimand, the pharmacological effect can be assessed as it only considers the time under treatment of the new drug (Leuchs et al., 2015, p. 585). Thus, the effect of the medication can be evaluated as if the treatment was taken as originally adviced (Mallinckrodt et al., 2017, p. 30). Therefore, the goal in an early development phase of the new medication should be oriented towards the evaluation of efficiency (Leuchs et al., 2015, p. 588). But the effectiveness of the therapeutic strategy applied in clinical practice can't be evaluated. The guideline on missing data in confirmatory clinical trials requires from a confirmatory clinical trial that it "should estimate the effect of the experimental intervention in the population of patients with greatest external validity and not the effect in the unrealistic scenario where all patients receive treatment with full compliance to the treatment schedule and with a complete follow-up as per protocol" (Committee for Medicinal Products for Human Use (CHMP) et al., 2010, pp. 3–4). This requires the estimation of the treatment-policy estimand for the submission of a new drug. But for this estimand, the collection of data after treatment discontinuation must already be intended during the planning phase of the study (Phillips et al., 2016, p. 7). Because of the follow-up of patients, it can be documented if patients suffered severe side effects after their treatment discontinuation. For the submission of the new drug, it is of interest if these side effects can be related to this medication (Permutt, 2017, p. 20). Furthermore, treatment discontinuation should be noticed as it often represents a proof of low efficacy (Permutt, 2016, p. 2866).

There are scenarios in clinical trials in which the validity of the treatment-policy estimand is limited.

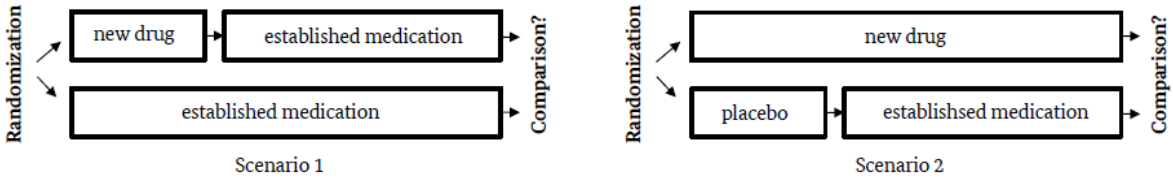

Figure 2: Scenarios showing limitations of treatment-policy estimand (Akacha et al., 2017b, p. 10)

In figure 2, two of these scenarios are shown. Both of them are cited from Akacha et al. (2017b, p. 10) and arise from a randomized controlled trial. In the first scenario, a new drug is compared to an already established medication. Patients in the treatment group receive the investigational treatment, but after the first dose, they stop the treatment and receive the established medication from now on. Patients in placebo group are only treated with the established medication. The analysis in this scenario is actually based on a comparison between the established medication and a delayed start of the established medication where patients are treated with the new drug before. In this scenario, it can't be proved that the new drug is worse than the already established one. Consequently, if the period under the new drug is very short, a comparison between both trial arms can't lead to valid conclusions. Scenario 2 shows a placebo controlled trial where patients in placebo group receive the already established treatment as rescue medication after taking one dose of placebo. From this constellation, it can not be concluded that the new drug shows no effect at all. If the period under placebo is very short, it can only be shown that the new drug is not superior to the rescue medication. These two scenarios may not seem to be very realistic, but the idea behind it shows that the period under the new drug and the placebo should be as long as possible.

2.2 Conditions for equivalence of null hypotheses of the hypothetical and treatment-policy estimand

As the hypothetical estimand $\big(\log(\Delta_{hyp})\big)$ and treatment-policy estimand $\big(\log(\Delta_{tp})\big)$ differ in the way of handling treatment discontinuation, certain conditions must exist to ensure the equivalence of their null hypotheses. In this thesis, the treatment effects are provided in logarithmized form. The null hypotheses are the same if:

$$H_0 : \log(\Delta_{hyp}) = \log(\Delta_{tp}) = 0 \tag{2.1}$$

Figure 3 shows the basic scenario of a clinical trial in asthma consisting of a treatment and placebo arm. The λ_{1_A} denotes the event rate of asthma exacerbations in the treatment arm before treatment discontinuation, whereas the $\lambda_{1_{PL}}$ represents the event rate in the placebo arm before treatment discontinuation under the investigational drug. The λ_{2_A} and the $\lambda_{2_{PL}}$ denote the event rates under rescue medication in the treatment or the placebo arm after treatment discontinuation. The λ_{TD_A} and the $\lambda_{TD_{PL}}$ are the event rates describing the time to the treatment discontinuation in treatment and placebo arm.

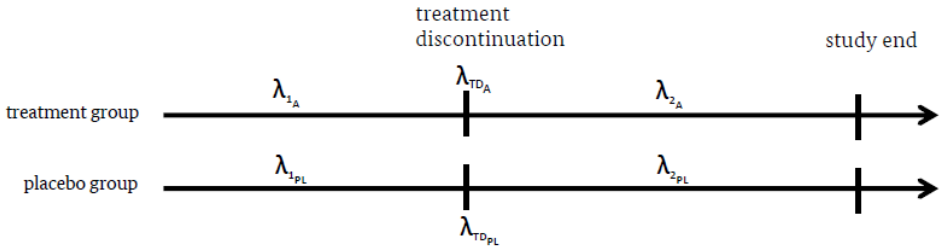

Figure 3: Conditions for equivalence of null hypotheses of the hypothetical and treatment-policy estimand

If $\lambda_{1_A} = \lambda_{1_{PL}}$, it holds:

$$H_0 : \log(\Delta_{hyp}) = 0 \tag{2.2}$$

If $\lambda_{1_A} = \lambda_{1_{PL}}$ and $\lambda_{2_A} = \lambda_{2_{PL}}$ and $\lambda_{TD_A} = \lambda_{TD_{PL}}$, then

$$H_0 : \log(\Delta_{tp}) = 0 \tag{2.3}$$

For the equivalence of both null hypotheses, $\lambda_{1_A} = \lambda_{1_{PL}}$ and $\lambda_{2_A} = \lambda_{2_{PL}}$ and $\lambda_{TD_A} = \lambda_{TD_{PL}}$ must be valid.

If $\lambda_{1_A} = \lambda_{1_{PL}} = \lambda_{2_A} = \lambda_{2_{PL}}$, which is a special case of the above stated parameter combination, $\lambda_{TD_A} = \lambda_{TD_{PL}}$ is not necessary to hold for the equivalence of both null hypotheses.

2.3 Consequence of the choice of estimand for the design of a clinical trial

As the hypothetical and the treatment-policy estimand address different treatment effects, the design of a clinical trial must be appropriate for the evaluation of the desired estimand. Figure 4 shows the relation between the aim of a clinical trial, the estimand to be chosen, and the statistical analysis methods to evaluate the treatment effect. First, the goal of the study must be determined. It depends on the different stakeholders who are involved in the study, e.g. the patient, the doctor, the pharmaceutical company, the pharmaceutical or the regulator (Akacha et al., 2017b, p. 7). In this context, it should be clarified how to deal with the occurrence of intercurrent events, e.g. treatment discontinuation before the end of the study (Mallinckrodt et al., 2017, p. 31). After all these decisions, the appropriate estimand can be chosen. The estimand affects the design of the study, as it determines whether data after treatment discontinuation should be collected in form of a follow-up of patients or not. The selection of the statistical methods

of analysis is limited to those which are able to evaluate the fixed estimand. Hence, the choice of an estimand influences the statistical evaluation of the trial. A problem will arise if the treatment-policy estimand is selected but patients are not followed up after their treatment discontinuation or if they discontinue their follow-up period before the end of the study because this leads to missing data (Little and Kang, 2015, p. 2381). The mechanisms underlying the occurrence and the consequences of missing data are briefly described in section 3.

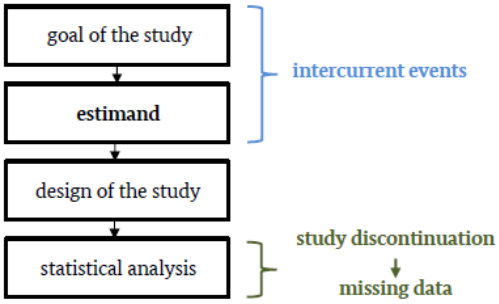

Figure 4: Consequences for the design of a clinical trial based on (Leuchs et al., 2015, p. 586)

In practice, the order of defining and analyzing a clinical trial illustrated in figure 4 is often reversed. Then, the estimand results from the chosen statistical methods of analysis (Mehrotra et al., 2016, p. 457). But this is not the correct way to define the design of the study in dependence of the methods of analysis that will be applied. The choice of the estimand determines the design and methods. Regulatory decision making must take into account the evaluated estimand in a submitted trial (National Research Council et al., 2010, p. 26).

3 Missing data mechanisms in a confirmatory clinical trial

This section deals with the missing data mechanisms in a confirmatory clinical trial. In general, there exist three different ways in which missing data are linked to the outcome variable: missing completely at random (MCAR), missing at random (MAR) and missing not at random (MNAR) (Akacha and Benda, 2010, p. 1636). These mechanisms are relevant to describe the treatment discontinuation and the study dropout in asthma trial in this thesis. Hereby the treatment discontinuation mechanisms can either occur MCAR or MNAR. For the hypothetical estimand, it leads to missing values whereas for the treatment-policy estimand it does not necessarily cause missing data because of the follow-up. Whereas the study dropout mechanism is simulated just under MCAR for those patients who have discontinued treatment. Study dropout leads to missing data. As for this thesis only MCAR and MNAR are used, only these two mechanisms are described. If data is missing completely at random, their probability of not being observed during a clinical trial is neither related to the observed values nor to the unobserved values of the dependent variable (Mallinckrodt, 2013, p. 10). Therefore, their missing can be caused by different circumstances, e.g. mistimed or too often requested clinic visits, relocation of a trial participant or any health reason if it is not linked to the response of the trial (Carpenter et al., 2002, p. 1047). Data that can be collected after a missing completely at random mechanism simply represent a random sample of all the data that could originally have been observed (Ibrahim et al., 2005, p. 333). But as Kenward (2013, p. 244) states, the missing completely at random assumption seems not to be a very realistic mechanism.

If data is missing not at random, the probability of missing values is related to the unobserved values of the dependent variable (Akacha and Benda, 2010, p. 1636). In an asthma trial, the MNAR mechanism arises if a patient's treatment discontinuation depends on the theoretical total number of events supposed the patient was observed until the end of the study. Therefore, the person would withdraw from treatment earlier if his potential number of events was higher. In this case, treatment discontinuation is linked to the therapy as the unobserved total number of events is influenced by the treatment if a treatment effect exists. In this context, the MNAR mechanism appears to describe a study dropout more realistically.

The occurrence of missing data leads to biased estimates of the treatment effect (National Research Council et al., 2010, p. 26). Thus, missing data after treatment discontinuation has a stronger impact on the estimate than treatment discontinuation itself (Leuchs et

al., 2014, p. 194). In this thesis, the bias of the treatment-policy estimand due to missing values will be explored. Furthermore it is of interest, if the mechanism of the treatment discontinuation, which does not necessarily lead to missing values has an impact on the bias, too. In this thesis, this will be explored for the hypothetical and treatment-policy estimand. For the treatment-policy estimand, biased estimates are expected even under the MCAR treatment discontinuation mechanism if study dropout occurs. These biased estimates are reasoned by the missing information about the actual treatment effect.

4 Endpoints and their true treatment effects

This chapter explains the three chosen endpoints which are of interest for this simulation study. In this context, the distribution of the underlying data of each endpoint is presented. Furthermore, the true treatment effect of each endpoint of the hypothetical estimand are shown. The true treatment effect of the first endpoint of the treatment-policy estimand is derived. The remaining treatment effects are calculated numerically. These treatment effects are needed for the bias calculations of the statistical methods.

4.1 Choice of endpoints and their underlying distributions

Recurrent events can either be recorded as the number of events or as the time between the single events (Lawless, 1987, p. 808). Therefore, data can be analyzed with respect to the counted number of events, or the time until the first event or the time between all occurring events respectively. The differing emphases of evaluation lead to several endpoints. For this study, three endpoints are chosen:

1. number of a patient's asthma exacerbations during one year
2. time to first exacerbation of each patient or time between all succeeding exacerbations
3. occurrence of at least one exacerbation per patient

Each of these endpoints show specific advantages. For the analysis of the first one, information about the whole number of exacerbations is used. Whereas it is not considered whether the time period to the first asthma exacerbation is decreased by the new drug ot not. This possible reduction of the time to first exacerbation is accounted for with the second endpoint instead. The third one represents the simplest one by just remarking the occurrence of at least one event.

4.1.1 Underlying distributions of the hypothetical estimand

The first endpoint represents a patient's number of exacerbations. To get this number, it is counted how often an event occurs. In the case of all patients suffering from the same severity of asthma, the time until the next exacerbation t_E is drawn from an Exponential distribution with probability density function $f(t_E)$ and cumulative distribution function $F(t_E)$ (Held and Sabanés Bové, 2014, p. 336):

$$t_E \sim \text{Exp}(\lambda_E) \tag{4.1}$$

$$f(t) = \lambda \cdot e^{-\lambda t} \tag{4.2}$$

$$F(t) = 1 - e^{-\lambda t} \tag{4.3}$$

If the time between succeeding events is exponentially distributed with parameter λ_E, the total number of events y during the study of one year follows a Poisson distribution, where Δt denotes the time period of interest (Johnson et al., 1993, p. 153):

$$\lambda_P = \lambda_E \cdot \Delta t \tag{4.4}$$

$$y \sim \text{Po}(\lambda_P) \tag{4.5}$$

The Poisson distribution is given as follows (Zucchini et al., 2009, p. 152):

$$f(y) = \begin{cases} \frac{\lambda_P^y}{y!} e^{-\lambda_P} & \text{for y = 0, 1, 2, ...} \\ 0 & \text{otherwise} \end{cases} \tag{4.6}$$

The number of exacerbations is carried out via a Poisson process (Lawless, 1987, p. 808). The resulting recurrent events are independent between patients (Metcalfe and Thompson, 2006, p. 167). Furthermore, as events occur with constant rate within an individual, a patient's exacerbation does not depend on any preceding event. This feature characterizes a homogeneous Poisson process (Jahn-Eimermacher, 2008, p. 4990). Its intensity function, which shows the "instantaneous probability of an event occurring at t" and depends on the process history $M(t)$, is defined by (Cook and Lawless, 2007, p. 10):

$$h\big(t|M(t)\big) = \lim_{\Delta t \to 0} \frac{\text{P}\big(\Delta Y(t) = 1|M(t)\big)}{\Delta t} \tag{4.7}$$

$Y(t)$ gives the number of exacerbations suffered during the time interval $[0, t]$ (Cook and Lawless, 2007, p. 9). To include the effect of covariates, e.g. the influence of the belonging to one trial arm, the intensity function is described as follows (Cook and Lawless, 2002, p. 144):

$$h\big(t|M(t), \mathbf{X}\big) = h_0\big(t; M(t)\big) g(\mathbf{X}) \tag{4.8}$$

with

$$g(\mathbf{X}) = e^{\beta^T \mathbf{X}} \tag{4.9}$$

$h_0\big(t; M(t)\big)$ represents the baseline hazard function.

This kind of count data has no overdispersion, as it holds (Fahrmeir et al., 2013, p. 294):

$$y \sim \text{Po}(\lambda_P), \tag{4.10}$$

$$\lambda_P = \text{E}(y) = \text{Var}(y) \tag{4.11}$$

In reality, patients taking part of a clinical trial show individual severities of asthma. Therefore, the rate λ_P may now vary between patients but is still kept constant over time within one person. In this thesis, it is assumed to follow a Gamma distribution (Held and Sabanés Bové, 2014, p. 336):

$$\lambda_P \sim \text{Ga}(\alpha, \beta); \quad \alpha, \beta > 0 \tag{4.12}$$

$$f(\lambda_P) = \frac{\beta^\alpha}{\Gamma(\alpha)} \lambda_E^{\alpha-1} e^{-\beta\lambda_E} \tag{4.13}$$

$$\text{E}(\lambda_P) = \frac{\alpha}{\beta} \Rightarrow \alpha = \text{E}(\lambda_P) \cdot \beta \tag{4.14}$$

$$\text{Var}(\lambda_P) = \frac{\alpha}{\beta^2} \tag{4.15}$$

In this case, the number of exacerbations y is no longer described by the Poisson distribution but by the Poisson-Gamma distribution (Held and Sabanés Bové, 2014, p. 335):

$$y \sim \text{PoGa}(\alpha, \beta, \vartheta) \tag{4.16}$$

$$f(y) = \left(\frac{\beta}{\vartheta + \beta}\right)^\alpha \frac{1}{\Gamma(\alpha)} \frac{\Gamma(\alpha + y)}{y!} \left(\frac{\vartheta}{\vartheta + \beta}\right)^y \tag{4.17}$$

This results in overdispersed data, as

$$\text{E}(y) = \vartheta\frac{\alpha}{\beta} = \lambda_P, \quad \alpha, \beta > 0 \tag{4.18}$$

$$\text{Var}(y) = \vartheta\frac{\alpha}{\beta}(1 + \frac{\vartheta}{\beta}) = \lambda_P(1 + \frac{\vartheta}{\beta}) \tag{4.19}$$

$$\Rightarrow \text{Var}(y) = \phi\text{E}(y) \tag{4.20}$$

$$\Rightarrow \phi = \frac{\text{Var}(y)}{\text{E(y)}} = \frac{\alpha\frac{\vartheta}{\beta}\left(1 + \frac{\vartheta}{\beta}\right)}{\vartheta\frac{\alpha}{\beta}} = 1 + \frac{\vartheta}{\beta} \tag{4.21}$$

$$\Rightarrow \beta = \frac{\vartheta}{\phi - 1} \tag{4.22}$$

where $\phi = (1 + \frac{\vartheta}{\beta})$ is called the overdisperison parameter whereby overdispersion only exists if $\phi > 1$ (Fahrmeir et al., 2013, p. 294).

The mixture of a Poisson distribution with a Gamma distribution leads to the Poisson-Gamma distribution shown above which can be transformed into the Negative-Binomial distribution. The density of a Poisson-Gamma distribution is derived as follows (Hardin and Hilbe, 2012, p. 244; Held and Sabanés Bové, 2014, p. 335):

$$f(y) = \int_0^\infty \frac{\lambda_P^y e^{-\lambda_P}}{y!} \frac{\beta^\alpha}{\Gamma(\alpha)} \lambda_P^{\alpha-1} e^{-\beta\lambda_P} d\lambda_P \tag{4.23}$$

$$= \frac{\beta^\alpha}{\Gamma(y+1)\Gamma(\alpha)} \int_0^\infty \lambda_P^y e^{-\lambda_P} \lambda_P^{\alpha-1} e^{-\beta\lambda_P} d\lambda_P \tag{4.24}$$

$$= \frac{\beta^\alpha}{\Gamma(y+1)\Gamma(\alpha)} \int_0^\infty \lambda_P^{y+\alpha-1} e^{-\lambda_P(1+\beta)} d\lambda_P \tag{4.25}$$

$$= \frac{\beta^\alpha}{\Gamma(y+1)\Gamma(\alpha)} \frac{\Gamma(y+\alpha)}{(1+\beta)^{y+\alpha}} \frac{(1+\beta)^{y+\alpha}}{\Gamma(y+\alpha)} \int_0^\infty \lambda_P^{y+\alpha-1} e^{-\lambda_P(1+\beta)} d\lambda_P \tag{4.26}$$

$$\text{Substituting } z = \lambda_P(1+\beta) \Rightarrow \lambda_P = \frac{z}{1+\beta}$$

$$dz = (1+\beta)d\lambda_P \Rightarrow d\lambda_P = \frac{dz}{1+\beta}$$

$$= \frac{\beta^\alpha}{\Gamma(y+1)\Gamma(\alpha)} \frac{\Gamma(y+\alpha)}{(1+\beta)^{y+\alpha}} \frac{(1+\beta)^{y+\alpha}}{\Gamma(y+\alpha)} \int_0^\infty \left(\frac{z}{1+\beta}\right)^{y+\alpha-1} e^{-z} \frac{dz}{1+\beta} \tag{4.27}$$

$$= \frac{\beta^\alpha}{\Gamma(y+1)\Gamma(\alpha)} \frac{\Gamma(y+\alpha)}{(1+\beta)^{y+\alpha}} \frac{(1+\beta)^{y+\alpha}}{\Gamma(y+\alpha)} \int_0^\infty \frac{z^{y+\alpha-1}}{(1+\beta)^{y+\alpha}} e^{-z} dz \tag{4.28}$$

$$= \frac{\beta^\alpha}{\Gamma(y+1)\Gamma(\alpha)} \frac{\Gamma(y+\alpha)}{(1+\beta)^{y+\alpha}} \frac{1}{\Gamma(y+\alpha)} \int_0^\infty z^{y+\alpha-1} e^{-z} dz \tag{4.29}$$

$$= \frac{\beta^\alpha}{\Gamma(y+1)\Gamma(\alpha)} \frac{\Gamma(y+\alpha)}{(1+\beta)^{y+\alpha}} \tag{4.30}$$

$$= \frac{\Gamma(y+\alpha)}{\Gamma(y+1)\Gamma(\alpha)} \frac{\beta^\alpha}{(1+\beta)^{y+\alpha}} \tag{4.31}$$

$$= \frac{\Gamma(y+\alpha)}{\Gamma(y+1)\Gamma(\alpha)} \frac{\beta^\alpha}{(1+\beta)^y(1+\beta)^\alpha} \tag{4.32}$$

$$= \frac{\Gamma(y+\alpha)}{\Gamma(y+1)\Gamma(\alpha)} \left(\frac{\beta}{1+\beta}\right)^\alpha \left(\frac{1}{1+\beta}\right)^y \tag{4.33}$$

$$= \left(\frac{\beta}{\vartheta+\beta}\right)^\alpha \frac{1}{\Gamma(\alpha)} \frac{\Gamma(\alpha+y)}{y!} \left(\frac{\vartheta}{\vartheta+\beta}\right)^y \tag{4.34}$$

In equation 4.34, it can be observed that ϑ equals 1, but this parameter will be kept in the following equations to use the same notation as Held and Sabanés Bové (2014, p. 335).

The first part of equation 4.33 equals the binomial coefficient (Johnson et al., 1993, p. 3; Hilbe, 2011, p. 189):

$$\binom{y+\alpha-1}{y} = \frac{(y+\alpha-1)!}{y!(y+\alpha-1-y)!} \tag{4.35}$$

$$= \frac{\Gamma(y+\alpha-1+1)}{\Gamma(y+1)\Gamma(y+\alpha-1-y+1)} \tag{4.36}$$

$$= \frac{\Gamma(y+\alpha)}{\Gamma(y+1)\Gamma(\alpha)} \tag{4.37}$$

Therefore, the density of the Negative-Binomial distribution can be derived, whereby $p = \left(\frac{\beta}{1+\beta}\right)$ denotes the probability of success in each trial (Hardin and Hilbe, 2012, p. 248):

$$f(y) = \binom{y+\alpha-1}{y} \left(\frac{\beta}{1+\beta}\right)^\alpha \left(\frac{1}{1+\beta}\right)^y \tag{4.38}$$

$$= \binom{y+\alpha-1}{y} p^\alpha (1-p)^y \tag{4.39}$$

Hence, the Poisson-Gamma distribution can be transformed into the Negative-Binomial distribution.

Recurrent event data are in case of individual event rates still simulated via a homogeneous Poisson process but with individual intensity function λ_{P_i} (Jahn-Eimermacher,

2008, p. 4991; Cook and Lawless, 2007, p. 76):

$$h_i\big(t|M_i(t), z_i\big) = \lim_{\Delta t \to 0} \frac{\mathrm{P}\big(\Delta Y_i(t) = 1|M_i(t), z_i\big)}{\Delta t}, \tag{4.40}$$

Hereby, z_i denotes the unobservable frailty of individual i (Lawless, 1995, p. 491). The term frailty paraphrases the patient's severity of the asthma disease, whereby a person with a higher frailty will suffer the exacerbation earlier than one with a lower frailty (Sölkner, 1996, p. 238; Therneau and Grambsch, 2000, p. 232).

The second endpoint considers the time to the first exacerbation or the time between all succeeding exacerbations t_E respectively. These periods of time are drawn from an Exponential distribution as already described above if patients suffer from the same severity of asthma. Otherwise, the parameter λ_E is drawn from a Gamma distribution which gives the following mixture distribution:

$$f(t_E) = \int_0^\infty \lambda_E e^{-\lambda t_E} \frac{\beta^\alpha}{\Gamma(\alpha)} \lambda_E^{\alpha-1} e^{-\beta\lambda_E} d\lambda_E \tag{4.41}$$

$$= \frac{\beta^\alpha}{\Gamma(\alpha)} \frac{\Gamma(\alpha+1)}{(t_E+\beta)^{\alpha+1}} \frac{(t_E+\beta)^{\alpha+1}}{\Gamma(\alpha+1)} \int_0^\infty \lambda_E^\alpha e^{-\lambda(t_E+\beta)} \tag{4.42}$$

$$\text{Substituting } z = \lambda_E(t_E+\beta) \Rightarrow \lambda_E = \frac{z}{t_E+\beta}$$

$$dz = (t_E+\beta)d\lambda_E \Rightarrow d\lambda_E = \frac{dz}{t_E+\beta}$$

$$= \frac{\beta^\alpha}{\Gamma(\alpha)} \frac{\Gamma(\alpha+1)}{(t_E+\beta)^{\alpha+1}} \frac{(t_E+\beta)^{\alpha+1}}{\Gamma(\alpha+1)} \int_0^\infty \frac{z^\alpha}{(t_E+\beta)^\alpha} e^{-z} \frac{dz}{t_E+\beta} \tag{4.43}$$

$$= \frac{\beta^\alpha}{\Gamma(\alpha)} \frac{\Gamma(\alpha+1)}{(t_E+\beta)^{\alpha+1}} \frac{(t_E+\beta)^{\alpha+1}}{\Gamma(\alpha+1)} \int_0^\infty \frac{z^\alpha}{(t_E+\beta)^{\alpha+1}} e^{-z} dz \tag{4.44}$$

$$= \frac{\beta^\alpha}{\Gamma(\alpha)} \frac{\Gamma(\alpha+1)}{(t_E+\beta)^{\alpha+1}} \frac{1}{\Gamma(\alpha+1)} \Gamma(\alpha+1) \tag{4.45}$$

$$= \frac{\beta^\alpha}{\Gamma(\alpha)} \frac{\Gamma(\alpha+1)}{(t_E+\beta)^{\alpha+1}} \tag{4.46}$$

$$= \frac{\alpha\beta^\alpha}{(t_E+\beta)^{\alpha+1}} \tag{4.47}$$

$$= \frac{\alpha\beta^\alpha}{(t_E+\beta)^{\alpha+1}} \cdot \beta \cdot \frac{1}{\beta} \tag{4.48}$$

$$= \frac{\alpha \frac{1}{\beta} \beta^{\alpha+1}}{(t_E + \beta)^{\alpha+1}} \tag{4.49}$$

$$= \frac{\alpha}{\beta} \left(\frac{\beta}{t_E + \beta} \right)^{\alpha+1} \tag{4.50}$$

$$= \frac{\alpha}{\beta} \left(1 + \frac{t_E}{\beta} \right)^{-(\alpha+1)} \tag{4.51}$$

This distribution equals the Pareto Type II distribution which is also known as the Lomax distribution (Siri et al., 2012, p. 143). It's cumulative density function can be derived by:

$$F(t_E) = \int_0^{t_E} \frac{\alpha}{\beta} \left(1 + \frac{x}{\beta} \right)^{-(\alpha+1)} dx \tag{4.52}$$

$$= -\left(1 + \frac{x}{\beta} \right)^{-\alpha} \Bigg|_0^{t_E} \tag{4.53}$$

$$= 1 - \left(1 + \frac{t_E}{\beta} \right)^{-\alpha} \tag{4.54}$$

As it can be observed, this mixture distribution no longer depends on the parameter λ_E but on the parameters of the Gamma distribution or the Gamma-Poisson distribution respectively.

The evaluation of the third endpoint is based on binary data. This binary data documents whether an event has occurred or not.

4.1.2 Underlying distributions for the treatment-policy estimand

For the definition of the underlying distribution for the treatment-policy estimand, the distribution of the time until treatment discontinuation t_{TD} must be known. In this thesis, it comes from an exponential distribution with parameter λ_{TD}. Referring to the count data models, the total number of events y under treatment-policy estimand is described by the sum of the number of events before treatment discontinuation y_1 and the number of events after treatment discontinuation y_2. In the following, the index 1 denotes the period before treatment discontinuation and the index 2 describes the period

after treatment discontinuation.

$$Y = Y_1 + Y_2 \tag{4.55}$$

This sum is described by the convolution of two discrete distributions (Johnson et al., 1993, p. 50):

$$f_Y(y) = f_{Y_1+Y_2} = P(Y = y) = \sum_{y_1=0}^{y} f_{y_1}(y) f_{y_2}(y - y_1) \tag{4.56}$$

For the Log-linear Poisson model y_1 and y_2 come from two Poisson distributions, and therefore y is calculated by:

$$f_Y(y) = \sum_{y_1=0}^{y} \frac{\lambda_{P_1 1}^{y}}{y_1!} e^{-\lambda_{P_1}} \frac{\lambda_{P_2}^{y-y_1}}{(y-y_1)!} e^{-\lambda_{P_2}} \tag{4.57}$$

$$= e^{-(\lambda_{P_1}+\lambda_{P_2})} \sum_{y_1=0}^{y} \frac{\lambda_{P_1 1}^{y}}{y_1!} \frac{\lambda_{P_2}^{y-y_1}}{(y-y_1)!} \tag{4.58}$$

With the use of the binomial theorem (Johnson et al., 1993, p. 3):

$$(a+b)^n = \sum_{j=0}^{n} \binom{n}{j} a^{n-j} b^j, \tag{4.59}$$

the final distribution results, whereby $n = y$, $a = \lambda_{P_2}$ and $b = \lambda_{P_1}$:

$$f_Y(y) = e^{-(\lambda_{P_1}+\lambda_{P_2})} \sum_{y_1=0}^{y} \frac{(\lambda_{P_1} + \lambda_{P_2})^y}{y!} \tag{4.60}$$

Therefore,

$$Y \sim \mathrm{Po}(\lambda_{P_1} + \lambda_{P_2}) \tag{4.61}$$

In analogy to, the convolution of the Negative-Binomial distribution underlying the Log-linear negative-binomial model is built. Hereby, α_1 denotes the parameter for the Negative-Binomial distribution describing the number of events y_1 before treatment discontinuation. The parameter α_2 specifies the Negative-Binomial distribution after treatment discontinuation. As β depends on the overdispersion parameter ϕ (see equation

5.60) which is constant over time, it holds that $p = p_1 = p_2$ (Furman, 2007, p. 170):

$$f_Y(y) = \sum_{y_1=0}^{y} \binom{y_1 + \alpha_1 - 1}{y_1} p^{\alpha_1}(1-p)^{y_1} \binom{y_2 + \alpha_2 - 1}{y_2} p^{\alpha_2}(1-p)^{y_2} \tag{4.62}$$

$$= \sum_{y_1=0}^{y} \binom{y_1 + \alpha_1 - 1}{y_1} \left(\frac{\beta}{1+\beta}\right)^{\alpha_1} \left(\frac{1}{1+\beta}\right)^{y_1} \cdot \binom{y_2 + \alpha_2 - 1}{y_2} \left(\frac{\beta}{1+\beta}\right)^{\alpha_2} \left(\frac{1}{1+\beta}\right)^{y_2} \tag{4.63}$$

$$= \sum_{y_1=0}^{y} \binom{y_1 + \alpha_1 - 1}{y_1} \left(\frac{\beta}{1+\beta}\right)^{\alpha_1} \left(\frac{1}{1+\beta}\right)^{y_1} \cdot \binom{y - y_1 + \alpha_2 - 1}{y - y_1} \left(\frac{\beta}{1+\beta}\right)^{\alpha_2} \left(\frac{1}{1+\beta}\right)^{y-y_1} \tag{4.64}$$

$$= \left(\frac{\beta}{1+\beta}\right)^{\alpha_1+\alpha_2} \left(\frac{1}{1+\beta}\right)^{y} \sum_{y_1=0}^{y} \binom{y_1 + \alpha_1 - 1}{y_1} \binom{y - y_1 + \alpha_2 - 1}{y - y_1} \tag{4.65}$$

With (Johnson et al., 1993, p. 3):

$$\binom{-n}{r} = (-1)^r \binom{n + r - 1}{r}, \tag{4.66}$$

the binomial coefficients can be transformed into:

$$\sum_{y_1=0}^{y} \binom{y_1 + \alpha_1 - 1}{y_1} \binom{y - y_1 + \alpha_2 - 1}{y - y_1} \tag{4.67}$$

$$= \sum_{y_1=0}^{y} (-1)^{y_1} \binom{-\alpha_1}{y_1} (-1)^{y-y_1} \binom{-\alpha_2}{y - y_1} \tag{4.68}$$

$$= (-1)^y \sum_{y_1=0}^{y} \binom{-\alpha_1}{y_1} \binom{-\alpha_2}{y - y_1} \tag{4.69}$$

With help of Vandermonde's theorem (Johnson et al., 1993, p. 3):

$$\binom{a+b}{n} = \sum_{j=0}^{n} \binom{a}{j} \binom{b}{n-j}, \tag{4.70}$$

the binomial coefficients can be merged.

$$(-1)^y \sum_{y_1=0}^{y} \binom{-\alpha_1}{y_1} \binom{-\alpha_2}{y-y_1} \tag{4.71}$$

$$=(-1)^y \binom{-\alpha_1-\alpha_2}{y} \tag{4.72}$$

$$=\binom{y+\alpha_1+\alpha_2-1}{y} \tag{4.73}$$

Finally, the convolution gives:

$$f_Y(y) = \binom{y+\alpha_1+\alpha_2-1}{y} \left(\frac{\beta}{1+\beta}\right)^{\alpha_1+\alpha_2} \left(\frac{1}{1+\beta}\right)^{y} \tag{4.74}$$

The underlying distribution of the second endpoint describing the time to the first event represents a Schuhl distribution which represents a mixture of two Exponential distributions (Johnson et al., 1994, p. 547). Because of the property of the Exponential distribution of being memoryless, $f(t_E)$ is the sum of the distribution of the time to the first event before treatment discontinuation $f_1(t_E)$ and of the distribution after treatment discontinuation $f_2(t_E)$. Both distributions are weighted by the probability of suffering the first event before or after the change of medication. In the case of scenarios without overdispersion, this distribution is given by:

$$f(t_E) = f_1(t_E)\mathrm{P}(t_{TD} > t_E) + f_2(t_E)\mathrm{P}(t_{TD} \leq t_E) \tag{4.75}$$

$$= \lambda_{E_1} e^{-\lambda_{E_1} t} e^{-\lambda_{TD} t} + \lambda_{E_2} e^{-\lambda_{E_2} t}(1 - e^{-\lambda_{TD} t}) \tag{4.76}$$

$$= \lambda_{E_1} e^{t(-\lambda_{E_1}-\lambda_{TD})} + \lambda_{E_2} e^{-\lambda_{E_2} t} - \lambda_{E_2} e^{t(-\lambda_{E_2}-\lambda_{TD})} \tag{4.77}$$

Otherwise, in scenarios reflecting the individual frailties of patients $f(t)$ represents a mixture of two Lomax distributions:

$$f(t_E) = f_1(t_E)\mathrm{P}(t_{TD} > t_E) + f_2(t_E)\mathrm{P}(t_{TD} \leq t_E) \tag{4.78}$$

$$= \frac{\alpha_1}{\beta}\left(1+\frac{t_E}{\beta}\right)^{-(\alpha_1+1)} e^{-\lambda_{TD} t} + \frac{\alpha_2}{\beta}\left(1+\frac{t_E}{\beta}\right)^{-(\alpha_2+1)} (1 - e^{-\lambda_{TD} t}) \tag{4.79}$$

The third endpoint evaluates again the occurrence of at least one event based on the distributions of endpoint 1.

These three endpoints differ with respect to the information they require for the analysis. A patient taking part of a confirmatory trial benefits of a shorter study duration. That could be achieved by focusing on the second or third endpoint. The analysis of the first endpoint takes additionally into account all further exacerbations after the first one. Consequently, methods for the analysis of the first endpoint or methods for the second endpoint considering the time between all succeeding events might estimate the treatment effect closer to the true one.

4.2 True treatment effects of the endpoints

The true treatment effect of each endpoint of the hypothetical estimand are shown. The true treatment effect of the first endpoint of the treatment-policy estimand is derived. The remaining treatment effects are calculated numerically. In general, the true treatment effect measures the true difference between both trial arms. Each difference is based on the recorded information of the chosen endpoint. All treatment effects are calculated on the log-scale to measure the treatment effect of the active group (A) in relation to the one of the placebo group (PL). Therefore, a better comparability of the treatment effects between the models is ensured.

4.2.1 True treatment effects for the hypothetical estimand

As for the hypothetical estimand, the time of treatment discontinuation is not taken into account and the mechanism of treatment discontinuation is irrelevant. First, the true logarithmized treatment effect $\log(\Delta_{true_{CD_{hyp}}})$ of the count data of the first endpoint for the hypothetical estimand is derived. This endpoint considers the total number of a patient's asthma exacerbations during one year. Therefore, the true treatment effect measures the true difference between the number of events in treatment and placebo arm based on their underlying distributions. As there is no distinction between data with and without overdispersion the Poisson distribution and the Negative-Binomial distribution holds:

$$E(y) = \lambda_{P_1} \tag{4.80}$$

Hereby, λ_{P_1} denotes the parameter of the Poisson distribution describing the number of events before treatment discontinuation. With equation 4.4 on page 13:

$$\lambda_{P_1} = \lambda_{E_1} \cdot \Delta t = \lambda_{E_1}, \tag{4.81}$$

as Δt equals one year for the hypothetical estimand. λ_{E_1} represents the parameter

of the exponential distribution describing the periods of time between the events until treatment discontinuation.

The logarithmized true treatment effect $\log(\Delta_{true_{CD_{hyp}}})$ of the count data of the first endpoint for the hypothetical estimand is calculated as follows:

$$\log(\Delta_{true_{CD_{hyp}}}) = \log\left(\frac{\mathrm{E}(y|A)}{\mathrm{E}(y|PL)}\right) \tag{4.82}$$

$$= \log\big(\mathrm{E}(y|A)\big) - \log\big(\mathrm{E}(y|PL)\big) \tag{4.83}$$

$$= \log\big(\lambda_{P_{1_A}}\big) - \log\big(\lambda_{P_{1_{PL}}}\big) \tag{4.84}$$

The true treatment effect $\log(\Delta_{true_{E_{hyp}}})$ for the second endpoint for the hypothetical estimand describes the ratio of the hazard rates of the treatment $h(t_E|A)$ and the placebo arm $h(t_E|PL)$:

$$\log(\Delta_{true_{E_{hyp}}}) = \log\left(\frac{h(t_E|A)}{h(t_E|PL)}\right) \tag{4.85}$$

$$= \log\big(h(t_E|A)\big) - \log\big(h(t_E|PL)\big) \tag{4.86}$$

For scenarios without overdispersion, $h(t_E)$ forms the hazard rate of the exponential distribution which is the fraction of the density divided by the survival function $\big(S(t_E)\big)$ (Klein and Moeschberger, 1997, p. 37):

$$h(t) = \frac{f(t_E)}{S(t_E)} = \frac{f(t_E)}{1 - F(t_E)} = \frac{\lambda_E e^{-\lambda_E t_E}}{e^{-\lambda_E t_E}} = \lambda_E \tag{4.87}$$

Consequently, $\log(\Delta_{true_{E_{hyp}}})$ represents:

$$\log(\Delta_{true_{E_{hyp}}}) = \log\big(h(t_E|A)\big) - \log\big(h(t_E|PL)\big) \tag{4.88}$$

$$= \log\big(\lambda_{E_{1_A}}\big) - \log\big(\lambda_{E_{1_{PL}}}\big) \tag{4.89}$$

For scenarios with overdispersion, the individual frailties z_i influence the hazard rate. This results in a conditional hazard rate $h_i(t_E|z_i)$ (Omori and Johnson, 1993, p. 910). The Gamma density, which is the distribution of the frailties, describes the frailties of the sample at the beginning of the trial (Wienke, 2010, p. 60). But, patients with higher frailty will experience their first event earlier. Afterwards, they are no longer taken into

account for the following course of the trial. Therefore, the distribution of the frailty term decreases with longer study duration. To derive the unconditional hazard rate $h(t_E)$, the unconditional survival function is needed and will be derived in the following. The conditional survival function $S(t_E|z_i)$ is defined as (Rodríguez, 2005, p. 1):

$$S(t_E|z) = S_0(t_E)^z, \tag{4.90}$$

with $S_0(t_E)$ denoting the baseline survival function.

$$S(t_E) = \int_0^\infty S(t_E|z) \cdot f(z)dz \tag{4.91}$$

$$= \int_0^\infty S_0(t_E)^z \cdot f(z)dz \tag{4.92}$$

$$= \int_0^\infty e^{(-\lambda_E t_E)^z} \frac{\beta^\alpha}{\Gamma(\alpha)} z^{\alpha-1} e^{-\beta z} dz \tag{4.93}$$

$$= \frac{\beta^\alpha}{\Gamma(\alpha)} \int_0^\infty z^{\alpha-1} e^{-(\beta+\lambda_E t_E)z} dz \tag{4.94}$$

$$\text{Substituting } v = (\beta + \lambda_E t_E)z \Rightarrow z = \frac{v}{\beta + \lambda_E t_E}$$

$$dv = (\beta + \lambda_E t_E)dz \Rightarrow dz = \frac{dv}{\beta + \lambda_E t_E}$$

$$= \frac{\beta^\alpha}{\Gamma(\alpha)} \int_0^\infty \left(\frac{v}{\beta + \lambda_E t_E}\right)^{\alpha-1} e^{-v} \frac{dv}{\beta + \lambda_E t_E} \tag{4.95}$$

$$= \frac{\beta^\alpha}{\Gamma(\alpha)} \int_0^\infty \frac{v^{\alpha-1}}{(\beta + \lambda_E t_E)^\alpha} e^{-v} dv \tag{4.96}$$

$$= \frac{\beta^\alpha}{\Gamma(\alpha)} \frac{1}{(\beta + \lambda_E t_E)^\alpha} \int_0^\infty v^{\alpha-1} e^{-v} dv \tag{4.97}$$

$$= \frac{\beta^\alpha}{\Gamma(\alpha)} \frac{\Gamma(\alpha)}{(\beta + \lambda_E t_E)^\alpha} \tag{4.98}$$

$$= \left(\frac{\beta}{\beta + \lambda_E t}\right)^\alpha \tag{4.99}$$

With the definition of the unconditional cumulative hazard function $H(t_E)$ (Klein and

Moeschberger, 1997, p. 27):

$$H(t_E) = -\log\big(S(t_E)\big), \tag{4.100}$$

$S(t_E)$ can be tranformed into $H(t_E)$.

$$H(t_E) = -\log\left(\left(\frac{\beta}{\beta + \lambda_E t_E}\right)^{\alpha}\right) \tag{4.101}$$

$$= -\alpha \cdot \log(\beta) + \alpha \cdot \log(\beta + \lambda_E t_E) \tag{4.102}$$

As $h(t_E)$ forms the first derivative of $H(t_E)$ with respect to t_E (Klein and Moeschberger, 1997, p. 27; Wienke, 2010, p. 74):

$$h(t_E) = \frac{\partial H(t_E)}{\partial t_E} \tag{4.103}$$

$$= \frac{\alpha\lambda_E}{\beta + \lambda_E t_E} \tag{4.104}$$

$$= \frac{\frac{\alpha}{\beta^2}\alpha\lambda_E}{\frac{\alpha}{\beta^2}\beta + \frac{\alpha}{\beta^2}\lambda_E t_E} \tag{4.105}$$

$$= \frac{\frac{\alpha^2}{\beta^2}\lambda_E}{\frac{\alpha}{\beta} + \frac{\alpha}{\beta^2}\lambda_E t_E} \tag{4.106}$$

$$= \frac{\mu^2\lambda_E}{\mu + \sigma^2\lambda_E t_E} \tag{4.107}$$

Hereby, μ represents the mean of the Gamma distribution (see equation 4.18) and σ^2 denotes its variance (see equation 4.19). For the calculation of the unconditional average hazard ratio (AHR), the unconditional hazard ratio of each trial arm is set into relation with the total hazard ratio. This procedure is published by Kalbfleisch & Prentice (1981, p. 106). Hereby, the total hazard function represents:

$$h_{\text{total}}(t_E) = h_A(t_E) + h_{PL}(t_E) \tag{4.108}$$

The AHR averages the unconditional hazard ratio over time and is therefore calculated

by (Schemper et al., 2009, p. 2475):

$$\text{AHR} = \frac{\int_0^1 \frac{h_A(t_E)}{h_{total}(t_E)} w(t_E) f_{A+PL}(t_E) dt_E}{\int_0^1 \frac{h_{PL}(t_E)}{h_{total}(t_E)} w(t_E) f_{A+PL}(t_E) dt_E} \tag{4.109}$$

Hereby, $w(t_E)$ denotes the weight function. With the choice of $w(t_E) = 1$, which is used in this thesis, the hazard ratios of each time point are included into AHR with the same weight. $f_{A+PL}(t_E)$ represents the density describing the time to the first event of both joint trial arms. This distribution is built analogously to the Schuhl distribution which represents a mixture of two exponential distributions (Johnson et al., 1994, p. 547).

The time until the first event t_E of both joint groups is described by the Lomax distribution, $f_{A+PL}(t_E)$ represents a mixture of the Lomax distributions of both trial arms:

$$f_{A+PL}(t_E) = \gamma_A \frac{\alpha_A \beta^{\alpha_A}}{(t_E + \beta)^{\alpha_A + 1}} + (1 - \gamma_A) \frac{\alpha_{PL} \beta^{\alpha_{PL}}}{(t_E + \beta)^{\alpha_{PL} + 1}} \tag{4.110}$$

Hereby, the notation of the Lomax distribution in equation 4.47 is used. γ_A denotes the proportion of patients in the treatment group and $\gamma_{PL} = 1 - \gamma_A$ represents the proportion of patients in the placebo group.

Finally, the true treatment effect on the log-scale $\log(\Delta_{true_{E_{hyp}}})$ is:

$$\log(\Delta_{true_{E_{hyp}}}) = \log(\text{AHR}) \tag{4.111}$$

For the bias calculations in this thesis, the AHR is estimated using the R-package 'AHR' (Brueckner, 2016). This package provides procedures to estimate the average hazard ratio as defined by Kalbfleisch and Prentice (1981). Hereby, the Kaplan-Meier estimator is chosen for the estimation of the survival function.

The true treatment effect $\log(\Delta_{true_{OC_{hyp}}})$ for the third endpoint which measures the occurrence of at least one event is represented by the logarithmized odds ratio (OR):

$$\log(\Delta_{true_{OR_{hyp}}}) = \log\left(\frac{\frac{\mathrm{P}_{(Y>0|A)}}{1-\mathrm{P}_{(Y>0|A)}}}{\frac{\mathrm{P}_{(Y>0|PL)}}{1-\mathrm{P}_{(Y>0|PL)}}}\right) \tag{4.112}$$

$$= \log\left(\frac{\mathrm{P}(Y>0|A)\cdot\big(1-\mathrm{P}(Y>0|PL)\big)}{\big(1-\mathrm{P}(Y>0|A)\big)\cdot\mathrm{P}(Y>0|PL)}\right) \tag{4.113}$$

$$= \log\left(\frac{\big(1-\mathrm{P}(Y=0|A)\big)\cdot\mathrm{P}(Y=0|PL)}{\mathrm{P}(Y=0|A)\cdot\big(1-\mathrm{P}(Y=0|PL)\big)}\right) \tag{4.114}$$

In the case of data with overdispersion, the calculation of the odds ratio is based on the Negative-Binomial distribution. Whereas for non-overdispersed data, the Poisson distribution is used.

4.2.2 True treatment effects for the treatment-policy estimand

4.2.2.1 True treatment effects for the treatment-policy estimand with treatment discontinuation MCAR

The true treatment effect on the log-scale $\log(\Delta_{true_{CD_{tp}}})$ for the treatment-policy estimand of the first endpoint with treatment discontinuation missing completely at random is calculated analogously to the hypothetical estimand. $\log(\Delta_{true_{CD_{tp_{MCAR}}}})$ measures the true difference between the number of events in treatment and placebo arm. Hereby, the number of events before treatment discontinuation y_1 and after treatment discontinuation y_2 must be combined. As $\mathrm{E}(Y_1)$ and $\mathrm{E}(Y_2)$ are independent (Johnson et al., 1993, p. 40):

$$\mathrm{E}(Y) = \mathrm{E}(Y_1+Y_2) = \mathrm{E}(Y_1)+\mathrm{E}(Y_2) \tag{4.115}$$

$$\mathrm{E}(Y) = \mathrm{E}(Y_1+Y_2|T_{TD}\leq 1)\mathrm{P}(T_{TD}\leq 1) + \mathrm{E}(Y_1|T_{TD}>1)\mathrm{P}(T_{TD}>1) \tag{4.116}$$

$$= \int_0^1 f(t_{TD})\cdot\big(\mathrm{E}(Y_1)+\mathrm{E}(Y_2)\big)dt_{TD} + \mathrm{E}(Y_1|T_{TD}>1)P(T_{TD}>1) \tag{4.117}$$

$$= \int_0^1 \lambda_{TD}e^{-\lambda_{TD}t_{TD}}\big(\lambda_{E_1}t_{TD}+\lambda_{E_2}(1-t_{TD})\big)dt_{TD} + \lambda_{E_1}e^{-\lambda_{TD}} \tag{4.118}$$

$$= \int_0^1 \lambda_{TD}\lambda_{E_1} t_{TD} e^{-\lambda_{TD} t_{TD}} + \lambda_{TD}\lambda_{E_2} e^{-\lambda_{TD} t_{TD}} - \lambda_{TD}\lambda_{E_2} t_{TD} e^{-\lambda_{TD} t_{TD}} dt_{TD} + \lambda_{E_1} e^{-\lambda_{TD}} \tag{4.119}$$

$$= \lambda_{E_1} e^{-\lambda_{TD}} + \Big(\frac{\lambda_{E_1}}{\lambda_{TD}} e^{-\lambda_{TD} t_{TD}}\Big)\Big|_0^1 + \Big(-\lambda_{E_2} e^{-\lambda_{TD} t_{TD}}\Big)\Big|_0^1 - \Big(-\lambda_{E_2} t_{TD} e^{-\lambda_{TD} t_{TD}}\Big)\Big|_0^1 - \Big(\frac{\lambda_{E_2}}{\lambda_{TD}} e^{-\lambda_{TD} t_{TD}}\Big)\Big|_0^1 + \lambda_{E_1} e^{-\lambda_{TD}} \tag{4.120}$$

$$= -\lambda_{E_1} e^{-\lambda_{TD}} - \frac{\lambda_{E_1}}{\lambda_{TD}} e^{-\lambda_{TD}} + \frac{\lambda_{E_1}}{\lambda_{TD}} - \lambda_{E_2} e^{-\lambda_{TD}} + \lambda_{E_2} + \lambda_{E_2} e^{-\lambda_{TD}} + \frac{\lambda_{E_2}}{\lambda_{TD}} e^{-\lambda_{TD}} - \frac{\lambda_{E_2}}{\lambda_{TD}} + \lambda_{E_1} e^{-\lambda_{TD}} \tag{4.121}$$

Consequently, the true treatment effect on the log-scale $\log(\Delta_{true_{CD_{tp_{MCAR}}}})$ is calculated by:

$$\begin{aligned}
&\log(\Delta_{true_{CD_{tp_{MCAR}}}}) = \\
&= \log\Big(-\lambda_{E_{1_A}} e^{-\lambda_{TD_A}} - \frac{\lambda_{E_{1_A}}}{\lambda_{TD_A}} e^{-\lambda_{TD_A}} + \frac{\lambda_{E_{1_A}}}{\lambda_{TD_A}} - \lambda_{E_{2_A}} e^{-\lambda_{TD_A}} + \lambda_{E_{2_A}} + \\
&\quad + \lambda_{E_{2_A}} e^{-\lambda_{TD_A}} + \frac{\lambda_{E_{2_A}}}{\lambda_{TD_A}} e^{-\lambda_{TD_A}} - \frac{\lambda_{E_{2_A}}}{\lambda_{TD_A}} + \lambda_{E_{1_A}} e^{-\lambda_{TD_A}}\Big) \\
&- \log\Big(-\lambda_{E_{1_{PL}}} e^{-\lambda_{TD_{PL}}} - \frac{\lambda_{E_{1_{PL}}}}{\lambda_{TD_{PL}}} e^{-\lambda_{TD_{PL}}} + \frac{\lambda_{E_{1_{PL}}}}{\lambda_{TD_{PL}}} - \lambda_{E_{2_{PL}}} e^{-\lambda_{TD_{PL}}} + \\
&\quad + \lambda_{E_{2_{PL}}}\lambda_{E_{2_{PL}}} e^{-\lambda_{TD_{PL}}} + \frac{\lambda_{E_{2_{PL}}}}{\lambda_{TD_{PL}}} e^{-\lambda_{TD_{PL}}} - \frac{\lambda_{E_{2_{PL}}}}{\lambda_{TD_{PL}}} + \lambda_{E_{1_{PL}}} e^{-\lambda_{TD_{PL}}}\Big)
\end{aligned} \tag{4.122}$$

Hereby, the index A symbolizes the parameters of the treatment group and PL denotes the parameters of the placebo group.

The underlying distribution of the second endpoint describing the time to the first event is the Schuhl distribution which was already used in equation 4.75. The AHR and therefore $\log(\Delta_{true_{E_{tp_{MCAR}}}})$ is again numerically calculated with the R-package 'AHR' (Brueckner, 2016). Furthermore, the logarithmized true values for the treatment-policy estimand under MCAR mechanism $\log(\Delta_{true_{OR_{tp_{MCAR}}}})$ are calculated numerically.

Finally, the hypothetical and treatment-policy estimand are compared graphically for each of the three endpoints. For this purpose, the difference between both estimands

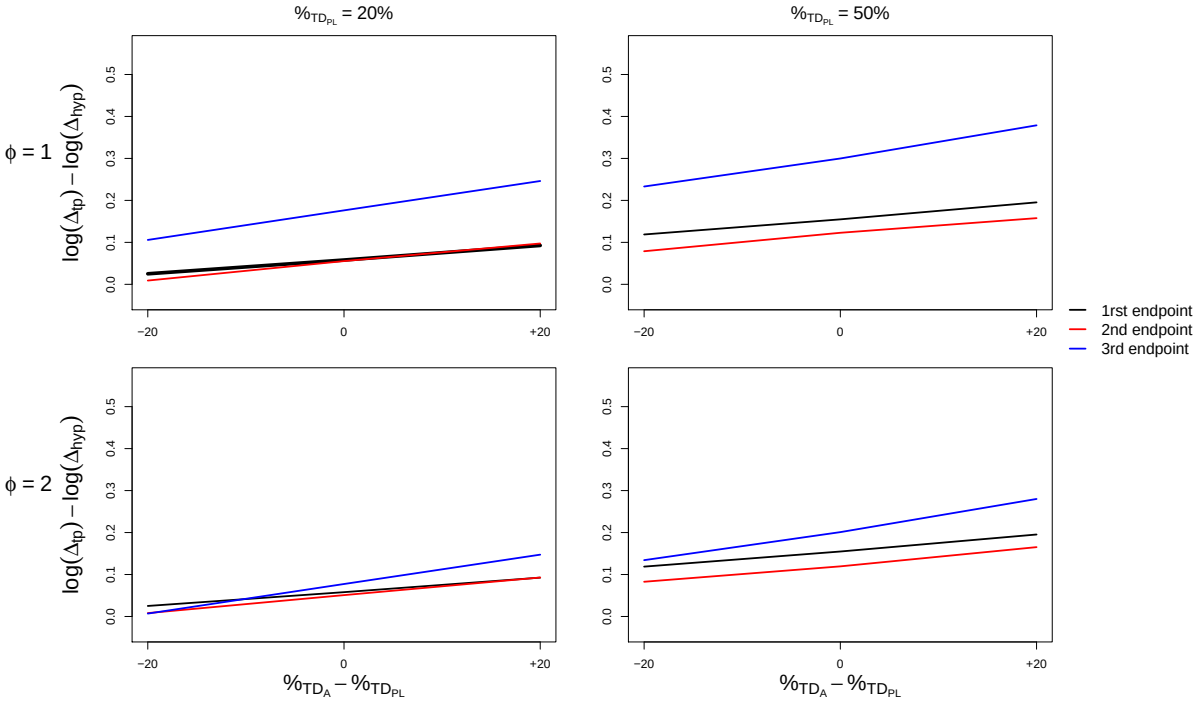

Figure 5: Difference between both logarithmized estimands $\big(\log(\Delta_{tp}) - log(\Delta_{hyp})\big)$ for each endpoint depending on the percentage of patients with treatment discontinuation in placebo group ($\%_{TD_{PL}}$) and on the difference between the percentage of patients with treatment discontinuation in both trial arms ($\%_{TD_A} - \%_{TD_{PL}}$) for missing completely at random

$(\log(\Delta_{tp}) - \log(\Delta_{hyp}))$ is calculated for the percentages that are used later in the context of the simulations. The according parameters for the simulation of these percentages are explained in section 6. In figure 5, the comparison is done under MCAR. Hereby, the difference between these estimands is calculated for a percentage of patients with treatment discontinuation in the placebo group ($\%_{TD_{PL}}$) of either 20% or 50%. The percentage of patients with treatment discontinuation in the treatment group ($\%_{TD_A}$) is determined by the discrepancy of patients with treatment discontinuation between both trial arms ($\%_{TD_A} - \%_{TD_{PL}}$), which can either be equal to -20%, 0 or +20%. In the first case, the percentage of patients with treatment discontinuation in placebo group is about 20% higher than in the treatment group, referred to the total population. In the last case, this is vice versa. With a difference of 0%, there are equal proportions of patients with treatment discontinuation in both groups. These constellations are explored either without overdispersion ($\phi = 1$) or with overdispersion ($\phi = 2$). Furthermore, if no treatment discontinuation occurs, 70% of the patients will suffer at least one asthma exacerbation in the placebo group and for 50% of the patients an event will happen in the treatment group. The effect of the rescue medication is located between the placebo and

the new drug. Hence, 60% of the patients would suffer at least one event if they were just treated with this alternative medication. The study duration is considered to be one year. In the context of such a visualization, the hypothetical estimand is held constant for each ϕ, as treatment discontinuation does not influence the treatment effect of the hypothetical estimand. But because of the changing percentages of treatment discontinuation, the treatment effect of the treatment-policy estimand varies. For each combination of ϕ and $\%_{TD_{PL}}$, the differences of all three endpoints are increasing monotonously with a larger proportion of patients with treatment discontinuation in the treatment group. To understand this phenomenon it must be reminded that the treatment effects are measured on the log-scale. Hence, these effects result in negative values. The more patients in treatment group change to rescue medication, the smaller the treatment-policy estimand becomes. This is reasoned by the treatment effect of the rescue medication which is smaller than the one of new drug. Less negative values, thus larger one will result on the log-scale. Therefore, the difference between the rising treatment-policy estimand and the hypothetical estimand increases.

4.2.2.2 True treatment effects for the treatment-policy estimand with treatment discontinuation MNAR

If treatment discontinuation is considered to appear missing not at random, patients with a higher hypothetical number of events will change their medication earlier. If the patient suffers at least one exacerbation, the λ_{TD} is multiplied with the individual hypothetical number of events, which either comes from a Poisson or from a Negative-Binomial distribution. Consequently, a 'continuous' Poisson distribution or a 'continuous' Poisson-Gamma distribution will result. If $\phi = 1$ and $y > 0$, then

$$\lambda_{TD} \cdot y \sim \text{ 'continuous' Po}(\lambda_{P_1}) \tag{4.123}$$

The distribution of t_{TD} can be expressed with the indicator function $I_B(y)$:

$$I_B(y) = \begin{cases} 1 & \text{for } y \in B \\ 0 & \text{for } y \notin B \end{cases} \tag{4.124}$$

$$f_P(t_{TD}) = \sum_{y=0}^{\infty} \frac{\lambda_{P_1}^y}{y!} e^{-\lambda_{P_1}} \Big(\lambda_{TD} e^{-\lambda_{TD} t_{TD}} I_{[0]}(y) + (\lambda_{TD} \cdot y) e^{-(\lambda_{TD} \cdot y) t_{TD}} I_{[1;\infty)}(y) \Big) \tag{4.125}$$

$$= \sum_{y=0}^{\infty} \Big(\frac{\lambda_{P_1}^y}{y!} e^{-\lambda_{P_1}} \lambda_{TD} e^{-\lambda_{TD} t_{TD}} I_{[0]}(y) + \\ + \frac{\lambda_{P_1}^y}{y!} e^{-\lambda_{P_1}} (\lambda_{TD} \cdot y) e^{-(\lambda_{TD} \cdot y) t_{TD}} I_{[1;\infty)}(y) \Big) \tag{4.126}$$

$$= \sum_{y=0}^{0} \frac{\lambda_{P_1}^y}{y!} e^{-\lambda_{P_1}} \lambda_{TD} e^{-\lambda_{TD} t_{TD}} I_{[0]}(y) + \\ + \sum_{y=1}^{\infty} \frac{\lambda_{P_1}^y}{y!} e^{-\lambda_{P_1}} (\lambda_{TD} \cdot y) e^{-(\lambda_{TD} \cdot y) t_{TD}} I_{[1;\infty)}(y) \tag{4.127}$$

$$= \frac{\lambda_{P_1}^y}{y!} e^{-\lambda_{P_1}} \lambda_{TD} e^{-\lambda_{TD} t_{TD}} I_{[0]}(y) + \\ + \sum_{y=1}^{\infty} \frac{\lambda_{P_1}^y}{y!} e^{-\lambda_{P_1}} (\lambda_{TD} \cdot y) e^{-(\lambda_{TD} \cdot y) t_{TD}} I_{[1;\infty)}(y) \tag{4.128}$$

Otherwise, if $\phi = 2$ and $y > 0$:

$$\lambda_{TD} \cdot y \sim \text{ 'continuous' PoGa}(\alpha_1, \beta, \vartheta) \tag{4.129}$$

Consequently, $f(t_{TD})$ is defined by:

$$f_{NB}(t_{TD}) = \sum_{y=0}^{\infty} \left(\frac{\beta}{\vartheta + \beta} \right)^{\alpha_1} \frac{1}{\Gamma(\alpha_1)} \frac{\Gamma(\alpha_1 + y)}{y!} \left(\frac{\vartheta}{\vartheta + \beta} \right)^{y} \cdot \\ \cdot \Big(\lambda_{TD} e^{-\lambda_{TD} t_{TD}} I_{[0]}(y) + (\lambda_{TD} \cdot y) e^{-(\lambda_{TD} \cdot y) t_{TD}} I_{[1;\infty)}(y) \Big) \tag{4.130}$$

$$= \sum_{y=0}^{0} \left(\frac{\beta}{\vartheta + \beta} \right)^{\alpha_1} \frac{1}{\Gamma(\alpha_1)} \frac{\Gamma(\alpha_1 + y)}{y!} \left(\frac{\vartheta}{\vartheta + \beta} \right)^{y} \\ \cdot \lambda_{TD} e^{-\lambda_{TD} t_{TD}} I_{[0]}(y) + \\ + \sum_{y=1}^{\infty} \left(\frac{\beta}{\vartheta + \beta} \right)^{\alpha_1} \frac{1}{\Gamma(\alpha_1)} \frac{\Gamma(\alpha_1 + y)}{y!} \left(\frac{\vartheta}{\vartheta + \beta} \right)^{y} \\ \cdot (\lambda_{TD} \cdot y) e^{-(\lambda_{TD} \cdot y) t_{TD}} I_{[1;\infty)}(y) \tag{4.131}$$

$$
\begin{aligned}
&= \left(\frac{\beta}{\vartheta+\beta}\right)^{\alpha_1} \frac{1}{\Gamma(\alpha_1)} \frac{\Gamma(\alpha_1+y)}{y!} \left(\frac{\vartheta}{\vartheta+\beta}\right)^{y} \\
&\quad \cdot \lambda_{TD} e^{-\lambda_{TD} t_{TD}} I_{[0]}(y) + \\
&\quad + \sum_{y=1}^{\infty} \left(\frac{\beta}{\vartheta+\beta}\right)^{\alpha_1} \frac{1}{\Gamma(\alpha_1)} \frac{\Gamma(\alpha_1+y)}{y!} \left(\frac{\vartheta}{\vartheta+\beta}\right)^{y} \\
&\quad \cdot (\lambda_{TD} \cdot y) e^{-(\lambda_{TD} \cdot y) t_{TD}} I_{[1;\infty)}(y)
\end{aligned} \tag{4.132}
$$

For the first endpoint, the treatment-policy estimand under MNAR $\log(\Delta_{true_{CD_{tp_{MNAR}}}})$ for scenarios without overdispersion is based on:

$$
\begin{aligned}
\mathrm{E}(Y) &= \mathrm{E}(Y_1 + Y_2 | T_{TD} \leq 1)\mathrm{P}(T_{TD} \leq 1) + \mathrm{E}(Y_1 | T_{TD} > 1)\mathrm{P}(T_{TD} > 1) && (4.133) \\
&= \int_0^1 f(t_{TD}) \cdot \big(\mathrm{E}(Y_1) + \mathrm{E}(Y_2)\big) dt_{TD} + \mathrm{E}(Y_1 | T_{TD} > 1)\mathrm{P}(T_{TD} > 1) && (4.134) \\
&= \int_0^1 \left(\frac{\lambda_{P_1}^{y}}{y!} e^{-\lambda_{P_1}} \lambda_{TD} e^{-\lambda_{TD} t_{TD}} I_{[0]}(y) + \right. \\
&\quad \left. + \sum_{y=1}^{\infty} \frac{\lambda_{P_1}^{y}}{y!} e^{-\lambda_{P_1}} (\lambda_{TD} \cdot y) e^{-(\lambda_{TD} \cdot y) t_{TD}} I_{[1;\infty)}(y) \right) \cdot && (4.135) \\
&\quad \cdot \big(\lambda_{E_1} t_{TD} + \lambda_{E_2}(1 - t_{TD})\big) dt_{TD} + \lambda_{E_1} e^{-\lambda_{TD}}
\end{aligned}
$$

Whereas for scenarios with overdispersion $\log(\Delta_{true_{CD_{tp_{MNAR}}}})$ is derived by:

$$
\begin{aligned}
\mathrm{E}(Y) &= \mathrm{E}(Y_1 + Y_2 | T_{TD} \leq 1)\mathrm{P}(T_{TD} \leq 1) + \mathrm{E}(Y_1 | T_{TD} > 1)\mathrm{P}(T_{TD} > 1) && (4.136) \\
&= \int_0^1 f(t_{TD}) \cdot \big(\mathrm{E}(Y_1) + E(Y_2)\big) dt_{TD} + \mathrm{E}(Y_1 | T_{TD} > 1)\mathrm{P}(T_{TD} > 1) && (4.137) \\
&= \int_0^1 \left(\left(\frac{\beta}{\vartheta+\beta}\right)^{\alpha_1} \frac{1}{\Gamma(\alpha_1)} \frac{\Gamma(\alpha_1+y)}{y!} \left(\frac{\vartheta}{\vartheta+\beta}\right)^{y} \cdot \lambda_{TD} e^{-\lambda_{TD} t_{TD}} I_{[0]}(y) + \right. \\
&\quad + \sum_{y=1}^{\infty} \left(\frac{\beta}{\vartheta+\beta}\right)^{\alpha_1} \frac{1}{\Gamma(\alpha_1)} \frac{\Gamma(\alpha_1+y)}{y!} \left(\frac{\vartheta}{\vartheta+\beta}\right)^{y} \cdot && (4.138) \\
&\quad \left. \cdot (\lambda_{TD} \cdot y) e^{-(\lambda_{TD} \cdot y) t_{TD}} I_{[1;\infty)}(y) \right) \cdot \\
&\quad \cdot \big(\lambda_{E_1} t_{TD} + \lambda_{E_2}(1 - t_{TD})\big) dt_{TD} + \lambda_{E_1} e^{-\lambda_{TD}}
\end{aligned}
$$

The logarithmized treatment-policy estimand under MNAR $\log(\Delta_{true_{CD_{tp_{MNAR}}}})$ is calcu-

lated numerically. Furthermore, the AHR for the second endpoint under MNAR mechanism and therefore $\log(\Delta_{true_{E_{tp_{MNAR}}}})$ is again calculated numerically with the R-package 'AHR' (Brueckner, 2016). The true values for the treatment-policy estimand under MNAR mechanism $\log(\Delta_{true_{OR_{tp_{MNAR}}}})$ are also calculated numerically.

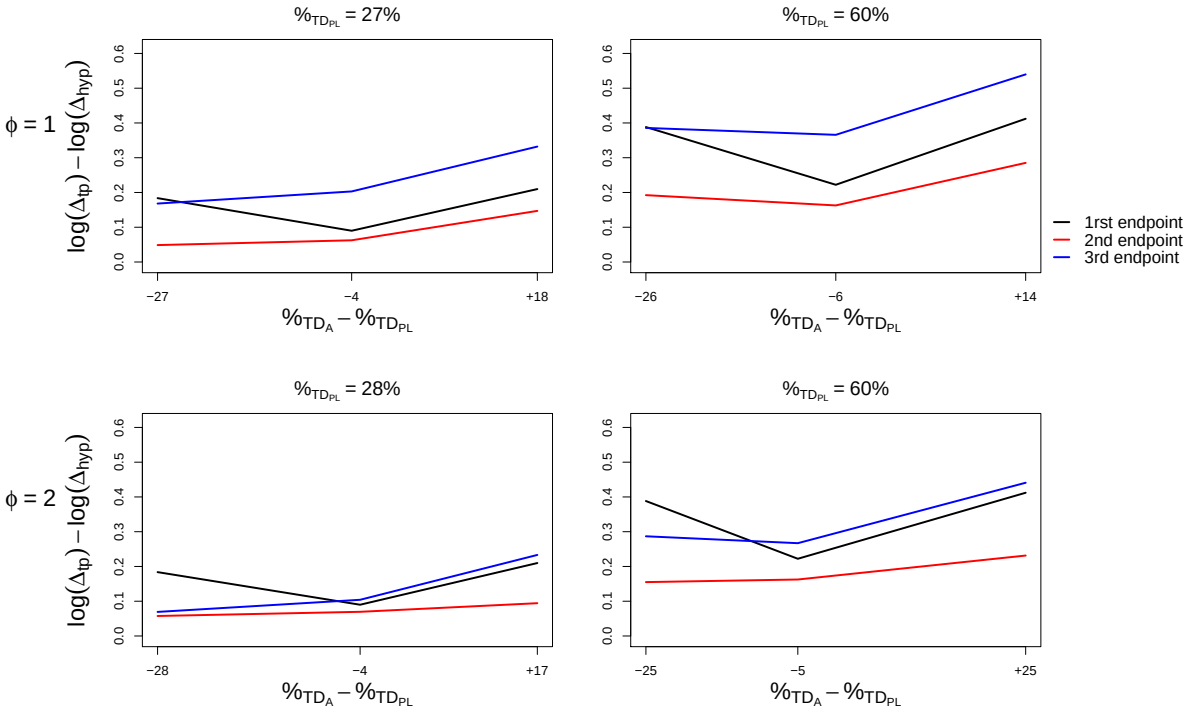

Figure 6: Difference between both logarithmized estimands $\big(\log(\Delta_{tp}) - \log(\Delta_{hyp})\big)$ for each endpoint depending on the percentage of patients with treatment discontinuation in placebo group ($\%_{TD_{PL}}$) and the difference between the percentage of patients with treatment discontinuation in both trial arms ($\%_{TD_A} - \%_{TD_{PL}}$) for missing not at random

In analogy to the MCAR mechanism, the differences between the hypothetical and treatment-policy estimand under MNAR are visualized in figure 6. For the simulations of these treatment effects, the same percentages, which are already explained in the context of figure 5, are used. Because of the MNAR treatment discontinuation mechanism, the resulting percentages are different from those under MCAR. If $\phi = 1$, the $\%_{TD_{PL}}$ denotes either 27% or 60%, and if $\phi = 2$, the $\%_{TD_{PL}}$ represents either 28% or 60%. Section 7.1 deals with the explanation of the resulting percentages under MCAR and MNAR. In figure 6, it becomes obvious that the difference $\log(\Delta_{tp}) - \log(\Delta_{hyp})$ is not increasing monotonously under the MNAR mechanism. The differences between both estimands in those scenarios with minimal divergence of treatment discontinuation are

lower than in scenarios with a maximal negative divergence, although the proportion of patients with treatment discontinuation in the treatment group increases. In scenarios without overdispersion, this phenomenon is more clearly developed than in scenarios with overdispersion.

5 Statistical methods

5.1 Count data models

For the evaluation of the three chosen endpoints adequate methods are needed. Two count data models are used in this thesis to analyse the first endpoint. In this context the number of asthma exacerbations y is considered as count variable.

In the case of no overdispersed data, all patients are believed to suffer from the same severity of asthma. Thus, exacerbations occur with the same rate λ_P, which is constant over the whole time period. A Log-linear Poisson model is appropriate, if the number of exacerbations of the individual patients is independent of each other (Fahrmeir et al., 2013, p. 293).

To predict the number of exacerbations of each patient i for the hypothetical estimand, the linear predictor

$$\eta_i = \beta_0 + \beta_1 x_i + \log(t_{TD_i}) \tag{5.1}$$

is linked to the rate λ_{P_i} in log-linear form (Fahrmeir et al., 2013, p. 293).

The following equation results (McCullagh and Nelder, 1987, p. 138):

$$\log(\lambda_{P_i}) = \eta_i = \beta_0 + \beta_1 x_i + \log(t_{TD_i}) \tag{5.2}$$

The binary independent variable x_i denotes a patient's belonging to one of both trial arms, in which $x_i = 0$ stands for the placebo group and $x_i = 1$ codes for treatment group. There aren't used any further predictor variables in the models of this thesis. The β represents the coefficient of the predictor variable. As for the hypothetical estimand, only the number of exacerbations until treatment discontinuation is used, there exists the need of an offset parameter. It is represented by the last term of this model. By involving a patient's individual time on treatment t_i into the model, it is taken into account that people with a longer time on treatment may have more events than those with a shorter one. The offset parameter enters the model as the natural logarithm of t_{TD_i} to be adequate to the linear predictor η_i (Hardin and Hilbe, 2012, p. 226).

For the treatment-policy estimand, the linear predictor for scenarios without study-dropout

$$\eta_i = \beta_0 + \beta_1 x_i \tag{5.3}$$

does not contain an offset parameter. As all patients are followed up after their treatment discontinuation until the end of the study, the number of events of all patients is recorded

for the same period of time of one year. Therefore, the model is built in the following way:

$$\log(\lambda_{P_i}) = \eta_i = \beta_0 + \beta_1 x_i \tag{5.4}$$

But for scenarios with study dropout, the linear predictor containing again an offset parameter is defined similarly to the hypothetical estimand

$$\eta_i = \beta_0 + \beta_1 x_i + \log(t_{SD_i}) \tag{5.5}$$

Here, t_{SD_i} denotes the individual time until study dropout of patient i. The model is described:

$$\log(\lambda_{P_i}) = \eta_i = \beta_0 + \beta_1 x_i + \log(t_{SD_i}) \tag{5.6}$$

If individual severities of the asthma disease of the patients are taken into account, the number of exacerbations y is no longer described by the Poisson distribution but by the Poisson-Gamma distribution, as already described in section 4.1. To master the heterogeneity between patients in the analysis, a Log-linear negative-binomial model is used. For the hypothetical estimand, the model contains an offset parameter analogous to the Log-linear Poisson model, to take into account the individual time on treatment t_{TD_i} (Nicholas et al., 2011, p. 1212):

$$\log(\lambda_{P_i}) = \eta_i = \beta_0 + \beta_1 x_i + \log(t_{TD_i}) \tag{5.7}$$

For the treatment-policy estimand, the model is built again similar to the Log-linear Poisson model. In the case of scenarios without study-dropout, the model does not contain an offset parameter:

$$\log(\lambda_{P_i}) = \eta_i = \beta_0 + \beta_1 x_i \tag{5.8}$$

Whereas in the case of scenarios with study-dropout, an offset parameter with the time until study-dropout t_{SD_i} joins the model:

$$\log(\lambda_{P_i}) = \eta_i = \beta_0 + \beta_1 x_i + \log(t_{SD_i}) \tag{5.9}$$

For all models explained above, the estimated treatment effect is representend by the parameter β_1 (Schneider et al., 2013, p. 5450).

The Negative-Binomial distribution is a common choice to describe between-patient het-

erogeneity (Friede, Schmidli, et al., 2010, p. 619). The resulting Log-linear negative-binomial models give an easily interpretable estimate of the treatment effect (Cook et al., 2009, p. 2618). Siri et al. (2012, p. 140) state a superiority of the Log-linear neagtive-binomial model to the Log-linear Poisson model in that way as it is possible to estimate the expected number of events and the overdispersion simultaneously. This thesis explores whether this advantage of the Log-linear negative-binomial model in considering heterogeneity between patients is reflected in a decreased type I error rate and bias or an increased power compared to the Log-linear Poisson model. In the context of missing data, Keene et al. (2007, p. 91) expect a lower bias of the estimated treatment effect under the MNAR mechanism with the Log-linear neagtive-binomial model. They see the advantage of this method compared to the Log-linear Poisson model again in the consideration of the individual frailties.

5.2 Time-To-First-Event Methods and extension

For the evaluation of the second endpoint five methods are compared. Four of them are used to analyse whether the time to the first asthma event t_E differs between both trials arms. Whereas the fifth method, a Shared gamma frailty model, takes into account the time between all succeeding exacerbations as an extension to the time-to-first-event methods. These methods do not vary between both estimands with respect to their model equations, but they just need different censoring variables. On the one hand, those patients are censored who do not suffer an exacerbation until their individual study dropout. Therefore, the censoring mechanism represents a common tool to deal with missing data (Keene et al., 2007, p. 96). The presented time-to-event methods are based on right censoring, as each observation after the end of the study or the study dropout is cut (Klein and Moeschberger, 1997, p. 56).

To begin with, it will be explored with the nonparametric Log rank test whether the new asthma drug is able to significantly extend the time to the first asthma exacerbation. Two assumptions must be satisfied. Firstly, proportional hazards have to exist, which means that the ratio of the risks to experience an event of both trial arms must be independent of time t (Ziegler et al., 2007, e41). Secondly, the time of censoring has to be random so that it is not influenced by the patient's time of event (Klein and Moeschberger, 1997, p. 61). These assumptions are valid for each of the following methods of the second endpoint.

The test statistics of the Log rank test $\chi^2_{logrank}$ is based on the squared difference between the number of observed O_j and expected E_j asthma exacerbations (Machin et al., 2006, p. 58). For each trial arm j, this difference is calculated and each of them is divided by

the expected number of events E_j. The test statistic is now calculated by the summation of these quotients which follows a χ^2 distribution with m-1 degrees of freedom (df), where m is the number of trial arms:

$$\chi^2_{Logrank} = \sum_{j=1}^{m} \frac{(O_j - E_j)^2}{E_j} \sim \chi^2_1 \tag{5.10}$$

In the case of one treatment arm and one placebo arm, then $df = 2 - 1 = 1$. Under the validity of the null hypothesis, the order of events is selected randomly. Hence, the time to the first exacerbation is independent of the trial arm (Ziegler et al., 2007, e39).

Next, the semiparametric Cox proportional hazards model is used to analyse the second endpoint. It estimates the influence of the trial arm under consideration of experiencing a patient's individual hazard $h(t|\mathbf{X})$ of an asthma exacerbation (Klein and Moeschberger, 1997, p. 229). The model is built as follows (Lee and Wang, 2013, p. 283):

$$h(t|\mathbf{X}) = h_0(t) \cdot g(\mathbf{X}), \tag{5.11}$$

with

$$g(\mathbf{X}) = e^{\beta^T \mathbf{X}} \tag{5.12}$$

The nonparametric part of the model in 5.11 contains the baseline hazard function $h_0(t)$, that describes the hazard of an event in the absence of any influencing variable. As it represents the nonparametric part, no assumption is made about its distribution. A person's belonging to one of both trial arms describes the parametric part $g(\mathbf{X})$, in which $x = 0$ codes for the placebo group and $x = 1$ represents the treatment arm. β denotes the vector with coefficients of the predictor variable. This notation is also valid for the Gamma frailty model and the Shared gamma frailty model. The hazard ratio (HR) is calculated as the ratio of the hazards of both trial arms (Klein and Moeschberger, 1997, p. 231):

$$\text{HR} = \frac{h_0(t) \cdot e^{1 \cdot \beta_1}}{h_0(t) \cdot e^{0 \cdot \beta_1}} = e^{\beta_1} \tag{5.13}$$

As the Cox proportional hazards model in this thesis only includes one explanatory variable, it does not differ from the Log rank test (Klein and Moeschberger, 1997, p. 229).

The third model for the evaluation of the second endpoint is the parametric Likelihood

ratio test. By comparing the distributions describing the time to the first event in both trial arms, it tests if they come from one common distribution or not (Lee and Wang, 2013, p. 229). In this thesis, the time to first event t_E follows an Exponential distribution with parameter λ_E (Held and Sabanés Bové, 2014, p. 336). For simplicity, the index E will be left for the description of the Likelihood ratio test:

$$t \sim \text{Exp}(\lambda) \tag{5.14}$$

$$f(t) = \lambda \cdot e^{-\lambda t} \tag{5.15}$$

$$F(t) = 1 - e^{-\lambda t} \tag{5.16}$$

The Likelihood ratio test works with the ratio of the likelihood of the one resulting sample after combining the trial arms $L(\widehat{\lambda_C})$ divided by the joint likelihood for the two groups $L(\widehat{\lambda_A}, \widehat{\lambda_{PL}})$ (Lee and Wang, 2013, p. 230). $\widehat{\lambda_C}$ denotes the maximum likelihood estimator (MLE) of the likelihood function under the assumption of one common Exponential distribution for the time to the first event for all patients. Whereas $\widehat{\lambda_A}$ and $\widehat{\lambda_{PL}}$ represent the MLEs for the likelihood function for the active ($\widehat{\lambda_A}$) and placebo ($\widehat{\lambda_{PL}}$) group based on different exponential distributions.

The likelihood function $L(\lambda)$ for right censored data in general is given by (Klein and Moeschberger, 1997, p. 69):

$$L(\lambda) \propto \prod_{i=1}^{N} f(t_i)^{\delta_i} S(t_i)^{1-\delta_i}, \tag{5.17}$$

where N denotes the total number of patients taking part in the clinical trial. δ_i represents the censoring status of individual i with a coding $\delta_i = 0$ for censored and $\delta_i = 1$ uncensored. $S(t_i)$ describes the survival function which looks in the case of an exponentially distributed time to first event random variable (Klein and Moeschberger, 1997, p. 22):

$$S(t) = 1 - F(t) \tag{5.18}$$

$$= 1 - (1 - e^{-\lambda t}) \tag{5.19}$$

$$= e^{-\lambda t} \tag{5.20}$$

Therefore, the likelihood function $L(\lambda)$ can be calculated, where r gives the number of

uncensored patients in the trial:

$$L(\lambda) \propto \prod_{i=1}^{N} f(t_i)^{\delta_i} S(t_i)^{1-\delta_i} \tag{5.21}$$

$$= \prod_{i=1}^{N} \left(\lambda \cdot e^{-\lambda t_i}\right)^{\delta_i} \left(e^{-\lambda t_i}\right)^{1-\delta_i} \tag{5.22}$$

$$= \prod_{i=1}^{N} \lambda^{\delta_i} \cdot e^{-\lambda t_i \delta_i} \cdot e^{-\lambda t_i (1-\delta_i)} \tag{5.23}$$

$$= \lambda^{\sum_{i=1}^{N} \delta_i} \cdot e^{\sum_{i=1}^{N} -\lambda t_i \delta_i} \cdot e^{\sum_{i=1}^{N} -\lambda t_i (1-\delta_i)} \tag{5.24}$$

$$= \lambda^{r} \cdot e^{-\lambda \sum_{i=1}^{N} t_i \delta_i - \lambda \sum_{i=1}^{N} t_i (1-\delta_i)} \tag{5.25}$$

$$= \lambda^{r} \cdot e^{-\lambda \sum_{i=1}^{N} t_i} \tag{5.26}$$

The log-likelihood function $l(\lambda)$ is given by:

$$l(\lambda) = \log\left(L(\lambda)\right) \tag{5.27}$$

$$= r \cdot \log(\lambda) - \lambda \sum_{i=1}^{N} t_i \tag{5.28}$$

The score function $s(\lambda)$ is derived:

$$s(\lambda) = \frac{\partial l(\lambda)}{\partial \lambda} \tag{5.29}$$

$$= \frac{r}{\lambda} - \sum_{i=1}^{N} t_i \tag{5.30}$$

Finally, the general MLE $(\widehat{\lambda})$ can be calculated:

$$s(\lambda) \stackrel{!}{=} 0 \tag{5.31}$$

$$\frac{r}{\lambda} - \sum_{i=1}^{N} t_i \stackrel{!}{=} 0 \tag{5.32}$$

$$\frac{r}{\lambda} = \sum_{i=1}^{N} t_i \tag{5.33}$$

$$\widehat{\lambda} = \frac{r}{\sum_{i=1}^{N} t_i} \tag{5.34}$$

The joint likelihood function for the two trial arms $L(\lambda_A, \lambda_P)$ is derived in analogy to $L(\lambda)$. Taking into account that the joint likelihood function is the product of the likelihood function of the treatment and placebo arm, where n_A gives the number of patients in the treatment group and n_{PL} contains the number of patients in the placebo group. r_A is the number of uncensored patients in the treatment arm, whereas r_{PL} stands for the number of uncensored patients in the placebo arm.

$$L\big(\lambda_A, \lambda_P\big) \propto \prod_{i=1}^{n_A} f(t_{A_i})^{\delta_{A_i}} S(t_{A_i})^{1-\delta_{A_i}} \cdot \prod_{i=1}^{n_{PL}} f(t_{PL_i})^{\delta_{PL_i}} S(t_{PL_i})^{1-\delta_{PL_i}} \tag{5.35}$$

$$\begin{aligned} &= \prod_{i=1}^{n_A} \big(\lambda_A \cdot e^{-\lambda_A t_{A_i}}\big)^{\delta_{A_i}} \big(- e^{-\lambda_A t_{A_i}}\big)^{1-\delta_{A_i}} \cdot \\ &\quad \prod_{i=1}^{n_{PL}} \big(\lambda_{PL} \cdot e^{-\lambda_{PL} t_{PL_i}}\big)^{\delta_{PL_i}} \big(- e^{-\lambda_{PL} t_{PL_i}}\big)^{1-\delta_{PL_i}} \end{aligned} \tag{5.36}$$

$$\begin{aligned} &= \lambda_A^{\sum_{i=1}^{n_A} \delta_{A_i}} \cdot e^{\sum_{i=1}^{n_A} -\lambda_A t_{A_i} \delta_{A_i}} \cdot e^{\sum_{i=1}^{n_A} -\lambda_A t_{A_i}(1-\delta_i)} \cdot \\ &\quad \lambda_{PL}^{\sum_{i=1}^{n_{PL}} \delta_i} \cdot e^{\sum_{i=1}^{n_{PL}} -\lambda_{PL} t_{PL_i} \delta_{PL_i}} \cdot e^{\sum_{i=1}^{n_{PL}} -\lambda_{PL} t_{PL_i}(1-\delta_{PL_i})} \end{aligned} \tag{5.37}$$

$$= \lambda_A^{r_A} \cdot e^{-\lambda_A \sum_{i=1}^{n_A} t_{A_i}} \cdot \lambda_{PL}^{r_{PL}} \cdot e^{-\lambda_{PL} \sum_{i=1}^{n_{PL}} t_{PL_i}} \tag{5.38}$$

$$= \lambda_A^{r_A} \cdot \lambda_{PL}^{r_{PL}} \cdot e^{-\lambda_A \sum_{i=1}^{n_A} t_{A_i} - \lambda_{PL} \sum_{i=1}^{n_{PL}} t_{PL_i}} \tag{5.39}$$

The log-likelihood function $l(\lambda_A, \lambda_{PL})$ can again be calculated by:

$$l(\lambda_A, \lambda_{PL}) = \log\big(L(\lambda_A, \lambda_{PL})\big) \tag{5.40}$$

$$= r_A \cdot \log(\lambda_A) + r_{PL} \cdot \log(\lambda_{PL}) - \lambda_A \sum_{i=1}^{n_A} t_{A_i} - \lambda_{PL} \sum_{i=1}^{n_{PL}} t_{PL_i} \tag{5.41}$$

The test statistic for the Likelihood ratio test is defined as (Lee and Wang, 2013, p. 230):

$$\chi^2_{LR} = -2 \cdot \log\left(\frac{L(\widehat{\lambda_C})}{L(\widehat{\lambda_A}, \widehat{\lambda_{PL}})}\right) \tag{5.42}$$

$$= -2 \cdot \left(\log\big(L(\widehat{\lambda_C})\big) - \log\big(L(\widehat{\lambda_A}, \widehat{\lambda_{PL}})\big)\right) \tag{5.43}$$

$$= -2 \cdot \left(l(\widehat{\lambda_C}) - l(\widehat{\lambda_A}, \widehat{\lambda_{PL}})\right) \tag{5.44}$$

$$
\begin{aligned}
&= -2 \cdot \left(\left(r \cdot \log \left(\frac{r}{\sum_{i=1}^{N} t_i} \right) - r \right) \right. \\
&\quad \left. - \left(r_A \cdot \log \left(\frac{r_A}{\sum_{i=1}^{n_A} t_{A_i}} \right) + r_{PL} \cdot \log \left(\frac{r_{PL}}{\sum_{i=1}^{n_{PL}} t_{PL_i}} \right) - r_A - r_{PL} \right) \right)
\end{aligned} \tag{5.45}
$$

$$
\begin{aligned}
&= -2 \cdot \left(r \cdot \log \left(\frac{r}{\sum_{i=1}^{N} t_i} \right) - r \right. \\
&\quad \left. - r_A \cdot \log \left(\frac{r_A}{\sum_{i=1}^{n_A} t_{A_i}} \right) - r_{PL} \cdot \log \left(\frac{r_{PL}}{\sum_{i=1}^{n_{PL}} t_{PL_i}} \right) + r_A + r_{PL} \right)
\end{aligned} \tag{5.46}
$$

Under the validity of the null hypothesis, the test statistic χ^2_{LR} follows a χ^2 distribution with $m - 1$ degrees of freedom:

$$\chi^2_{LR} \sim \chi^2_1 \tag{5.47}$$

The fourth method for analyzing the second endpoint is a parametric Gamma frailty model. This model compares, similar to the semiparametric Cox proportional hazards model, the influence of the two trials arms on the hazard $h(t|\mathbf{X}, \mathbf{Z})$ of the occurrence of the first event (Therneau and Grambsch, 2000, p. 231). But furthermore, it is able to take into account the impact of the frailty $\mathbf{Z}$ on the time to the first exacerbation. Therefore, the Gamma frailty model considers the heterogeneity between patients in a clinical trial which is constant within one individual. As explained in section 5.1, the frailty is drawn from a Gamma distribution in this thesis. For this reason, the Gamma frailty model is used. Belonging to the parametric methods, a distribution for the baseline hazard $h_0(t)$ is assumed. In this thesis the Exponential distribution is chosen, as for the Likelihood ratio test on page 39 already described.

The model is built as follows (Wienke, 2010, p. 58):

$$h(t|\mathbf{X}, \mathbf{Z}) = \mathbf{Z} \cdot h_0(t) \cdot g(\mathbf{X}) \tag{5.48}$$

Similar to equation 5.12, $g(\mathbf{X})$ is defined as:

$$g(\mathbf{X}) = e^{\beta^T \mathbf{X}} \tag{5.49}$$

It represents again the trial arms, whereby $x = 0$ defines the belonging to the placebo group and $x = 1$ represents the treatment arm.

The last model which is used for evaluating the second endpoint is a parametric Shared

gamma frailty model. Similar to the parametric Gamma frailty model, the individual frailty of each patient is taken into account. In general, the Shared gamma frailty model considers all observations within one cluster to have the same frailty in common (Duchateau and Janssen, 2007, p. 44). In the case of recurrent events, frailty Z is shared over all exacerbations of one patient (Box-Steffensmeier and De Boef, 2006, p. 3520). Therefore, the time between all events of each person are taken into account whereby each patient forms one cluster with individual frailty. The frailty describes analogous to the Gamma frailty model the heterogeneity between patients but is constant within one patient.

The model of the Shared gamma frailty model is equivalent to the one of the Gamma frailty model in equation 5.48 (Munda et al., 2012, p. 3):

$$h(t|\mathbf{X}, \mathbf{Z}) = \mathbf{Z} \cdot h_0(t) \cdot g(\mathbf{X}), \tag{5.50}$$

with

$$g(\mathbf{X}) = e^{\beta^T \mathbf{X}} \tag{5.51}$$

5.3 χ^2-test of Independence

For the third endpoint it is of interest if there exists a dependence between the belonging to one trial arm j and the occurrence of at least one asthma exacerbation a. This results in binary data for the occurrence or no occurrence of at least one asthma exacerbation and the group respectively, which is analysed by the help of the χ^2-test of independence. In this context, a completer evaluation is performed by only considering those, who did not discontinue their treatment before the end of the study. For an appropriate use of this test, the expected counts of E_{aj} of all possible combinations of the trial arm j and the occurrence of at least one asthma exacerbation a must be equal or even larger 5 (Bamberg et al., 2011, p. 189):

$$E_{aj} \geq 5 \tag{5.52}$$

The test statistic for the χ^2-test of independence $\chi^2_{Ind.}$ is calculated in analogy to the test statistic of the Log rank test which is given in equation 5.10. It is the sum over all squared differences between the observed O_{aj} and the expected E_{aj} counts which are respectively divided by its expected count E_{aj}. Under the validity of the null hypotheses with the assumption of the independence of the belonging to one trial arm and the occurrence of at least one exacerbation, the test statistic $\chi^2_{Ind.}$ follows approximately a χ^2-distribution with (m-1)(s-1) degrees of freedom (Zucchini et al., 2009, p. 338). Hereby

m denotes again the number of trial arms and s contains the two possibilities of either one occurrence of an event or no occurrence. Hence, the degrees of freedom result:

$$\chi^2_{Ind.} = \sum_{j=1}^{m} \sum_{a=0}^{s} \frac{(O_{aj} - E_{aj})^2}{E_{aj}} \sim \chi^2_1 \tag{5.53}$$

The variety of these methods is oriented towards Benda et al. (2010, p. 312), who use among others the Log-linear Poisson model, Log-linear negative-binomial model and the Cox proportional hazards model to evaluate recurrent events or the time to the first event respectively. In this thesis, the Gamma frailty model, Shared gamma frailty and the Likelihood ratio test are chosen to include parametric methods into the evaluation of the second endpoint. Furthermore, this thesis explores whether the results of the Shared gamma frailty model are comparable to the results of the methods of the first endpoint as it takes into account the time between all succeeding events. The Log rank test completes the selection of the basic time-to-first-event methods. The χ^2-test is taken into account to investigate whether a simple binary evaluation of a clinical trial is sufficient.

5.4 Comparability of the statistical methods

To ensure the comparability of the parametric statistical methods explained above, their null hypotheses are transformed into each other. Firstly, this is done for the hypothetical estimand and secondly for the treatment-policy estimand. In general, the parametric methods state the equivalence of the distribution in the treatment and placebo group in their null hypotheses. So, the treatment effect is zero because the hypothetical and treatment-policy effects are logarithmized in this thesis. No treatment effect exists if the parameters specifying the according distributions are the same in both trial arms. In this section, the hypotheses of the Log-linear Poisson model represent the starting point. The hypotheses of the other models are transformed into this initial one.

5.4.1 Comparability of the statistical methods for the hypothetical estimand

For the logarithmized hypothetical estimand $\big(\log(\Delta_{hyp})\big)$ it holds:

$$H_{0_{hyp}} : \log(\Delta_{hyp}) = 0 \tag{5.54}$$

By using the Poisson Model, the Poisson distribution is assumed to describe the number of events depending on the parameter λ_{P_A} in the treatment group and on $\lambda_{P_{PL}}$ in the placebo group. To have the same Poisson distribution in both trial arms, the following

equation must hold:

$$H_{0_{Pois_{hyp}}} : \lambda_{P_A} = \lambda_{P_{PL}} \tag{5.55}$$

The null hypothesis of the Log-linear negative-binomial model $H_{0_{NB_{hyp}}}$ is given by:

$$H_{0_{NB_{hyp}}} : \alpha_A = \alpha_{PL} \ \wedge \ \beta_A = \beta_{PL} \tag{5.56}$$

To link the null hypothesis of the Log-linear Poisson model with the one of the Log-linear negative-binomial model, the parameters α, β and ϑ must be defined with the help of the Gamma distribution and Poisson-Gamma distribution given in section 4.1:

$$E(\lambda_P) = \frac{\alpha}{\beta} \tag{5.57}$$

$$\Rightarrow \alpha = E(\lambda_P) \cdot \beta \tag{5.58}$$

$$\phi = \frac{Var(y)}{E(y)} = \frac{\alpha\frac{\vartheta}{\beta}\left(1+\frac{\vartheta}{\beta}\right)}{\vartheta\frac{\alpha}{\beta}} = 1 + \frac{\vartheta}{\beta} = \phi \tag{5.59}$$

$$\Rightarrow \beta = \frac{\vartheta}{\phi - 1} \tag{5.60}$$

The null hypothesis of the count data models will now be transformed into the null hypothesis of the parametric time-to-first-event methods. For this connection, the relationship

$$\lambda_P = \lambda_E \cdot \Delta t, \tag{5.61}$$

already given in equation 4.4 between λ_P and λ_E must be remembered.

The Likelihood ratio test compares both exponential distributions of treatment and placebo arm describing the time to the first event t_E. These distributions depend on the parameters λ_{E_A} respectively $\lambda_{E_{PL}}$. The null hypothesis of the Likelihood ratio test $H_{0_{LR_{hyp}}}$ is therefore (Lee and Wang, 2013, p. 229):

$$H_{0_{LR_{hyp}}} : \lambda_{E_A} = \lambda_{E_{PL}} = \frac{\lambda_{P_A}}{\Delta t} = \frac{\lambda_{P_{PL}}}{\Delta t} \tag{5.62}$$

To compare the null hypothesis of the Gamma frailty model, the Lomax distributions in treatment and placebo arm resulting of the mixture of the exponential distribution describing the time to the first event t_E and the Gamma distribution describing λ_E must be equal:

Therefore, the null hypothesis of the Gamma frailty model $H_{0_{GF_{hyp}}}$ is equivalent to the one of the Log-linear negative-binomial model:

$$H_{0_{GF_{hyp}}} : \alpha_A = \alpha_{PL} \ \wedge \ \beta_A = \beta_{PL} \tag{5.63}$$

In consequence, the Shared gamma frailty model which evaluates the time between all succeeding events possesses the same null hypothesis $H_{0_{SGF_{hyp}}}$:

$$H_{0_{SGF_{hyp}}} : \alpha_A = \alpha_{PL} \ \wedge \ \beta_A = \beta_{PL} \tag{5.64}$$

5.4.2 Comparability of the statistical methods for the treatment-policy estimand

The null hypotheses for the logarithmized treatment-policy estimand $\big(\log(\Delta_{tp})\big)$is given by:

$$H_{0_{tp}} : \log(\Delta_{tp}) = 0 \tag{5.65}$$

Consequently, both parameters λ_{P_1} and λ_{P_2} must be equal in treatment and placebo arm under validity of the null hypothesis $H_{0_{Pois_{tp}}}$:

$$H_{0_{Pois_{tp}}} : \lambda_{P_{1_A}} = \lambda_{P_{1_{PL}}} \ \wedge \ \lambda_{P_{2_A}} = \lambda_{P_{2_{PL}}} \tag{5.66}$$

The null hypothesis of the Log-linear negative-binomial model $H_{0_{NB_{tp}}}$ is therefore defined:

$$H_{0_{NB_{tp}}} : \alpha_{1_A} = \alpha_{1_{PL}} \ \wedge \ \alpha_{2_A} = \alpha_{2_{PL}} \ \wedge \ \beta_A = \beta_{PL} \tag{5.67}$$

The Likelihood ratio test compares the Schuhl distributions of both trial arms depending on the parameters λ_{E_1} and λ_{E_2}. Because of the relationships

$$\lambda_{P_1} = \lambda_{E_1} \cdot t_{TD} \tag{5.68}$$

and

$$\lambda_{P_2} = \lambda_{E_2} \cdot (1 - t_{TD}), \tag{5.69}$$

it becomes obvious that t_{TD} must be drawn from the same distributions in treatment and placebo arm, to ensure comparability between the null hypothesis for count data models and those for the second endpoint.

Therefore, $H_{0_{LR_{tp}}}$ equals:

$$H_{0_{LR_{tp}}} = \lambda_{E_{1_A}} = \lambda_{E_{1_{PL}}} \ \wedge \ \lambda_{E_{2_A}} = \lambda_{E_{2_{PL}}} \ \wedge \ \lambda_{TD_A} = \lambda_{TD_{PL}} \tag{5.70}$$

In general:

$$\lambda_{E_1} = \frac{\lambda_{P_1}}{t_{TD}} \tag{5.71}$$

$$\lambda_{E_2} = \frac{\lambda_{P_2}}{1 - t_{TD}} \tag{5.72}$$

The Gamma frailty model additionally takes into consideration the individual frailty of the patients. Its null hypothesis $H_{0_{GF_{tp}}}$ is therefore:

$$H_{0_{GF_{tp}}} : \alpha_{1_A} = \alpha_{1_{PL}} \ \wedge \ \alpha_{2_A} = \alpha_{2_{PL}} \ \wedge \ \beta_A = \beta_{PL} \ \wedge \ \lambda_{TD_A} = \lambda_{TD_{PL}} \tag{5.73}$$

With the general definition of

$$\alpha_1 = \frac{\lambda_{P_{1_A}}}{t_{TD}} \cdot \beta \tag{5.74}$$

$$\alpha_2 = \frac{\lambda_{P_{1_A}}}{1 - t_{TD}} \cdot \beta, \tag{5.75}$$

it becomes obvious, that t_{TD} must again be described by the same distributions. The same is valid for the Shared gamma frailty model:

$$H_{0_{SGF_{tp}}} : \alpha_{1_A} = \alpha_{1_{PL}} \ \wedge \ \alpha_{2_A} = \alpha_{2_{PL}} \ \wedge \ \beta_A = \beta_{PL} \ \wedge \ \lambda_{TD_A} = \lambda_{TD_{PL}} \tag{5.76}$$

5.5 Type I error rate, Power and Bias to assess statistical methods

To assess the quality of the chosen statistical methods, their two-sided type I error rate, power and bias are evaluated. The aim of this thesis is to explore in scenarios without treatment effect whether the methods reject their null hypotheses only in maximal 5% of the testing decisions. If the null hypothesis is rejected in scenarios without treatment effect, a new ineffective drug will be approved (Zucchini et al., 2009, p. 244). Therefore, the type I error rate of each method may not be larger than 5% for each scenario. The corresponding 95%-confidence intervals of the calculated type I error rate ($\hat{\alpha}$) is

calculated by:

$$\left[\hat{\alpha} - 1.96\sqrt{\frac{\hat{\alpha}(1-\hat{\alpha})}{N_{conv}}} \;;\; \hat{\alpha} + 1.96\sqrt{\frac{\hat{\alpha}(1-\hat{\alpha})}{N_{conv}}}\right] \tag{5.77}$$

N_{conv} denotes the number of converged models.

Those methods which do not show a larger type I error rate than 5% are compared with respect to their power. Hereby, these methods are used to analyze the equivalent scenarios with treatment effect to investigate the method with the largest power. Whereby power is defined as follows:

$$Power = 1 \; - \; \beta\text{-}error \tag{5.78}$$

The β-error arises if the treatment effect is not found by the statistical method, and therefore the null hypothesis is not rejected (Zucchini et al., 2009, p. 244). Hence, the power denotes the capability to recognize a effective drug.

The third criterion is the bias which measures an average difference between the estimated and true treatment effect of each method (Kneib and Hambuckers, winter term 2017/2018, p. 26). Usually, the bias is calculated on the logarithmized scale (see e.g. Wan et al., 2015, p. 2240). This is also done in this thesis to provide an improved comparability between the models. The calculation of the bias on this log-scale ensures the measurement of the relative difference between the true and estimated treatment effect.

$$Bias = \log\left(\frac{\Delta_{estimated}}{\Delta_{true}}\right) = \log(\Delta_{estimated}) - \log(\Delta_{true}) \tag{5.79}$$

If the bias is zero, the method will estimate the true treatment effect. With a bias $>$ 0, $\log(\Delta_{estimated})$ is too small, and therefore the true treatment effect is underestimated. Whereas the true treatment effect is overestimated if bias < 0. In this case, $\log(\Delta_{estimated})$ is too large. It must be remembered that the bias is measured on the logarithmized scale. Consequently, a treatment effect which is smaller than zero depicts that the medication in the active group is more effective than the one in the placebo group. Otherwise, a treatment effect larger than zero indicates a superiority of the medication strategy in the placebo group compared to the one in the active group. This can especially be the case for the treatment-policy estimand. If the new drug is ineffective and if there is a higher percentage of patients with treatment discontinuation in the placebo group than in the treatment group, there are more patients medicated with an effective rescue medication in the placebo group.

6 Simulations

After the description of adequate statistical methods for analyzing the samples, the simulation of this data now comes to the fore. Firstly, the chosen scenarios are defined. Secondly, the generation of data with the according parameters is presented. The function of simulation for the scenarios with treatment discontinuation mechanism MCAR is shown in the appendix B.1 on page 145. The function of simulation for the scenarios with MNAR is presented in the appendix B.2 on page 151.

6.1 Choice of scenarios and parameters

For the evaluation of the hypothetical estimand, scenarios based on six varying parameters are chosen:

1. The first parameter encompasses the treatment discontinuation in the placebo group. Patients in the placebo group refuse the participation in the trial at some point of time as they recognize their belonging to the placebo group. Furthermore, they could need an effective medication because of the severity of their disease. The percentage of treatment discontinuation ($\%_{TD_{PL}}$) changes between either 20% or 50% of patients in the placebo group who discontinue their treatment early. Hence, this parameter contains two different variations. Hereby, the choice of the lowest discontinuation proportion of 20% is oriented towards the real study 'Dose Ranging Efficacy And safety with Mepolizumab (DREAM) in severe asthma' (Pavord et al., 2012, p. 651). In the study, a dropout percentage of 18% is observed which includes treatment discontinuation and study dropout for any other reason (Keene et al., 2014, p. 256). In this thesis, the referred dropout is rounded to 20% and is used as percentage for the treatment discontinuation only. It will be explored if such a low proportion leads to a difference between both estimands. Whereas with a high treatment discontinuation percentage of 50% an obvious distinction between the two treatment effects is expected.

2. With the treatment discontinuation in the placebo group, the treatment discontinuation in the treatment group is closely related. It is chosen in that way to generate a difference in the treatment discontinuation of -20%, 0%, and +20% between both trial arms. For each of the two possible percentages of treatment discontinuation in the placebo group, the percentage of treatment discontinuation in the treatment group ($\%_{TD_A}$) varies between -20%, 0% and +20% of patients in the treatment group. If the treatment discontinuation in the treatment arm reaches a very high percentage, patients suffer from side effects of the new drug and stop their med-

ication. Otherwise, if the new drug shows a great treatment effect, only a low percentage of patients in the treatment group will discontinue their treatment.

3. Data are simulated with two different treatment discontinuation mechanisms which are either missing completely at random or missing not at random. For MCAR, patients are randomly chosen to interrupt their medication. Whereas for MNAR, patients who would potentially have a high number of exacerbations at the end of the study, discontinue treatment earlier.

4. The fourth parameter describes the percentage of patients with at least one asthma exacerbation in the placebo group ($\%_{E_{PL}}$). The DREAM study graphically shows the proportion of patients without any exacerbation before treatment discontinuation which approximately are 34% (which are 66% of patients with at least one event during the whole period) (Pavord et al., 2012, p. 654). For a close reproduction of such a trial, it is simulated that 70% of the patients in the placebo group potentially suffer an exacerbation during the whole year. This yields a proportion of 64% of patients with at least one asthma exacerbation until the treatment discontinuation.

5. Two different values are used for the simulation of the percentage of patients with at least one asthma exacerbation in the treatment group ($\%_{E_A}$). If no treatment effect exists, it represents 70%, too. In the case that the new drug shows an impact on the number of asthma exacerbations, only 50% of the patients in the treatment group suffer at least one event if no treatment discontinuation occurs in the active group. This value for the simulation of the treatment effect represents approximately the mean of the active groups of patients with at least one event (Pavord et al., 2012, p. 654).

6. The last parameter represents the overdispersion which arises with the simulation of independent event rates. The overdispersion parameter ϕ can either take the value 1 or 2. In the first case, if $\phi = 1$, overdispersion does not exist what results in equation 4.11. Otherwise, if $\phi = 2$, data with a strong overdispersion are simulated. In this case, a high type I error rate of the Log-linear Poisson model is expected.

Therefore, for the hypothetical estimand 2 x 3 x 2 x 1 x 2 x 2 = 48 scenarios are chosen.

For the treatment-policy estimand, two further parameters are used:

7. The percentage of patients who suffer at least one exacerbation after their treatment discontinuation is 60%. But only those patients, whose treatment discontinuation takes place before the end of the trial, can suffer an event after their

treatment discontinuation. In this case, the rescue medication which is administered to those who discontinue their treatment is less effective than the new drug but more than a placebo.

8. The last parameter describes the percentage of patients with study dropout after their treatment dropout ($\%_{SD}$) whereby three variations exist. First, no study dropout occurs, therefore data of all patients until the end of the study can be collected. Second, 50% of those patients, who discontinue treatment, stop the participation in the study at the same time. At last, 100% of the patients with treatment discontinuation will drop out of the study.

There result 2 x 3 x 2 x 1 x 2 x 2 x 1 x 3 = 144 scenarios for the treatment-policy estimand.

As already described in section 2.1, for the hypothetical estimand only data until treatment discontinuation is considered. Whereas for the hypothetical estimand, all data until the end of the study respectively until the study dropout are taken into account. The hypothetical estimand is evaluated on the basis of the samples, that include the scenarios with the percentage of patients, who suffer at least one exacerbation after their treatment discontinuation of 60%. To reach a power of 90% in the evaluation of the hypothetical estimand with the Negative-binomial model in a scenario with treatment discontinuation of 20% in placebo and treatment group, 50% of patients with at least one exacerbation in the placebo group, 70% of patients with at least one exacerbation in the treatment group and an overdispersion of 2, a sample size of 330 patients for all scenarios is chosen. Each scenario is simulated 10.000 times.

6.2 Generation of data

In the MCAR case patients are randomly assigned to either placebo or treatment group at first. Then, time to treatment discontinuation t_{TD} is simulated with parameter λ_{TD}:

$$t_{TD} \sim \text{Exp}(\lambda_{TD}) \tag{6.1}$$

The density function and cumulative probability function of the exponential distribution are already given in equation 4.2 and 4.3 in other context. $\lambda_{TD_{PL}}$ must be derived from the parameters given in section 6.1. This calculation is exemplified for 20% of patients with treatment discontinuation. All other parameter values described in section 6.1 are achieved the same way. Table 2 shows $\lambda_{TD_{PL}}$.

$$F(t_{TD} \geq 1) = 0.8 \tag{6.2}$$
$$1 - F(t_{TD} \leq 1) = 0.8 \tag{6.3}$$
$$1 - (1 - e^{-\lambda_{TD_{PL}}}) = 0.8 \tag{6.4}$$
$$e^{-\lambda_{TD_{PL}}} = 0.8 \tag{6.5}$$
$$-\lambda_{TD_{PL}} = \log(0.8) \tag{6.6}$$
$$\lambda_{TD_{PL}} = -\log(0.8) \tag{6.7}$$
$$\lambda_{TD_{PL}} = 0.22 \tag{6.8}$$

Table 2: $\lambda_{TD_{PL}}$ for proportions of patients with treatment discontinuation in placebo group

$\%_{TD_{PL}}$	0.2	0.5
$\lambda_{TD_{PL}}$	0.22	0.69

In table 3 the λ_{TD_A} is listed. As mentioned in section 6.1, the percentage of treatment discontinuation in treatment group $\%_{TD_A}$ depends on the percentage of treatment discontinuation in the placebo group $\%_{TD_{PL}}$. $\%_{TD_A}$ is calculated from the sum of $\%_{TD_{PL}}$ plus either -20%, 0% or +20%. If $\%_{TD_A}$ becomes 0, no treatment discontinuation is simulated for the according scenarios. Therefore, λ_{TD_A} is not specified in this case.

Table 3: λ_{TD_A} for proportions of patients with treatment discontinuation in treatment group

$\%_{TD_A} - \%_{TD_{PL}}$ \ $\%_{TD_{PL}}$	0.2	0.5
-0.2	---	0.36
±0.0	0.22	0.69
+0.2	0.51	1.20

Third, time between events until treatment discontinuation is simulated. Hereby, the time until the next exacerbation t_E is drawn from an Exponential distribution:

$$t_E \sim \text{Exp}(\lambda_E) \tag{6.9}$$

In case of overdispersed data, both parameters α and β of the Poisson-Gamma distribution need to be determined. These are derived exemplarily for 50% of patients with at least one exacerbation in the placebo group and an overdispersion parameter $\phi = 2$ and $\vartheta = 1$ below. As $\lambda_{E_{PL}} \sim \text{Ga}(\alpha, \beta)$:

$$\text{E}(\lambda_{E_{PL}}) = \frac{\alpha}{\beta} = 0.69 \tag{6.10}$$

$$\Rightarrow \alpha = \text{E}(\lambda_{E_{PL}}) \cdot \beta = 0.69 \cdot \beta \tag{6.11}$$

$$\tag{6.12}$$

$$\phi = \frac{\text{Var}(y)}{\text{E}(y)} = \frac{\alpha\frac{\vartheta}{\beta}\left(1 + \frac{\vartheta}{\beta}\right)}{\vartheta\frac{\alpha}{\beta}} = 1 + \frac{\vartheta}{\beta} = 2 \tag{6.13}$$

$$\Rightarrow \beta = \frac{\vartheta}{\phi - 1} = 1 \tag{6.14}$$

$$\Rightarrow \alpha = 0.69 \tag{6.15}$$

In table 4, $\lambda_{E_{1_{PL}}}$ and $\lambda_{E_{1_A}}$ are given. α and β are calculated for $\phi = 2$. If $\phi = 1$, no overdispersion exists, hence α_1 and β are not needed. The index 1 denotes the period until treatment discontinuation.

Table 4: $\lambda_{E_{1_{PL}}}$, $\lambda_{E_{1_A}}$ and parameters of Poisson-Gamma distribution for proportions of patients with at least one event in placebo or trial arm

$\%_{E_{1_{PL}}}$	0.5	$\%_{E_{1_A}}$	0.5	0.7
$\lambda_{E_{1_{PL}}}$	0.69	$\lambda_{E_{1_A}}$	0.69	1.20
$\alpha_{1_{PL}}$	0.69	α_A	0.69	1.20
β	1.00	β	1.00	1.00

Fourth, the time between the events after treatment discontinuation follow an Exponential distribution in analogy to the time between the events before treatment discontinuation. In table 5, λ_{E_2} and the parameters of the Poisson-Gamma distribution α and β are shown:

Table 5: λ_{E_2} and parameters of Poisson-Gamma distribution for proportions of patients with at least one event after treatment discontinuation

$\%_{E_2}$	0.6
λ_{E_2}	0.92
α	0.92
β	1.00

In the case of scenarios with treatment discontinuation missing not at random, the time between events until the end of the study is simulated first. Those patients with a higher number of exacerbation will stop their treatment earlier. To calculate the individual λ_{TD}, the general $\lambda_{TD_{PL}}$, given in table 2, and λ_{TD_A}, in table 3, are multiplied by the patient's number of events. If a person suffers no exacerbation until the end of the study, the original λ_{TD} is still valid for this patient. With help of the individual λ_{TD}, the time until treatment discontinuation is simulated. Afterwards, the time between events until treatment discontinuation are chosen. Events which occur later are not counted any more.

Finally, a censoring variable is generated. It is needed for the evaluation of the second endpoint. The definition of the censoring variable is described in section 5.2.

For scenarios that include a study dropout of 50% of the patients that discontinue treatment during the time of the study, the time until study dropout t_{SD} is simulated afterwards. Hereby, t_{SD} is the sum of the time until treatment discontinuation t_{TD} and the time between treatment discontinuation and study dropout t_{SD-TD}:

$$t_{TD} \sim \text{Exp}(\lambda_{TD}),\ t_{SD-TD} \sim \text{Exp}(\lambda_{SD-TD}) \tag{6.16}$$

$$t_{SD} = t_{TD} + t_{SD-TD} \tag{6.17}$$

The distribution of t_{SD} can be derived by the convolution of two exponentially distributed random variables. For simplicity, this substitution holds:

$$t_{SD} = Z,\ t_{TD} = X,\ t_{SD-TD} = Y,\ \lambda_{TD} = \lambda_X,\ \lambda_{SD-TD} = \lambda_Y \tag{6.18}$$

$$Z = X + Y \tag{6.19}$$

Then, the convolution gives:

$$f(z) = \int_0^z f(x)f(z-x)dx \tag{6.20}$$

$$= \int_0^z \lambda_X e^{-\lambda_X x} \cdot \lambda_Y e^{-\lambda_Y (z-x)} dx \tag{6.21}$$

$$= \lambda_X \lambda_Y e^{-\lambda_Y z} \int_0^z e^{x(-\lambda_X + \lambda_Y)} dx \tag{6.22}$$

$$= \lambda_X \lambda_Y e^{-\lambda_Y z} \left(\frac{e^{x(-\lambda_X + \lambda_Y)}}{-\lambda_X + \lambda_Y} \right) \Bigg|_0^z \tag{6.23}$$

$$= \lambda_X \lambda_Y e^{-\lambda_Y z} \left(\frac{e^{z(-\lambda_X + \lambda_Y)}}{-\lambda_X + \lambda_Y} - \frac{1}{-\lambda_X + \lambda_Y} \right) \tag{6.24}$$

$$= \frac{\lambda_X \lambda_Y}{-\lambda_X + \lambda_Y} \left(e^{-\lambda_X z} - e^{-\lambda_Y z} \right) \tag{6.25}$$

Hence, the density describing the time until study dropout t_{SD} is defined as:

$$f(t_{SD}) = \frac{\lambda_{TD} \lambda_{SD-TD}}{-\lambda_{TD} + \lambda_{SD-TD}} \left(e^{-\lambda_{TD} t_{SD}} - e^{-\lambda_{SD-TD} t_{SD}} \right) \tag{6.26}$$

The cumulative density function is derived in general by:

$$F(z) = \int_0^z f(z) dz \tag{6.27}$$

$$= \int_0^z \frac{\lambda_X \lambda_Y}{-\lambda_X + \lambda_Y} \left(e^{-\lambda_X z} - e^{-\lambda_Y z} \right) dz \tag{6.28}$$

$$= \frac{\lambda_X \lambda_Y}{-\lambda_X + \lambda_Y} \left(\int_0^z e^{-\lambda_X z} dz - \int_0^t e^{-\lambda_Y z} dz \right) \tag{6.29}$$

$$= \frac{\lambda_X \lambda_Y}{-\lambda_X + \lambda_Y} \left(-\frac{e^{-\lambda_X z}}{\lambda_X} \Bigg|_0^z - \left(-\frac{e^{-\lambda_Y z}}{\lambda_Y} \right) \Bigg|_0^z \right) \tag{6.30}$$

$$= \frac{\lambda_X \lambda_Y}{-\lambda_X + \lambda_Y} \left(-\frac{e^{-\lambda_X z}}{\lambda_X} + \frac{1}{\lambda_X} + \frac{e^{-\lambda_Y z}}{\lambda_Y} - \frac{1}{\lambda_Y} \right) \tag{6.31}$$

$$= \frac{\lambda_Y(-e^{-\lambda_X z}+1)+\lambda_X(-e^{-\lambda_Y z}-1)}{-\lambda_X+\lambda_Y} \tag{6.32}$$

Consequently, the cumulative density function of $f(t_{SD})$ is given by:

$$F(t_{SD}) = \frac{\lambda_{SD-TD}(-e^{-\lambda_{TD}t_{SD}}+1)+\lambda_{TD}(-e^{-\lambda_{SD-TD}t_{SD}}-1)}{-\lambda_{TD}+\lambda_{SD-TD}} = P \tag{6.33}$$

Hereby, for each possible $\%_{TD}$ is $\%_{SD} = 0.05 \cdot \%_{TD}$. To ensure a study dropout of 50% of the patients with treatment discontinuation in each treatment and placebo group, $F(t_{SD})$ must be solved with respect to λ_{SD-TD}. As this is not possible in closed form, $e^{-\lambda_{SD-TD}}$ is approximated by a Taylor series to the fourth degree:

$$-e^{-\lambda_{SD-TD}t_{SD}} = 1-\lambda_{SD-TD}+\frac{\lambda_{SD-TD}^2}{2!}-\frac{\lambda_{SD-TD}^3}{3!}+\frac{\lambda_{SD-TD}^4}{4!} \tag{6.34}$$

This approximation is now set into $F(t_{SD})$ and further simplified:

$$\begin{aligned}&\lambda_{SD-TD}(-e^{-\lambda_{TD}}+1)+\lambda_{TD}\Big(-\lambda_{SD-TD}+\frac{\lambda_{SD-TD}^2}{2}-\frac{\lambda_{SD-TD}^3}{6}+\frac{\lambda_{SD-TD}^4}{24}\Big)\\ &\quad = P(\lambda_{SD-TD}-\lambda_{TD})\end{aligned} \tag{6.35}$$

$$\begin{aligned}&\frac{\lambda_{TD}}{24}\cdot\lambda_{SD-TD}^4-\frac{\lambda_{TD}}{6}\cdot\lambda_{SD-TD}^3+\frac{\lambda_{TD}}{2}\cdot\lambda_{SD-TD}^2\\ &\quad+(-e^{-\lambda_{TD}}+1-\lambda_{TD}-P)\cdot\lambda_{SD-TD}+\lambda_{TD}P=0\end{aligned} \tag{6.36}$$

This equation to the power of four can be solved with respect to λ_{SD-TD} using Ferrari's formula. As these solutions represent just an approximation, they are adjusted by hand to reach the correct value of $\%_{SD}$. In table 6, the solutions for λ_{SD-TD} used for the simulations are shown:

Table 6: values for λ_{SD-TD}

$\%_{TD}$	$\%_{SD}$	λ_{SD-TD}
0.2	0.2·0.5 = 0.1	1.52
0.3	0.3·0.5 = 0.15	1.48
0.4	0.4·0.5 = 0.2	1.44
0.5	0.5·0.5 = 0.25	1.39
0.7	0.7·0.5 = 0.35	1.27

The censoring variable and the number of events until study dropout is adjusted at the end.

7 Results

The following section deals with the presentation of the results. First, the actual parameters of the simulated scenarios are provided. Second, the results of the evaluation of these scenarios are shown for the hypothetical estimand and the treatment-policy estimand.

The simulation and analysis of scenarios is implemented with the statistical software `R` version 3.4.3 (R Core Team, 2017). It is used with the interface `RStudio` version 1.1.383 (RStudio Team, 2017). The Log-linear Poisson model and χ^2-test are calculated with the functions of the basic software. The Log-linear negative-binomial model is implemented in the `MASS`-package (Ripley et al., 2017). The Log rank test and the Cox proportional hazards model are evaluated using the `survival`-package (Therneau and Lumley, 2017). The package `parfm` contains the functions for the Gamma frailty model and the Shared gamma frailty model (Munda et al., 2017). The function for the calculation of the Likelihood ratio test is not yet implemented in `R`.

7.1 Description of the simulated scenarios

7.1.1 Description of the simulated scenarios for the hypothetical estimand

To begin with, the simulated scenarios for the hypothetical estimand are described. In table 39 on page 136 the actual parameters for scenarios of the hypothetical estimand are given. In figure 7 on page 60 these parameters are visualized. For the scenarios 1-12 which contain an overdispersion parameter $\phi = 1$ and treatment discontinuation mechanism MCAR, the percentage of patients with treatment discontinuation in placebo ($\%_{TD_{PL}}$) of either 20% or 50% and in treatment group ($\%_{E_{1_A}}$) with either a difference of -20%, 0% or +20% compared to placebo group represent exactly the planned scenarios. The percentage of patients with at least one event is lower than the in table 4 described $\%_{E_{1_{PL}}}$ and $\%_{E_{1_A}}$. This is caused by the fact that table 4 gives the parameters for a period of one year but in the simulated scenarios patients may drop out of the treatment or placebo group earlier. Hence, $\%_{E_{1_{PL}}}$ is either 64% or 54% of the patients, which depends of the lower or higher $\%_{TD_{PL}}$. In the case of a treatment effect, the $\%_{E_{1_A}}$ changes from 50% to 31% again influenced by a low or high $\%_{TD_A}$. If no treatment effect is simulated, $\%_{E_{1_A}}$ lies in the range between 70% and 45%. This phenomenon of a lower percentage of patients with at least one event is valid for all scenarios of either hypothetical or treatment-policy estimand.

Scenarios 13-24 again are simulated with $\phi = 1$ but with the treatment discontinuation mechanism MNAR. Therefore, $\%_{TD_{PL}}$ and $\%_{TD_A}$ give other values than those described

in the scenarios before. $\%_{TD_{PL}}$ represents now either 27% or 60%. They are therefore a bit higher than in the scenarios with MCAR, as patients with a higher number of events discontinue treatment earlier. In the treatment arm for scenarios 13-18 with treatment effect, the aimed differences of $\%_{TD_A}$ compared to placebo group are are not achieved. Because of the treatment effect, the time until treatment discontinuation is influenced in different way in both trial arms. Thus, $\%_{TD_A}$ changes between 0%, 23% and 45 % in the case of low $\%_{TD_{PL}}$ and between 34%, 54% and 74% for the higher $\%_{TD_{PL}}$. But for the scenarios 19-24, the planned differences $\%_{TD_{PL}}$-$\%_{TD_A}$ are approximately achieved. $\%_{E_{1_{PL}}}$ and $\%_{E_{1_A}}$ differ, compared to scenarios 1-12 as the periods of time until treatment discontinuation are not the same. The $\%_{E_{1_{PL}}}$ represents either 62% or 61% respectively, or 48%. Here $\%_{E_{1_A}}$ lies in the range of 50% and 29% for scenarios with treatment effect and between 70% and 40%, if no treatment effect occurs.

Next, the scenarios 25-36 are presented which are based on $\phi = 2$ and the treatment discontinuation mechanism MCAR. The $\%_{TD_{PL}}$ and the $\%_{TD_A}$ denote exactly the same values as in scenarios 1-12. The $\%_{E_{1_{PL}}}$ and the $\%_{E_{1_A}}$ are lower than in the corresponding scenarios without overdispersion. The $\%_{E_{1_{PL}}}$ is either 52% for the lower $\%_{TD_{PL}}$ and 44% for the higher $\%_{TD_{PL}}$. In scenarios with treatment effect 25-30, the $\%_{E_{1_{PL}}}$ ranges between 38% and 25% and in the scenarios 31-36 without treatment effect between 57% and 38%.

Finally, the scenarios 37-48 contain overdispersed data with $\phi = 2$, but with treatment discontinuation mechanism MNAR. Again, the $\%_{TD_{PL}}$ and the $\%_{TD_A}$ represent comparable parameter values to scenarios 13-24. In anaLogy to these scenarios, the $\%_{E_{1_{PL}}}$ and the $\%_{E_{1_A}}$ are lower than the percentage of the corresponding scenarios with treatment discontinuation mechanism MCAR, as patients with higher number of events discontinue treatment earlier. Hereby, the $\%_{E_{1_{PL}}}$ denotes either 49% or 39% or 38% respectively. For the scenarios 37-42 with treatment effect, varies in the range between 38% and 23%. For the scenarios 43-48 which do not contain a treatment effect, the $\%_{E_{1_A}}$ is in the interval from 57% to 33% .

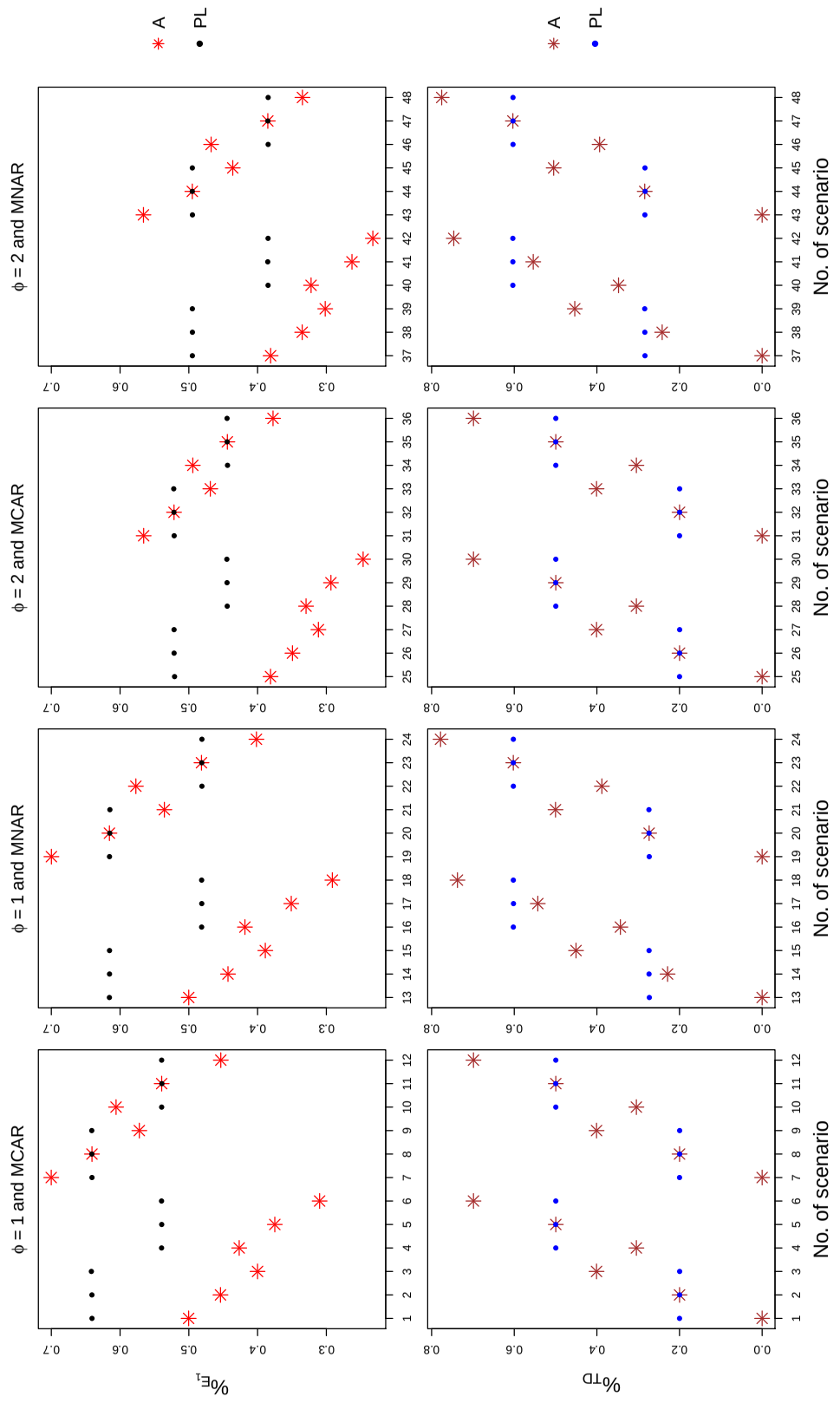

Figure 7: The percentage of patients with at least one event before treatment discontinuation ($\%_{E_1}$) and the percentage of patients with treatment discontinuation ($\%_{TD}$) in treatment (A) and placebo (PL) group for each scenario under missing completely at random (MCAR) and missing not at random (MNAR) for hypothetical estimand

7.1.2 Description of the simulated scenarios for the treatment-policy estimand

In table 40 starting on page 138, the resulting percentages of patients with treatment discontinuation in treatment ($\%_{TD_A}$) and placebo group ($\%_{TD_{PL}}$) are documented. Furthermore, this table contains the percentage of patients with study dropout after treatment discontinuation in treatment ($\%_{SD_A}$) and placebo group ($\%_{SD_{PL}}$), the percentage of patients with at least one asthma exacerbation before treatment discontinuation in treatment ($\%_{E_{1_A}}$) and placebo group ($\%_{E_{1_{PL}}}$), the percentage of patients with at least one event after treatment discontinuation in treatment ($\%_{E_{2_A}}$) and placebo group ($\%_{E_{2_{PL}}}$), the percentage of patients with at least one exacerbation during the whole study in treatment ($\%_{E_{total_A}}$) and placebo group ($\%_{E_{total_{PL}}}$), the mechanism of treatment discontinuation and the overdispersion parameter ϕ for treatment-policy estimand. The parameters concerning treatment discontinuation and study discontinuation are plotted in figure 8 on page 63. Furthermore, figure 9 on page 64 visualizes the percentages of the patients with at least one event before and after treatment discontinuation.

Scenarios 1-36 have $\phi = 1$ and the treatment discontinuation mechanism MCAR in common. The $\%_{TD_{PL}}$ is either 20% or 50%, and the $\%_{TD_A}$ is either the difference of -20%, 0% or +20%, compared to placebo group. The $\%_{SD_A}$ and $\%_{SD_{PL}}$ are zero for the scenarios 1-12, and one for the scenarios 25-36. In the scenarios 13-24, approximately 50% of those patients with treatment discontinuation stop their study participation early. The exact percentage of study dropout are not always reached, as the calculated parameters explained in section 6.1 are just approximated via Taylor Series Expansion. The $\%_{E_{1_{PL}}}$ and the $\%_{E_{1_A}}$ represent the same values as the parameters for the hypothetical estimand for scenarios without overdispersion and MCAR. The $\%_{E_{2_{PL}}}$ is either 7% in case of the lower percentage of treatment discontinuation in placebo group, or 19%, if 50% of the patients discontinue treatment. The $\%_{E_{2_A}}$ changes between 0% and 28%, which depends on the lowest percentage of treatment discontinuation of 0% and the highest one of 70%. The $\%_{E_{total_{PL}}}$ summarizes the events before and after treatment discontinuation. Hereby, either 69% or 67% of the patients suffer at least one exacerbation in the placebo group. Whereas in the treatment group and event occurs for only 50% of the patients in the lowest case, and up to 70% of the patients in the highest case.

The scenarios 37-72 share $\phi = 1$ and the treatment discontinuation mechanism MNAR. The $\%_{TD_{PL}}$, the $\%_{TD_A}$, the $\%_{E_{1_A}}$ and the $\%_{E_{1_{PL}}}$ reflect the same values as in the corresponding scenarios of the hypothetical estimand with $\phi = 1$ and MNAR. In scenario 37-48 0% of the patients drop out of the study, in scenario 49-60, approximately 50% of the patients with treatment discontinuation drop out of the study and in scenario 61-72

all patients with treatment discontinuation drop out of the study before its official end. It must be taken into account that the higher the $\%_{TD}$, the more the $\%_{SD}$ deviates from 50% in the scenarios 49-60. It appears to be noticeable that both the $\%_{E2_{PL}}$ and the $\%_{E2_A}$ are higher than in the MCAR. This is reasoned by the fact that under MNAR, the treatment discontinuation tends to be earlier. Therefore, the percentage of patients with at least one event after treatment discontinuation can increase. $\%_{E_{total_{PL}}}$ and $\%_{E_{total_A}}$ have approximately the same values as in the MCAR case.

Next, the scenarios 73-108 with $\phi = 2$ and the treatment discontinuation mechanism MCAR are considered. Again, the $\%_{TD_{PL}}$, the $\%_{TD_A}$, the $\%_{E1_A}$ and the $\%_{E1_{PL}}$ are the same values as in the corresponding scenarios of the hypothetical estimand. The percentage of study dropout corresponds to the values of the scenarios with $\phi = 1$ and MCAR. In anaLogy to the percentages of patients with at least one event before treatment discontinuation, both the $\%_{E2_{PL}}$ and the $\%_{E2_A}$ are a little lower in scenarios with overdispersion compared to those without. Hereby, the $\%_{E2_{PL}}$ is either 6% or 15%, and the $\%_{E2_A}$ changes in the range between 0% and 23%. Consequently, the $\%_{E_{total_{PL}}}$ and the $\%_{E_{total_A}}$ are also lower than in the MCAR case. The $\%_{E_{total_{PL}}}$ equals 56%, and the $\%_{E_{total_A}}$ lies between 38% and 44% for the scenarios with treatment effect. In scenarios without treatment effect, the $\%_{E_{total_A}}$ is between 55% and 57% for the scenarios without treatment effect.

At last, the scenarios 109-144 are described, which have $\phi = 2$ and MNAR in common. The anaLogy between hypothetical and treatment-policy estimand can be continued. The $\%_{TD_{PL}}$, the $\%_{TD_A}$, the $\%_{E1_A}$ and the $\%_{E1_{PL}}$ adopt the same values as in the scenarios of the hypothetical estimand with overdispersion and MNAR. Both the $\%_{SD_{PL}}$ and the $\%_{SD_A}$ are comparable to those values of study dropout for scenarios with $\phi = 1$ and MNAR of the treatment-policy estimand. The $\%_{E2_{PL}}$ and the $\%_{E2_A}$ are both again higher than in the MCAR scenarios. This phenomenon has already been observed and explained for the scenarios without overdispersion. The values for the $\%_{E_{total_{PL}}}$ and the $\%_{E_{total_A}}$ are again comparable to those of the scenarios with $\phi = 2$ and MCAR. This is in analogy to the scenarios without overdispersion.

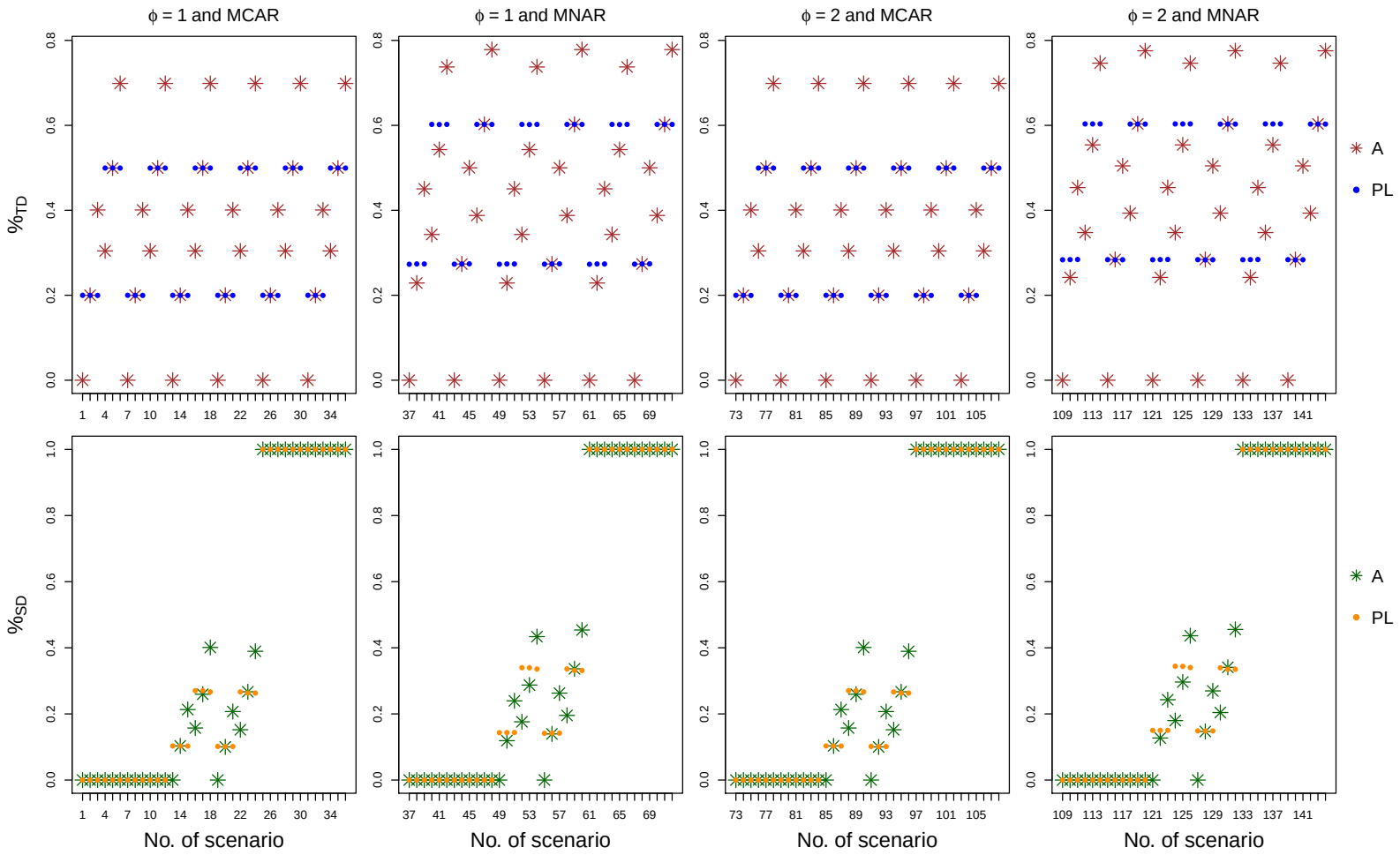

Figure 8: The percentage of patients with treatment discontinuation ($\%_{TD}$) and the percentage of patients with study dropout ($\%_{SD}$) in treatment (A) and placebo (PL) group for each scenario under missing completely at random (MCAR) and missing not at random (MNAR) for treatment-policy estimand

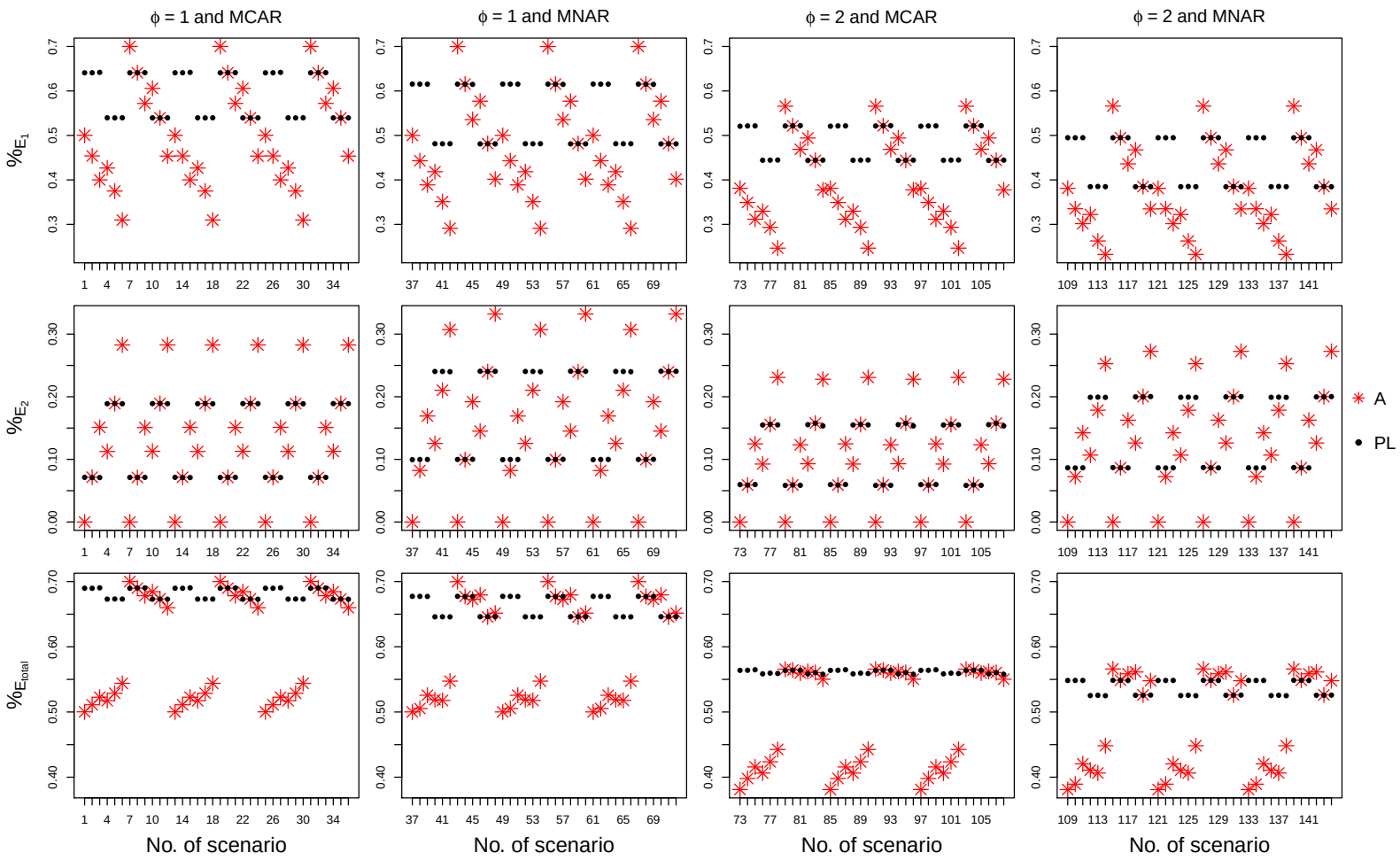

Figure 9: The percentage of patients with at least one event before treatment discontinuation ($\%_{E_1}$), the percentage of patients with at least one event after treatment discontinuation ($\%_{E_2}$) and the percentage of patients with at least one event during the whole study ($\%_{E_{total}}$) in treatment (A) and placebo (PL) group for each scenario under missing completely at random (MCAR) and missing not at random (MNAR) for treatment-policy estimand

7.2 Results for hypothetical estimand

This section presents the results for the hypothetical estimand seperated by both treatment discontinuation mechanisms MCAR and MNAR.

7.2.1 Results for the hypothetical estimand with treatment discontinuation mechanism MCAR

The results for the hypothetical estimand with treatment discontinuation mechanism MCAR, $\phi = 1$ and $\%_{TD_{PL}} = 20\%$ are listed in table 7 on page 69 and for $\phi = 1$ with $\%_{TD_{PL}} = 50\%$ in table 8 on page 70. Furthermore, the results for the hypothetical estimand with MCAR, $\phi = 2$ and $\%_{TD_{PL}} = 20\%$ are given in table 9 and for $\phi = 12$ and $\%_{TD_{PL}} = 50\%$ in table 10 on page 72. In table 9, the Log-linear negative-binomial model reaches a power of 0.903, if $\%_{TD_A} - \%_{TD_{PL}} = 0\%$. This scenario is used as the reference scenariot the power of the Log-linear negative-binomial model was used for the determination of the sample size. In general, power increases, if fewer patients discontinue treatment than in the reference scenario. Otherwise, if more people have a treatment discontinuation than in the reference scenario, the power of each method will be less than the corresponding power of this method in the reference scenario. Figure 10 on page 73 visualizes the bias of the evaluated methods for the treatment discontinuation mechanism MCAR. For the Log rank test, neither in context of the hypothetical estimand nor of the treatment-policy estimand bias calculations are performed.

It must be noted that more than 40% of the Log-linear negative-binomial models do not converge in scenarios without overdispersion. Therefore, these results must be treated with caution and are not interpreted. However, they are listed in the corresponding tables and figures. This high percentage of not converged models for the Log-linear negative-binomial model in scenarios without overdispersion also appears for the treatment-policy estimand. But for scenarios with overdispersion, there are no problems concerning the convergence of the Log-linear negative-binomial model. The remaining methods reveal always reveal a high number of converged models.

In general, it can be observed that only a small number of evaluation methods present an type I error rate lower or equal to 5% in scenarios without treatment effect (no tr. effect). Additionally, the 95%-confidence intervals (CI) for the calculated type I error rates are provided to show if they include an acceptable type I error rate. But those methods whose calculated type I error rate is lower or equal than 5% are considered to recognize ineffective drugs in at least 5%. Starting with the scenarios without overdispersion, $\%_{TD_{PL}} = 20\%$ and $\%_{TD_A} - \%_{TD_{PL}} = 0\%$, the Likelihood ratio test, the Gamma frailty model, the Shared gamma frailty model and the χ^2-test fulfill this requirement without

any bias. The Shared gamma frailty model reaches the highest power of these models with a value of 0.994 without any bias in the scenario with treatment effect (with tr. effect). If $\%_{TD_A} - \%_{TD_{PL}} = -20\%$, only the Log-linear Poisson model and the Shared gamma frailty model have a type I error rate smaller than 5%. Their type I error rates represent 0.049 with a bias of 0.11 for the Log-linear Poisson model and 0.045 with zero bias for the Shared gamma frailty model. Their power represents 0.996, whereby the Shared gamma frailty model provides an estimate with zero bias. Otherwise, if $\%_{TD_A} - \%_{TD_{PL}} = +20\%$, only the Shared gamma frailty model shows an acceptable type I error rate with a bias of -0.02. Its power reaches 0.920 with a bias of -0.04. For $\%_{TD_{PL}} = 50\%$, the scenario with $\%_{TD_A} - \%_{TD_{PL}} = 0\%$ is the one with the largest number of models that return an acceptable the type I error rate. Hereby, again the Likelihood ratio test, the Gamma frailty model, the Shared gamma frailty model and the χ^2-test show a correct type I error rate with zero bias. The Shared gamma frailty model again provides the largest power of 0.983 without any bias. For the scenarios with $\%_{TD_A} - \%_{TD_{PL}} = -20\%$, again the Log-linear Poisson model and the Shared gamma frailty model are the unique correct models with an type I error rate of 0.048 and a bias of 0.15 for the Log-linear Poisson model and 0.045 with a bias of -0.01 for the Shared gamma frailty model. The power of the Log-linear Poisson model in this scenario is 0.936 and the one of the Shared gamma frailty model represents 0.934. These values are both a bit lower than in the case of $\%_{TD_{PL}} = 20\%$. Their corresponding bias reveals 0.15 and -0.02. If $\%_{TD_A} - \%_{TD_{PL}} = +20\%$, now even for two models an acceptable type I error rate is calculated. These are the Log-linear Poisson model with a type I error rate of 0.044 with a bias of -0.22 and the Shared gamma frailty model with an type I error rate of 0.043 and a bias of -0.04. The last one provides the larger power of 0.820 and the smaller bias of -0.08.

Now, the results of the scenarios with $\phi = 2$ are presented. Hereby, much fewer models show an acceptable type I error rate. Instead, the type I error rate increases, if overdispersion occurs. This is especially predominant for the Log-linear Poisson model, where the type I error rate rises heavily compared to the scenarios without overdispersion. If $\%_{TD_{PL}} = 20\%$, $\%_{TD_A} - \%_{TD_{PL}} = 0\%$, the χ^2-test returns a type I error rate of 0.042 with a power of 0.871. The according bias reveals a value of 0.04. For the other values of difference of treatment discontinuation between both trial arms, the Log-linear negative-binomial model and the Shared gamma frailty model provide a type I error rate which is smaller than 5%, if $\%_{TD_{PL}} = 20\%$ and even if $\%_{TD_{PL}} = 50\%$. These models give exactly the same values for the type I error rate and power. But the bias of the Shared gamma frailty model is much smaller than the one of the Log-linear negative-binomial model. This is true for all scenarios under MCAR for the hypothetical estimand with any dif-

ference between the proportions of patients with treatment discontinuation in both trial arms. Comparing the Log-linear negative-binomial model with the Log-linear Poisson model, it can be observed, that both models yield the same bias but different values for the type I error rate and power. This aspect is valid for scenarios with overdispersion either for the hypothetical estimand or for the treatment-policy estimand. Finally, the scenario with $\%_{TD_{PL}} = 50\%$ and $\%_{TD_A} - \%_{TD_{PL}} = 0\%$ is presented. The Log-linear negative-binomial model, the Log rank test, the Cox proportional hazards model, the Gamma frailty model, the Shared gamma frailty model and the χ^2-test provide a type I error rate smaller than 5%, each with a bias of zero. The Log-linear negative-binomial model and the Shared gamma frailty model reach the highest power of 0.854 without any bias or a very small one respectively.

As already mentioned, in figure 10, bias of all methods can be easily compared for each scenario. The bias of the Log-linear negative-binomial model is not plotted for scenarios without overdispersion. This is reasoned by the high number of not converged models. In the case of interpretable results of the Log-linear negative-binomial model, the bias of the Log-linear negative-binomial model is the same as the one of the Log-linear Poisson model. In each scenario without treatment effect, all methods estimated the true treatment effect without any bias, if $\%_{TD_A} - \%_{TD_{PL}} = 0\%$. But if $\phi = 1$, $\%_{TD_{PL}} = 20\%$ and $\%_{TD_A} - \%_{TD_{PL}} = -20\%$, only the Shared gamma frailty model provides a bias of zero. The Log-linear Poisson model and the χ^2-test underestimate the true treatment effect. Whereas the Cox proportional hazards model, the Likelihood ratio test and the Gamma frailty model estimate overestimate the treatment effect. If $\%_{TD_A} - \%_{TD_{PL}} = +20\%$, the direction of the bias of each method is reversed. The Shared gamma frailty model is no longer zero but it slightly overestimates the hypothetical estimand. The χ^2-test gives the largest absolute value of the bias which is also valid for all other scenarios without treatment effect under the hypothetical estimand. If $\phi = 1$, $\%_{TD_{PL}} = 50\%$ and $\%_{TD_A} - \%_{TD_{PL}} = -20\%$, all models but the Cox proportional hazards model and the Shared gamma frailty model keep the direction of the bias, compared to $\phi = 1$, $\%_{TD_{PL}} = 20\%$ and $\%_{TD_A} - \%_{TD_{PL}} = -20\%$. Now, the bias of the Cox proportional hazards model is zero and the Shared gamma frailty model slightly overestimates the true treatment effect with a bias of -0.01. For the combination of $\phi = 1$, $\%_{TD_{PL}} = 50\%$ and $\%_{TD_A} - \%_{TD_{PL}} = +20\%$, the models keep the direction of their bias as described for the scenario under $\phi = 1$, $\%_{TD_{PL}} = 20\%$ and $\%_{TD_A} - \%_{TD_{PL}} = +20\%$, but the absolute value of the bias increases now.

The bias of the scenarios under $\phi = 2$, $\%_{TD_{PL}} = 20\%$ behaves like the bias of the corresponding scenarios without overdispersion. For $\phi = 2$, $\%_{TD_{PL}} = 50\%$ and $\%_{TD_A} -$

$\%_{TD_{PL}} = -20\%$, the bias of the Cox proportional hazards model is again zero, as it was in the corresponding scenario without overdispersion. The remaining models also provide a bias larger than zero. In the case of $\phi = 2$, $\%_{TD_{PL}} = 50\%$ and $\%_{TD_A} - \%_{TD_{PL}} = +20\%$, the direction of the bias stays in the same direction compared to the according scenario without overdispersion.

Describing the bias of scenarios with treatment effect, the Log-linear Poisson model and the Log-linear negative-binomial model (if $\phi = 2$) estimate the true treatment effect if there is no difference in the percentage of patients with treatment discontinuation between both trials arms. For scenarios with and without overdispersion, the Shared gamma frailty model additionally estimates a bias of zero or close to zero, if $\%_{TD_A} - \%_{TD_{PL}} = 0\%$. If $\%_{TD_A} - \%_{TD_{PL}} = -20\%$ without overdispersion and $\%_{TD_{PL}} = 20\%$, the Log-linear Poisson model and the χ^2-test again provide a bias larger zero. The Cox proportional hazards model, the Likelihood ratio test and the Gamma frailty model overestimate the true treatment effect. If $\%_{TD_{PL}} = 50\%$ and $\%_{TD_A} - \%_{TD_{PL}} = -20\%$ in the scenario without overdispersion, all models except the Shared gamma frailty model have a positive bias.

For the scenarios with overdispersion, $\%_{TD_{PL}} = 20\%$ and $\%_{TD_A} - \%_{TD_{PL}} = -20\%$, the Log-linear negative-binomial model, the Log-linear Poisson model and the χ^2-test underestimate the hypothetical estimand. The remaining models give a negative bias except the Shared gamma frailty model. It returns an unbiased estimate. But if $\phi = 2$, $\%_{TD_{PL}} = 50\%$ and $\%_{TD_A} - \%_{TD_{PL}} = -20\%$, only the Likelihood ratio test, Gamma frailty model and Shared gamma frailty model underestimate the true treatment effect. In analogy to the scenarios without treatment effect, the bias of all scenarios with treatment effect and with $\%_{TD_A} - \%_{TD_{PL}} = +20\%$ shows the reverse direction compared to $\%_{TD_A} - \%_{TD_{PL}} = -20\%$. Except if $\phi = 1$ and $\%_{TD_{PL}} = 50\%$, the direction of the bias of the Cox proportional hazards model, the Likelihood ratio test, the Gamma frailty model and the one of the Shared gamma frailty model is not turned. If $\phi = 2$ and $\%_{TD_{PL}} = 50\%$, the Cox proportional hazards model and the Shared gamma frailty model do not change the direction of their bias between the negative and positive difference of treatment discontinuation in both trial arms. In general, the bias for $\%_{TD_A} - \%_{TD_{PL}} = +20\%$ conducts absolutely larger values for most of the models.

Table 7: Comparison of methods estimating the hypothetical estimand for scenarios with MCAR, $\phi = 1$ and $\%_{TD_{PL}} = 20\%$

		$\%_{TD_{PL}} = 20\%$								
		$\%_{TD_A} - \%_{TD_{PL}} = -20\%$			$\%_{TD_A} - \%_{TD_{PL}} = 0\%$			$\%_{TD_A} - \%_{TD_{PL}} = +20\%$		
		type I error rate [CI]	bias	N_{conv}	type I error rate [CI]	bias	N_{conv}	type I error rate [CI]	bias	N_{conv}
no tr. effect	GLM NB	0.039 [0.034;0.044]	0.11	5783	0.045 [0.04;0.05]	0.00	5800	0.046 [0.041;0.051]	-0.14	5846
	GLM Poisson	0.049 [0.045;0.053]	0.11	10000	0.051 [0.047;0.055]	0.00	10000	0.051 [0.047;0.055]	-0.14	10000
	Log rank test	0.066 [0.061;0.071]	-	10000	0.051 [0.047;0.055]	-	10000	0.314 [0.305;0.323]	-	10000
	Cox model	0.065 [0.06;0.07]	-0.05	10000	0.051 [0.047;0.055]	0.00	10000	0.313 [0.304;0.322]	0.13	10000
	LR test	0.064 [0.059;0.069]	-0.04	10000	0.047 [0.043;0.051]	0.00	10000	0.237 [0.229;0.245]	0.08	10000
	GF model	0.060 [0.055;0.065]	-0.04	9999	0.044 [0.04;0.048]	0.00	10000	0.172 [0.165;0.179]	0.07	9778
	SGF model	0.045 [0.041;0.049]	0.00	9982	0.047 [0.043;0.051]	0.00	10000	0.049 [0.045;0.053]	-0.02	9985
	χ^2-test	0.179 [0.171;0.187]	0.27	10000	0.039 [0.035;0.043]	0.00	10000	0.389 [0.379;0.399]	-0.34	10000
		power	bias	N_{conv}	power	bias	N_{conv}	power	bias	N_{conv}
with tr. effect	GLM NB	0.996	0.11	5971	0.995	0.00	5936	0.923	-0.14	5877
	GLM Poisson	0.996	0.11	10000	0.995	0.00	10000	0.920	-0.14	10000
	Log rank test	0.988	-	10000	0.948	-	10000	0.837	-	10000
	Cox model	0.988	-0.05	10000	0.948	0.04	10000	0.837	0.26	10000
	LR test	0.988	-0.05	10000	0.946	0.05	10000	0.805	0.19	10000
	GF model	0.986	-0.05	10000	0.942	0.04	10000	0.807	0.14	9225
	SGF model	0.996	0.00	10000	0.994	0.00	9999	0.920	-0.04	9999
	χ^2-test	0.697	0.27	10000	0.917	0.08	10000	0.818	-0.27	10000

Table 8: Comparison of methods estimating the hypothetical estimand for scenarios with MCAR, $\phi = 1$ and $\%_{TD_{PL}} = 50\%$

		$\%_{TD_{PL}} = 50\%$								
		$\%_{TD_A} - \%_{TD_{PL}} = -20\%$			$\%_{TD_A} - \%_{TD_{PL}} = 0\%$			$\%_{TD_A} - \%_{TD_{PL}} = +20\%$		
		type I error rate [CI]	bias	N_{conv}	type I error rate [CI]	bias	N_{conv}	type I error rate [CI]	bias	N_{conv}
no tr. effect	GLM NB	0.044 [0.039;0.049]	0.14	5857	0.047 [0.042;0.052]	0.00	6008	0.042 [0.037;0.047]	-0.22	5865
	GLM Poisson	0.048 [0.044;0.052]	0.15	10000	0.052 [0.048;0.056]	0.00	10000	0.044 [0.04;0.048]	-0.22	10000
	Log rank test	0.303 [0.294;0.312]	-	10000	0.053 [0.049;0.057]	-	10000	0.519 [0.509;0.529]	-	10000
	Cox model	0.302 [0.293;0.311]	0.00	10000	0.052 [0.048;0.056]	0.00	10000	0.518 [0.508;0.528]	0.21	10000
	LR test	0.224 [0.216;0.232]	-0.03	10000	0.042 [0.038;0.046]	0.00	10000	0.359 [0.35;0.368]	0.12	10000
	GF model	0.191 [0.183;0.199]	-0.04	9802	0.04 [0.036;0.044]	0.00	9998	0.286 [0.277;0.295]	0.11	9491
	SGF model	0.045 [0.041;0.049]	-0.01	9993	0.048 [0.044;0.052]	0.00	9995	0.043 [0.039;0.047]	-0.04	9994
	χ^2-test	0.769 [0.761;0.777]	0.25	10000	0.043 [0.039;0.047]	0.00	10000	0.736 [0.727;0.745]	-0.50	10000
		power	bias	N_{conv}	power	bias	N_{conv}	power	bias	N_{conv}
with tr. effect	GLM NB	0.933	0.15	5963	0.981	0.00	6038	0.819	-0.22	5858
	GLM Poisson	0.936	0.15	10000	0.984	0.00	10000	0.822	-0.22	10000
	Log rank test	0.863	-	10000	0.876	-	10000	0.713	-	10000
	Cox model	0.863	0.08	10000	0.874	0.10	10000	0.712	0.45	10000
	LR test	0.858	0.05	10000	0.858	0.13	10000	0.668	0.33	10000
	GF model	0.880	0.01	9449	0.847	0.13	9996	0.707	0.27	8544
	SGF model	0.934	-0.02	9995	0.983	0.00	10000	0.820	-0.08	9992
	χ^2-test	0.325	0.30	10000	0.837	0.17	10000	0.642	-0.36	10000

Table 9: Comparison of methods estimating the hypothetical estimand for scenarios with MCAR, $\phi = 2$, and $\%_{TD_{PL}} = 20\%$

		$\%_{TD_{PL}} = 20\%$								
		$\%_{TD_A} - \%_{TD_{PL}} = -20\%$			$\%_{TD_A} - \%_{TD_{PL}} = 0\%$			$\%_{TD_A} - \%_{TD_{PL}} = +20\%$		
		type I error rate [CI]	bias	N_{conv}	type I error rate [CI]	bias	N_{conv}	type I error rate [CI]	bias	N_{conv}
no tr. effect	GLM NB	0.05 [0.046;0.054]	0.11	10000	0.051 [0.047;0.055]	0.00	10000	0.049 [0.045;0.053]	-0.15	10000
	GLM Poisson	0.163 [0.156;0.17]	0.11	10000	0.16 [0.153;0.167]	0.00	10000	0.142 [0.135;0.149]	-0.15	10000
	Log rank test	0.077 [0.072;0.082]	-	10000	0.052 [0.048;0.056]	-	10000	0.299 [0.29;0.308]	-	10000
	Cox model	0.077 [0.072;0.082]	-0.06	10000	0.052 [0.048;0.056]	0.00	10000	0.298 [0.289;0.307]	0.16	10000
	LR test	0.093 [0.087;0.099]	-0.07	10000	0.061 [0.056;0.066]	0.00	10000	0.286 [0.277;0.295]	0.12	10000
	GF model	0.072 [0.067;0.077]	-0.06	10000	0.052 [0.048;0.056]	0.00	10000	0.211 [0.203;0.219]	0.10	9649
	SGF model	0.05 [0.046;0.054]	0.00	10000	0.051 [0.047;0.055]	0.00	10000	0.049 [0.045;0.053]	-0.03	10000
	χ^2-test	0.11 [0.104;0.116]	0.18	10000	0.042 [0.038;0.046]	0.00	10000	0.322 [0.313;0.331]	-0.30	10000
		power	bias	N_{conv}	power	bias	N_{conv}	power	bias	N_{conv}
with tr. effect	GLM NB	0.915	0.11	10000	0.903	0.00	10000	0.836	-0.14	10000
	GLM Poisson	0.971	0.11	10000	0.968	0.00	10000	0.897	-0.14	10000
	Log rank test	0.961	-	10000	0.890	-	10000	0.831	-	10000
	Cox model	0.960	-0.06	10000	0.889	0.04	10000	0.830	0.28	10000
	LR test	0.970	-0.09	10000	0.898	0.03	10000	0.815	0.20	10000
	GF model	0.953	-0.18	10000	0.884	-0.05	10000	0.828	0.09	9472
	SGF model	0.915	0.00	10000	0.903	0.00	10000	0.836	-0.04	10000
	χ^2-test	0.692	0.18	10000	0.871	0.04	10000	0.806	-0.25	10000

Table 10: Comparison of methods estimating the hypothetical estimand for scenarios with MCAR, $\phi = 2$, and $\%_{TD_{PL}} = 50\%$

		$\%_{TD_{PL}} = 50\%$								
		$\%_{TD_A} - \%_{TD_{PL}} = -20\%$			$\%_{TD_A} - \%_{TD_{PL}} = 0\%$			$\%_{TD_A} - \%_{TD_{PL}} = +20\%$		
		type I error rate [CI]	bias	N_{conv}	type I error rate [CI]	bias	N_{conv}	type I error rate [CI]	bias	N_{conv}
no tr. effect	GLM NB	0.05 [0.046;0.054]	0.15	10000	0.05 [0.046;0.054]	0.00	10000	0.047 [0.043;0.051]	-0.22	10000
	GLM Poisson	0.139 [0.132;0.146]	0.15	10000	0.15 [0.143;0.157]	0.00	10000	0.117 [0.111;0.123]	-0.22	10000
	Log rank test	0.364 [0.355;0.373]	-	10000	0.049 [0.045;0.053]	-	10000	0.513 [0.503;0.523]	-	10000
	Cox model	0.362 [0.353;0.371]	0.00	10000	0.048 [0.044;0.052]	0.00	10000	0.511 [0.501;0.521]	0.25	9999
	LR test	0.356 [0.347;0.365]	-0.05	10000	0.051 [0.047;0.055]	0.00	10000	0.455 [0.445;0.465]	0.17	9999
	GF model	0.29 [0.281;0.299]	-0.05	9726	0.044 [0.04;0.048]	0.00	10000	0.362 [0.352;0.372]	0.14	9300
	SGF model	0.05 [0.046;0.054]	-0.01	10000	0.049 [0.045;0.053]	0.00	10000	0.047 [0.043;0.051]	-0.05	10000
	χ^2-test	0.558 [0.548;0.568]	0.14	10000	0.038 [0.034;0.042]	0.00	10000	0.599 [0.589;0.609]	-0.45	10000
		power	bias	N_{conv}	power	bias	N_{conv}	power	bias	N_{conv}
with tr. effect	GLM NB	0.834	0.15	10000	0.854	0.00	10000	0.723	-0.22	10000
	GLM Poisson	0.906	0.15	10000	0.941	0.00	10000	0.794	-0.22	10000
	Log rank test	0.861	-	10000	0.794	-	10000	0.734	-	10000
	Cox model	0.860	0.08	10000	0.792	0.11	10000	0.733	0.47	10000
	LR test	0.862	0.01	10000	0.794	0.10	10000	0.697	0.35	9999
	GF model	0.874	-0.07	9578	0.778	0.07	9997	0.741	0.26	8871
	SGF model	0.835	-0.03	10000	0.854	-0.01	10000	0.724	-0.08	10000
	χ^2-test	0.376	0.18	10000	0.783	0.09	10000	0.661	-0.36	10000

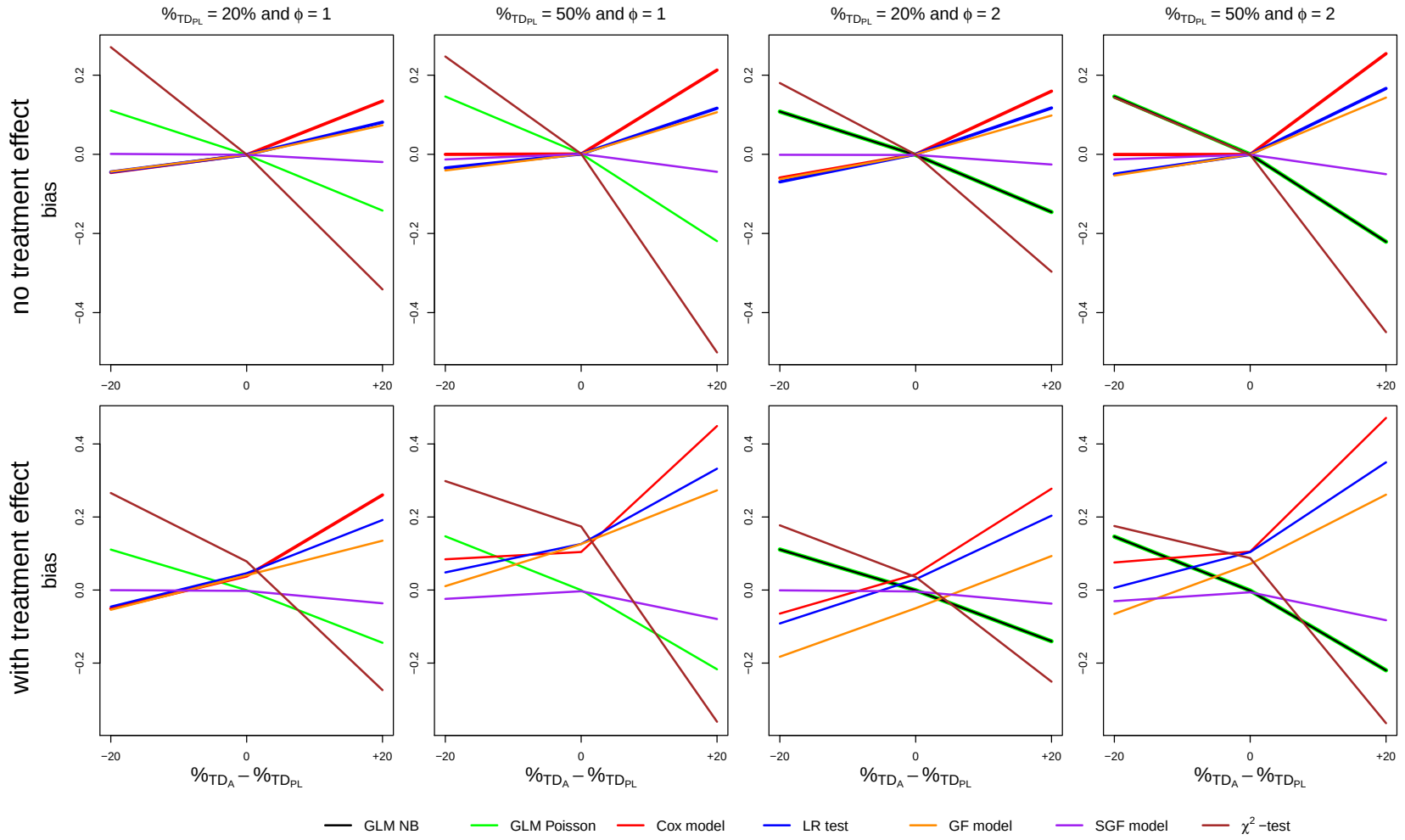

Figure 10: Bias of the Log-linear negative-binomial model (GLM NB), Log-linear Poisson model (GLM Poisson), Cox proportional hazards model (Cox model), Likelihood ratio test (LR test), Gamma frailty model (GF model), Shared gamma frailty model (SGF model) and χ^2-test for hypothetical estimand depending on the percentage of patients with treatment discontinuation in placebo group ($\%_{TD_{PL}}$), the difference of patients with treatment discontinuation between both trial arms ($\%_{TD_A} - \%_{TD_{PL}}$) and ϕ under MCAR

7.2.2 Results for the hypothetical estimand with treatment discontinuation mechanism MNAR

Now, the results for the hypothetical estimand with treatment discontinuation mechanism MNAR are presented. As already described in section 7.1, the values of the percentage received after the simulation process with MNAR differ from those under the MCAR mechanism. In table 11 on page 77, the results for scenarios without overdispersion and $\%_{TD_{PL}} = 27\%$ are listed. The evaluated values for scenarios without overdispersion and $\%_{TD_{PL}} = 60\%$ are tabulated in table 12 on page 78. Table 13 on page 79 summarizes the results for the scenarios with $\phi = 2$ and $\%_{TD_{PL}} = 28\%$ and table 14 on page 80 contains the results for the scenarios with $\phi = 2$ and $\%_{TD_{PL}} = 60\%$. Figure 11 on page 81 visualizes the bias for each method under the MNAR mechanism for the hypothetical estimand.

Under the treatment discontinuation mechanism MNAR, the tested methods yield an acceptable type I error rate of lower than 5% only if $\%_{TD_A} - \%_{TD_{PL}} = 0\%$. There exists one single exception if $\phi = 1$, $\%_{TD_{PL}} = 27\%$ and $\%_{TD_A} - \%_{TD_{PL}} = -27\%$. In this scenario, the Gamma frailty model returns an acceptable type I error rate of 0.048 with a bias of -0.03. Its power conducts 0.980 with a bias of -0.04. Now, the description of scenarios without any difference between the percentage of patients with treatment discontinuation in both trial arms follows. For the scenarios with $\phi = 1$, $\%_{TD_{PL}} = 27\%$ and $\%_{TD_A} - \%_{TD_{PL}} = 0\%$, all interpretable methods show a correct type I error rate, each with zero bias. Hereby, the Log-linear Poisson model and the Shared gamma frailty model reach the highest and same value for the power of 0.988 with a bias of 0.05 and 0.02 respectively. If $\%_{TD_A} - \%_{TD_{PL}} = +23\%$, the methods show very high type I error rates. This phenomenon can also be observed if $\phi = 1$ and $\%_{TD_{PL}} = 60\%$. Hereby, these high type I error rates appear both under $\%_{TD_A} - \%_{TD_{PL}} = -26\%$ and $\%_{TD_A} - \%_{TD_{PL}} = +14\%$. Only for $\%_{TD_A} - \%_{TD_{PL}} = 0\%$, the Log-linear Poisson model, the Likelihood ratio test, the Gamma frailty model, the Shared gamma frailty model and the χ^2-test lead to a wrong test decision under scenarios with treatment effect in less than 5%, each with a bias of zero. In analogy to $\phi = 1$, $\%_{TD_{PL}} = 27\%$ and $\%_{TD_A} - \%_{TD_{PL}} = 0\%$, now with $\%_{TD_{PL}} = 60\%$, the Log-linear Poisson model and the Shared gamma frailty model have the highest power of 0.921 with a bias of 0.12 and 0.05 respectively. Continuing with the scenarios based on $\phi = 2$, $\%_{TD_{PL}} = 28\%$ and $\%_{TD_A} - \%_{TD_{PL}} = 0\%$, the Log rank test, the Cox proportional hazards model, the Gamma frailty model and the χ^2-test show a type I error rate, which is lower than 5%. Hereby, the bias of these methods is zero again. The Log rank test possesses the largest power of these methods with a value of 0.871. For the scenarios with $\phi = 2$, $\%_{TD_{PL}} = 60\%$ and $\%_{TD_A} - \%_{TD_{PL}} = 0\%$, the Log rank test, the Cox proportional

hazards model, the Likelihood ratio test, the Gamma frailty model and the χ^2-test yield to acceptable type I error rates. Their bias is zero again. The Log rank test shows the largest power of these models with a value of 0.741. For the other differences between the percentages of patients with treatment discontinuation between treatment and placebo arm under $\phi = 2$, the methods again reach very high type I error rates. The discrepancies between the Log rank test and the Cox proportional hazards model are due to different implementations in the statistical software.

Figure 11 is referred, where the bias of all methods for the scenarios under the hypothetical estimand with treatment discontinuation mechanism MNAR are plotted. The bias of all methods in scenarios without treatment effect and no difference between the percentages of treatment discontinuation in treatment and placebo group equals zero. This was already the case for the scenarios under the MCAR mechanism. Concerning still the scenarios without treatment effect, the scenario with $\phi = 1$, $\%_{TD_{PL}} = 27\%$ and $\%_{TD_A} - \%_{TD_{PL}} = -27\%$, the models except the Shared gamma frailty model behave like in the according scenario under MCAR. The Shared gamma frailty model underestimates the true treatment effect under MNAR. This is comparable to the scenario with $\phi = 1$, $\%_{TD_{PL}} = 27\%$ and $\%_{TD_A} - \%_{TD_{PL}} = +23\%$. Here, the Shared gamma frailty model provides a bias larger than zero, too.

In the case of $\phi = 1$, $\%_{TD_{PL}} = 60\%$ and $\%_{TD_A} - \%_{TD_{PL}} = -21\%$, all methods yield a positive bias. The Log-linear Poisson model and the χ^2-test show the largest absolute values. The bias of these two methods is reversed under the scenario $\phi = 1$, $\%_{TD_{PL}} = 60\%$ and $\%_{TD_A} - \%_{TD_{PL}} = +18\%$. The remaining models keep the direction of their bias but their absolute value increases.

In the scenario $\phi = 2$, $\%_{TD_{PL}} = 28\%$ and $\%_{TD_A} - \%_{TD_{PL}} = -28\%$, the same constellation as in the corresponding scenario without overdispersion can be observed. The same is valid for $\phi = 2$, $\%_{TD_{PL}} = 28\%$ and $\%_{TD_A} - \%_{TD_{PL}} = +22\%$. The Shared gamma frailty model now keeps the positive direction of its bias. The bias of the Log-linear negative-binomial model is again comparable to the one of the Log-linear Poisson model under the MNAR mechanism as it was already the case under MCAR. The scenarios with $\phi = 2$, $\%_{TD_{PL}} = 60\%$ and $\%_{TD_A} - \%_{TD_{PL}} = -21\%$ and $\%_{TD_A} - \%_{TD_{PL}} = +18\%$ respectively are considered next. They are absolutely comparable to the according one without overdispersion. But in the scenario $\phi = 2$, $\%_{TD_{PL}} = 60\%$ and $\%_{TD_A} - \%_{TD_{PL}} = +18\%$, the Log-linear negative-binomial model and the Log-linear Poisson model provide a positive bias.

Now the bias of the scenarios with treatment effect is described. In general, these values

are comparable to the appropriate scenarios under MCAR. The bias of the Shared gamma frailty model is smaller than the one of the count data models, if they do not deliver an unbiased estimate. In the case of $\phi = 1$, $\%_{TD_{PL}} = 27\%$ and $\%_{TD_A} - \%_{TD_{PL}} = -27\%$ and $\%_{TD_A} - \%_{TD_{PL}} = +18\%$ respectively, the Shared gamma frailty model does not longer show a bias of zero respectively a negative bias but a positive one. This is also true for $\%_{TD_A} - \%_{TD_{PL}} = -4\%$, where additionally the count data models now provide a bias larger than zero. In the next scenarios, which are based on $\phi = 1$, $\%_{TD_{PL}} = 60\%$ and $\%_{TD_A} - \%_{TD_{PL}} = -26\%$ and $\%_{TD_A} - \%_{TD_{PL}} = +14\%$ respectively, the Shared gamma frailty model again shows a positive bias.

This is also valid for the scenario $\phi = 2$, $\%_{TD_{PL}} = 28\%$, $\%_{TD_A} - \%_{TD_{PL}} = -28\%$ and in the scenarios with $\phi = 2$, $\%_{TD_{PL}} = 28\%$, $\%_{TD_A} - \%_{TD_{PL}} = +17\%$ additionally the Log-linear Poisson model underestimates the true treatment effect. The last discrepancy between the bias of the scenarios under MCAR and MNAR lies in the scenarios $\phi = 2$, $\%_{TD_{PL}} = 60\%$ and $\%_{TD_A} - \%_{TD_{PL}} = -25\%$ and $\phi = 2$, $\%_{TD_{PL}} = 60\%$, $\%_{TD_A} - \%_{TD_{PL}} = +15\%$. In these scenarios both count data models and the Shared gamma frailty model underestimate the true treatment effect.

Table 11: Comparison of methods estimating the hypothetical estimand for scenarios with MNAR, $\phi = 1$, and $\%_{TD_{PL}} = 27\%$

		$\%_{TD_{PL}} = 27\%$								
		$\%_{TD_A} - \%_{TD_{PL}} = -27\%$			$\%_{TD_A} - \%_{TD_{PL}} = 0\%$			$\%_{TD_A} - \%_{TD_{PL}} = +23\%$		
		type I error rate [CI]	bias	N_{conv}	type I error rate [CI]	bias	N_{conv}	type I error rate [CI]	bias	N_{conv}
no tr. effect	GLM NB	0.089 [0.081;0.097]	0.23	5035	0.037 [0.031;0.043]	0.00	4349	0.08 [0.072;0.088]	-0.10	5000
	GLM Poisson	0.098 [0.092;0.104]	0.23	10000	0.044 [0.04;0.048]	0.00	10000	0.085 [0.08;0.09]	-0.11	10000
	Log rank test	0.055 [0.051;0.059]	-	10000	0.049 [0.045;0.053]	-	10000	0.391 [0.381;0.401]	-	10000
	Cox model	0.054 [0.05;0.058]	-0.03	10000	0.048 [0.044;0.052]	0.00	10000	0.391 [0.381;0.401]	0.21	10000
	LR test	0.051 [0.047;0.055]	-0.03	10000	0.046 [0.042;0.05]	0.00	10000	0.300 [0.291;0.309]	0.14	10000
	GF model	0.048 [0.044;0.052]	-0.03	10000	0.043 [0.039;0.047]	0.00	10000	0.212 [0.204;0.22]	0.13	9642
	SGF model	0.093 [0.087;0.099]	0.07	9995	0.043 [0.039;0.047]	0.00	9999	0.081 [0.076;0.086]	0.04	9996
	χ^2-test	0.329 [0.32;0.338]	0.38	10000	0.039 [0.035;0.043]	0.00	10000	0.507 [0.497;0.517]	-0.41	10000
		$\%_{TD_A} - \%_{TD_{PL}} = -27\%$			$\%_{TD_A} - \%_{TD_{PL}} = -4\%$			$\%_{TD_A} - \%_{TD_{PL}} = +18\%$		
		power	bias	N_{conv}	power	bias	N_{conv}	power	bias	N_{conv}
with tr. effect	GLM NB	0.983	0.23	5015	0.987	0.05	4446	0.869	-0.04	4995
	GLM Poisson	0.980	0.23	10000	0.988	0.05	10000	0.867	-0.06	10000
	Log rank test	0.982	-	10000	0.940	-	10000	0.815	-	10000
	Cox model	0.982	-0.03	10000	0.939	0.05	10000	0.814	0.32	10000
	LR test	0.981	-0.03	10000	0.937	0.05	10000	0.780	0.24	10000
	GF model	0.980	-0.04	9999	0.932	0.05	9999	0.785	0.18	9048
	SGF model	0.980	0.07	9999	0.988	0.02	10000	0.867	0.04	9997
	χ^2-test	0.529	0.37	10000	0.864	0.14	10000	0.718	-0.22	10000

Table 12: Comparison of methods estimating the hypothetical estimand for scenarios with MNAR, $\phi = 1$, and $\%_{TD_{PL}} = 60\%$

		$\%_{TD_{PL}} = 60\%$								
		$\%_{TD_A} - \%_{TD_{PL}} = -21\%$			$\%_{TD_A} - \%_{TD_{PL}} = 0\%$			$\%_{TD_A} - \%_{TD_{PL}} = +18\%$		
		type I error rate [CI]	bias	N_{conv}	type I error rate [CI]	bias	N_{conv}	type I error rate [CI]	bias	N_{conv}
no tr. effect	GLM NB	0.247 [0.234;0.26]	0.40	4467	0.037 [0.031;0.043]	0.00	3465	0.193 [0.181;0.205]	-0.03	4318
	GLM Poisson	0.252 [0.243;0.261]	0.37	10000	0.041 [0.037;0.045]	0.00	10000	0.204 [0.196;0.212]	-0.08	10000
	Log rank test	0.327 [0.318;0.336]	-	10000	0.053 [0.049;0.057]	-	10000	0.622 [0.612;0.632]	-	10000
	Cox model	0.325 [0.316;0.334]	0.08	10000	0.053 [0.049;0.057]	0.00	10000	0.621 [0.611;0.631]	0.32	10000
	LR test	0.227 [0.219;0.235]	0.03	10000	0.041 [0.037;0.045]	0.00	10000	0.451 [0.441;0.461]	0.20	10000
	GF model	0.191 [0.183;0.199]	0.02	9693	0.038 [0.034;0.042]	0.00	10000	0.379 [0.369;0.389]	0.19	9244
	SGF model	0.244 [0.236;0.252]	0.15	9991	0.041 [0.037;0.045]	0.00	9997	0.199 [0.191;0.207]	0.11	9983
	χ^2-test	0.858 [0.851;0.865]	0.34	10000	0.043 [0.039;0.047]	0.00	10000	0.782 [0.774;0.79]	-0.54	10000
		$\%_{TD_A} - \%_{TD_{PL}} = -26\%$			$\%_{TD_A} - \%_{TD_{PL}} = -6\%$			$\%_{TD_A} - \%_{TD_{PL}} = +14\%$		
		power	bias	N_{conv}	power	bias	N_{conv}	power	bias	N_{conv}
with tr. effect	GLM NB	0.737	0.43	4138	0.925	0.11	3292	0.576	0.02	4066
	GLM Poisson	0.721	0.41	10000	0.921	0.12	10000	0.558	-0.00	10000
	Log rank test	0.854	-	10000	0.839	-	10000	0.724	-	10000
	Cox model	0.853	0.15	10000	0.837	0.12	10000	0.723	0.55	10000
	LR test	0.836	0.10	10000	0.811	0.15	10000	0.641	0.41	10000
	GF model	0.874	0.06	9283	0.801	0.15	9992	0.718	0.36	8303
	SGF model	0.719	0.15	9991	0.921	0.05	9997	0.557	0.10	9988
	χ^2-test	0.282	0.49	10000	0.630	0.30	10000	0.615	-0.24	10000

Table 13: Comparison of methods estimating the hypothetical estimand for scenarios with MNAR, $\phi = 2$, and $\%_{TD_{PL}} = 28\%$

		$\%_{TD_{PL}} = 28\%$								
		$\%_{TD_A} - \%_{TD_{PL}} = -28\%$			$\%_{TD_A} - \%_{TD_{PL}} = 0\%$			$\%_{TD_A} - \%_{TD_{PL}} = +22\%$		
		type I error rate [CI]	bias	N_{conv}	type I error rate [CI]	bias	N_{conv}	type I error rate [CI]	bias	N_{conv}
no tr. effect	GLM NB	0.122 [0.116;0.128]	0.28	10000	0.055 [0.051;0.059]	0.00	10000	0.113 [0.107;0.119]	-0.07	9999
	GLM Poisson	0.315 [0.306;0.324]	0.31	10000	0.121 [0.115;0.127]	0.00	10000	0.265 [0.256;0.274]	-0.04	10000
	Log rank test	0.073 [0.068;0.078]	-	10000	0.045 [0.041;0.049]	-	10000	0.375 [0.366;0.384]	-	10000
	Cox model	0.072 [0.067;0.077]	-0.03	10000	0.045 [0.041;0.049]	0.00	10000	0.374 [0.365;0.383]	0.26	9999
	LR test	0.086 [0.081;0.091]	-0.04	10000	0.054 [0.05;0.058]	0.00	10000	0.362 [0.353;0.371]	0.21	9999
	GF model	0.067 [0.062;0.072]	-0.03	10000	0.045 [0.041;0.049]	0.00	10000	0.271 [0.262;0.28]	0.19	9517
	SGF model	0.12 [0.114;0.126]	0.11	10000	0.052 [0.048;0.056]	0.00	10000	0.11 [0.104;0.116]	0.08	10000
	χ^2-test	0.229 [0.221;0.237]	0.29	10000	0.042 [0.038;0.046]	0.00	10000	0.434 [0.424;0.444]	-0.35	10000
		$\%_{TD_A} - \%_{TD_{PL}} = -28\%$			$\%_{TD_A} - \%_{TD_{PL}} = -4\%$			$\%_{TD_A} - \%_{TD_{PL}} = +17\%$		
		power	bias	N_{conv}	power	bias	N_{conv}	power	bias	N_{conv}
with tr. effect	GLM NB	0.751	0.27	10000	0.886	0.03	10000	0.658	-0.02	10000
	GLM Poisson	0.840	0.31	10000	0.946	0.05	10000	0.720	0.02	10000
	Log rank test	0.940	-	10000	0.871	-	10000	0.799	-	10000
	Cox model	0.939	-0.04	10000	0.869	0.05	10000	0.798	0.36	10000
	LR test	0.951	-0.07	10000	0.882	0.03	10000	0.776	0.27	10000
	GF model	0.931	-0.15	10000	0.865	-0.04	10000	0.792	0.17	9399
	SGF model	0.748	0.11	10000	0.883	0.00	10000	0.655	0.06	10000
	χ^2-test	0.514	0.28	10000	0.810	0.08	10000	0.702	-0.20	10000

Table 14: Comparison of methods estimating the hypothetical estimand for scenarios with MNAR, $\phi = 2$, and $\%_{TD_{PL}} = 60\%$

		$\%_{TD_{PL}} = 60\%$								
		$\%_{TD_A} - \%_{TD_{PL}} = -21\%$			$\%_{TD_A} - \%_{TD_{PL}} = 0\%$			$\%_{TD_A} - \%_{TD_{PL}} = +18\%$		
		type I error rate [CI]	bias	N_{conv}	type I error rate [CI]	bias	N_{conv}	type I error rate [CI]	bias	N_{conv}
no tr. effect	GLM NB	0.35 [0.341;0.359]	0.47	9966	0.058 [0.053;0.063]	0.00	9970	0.3 [0.291;0.309]	0.03	9914
	GLM Poisson	0.604 [0.594;0.614]	0.52	10000	0.086 [0.081;0.091]	0.00	10000	0.491 [0.481;0.501]	0.07	10000
	Log rank test	0.396 [0.386;0.406]	-	10000	0.049 [0.045;0.053]	-	10000	0.618 [0.608;0.628]	-	10000
	Cox model	0.394 [0.384;0.404]	0.12	9999	0.048 [0.044;0.052]	0.00	10000	0.617 [0.607;0.627]	0.41	9999
	LR test	0.361 [0.352;0.37]	0.06	9999	0.048 [0.044;0.052]	0.00	10000	0.553 [0.543;0.563]	0.30	9999
	GF model	0.308 [0.299;0.317]	0.06	9600	0.043 [0.039;0.047]	0.00	9999	0.475 [0.465;0.485]	0.29	9238
	SGF model	0.34 [0.331;0.349]	0.24	10000	0.054 [0.05;0.058]	0.00	10000	0.294 [0.285;0.303]	0.20	10000
	χ^2-test	0.768 [0.76;0.776]	0.26	10000	0.038 [0.034;0.042]	0.00	10000	0.696 [0.687;0.705]	-0.45	10000
		$\%_{TD_A} - \%_{TD_{PL}} = -25\%$			$\%_{TD_A} - \%_{TD_{PL}} = -5\%$			$\%_{TD_A} - \%_{TD_{PL}} = +15\%$		
		power	bias	N_{conv}	power	bias	N_{conv}	power	bias	N_{conv}
with tr. effect	GLM NB	0.353	0.50	9977	0.792	0.08	9971	0.286	0.10	9934
	GLM Poisson	0.365	0.56	10000	0.842	0.10	10000	0.284	0.15	10000
	Log rank test	0.811	-	10000	0.741	-	10000	0.706	-	10000
	Cox model	0.809	0.13	10000	0.738	0.08	10000	0.704	0.57	10000
	LR test	0.804	0.05	10000	0.738	0.08	10000	0.654	0.43	10000
	GF model	0.819	-0.01	9537	0.719	0.05	9999	0.707	0.36	8894
	SGF model	0.343	0.17	10000	0.782	-0.04	10000	0.278	0.12	9999
	χ^2-test	0.260	0.38	10000	0.623	0.18	10000	0.573	-0.21	10000

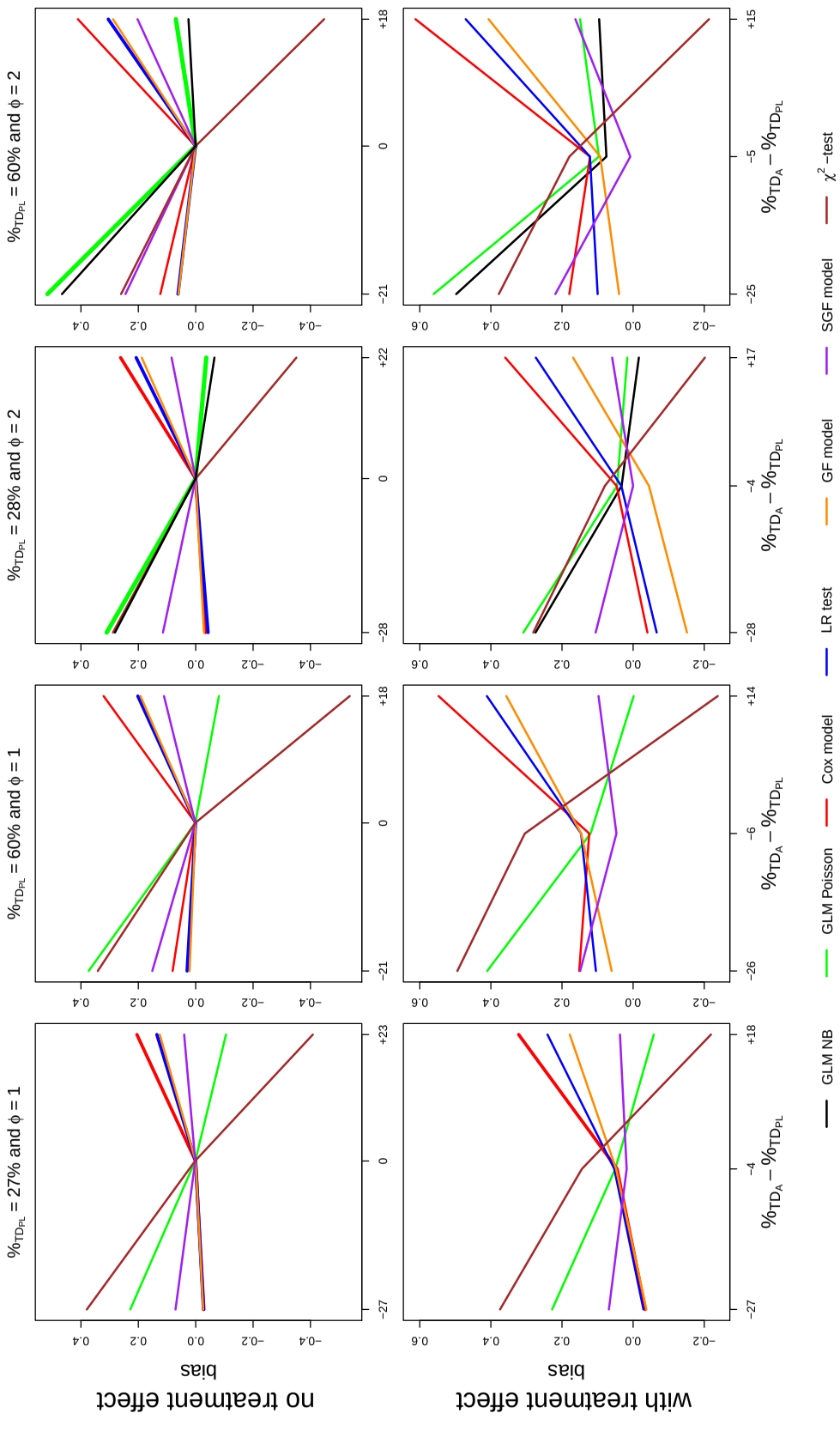

Figure 11: Bias of the Log-linear negative-binomial model (GLM NB), Log-linear Poisson model (GLM Poisson), Cox proportional hazards model (Cox model), Likelihood ratio test (LR test), Gamma frailty model (GF model), Shared gamma frailty model (SGF model) and χ^2-test for hypothetical estimand depending on the percentage of patients with treatment discontinuation in placebo group ($\%_{TD_{PL}}$), the difference of patients with treatment discontinuation between both trial arms ($\%_{TD_A} - \%_{TD_{PL}}$) and ϕ under MNAR

7.3 Results for the treatment-policy estimand

The results for the treatment-policy estimand are showed in this section. In analogy to the presentation of the results for the hypothetical estimand, the following section is divided by the two treatment discontinuation mechanisms MCAR and MNAR. Furthermore, for each mechanism the results are provided separately for each of the three percentages of study dropout. To compare the bias of the methods across the three different possibilities for the study dropout, figure 12 shows the bias for all scenarios under MCAR and figure 13 visualizes bias under MNAR at the end of section 7.3.1 and 7.3.2 respectively.

7.3.1 Results for the treatment-policy estimand with treatment discontinuation mechanism MCAR

7.3.1.1 Results for the treatment-policy estimand without study dropout with treatment discontinuation mechanism MCAR

At first, the results for the treatment-policy under MCAR without study dropout are presented. Table 15 on page 85 shows the results for the scenarios with $\phi = 1$ and $\%_{TD_{PL}} = 20\%$; in table 16 on page 86 the results for $\phi = 1$ and $\%_{TD_{PL}} = 50\%$ are listed. Furthermore, the results for scenarios with $\phi = 2$ and $\%_{TD_{PL}} = 20\%$ are tabulated on page 87 in table 17; table 18 on page 88 contains the results for the scenarios $\phi = 2$ and $\%_{TD_{PL}} = 50\%$.

In general, it must be stated, that compared to the hypothetical estimand, much fewer methods observe a type I error rate lower than 5%. This is valid for all percentages of study dropout either under MCAR or MNAR. For scenarios without any difference between the percentage of patients treatment discontinuation between both trial arms, a correct evaluation method can be found in each scenario. But if there is a difference between treatment discontinuation in treatment and placebo group, in most cases no method can hold an acceptable type I error rate. The statistical test, which provides an acceptable type I error rate the most is the χ^2-test. For the scenarios with $\phi = 1$, $\%_{TD_{PL}} = 20\%$ and $\%_{TD_A} - \%_{TD_{PL}} = 0\%$, the Log-linear Poisson model, the Gamma frailty model, the Shared gamma frailty model and the χ^2-test provide a type I error rate which is lower than 5% each with a bias of zero. Hereby, the Log-linear Poisson model and the Shared gamma frailty model have the larger power of 0.989 and 0.988 again with a bias of zero. If, $\phi = 1$ and $\%_{TD_{PL}} = 20\%$ and $\%_{TD_A} - \%_{TD_{PL}} = -20\%$, only the χ^2-test returns an acceptable type I error rate with a bias of 0.04. Its power represents 0.929 with a bias of -0.06. In the case of $\phi = 1$ and $\%_{TD_{PL}} = 20\%$ and $\%_{TD_A} - \%_{TD_{PL}} = +20\%$, no method is valid for an evaluation respecting the type I error

rate. The methods, which evaluated these three scenarios, always possess a power larger the one of the reference scenario as no study dropout occurs here. For the scenarios with $\phi = 1$ and $\%_{TD_{PL}} = 50\%$ only under $\%_{TD_A} - \%_{TD_{PL}} = 0\%$, the Shared gamma frailty model and the χ^2-test contain an acceptable type I error rate with unbiased estimate of the treatment effect. The Shared gamma frailty model reveals the larger power of 0.727 with a smaller bias of 0.03. If $\phi = 2$ with $\%_{TD_{PL}} = 20\%$, only the χ^2-test returns a type I error rate smaller than 5% under $\%_{TD_A} - \%_{TD_{PL}} = 0\%$. In the scenario with $\phi = 2$, $\%_{TD_{PL}} = 50\%$ and $\%_{TD_A} - \%_{TD_{PL}} = 0\%$, the Log rank test, the Cox proportional hazards model, the Gamma frailty model and the χ^2-test respect the type I error rate of lower of equal to 5%. In this context, the Log rank test and the Cox proportional hazards model provide the largest power of 0.762 and 0.760 without any bias. In the scenarios with overdispersion, the calculated type I error rate is generally higher than that of the scenarios without overdispersion. Especially if the difference between the percentage of patients with treatment discontinuation in treatment and placebo arm is positive, the type I error rates are even larger.

Referring to the bias of scenarios without treatment effect, which are plotted in figure 12, the bias of all methods of scenarios without study dropout and $\%_{TD_A} - \%_{TD_{PL}} = 0\%$ is zero. In scenarios with $\phi = 1$, the Log-linear Poisson model is the single model, which is unbiased for each difference of patients with treatment discontinuation between treatment and placebo arm. The χ^2-test is the one with the largest absolute bias of 0.04, if $\%_{TD_A} - \%_{TD_{PL}} = -20\%$, and of -0.03, if $\%_{TD_A} - \%_{TD_{PL}} = +20\%$ in the scenario without overdispersion. Again, it can be notified, that the direction of the bias is changing with positive respectively negative difference of $\%_{TD_A} - \%_{TD_{PL}}$. In analogy to the hypothetical estimand, the bias of the Log-linear negative-binomial model is not showed in figure 12, if $\phi = 1$. Its results are not interpretable because of the low number of converged models. The bias of the scenarios without treatment effect under $\phi = 2$ with $\%_{TD_{PL}} = 20\%$ and $\%_{TD_{PL}} = 50\%$ are similar. Hereby, the absolute bias of the methods is almost zero if $\%_{TD_A} - \%_{TD_{PL}} = -20\%$.

The treatment effect of scenarios with treatment effect and $\phi = 1$ is estimated unbiased with the Log-linear Poisson model. The remaining methods manifest slightly biased estimates. In this context, the χ^2-test provides the largest negative bias compared to the other models. If $\phi = 2$, the Log-linear negative-binomial model and the Log-linear Poisson model have the same bias, which is no longer zero. The Gamma frailty model reveals strong overestimated treatment effect. The bias of the Shared gamma frailty model, the Log-linear negative-binomial model and the Log-linear Poisson model in the scenario $\phi = 2$, $\%_{TD_{PL}} = 20\%$ and $\%_{TD_A} - \%_{TD_{PL}} = -20\%$ as well as the

bias of the χ^2-test in the scenario $\phi = 2$, $\%_{TD_{PL}} = 20\%$ and $\%_{TD_A} - \%_{TD_{PL}} = +20\%$ are slightly positive or zero respectively. Furthermore, the bias of the scenarios with $\%_{TD_A} - \%_{TD_{PL}} = +20\%$ is absolutely larger than those with $\%_{TD_A} - \%_{TD_{PL}} = -20\%$. In the scenario with $\phi = 2$, $\%_{TD_{PL}} = 50\%$ and $\%_{TD_A} - \%_{TD_{PL}} = -20\%$, the Log-linear negative-binomial model provides an unbiased estimate. The Log-linear Poisson model and the Shared gamma frailty model return a small positive bias. If, $\phi = 2$, $\%_{TD_{PL}} = 50\%$ and $\%_{TD_A} - \%_{TD_{PL}} = 0\%$, the Cox proportional hazards model is the unique method to estimate an unbiased treatment effect.

Table 15: Comparison of methods estimating the treatment-policy estimand for scenarios without study dropout, with MCAR, $\phi = 1$, and $\%_{TD_{PL}} = 20\%$

		$\%_{TD_{PL}} = 20\%$								
		$\%_{TD_A} - \%_{TD_{PL}} = -20\%$			$\%_{TD_A} - \%_{TD_{PL}} = 0\%$			$\%_{TD_A} - \%_{TD_{PL}} = +20\%$		
		type I error rate [CI]	bias	N_{conv}	type I error rate [CI]	bias	N_{conv}	type I error rate [CI]	bias	N_{conv}
no tr. effect	GLM NB	0.050 [0.044;0.056]	0.00	5802	0.044 [0.039;0.049]	0.00	5924	0.138 [0.129;0.147]	0.00	5881
	GLM Poisson	0.057 [0.052;0.062]	0.00	10000	0.048 [0.044;0.052]	0.00	10000	0.141 [0.134;0.148]	0.00	10000
	Log rank test	0.057 [0.052;0.062]	-	10000	0.051 [0.047;0.055]	-	10000	0.094 [0.088;0.1]	-	10000
	Cox model	0.056 [0.051;0.061]	0.02	10000	0.051 [0.047;0.055]	0.00	10000	0.093 [0.087;0.099]	-0.02	10000
	LR test	0.057 [0.052;0.062]	0.02	10000	0.053 [0.049;0.057]	0.00	10000	0.097 [0.091;0.103]	-0.02	10000
	GF model	0.054 [0.05;0.058]	0.02	10000	0.046 [0.042;0.05]	0.00	10000	0.09 [0.084;0.096]	-0.02	10000
	SGF model	0.053 [0.049;0.057]	0.02	9940	0.045 [0.041;0.049]	0.00	9963	0.137 [0.13;0.144]	-0.02	9954
	χ^2-test	0.044 [0.04;0.048]	0.04	10000	0.039 [0.035;0.043]	0.00	10000	0.083 [0.078;0.088]	-0.03	10000
		power	bias	N_{conv}	power	bias	N_{conv}	power	bias	N_{conv}
with tr. effect	GLM NB	0.994	0.00	6094	0.987	0.00	6153	0.938	-0.00	6030
	GLM Poisson	0.995	0.00	10000	0.989	0.00	10000	0.943	0.00	10000
	Log rank test	0.959	-	10000	0.941	-	10000	0.882	-	10000
	Cox model	0.959	0.01	10000	0.940	-0.01	10000	0.881	-0.02	10000
	LR test	0.959	0.01	10000	0.941	-0.01	10000	0.882	-0.02	10000
	GF model	0.956	-0.01	10000	0.936	-0.02	10000	0.874	-0.04	10000
	SGF model	0.994	0.02	10000	0.988	0.00	10000	0.940	-0.01	10000
	χ^2-test	0.929	-0.06	10000	0.894	-0.09	10000	0.819	-0.11	10000

Table 16: Comparison of methods estimating the treatment-policy estimand for scenarios without study dropout, with MCAR, $\phi = 1$, and $\%_{TD_{PL}} = 50\%$

		$\%_{TD_{PL}} = 50\%$								
		$\%_{TD_A} - \%_{TD_{PL}} = -20\%$			$\%_{TD_A} - \%_{TD_{PL}} = 0\%$			$\%_{TD_A} - \%_{TD_{PL}} = +20\%$		
		type I error rate [CI]	bias	N_{conv}	type I error rate [CI]	bias	N_{conv}	type I error rate [CI]	bias	N_{conv}
no tr. effect	GLM NB	0.111 [0.103;0.119]	0.00	5958	0.049 [0.044;0.054]	0.00	6138	0.158 [0.149;0.167]	0.00	6028
	GLM Poisson	0.118 [0.112;0.124]	0.00	10000	0.053 [0.049;0.057]	0.00	10000	0.165 [0.158;0.172]	0.00	10000
	Log rank test	0.086 [0.081;0.091]	-	10000	0.055 [0.051;0.059]	-	10000	0.106 [0.1;0.112]	-	10000
	Cox model	0.085 [0.08;0.09]	0.02	10000	0.054 [0.05;0.058]	0.00	10000	0.104 [0.098;0.11]	-0.01	10000
	LR test	0.089 [0.083;0.095]	0.02	10000	0.059 [0.054;0.064]	0.00	10000	0.112 [0.106;0.118]	-0.01	10000
	GF model	0.08 [0.075;0.085]	0.02	10000	0.053 [0.049;0.057]	0.00	10000	0.101 [0.095;0.107]	-0.01	10000
	SGF model	0.113 [0.107;0.119]	0.02	9966	0.049 [0.045;0.053]	0.00	9991	0.159 [0.152;0.166]	-0.02	9985
	χ^2-test	0.072 [0.067;0.077]	0.04	10000	0.041 [0.037;0.045]	0.00	10000	0.088 [0.082;0.094]	-0.03	10000
		power	bias	N_{conv}	power	bias	N_{conv}	power	bias	N_{conv}
with tr. effect	GLM NB	0.916	0.00	6232	0.930	-0.00	6239	0.806	0.00	6217
	GLM Poisson	0.922	0.00	10000	0.933	0.00	10000	0.817	0.00	10000
	Log rank test	0.849	-	10000	0.833	-	10000	0.728	-	10000
	Cox model	0.848	0.03	10000	0.832	0.01	10000	0.726	0.01	10000
	LR test	0.851	0.02	10000	0.833	0.01	10000	0.728	0.01	10000
	GF model	0.840	0.00	10000	0.824	-0.01	10000	0.716	0.00	10000
	SGF model	0.918	0.04	10000	0.928	0.03	10000	0.809	0.03	9999
	χ^2-test	0.766	-0.04	10000	0.727	-0.07	10000	0.607	-0.08	10000

Table 17: Comparison of methods estimating the treatment-policy estimand for scenarios without study dropout, with MCAR, $\phi = 2$, and $\%_{TD_{PL}} = 20\%$

		$\%_{TD_{PL}} = 20\%$								
		$\%_{TD_A} - \%_{TD_{PL}} = -20\%$			$\%_{TD_A} - \%_{TD_{PL}} = 0\%$			$\%_{TD_A} - \%_{TD_{PL}} = +20\%$		
		type I error rate [CI]	bias	N_{conv}	type I error rate [CI]	bias	N_{conv}	type I error rate [CI]	bias	N_{conv}
no tr. effect	GLM NB	0.074 [0.069;0.079]	0.01	10000	0.056 [0.051;0.061]	0.00	10000	0.224 [0.216;0.232]	-0.02	9993
	GLM Poisson	0.2 [0.192;0.208]	0.01	10000	0.161 [0.154;0.168]	0.00	10000	0.329 [0.32;0.338]	-0.02	10000
	Log rank test	0.076 [0.071;0.081]	-	10000	0.052 [0.048;0.056]	-	10000	0.209 [0.201;0.217]	-	10000
	Cox model	0.075 [0.07;0.08]	0.00	10000	0.051 [0.047;0.055]	0.00	10000	0.208 [0.2;0.216]	-0.02	10000
	LR test	0.093 [0.087;0.099]	0.00	10000	0.064 [0.059;0.069]	0.00	10000	0.224 [0.216;0.232]	-0.03	10000
	GF model	0.07 [0.065;0.075]	0.00	10000	0.053 [0.049;0.057]	0.00	10000	0.179 [0.171;0.187]	-0.03	10000
	SGF model	0.074 [0.069;0.079]	0.04	10000	0.056 [0.051;0.061]	0.00	10000	0.224 [0.216;0.232]	-0.09	9995
	χ^2-test	0.067 [0.062;0.072]	0.00	10000	0.041 [0.037;0.045]	0.00	10000	0.221 [0.213;0.229]	0.02	10000
		power	bias	N_{conv}	power	bias	N_{conv}	power	bias	N_{conv}
with tr. effect	GLM NB	0.872	0.01	10000	0.889	-0.01	10000	0.825	-0.03	9988
	GLM Poisson	0.951	0.01	10000	0.958	-0.01	10000	0.891	-0.03	10000
	Log rank test	0.901	-	10000	0.884	-	10000	0.829	-	10000
	Cox model	0.900	-0.01	10000	0.883	-0.01	10000	0.828	-0.03	10000
	LR test	0.915	-0.03	10000	0.898	-0.02	10000	0.837	-0.05	10000
	GF model	0.896	-0.15	10000	0.877	-0.14	10000	0.820	-0.17	9996
	SGF model	0.872	0.03	10000	0.889	-0.01	10000	0.825	-0.09	9986
	χ^2-test	0.874	-0.01	10000	0.841	-0.01	10000	0.804	0.00	10000

Table 18: Comparison of methods estimating the treatment-policy estimand for scenarios without study dropout, with MCAR, $\phi = 2$, and $\%_{TD_{PL}} = 50\%$

		$\%_{TD_{PL}} = 50\%$								
		$\%_{TD_A} - \%_{TD_{PL}} = -20\%$			$\%_{TD_A} - \%_{TD_{PL}} = 0\%$			$\%_{TD_A} - \%_{TD_{PL}} = +20\%$		
		type I error rate [CI]	bias	N_{conv}	type I error rate [CI]	bias	N_{conv}	type I error rate [CI]	bias	N_{conv}
no tr. effect	GLM NB	0.266 [0.257;0.275]	0.02	9966	0.052 [0.048;0.056]	0.00	10000	0.345 [0.336;0.354]	-0.03	9906
	GLM Poisson	0.399 [0.389;0.409]	0.02	10000	0.139 [0.132;0.146]	0.00	10000	0.443 [0.433;0.453]	-0.03	10000
	Log rank test	0.262 [0.253;0.271]	-	10000	0.049 [0.045;0.053]	-	10000	0.329 [0.32;0.338]	-	10000
	Cox model	0.26 [0.251;0.269]	-0.01	10000	0.048 [0.044;0.052]	0.00	10000	0.327 [0.318;0.336]	-0.04	10000
	LR test	0.286 [0.277;0.295]	-0.01	10000	0.059 [0.054;0.064]	0.00	10000	0.349 [0.34;0.358]	-0.04	9999
	GF model	0.222 [0.214;0.23]	-0.01	10000	0.048 [0.044;0.052]	0.00	10000	0.273 [0.264;0.282]	-0.05	9997
	SGF model	0.262 [0.253;0.271]	0.01	9986	0.052 [0.048;0.056]	0.00	10000	0.345 [0.336;0.354]	-0.13	9955
	χ^2-test	0.283 [0.274;0.292]	0.00	10000	0.04 [0.036;0.044]	0.00	10000	0.347 [0.338;0.356]	0.02	10000
		power	bias	N_{conv}	power	bias	N_{conv}	power	bias	N_{conv}
with tr. effect	GLM NB	0.696	0.00	9956	0.788	-0.02	9999	0.695	-0.05	9906
	GLM Poisson	0.799	0.01	10000	0.891	-0.02	10000	0.763	-0.05	10000
	Log rank test	0.744	-	10000	0.762	-	10000	0.688	-	10000
	Cox model	0.742	-0.01	10000	0.760	0.00	10000	0.686	-0.03	10000
	LR test	0.755	-0.03	10000	0.779	-0.01	10000	0.698	-0.05	9999
	GF model	0.749	-0.14	9995	0.755	-0.13	10000	0.679	-0.16	9995
	SGF model	0.694	0.02	9978	0.788	-0.01	10000	0.693	-0.10	9950
	χ^2-test	0.694	-0.01	10000	0.674	-0.02	10000	0.644	-0.01	10000

7.3.1.2 Results for the treatment-policy estimand with study dropout of 50% of patients with treatment discontinuation under MCAR

The results for the scenarios under MCAR and with a study dropout of 50% of patients with treatment discontinuation, $\phi = 1$ and $\%_{TD_{PL}} = 20\%$ are listed in table 19 on page 91. Furthermore, the results for $\phi = 1$ and $\%_{TD_{PL}} = 50\%$ are given on page 92 in table 20. Table 21 on page 93 reveals the results of the scenarios with $\phi = 2$ and $\%_{TD_{PL}} = 20\%$. Finally, scenarios with $\phi = 2$ and $\%_{TD_{PL}} = 50\%$ are presented in table 22 on page 94, each under MCAR and a study dropout of 50% of the patients with treatment discontinuation.

In analogy to the scenarios without study dropout, the single scenario with a difference between dropout rate of the treatment and placebo arm, where a method holds the type I error rate of lower or equal to 5%, is the scenario with $\phi = 1$ and $\%_{TD_{PL}} = 20\%$ and $\%_{TD_A} - \%_{TD_{PL}} = -20\%$. Here, the Log-linear Poisson model and the Shared gamma frailty model show a type I error rate of 0.048 and 0.044 with a power of 0.998 and 0.996 both with a bias of zero. The scenario under $\phi = 1$, $\%_{TD_{PL}} = 20\%$ and $\%_{TD_A} - \%_{TD_{PL}} = 0\%$ is acceptably evaluated by the Log-linear Poisson model, the Gamma frailty model, the Shared gamma frailty model and the χ^2-test. The largest power is provided by the Log-linear Poisson model and the Shared gamma frailty model with a value of 0.989 and 0.990. In this context, the Log-linear Poisson model estimates the treatment effect without any bias. No method respects a type I error rate of lower or equal to 5% in a scenario with $\%_{TD_A} - \%_{TD_{PL}} = +20\%$. The results for the scenarios with $\phi = 1$, $\%_{TD_{PL}} = 50\%$ behave like the corresponding scenarios without study dropout. The Shared gamma frailty model and the χ^2-test return an acceptable type I error rate of 0.05 and 0.037 both with zero bias, if $\%_{TD_A} - \%_{TD_{PL}} = 0\%$. Hereby, the Shared gamma frailty model reveals the larger power of 0.947 with a small bias of -0.01. For the scenarios with $\phi = 2$ and $\%_{TD_{PL}} = 20\%$, one adequate method of evaluation was found under $\%_{TD_A} - \%_{TD_{PL}} = 0\%$. This is the χ^2-test with a type I error rate of 0.039 without any bias and a power of 0.848 with a bias of -0.02. The results of the last scenarios with $\phi = 2$ and $\%_{TD_{PL}} = 50\%$ are similar to those without study dropout. Furthermore, it is remarkable that the analyses with the Log-linear negative-binomial model and the Shared gamma frailty model are not equal any more for scenarios with study dropout of 50% patients with treatment discontinuation under MCAR.

The bias of the scenarios with study dropout of 50% of patients with treatment discontinuation is visualized in figure 12, too. In analogy to the scenarios without study dropout, the bias of scenarios without treatment effect is zero, if $\%_{TD_A} - \%_{TD_{PL}} = 0\%$. But if $\phi = 2$, $\%_{TD_{PL}} = 50\%$, the Cox proportional hazards model, the Likelihood ratio test,

the Gamma frailty model and the Shared gamma frailty model reveal slightly negatively biased results. The estimated treatment effects of the scenarios without overdispersion are equal in the way that the χ^2-test estimates them heavily biased compared to the other methods, if there is a difference between the percentage of patients with treatment discontinuation between both trial arms. The remaining methods are just slightly biased with a reversed direction of the bias for the two differences of treatment discontinuation between both trial arms. The scenarios with $\phi = 2$ display a larger variance of the bias, if $\%_{TD_A} - \%_{TD_{PL}} = -20\%$ or $\%_{TD_A} - \%_{TD_{PL}} = +20\%$, compared to those without overdispersion. In particular, the χ^2-test manifests the most biased estimates in these scenarios.

Turning over to the scenarios with treatment effect and without overdispersion, only the Log-linear Poisson model returns unbiased estimates of the treatment-policy estimand, if $\%_{TD_A} - \%_{TD_{PL}} = 0\%$. In general, the χ^2-test again shows the strongest bias of all methods. But for each method of evaluation, the absolute value of the bias is larger, if $\%_{TD_A} - \%_{TD_{PL}} = +20\%$ compared to $\%_{TD_A} - \%_{TD_{PL}} = -20\%$. Furthermore, the direction of the bias is reversed between both differences of percentages of treatment discontinuation. The bias of the scenarios with $\phi = 2$ are quite similar to the corresponding scenarios without study dropout. The Gamma frailty model heavily overestimates the true treatment effect again. This is also valid for the χ^2-test, if $\phi = 2$ and $\%_{TD_A} - \%_{TD_{PL}} = +20\%$.

Table 19: Comparison of methods estimating the treatment-policy estimand for scenarios with study dropout of 50% of patients, with MCAR, $\phi = 1$, and $\%_{TD_{PL}} = 20\%$

		$\%_{TD_{PL}} = 20\%$								
		$\%_{TD_A} - \%_{TD_{PL}} = -20\%$			$\%_{TD_A} - \%_{TD_{PL}} = 0\%$			$\%_{TD_A} - \%_{TD_{PL}} = +20\%$		
		type I error rate [CI]	bias	N_{conv}	type I error rate [CI]	bias	N_{conv}	type I error rate [CI]	bias	N_{conv}
no tr. effect	GLM NB	0.044 [0.039;0.049]	0.00	6083	0.045 [0.04;0.05]	0.00	6418	0.167 [0.158;0.176]	-0.01	6756
	GLM Poisson	0.048 [0.044;0.052]	0.00	10000	0.050 [0.046;0.054]	0.00	10000	0.132 [0.125;0.139]	0.00	10000
	Log rank test	0.056 [0.051;0.061]	-	10000	0.052 [0.048;0.056]	-	10000	0.084 [0.079;0.089]	-	10000
	Cox model	0.055 [0.051;0.059]	0.01	10000	0.051 [0.047;0.055]	0.00	10000	0.083 [0.078;0.088]	-0.02	10000
	LR test	0.056 [0.051;0.061]	0.01	10000	0.052 [0.048;0.056]	0.00	10000	0.084 [0.079;0.089]	-0.02	10000
	GF model	0.052 [0.048;0.056]	0.01	10000	0.048 [0.044;0.052]	0.00	10000	0.081 [0.076;0.086]	-0.02	9997
	SGF model	0.044 [0.04;0.048]	0.00	9983	0.048 [0.044;0.052]	0.00	9996	0.06 [0.055;0.065]	-0.01	9987
	χ^2-test	0.071 [0.066;0.076]	0.12	10000	0.039 [0.035;0.043]	0.00	10000	0.242 [0.234;0.25]	-0.11	10000
		power	bias	N_{conv}	power	bias	N_{conv}	power	bias	N_{conv}
with tr. effect	GLM NB	0.997	0.00	6377	0.987	0.00	6718	0.811	0.03	6841
	GLM Poisson	0.998	0.00	10000	0.989	0.00	10000	0.839	0.00	10000
	Log rank test	0.961	-	10000	0.945	-	10000	0.877	-	10000
	Cox model	0.960	0.01	10000	0.945	-0.02	10000	0.876	-0.04	10000
	LR test	0.961	0.00	10000	0.945	-0.02	10000	0.884	-0.04	10000
	GF model	0.959	-0.01	9990	0.943	-0.04	9989	0.881	-0.05	9982
	SGF model	0.996	0.00	10000	0.990	-0.02	10000	0.871	-0.01	10000
	χ^2-test	0.870	0.02	10000	0.901	-0.09	10000	0.898	-0.21	10000

Table 20: Comparison of methods estimating the treatment-policy estimand for scenarios with study dropout of 50% of patients, with MCAR, $\phi = 1$, and $\%_{TD_{PL}} = 50\%$

		$\%_{TD_{PL}} = 50\%$								
		$\%_{TD_A} - \%_{TD_{PL}} = -20\%$			$\%_{TD_A} - \%_{TD_{PL}} = 0\%$			$\%_{TD_A} - \%_{TD_{PL}} = +20\%$		
		type I error rate [CI]	bias	N_{conv}	type I error rate [CI]	bias	N_{conv}	type I error rate [CI]	bias	N_{conv}
no tr. effect	GLM NB	0.129 [0.121;0.137]	-0.01	7146	0.046 [0.041;0.051]	0.00	7612	0.259 [0.249;0.269]	0.00	7837
	GLM Poisson	0.112 [0.106;0.118]	0.00	10000	0.055 [0.051;0.059]	0.00	10000	0.211 [0.203;0.219]	0.00	10000
	Log rank test	0.073 [0.068;0.078]	-	10000	0.055 [0.051;0.059]	-	10000	0.089 [0.083;0.095]	-	10000
	Cox model	0.072 [0.067;0.077]	0.00	10000	0.053 [0.049;0.057]	0.00	10000	0.088 [0.082;0.094]	-0.02	10000
	LR test	0.075 [0.07;0.08]	0.00	10000	0.058 [0.053;0.063]	0.00	10000	0.089 [0.083;0.095]	-0.02	10000
	GF model	0.07 [0.065;0.075]	0.00	9998	0.053 [0.049;0.057]	0.00	9999	0.085 [0.08;0.09]	-0.03	10000
	SGF model	0.053 [0.049;0.057]	-0.01	9998	0.05 [0.046;0.054]	0.00	10000	0.058 [0.053;0.063]	0.00	9999
	χ^2-test	0.307 [0.298;0.316]	0.15	10000	0.037 [0.033;0.041]	0.00	10000	0.411 [0.401;0.421]	-0.12	10000
		power	bias	N_{conv}	power	bias	N_{conv}	power	bias	N_{conv}
with tr. effect	GLM NB	0.863	0.02	7421	0.922	0.00	8064	0.679	0.04	7917
	GLM Poisson	0.883	0.00	10000	0.938	0.00	10000	0.699	0.00	10000
	Log rank test	0.860	-	10000	0.870	-	10000	0.743	-	10000
	Cox model	0.859	0.00	10000	0.868	-0.03	10000	0.741	-0.02	10000
	LR test	0.863	0.00	10000	0.871	-0.03	10000	0.750	-0.02	10000
	GF model	0.859	-0.02	9989	0.865	-0.05	9995	0.745	-0.04	9992
	SGF model	0.896	0.00	10000	0.947	-0.01	10000	0.714	0.03	10000
	χ^2-test	0.636	0.05	10000	0.746	-0.06	10000	0.746	-0.22	10000

Table 21: Comparison of methods estimating the treatment-policy estimand for scenarios with study dropout of 50% of patients, with MCAR, $\phi = 2$, and $\%_{TD_{PL}} = 20\%$

		$\%_{TD_{PL}} = 20\%$								
		$\%_{TD_A} - \%_{TD_{PL}} = -20\%$			$\%_{TD_A} - \%_{TD_{PL}} = 0\%$			$\%_{TD_A} - \%_{TD_{PL}} = +20\%$		
		type I error rate [CI]	bias	N_{conv}	type I error rate [CI]	bias	N_{conv}	type I error rate [CI]	bias	N_{conv}
no tr. effect	GLM NB	0.071 [0.066;0.076]	0.00	10000	0.055 [0.051;0.059]	0.00	10000	0.213 [0.205;0.221]	-0.01	10000
	GLM Poisson	0.184 [0.176;0.192]	0.01	10000	0.159 [0.152;0.166]	0.00	10000	0.314 [0.305;0.323]	-0.02	10000
	Log rank test	0.059 [0.054;0.064]	-	10000	0.053 [0.049;0.057]	-	10000	0.187 [0.179;0.195]	-	10000
	Cox model	0.058 [0.053;0.063]	0.03	10000	0.052 [0.048;0.056]	-0.01	10000	0.186 [0.178;0.194]	-0.02	10000
	LR test	0.072 [0.067;0.077]	-0.01	10000	0.065 [0.06;0.07]	-0.01	10000	0.189 [0.181;0.197]	0.01	10000
	GF model	0.059 [0.054;0.064]	-0.01	10000	0.053 [0.049;0.057]	-0.01	9999	0.163 [0.156;0.17]	0.00	9999
	SGF model	0.061 [0.056;0.066]	0.01	10000	0.054 [0.05;0.058]	-0.01	10000	0.189 [0.181;0.197]	-0.04	10000
	χ^2-test	0.064 [0.059;0.069]	0.06	10000	0.039 [0.035;0.043]	0.00	10000	0.206 [0.198;0.214]	-0.08	10000
		power	bias	N_{conv}	power	bias	N_{conv}	power	bias	N_{conv}
with tr. effect	GLM NB	0.920	0.00	10000	0.874	-0.01	10000	0.850	-0.02	10000
	GLM Poisson	0.972	0.01	10000	0.958	-0.01	10000	0.901	-0.03	10000
	Log rank test	0.915	-	10000	0.892	-	10000	0.830	-	10000
	Cox model	0.912	0.03	10000	0.890	0.00	10000	0.828	-0.03	10000
	LR test	0.930	-0.03	10000	0.906	-0.04	10000	0.838	-0.05	10000
	GF model	0.907	-0.15	10000	0.884	-0.15	9999	0.826	-0.16	9999
	SGF model	0.899	0.01	10000	0.888	-0.02	10000	0.834	-0.07	10000
	χ^2-test	0.833	0.06	10000	0.848	-0.02	10000	0.782	-0.12	10000

Table 22: Comparison of methods estimating the treatment-policy estimand for scenarios with study dropout of 50% of patients, with MCAR, $\phi = 2$, and $\%_{TD_{PL}} = 50\%$

		$\%_{TD_{PL}} = 50\%$								
		$\%_{TD_A} - \%_{TD_{PL}} = -20\%$			$\%_{TD_A} - \%_{TD_{PL}} = 0\%$			$\%_{TD_A} - \%_{TD_{PL}} = +20\%$		
		type I error rate [CI]	bias	N_{conv}	type I error rate [CI]	bias	N_{conv}	type I error rate [CI]	bias	N_{conv}
no tr. effect	GLM NB	0.239 [0.231;0.247]	0.00	9998	0.055 [0.051;0.059]	0.00	10000	0.33 [0.321;0.339]	-0.01	9997
	GLM Poisson	0.348 [0.339;0.357]	0.02	10000	0.141 [0.134;0.148]	0.00	10000	0.417 [0.407;0.427]	-0.03	10000
	Log rank test	0.192 [0.184;0.2]	-	10000	0.048 [0.044;0.052]	-	10000	0.282 [0.273;0.291]	-	10000
	Cox model	0.19 [0.182;0.198]	0.03	10000	0.047 [0.043;0.051]	0.00	10000	0.28 [0.271;0.289]	-0.04	10000
	LR test	0.197 [0.189;0.205]	-0.03	10000	0.059 [0.054;0.064]	0.00	10000	0.277 [0.268;0.286]	-0.03	9999
	GF model	0.166 [0.159;0.173]	-0.02	9999	0.05 [0.046;0.054]	0.00	10000	0.236 [0.228;0.244]	-0.05	9998
	SGF model	0.185 [0.177;0.193]	-0.02	10000	0.051 [0.047;0.055]	0.00	10000	0.274 [0.265;0.283]	-0.09	9999
	χ^2-test	0.260 [0.251;0.269]	0.06	10000	0.038 [0.034;0.042]	0.00	10000	0.339 [0.33;0.348]	-0.12	10000
		power	bias	N_{conv}	power	bias	N_{conv}	power	bias	N_{conv}
with tr. effect	GLM NB	0.812	-0.01	9995	0.782	-0.02	10000	0.735	-0.02	9994
	GLM Poisson	0.879	0.01	10000	0.899	-0.02	10000	0.797	-0.05	10000
	Log rank test	0.781	-	10000	0.789	-	10000	0.709	-	10000
	Cox model	0.779	0.04	10000	0.786	0.02	10000	0.707	-0.02	10000
	LR test	0.807	-0.06	10000	0.807	-0.05	10000	0.715	-0.07	9997
	GF model	0.782	-0.16	9998	0.786	-0.15	9999	0.699	-0.18	10000
	SGF model	0.764	-0.03	10000	0.801	-0.03	10000	0.709	-0.11	10000
	χ^2-test	0.601	0.03	10000	0.686	-0.03	10000	0.617	-0.19	10000

7.3.1.3 Results for the treatment-policy estimand with study dropout of 100% with treatment discontinuation under MCAR

The results for this section are based on MCAR and on a study dropout of 100% of patients with treatment discontinuation. They are provided in table 23 on page 96, if $\phi = 1$ and $\%_{TD_{PL}} = 20\%$. If $\phi = 1$ and $\%_{TD_{PL}} = 50\%$, table 24 on page 97 contains the relevant results. Furthermore, for the scenarios $\phi = 2$ and $\%_{TD_{PL}} = 20\%$, the results are provided in table 25 on page 98. The results for the scenarios with $\phi = 2$ and $\%_{TD_{PL}} = 50\%$ are collected in table 26 on page 99.

The type I error rate and power of methods, which are used for the evaluation of scenarios with study dropout of 100%, are equal to the results of the scenarios of the hypothetical estimand under MCAR. Just bias calculations differ. These are presented in figure 12 which reveals, that the bias increases the more patients drop out of the study in the scenarios with $\%_{TD_A} - \%_{TD_{PL}} = -20\%$ or $\%_{TD_A} - \%_{TD_{PL}} = +20\%$. Additionally, in the scenarios with a study dropout of 100% of patients with treatment discontinuation, the absolute value of the bias is larger, if $\%_{TD_A} - \%_{TD_{PL}} = +20\%$ compared to $\%_{TD_A} - \%_{TD_{PL}} = -20\%$.

The bias of the scenarios without treatment effect and without difference between the percentage of patients with treatment discontinuation in both trial groups is equivalent to those of the corresponding scenarios without treatment effect, $\%_{TD_A} - \%_{TD_{PL}} = 0\%$ and study dropout of 50% of the patients with treatment discontinuation. In the case of study dropout of 100% of patients with treatment discontinuation, the Shared gamma frailty model manifests the smallest bias, whereas the χ^2-test provides the most biased estimates.

Concerning the scenarios with treatment effect, only the Likelihood ratio test and the Gamma frailty model provide a unbiased treatment effect estimate under $\%_{TD_A} - \%_{TD_{PL}} = 0\%$. If $\phi = 2$, $\%_{TD_{PL}} = 20\%$ and $\%_{TD_A} - \%_{TD_{PL}} = +20\%$, just the Gamma frailty model returns a correct estimation of the treatment-policy estimand. In the remaining scenarios, calculated treatment effects are always biased. Hereby, the estimates of the χ^2-test are heavily overestimated, if $\%_{TD_A} - \%_{TD_{PL}} = +20\%$.

Table 23: Comparison of methods estimating the treatment-policy estimand for scenarios with study dropout of 100% of patients, with MCAR, $\phi = 1$, and $\%_{TD_{PL}} = 20\%$

		$\%_{TD_{PL}} = 20\%$								
		$\%_{TD_A} - \%_{TD_{PL}} = -20\%$			$\%_{TD_A} - \%_{TD_{PL}} = 0\%$			$\%_{TD_A} - \%_{TD_{PL}} = +20\%$		
		type I error rate [CI]	bias	N_{conv}	type I error rate [CI]	bias	N_{conv}	type I error rate [CI]	bias	N_{conv}
no tr. effect	GLM NB	0.039 [0.034;0.044]	0.09	5783	0.045 [0.04;0.05]	0.00	5800	0.046 [0.041;0.051]	-0.11	5846
	GLM Poisson	0.049 [0.045;0.053]	0.09	10000	0.051 [0.047;0.055]	0.00	10000	0.051 [0.047;0.055]	-0.11	10000
	Log rank test	0.066 [0.061;0.071]	-	10000	0.051 [0.047;0.055]	-	10000	0.314 [0.305;0.323]	-	10000
	Cox model	0.065 [0.06;0.07]	-0.05	10000	0.051 [0.047;0.055]	0.00	10000	0.313 [0.304;0.322]	0.14	10000
	LR test	0.064 [0.059;0.069]	-0.05	10000	0.047 [0.043;0.051]	0.00	10000	0.237 [0.229;0.245]	0.09	10000
	GF model	0.06 [0.055;0.065]	-0.05	9999	0.044 [0.04;0.048]	0.00	10000	0.172 [0.165;0.179]	0.08	9778
	SGF model	0.045 [0.041;0.049]	0.00	9982	0.047 [0.043;0.051]	0.00	10000	0.049 [0.045;0.053]	-0.01	9985
	χ^2-test	0.179 [0.171;0.187]	0.26	10000	0.039 [0.035;0.043]	0.00	10000	0.389 [0.379;0.399]	-0.32	10000
		power	bias	N_{conv}	power	bias	N_{conv}	power	bias	N_{conv}
with tr. effect	GLM NB	0.996	0.09	5971	0.995	-0.06	5936	0.923	-0.23	5877
	GLM Poisson	0.996	0.09	10000	0.995	-0.06	10000	0.920	-0.24	10000
	Log rank test	0.988	-	10000	0.948	-	10000	0.837	-	10000
	Cox model	0.988	-0.06	10000	0.948	-0.02	10000	0.837	0.16	10000
	LR test	0.988	-0.05	10000	0.946	-0.01	10000	0.805	0.09	10000
	GF model	0.986	-0.06	10000	0.942	-0.01	10000	0.807	0.04	9225
	SGF model	0.996	-0.01	10000	0.994	-0.06	9999	0.920	-0.13	9999
	χ^2-test	0.697	0.16	10000	0.917	-0.10	10000	0.818	-0.52	10000

Table 24: Comparison of methods estimating the treatment-policy estimand for scenarios with study dropout of 100% of patients, with MCAR, $\phi = 1$, and $\%_{TD_{PL}} = 50\%$

		$\%_{TD_{PL}} = 50\%$								
		$\%_{TD_A} - \%_{TD_{PL}} = -20\%$			$\%_{TD_A} - \%_{TD_{PL}} = 0\%$			$\%_{TD_A} - \%_{TD_{PL}} = +20\%$		
		type I error rate [CI]	bias	N_{conv}	type I error rate [CI]	bias	N_{conv}	type I error rate [CI]	bias	N_{conv}
no tr. effect	GLM NB	0.044 [0.039;0.049]	0.11	5857	0.047 [0.042;0.052]	0.00	6008	0.042 [0.037;0.047]	-0.18	5865
	GLM Poisson	0.048 [0.044;0.052]	0.12	10000	0.052 [0.048;0.056]	0.00	10000	0.044 [0.04;0.048]	-0.18	10000
	Log rank test	0.303 [0.294;0.312]	-	10000	0.053 [0.049;0.057]	-	10000	0.519 [0.509;0.529]	-	10000
	Cox model	0.302 [0.293;0.311]	-0.01	10000	0.052 [0.048;0.056]	0.00	10000	0.518 [0.508;0.528]	0.23	10000
	LR test	0.224 [0.216;0.232]	-0.04	10000	0.042 [0.038;0.046]	-0.00	10000	0.359 [0.35;0.368]	0.13	10000
	GF model	0.191 [0.183;0.199]	-0.05	9802	0.04 [0.036;0.044]	0.00	9998	0.286 [0.277;0.295]	0.12	9491
	SGF model	0.045 [0.041;0.049]	-0.02	9993	0.048 [0.044;0.052]	0.00	9995	0.043 [0.039;0.047]	-0.03	9994
	χ^2-test	0.769 [0.761;0.777]	0.23	10000	0.043 [0.039;0.047]	0.00	10000	0.736 [0.727;0.745]	-0.47	10000
		power	bias	N_{conv}	power	bias	N_{conv}	power	bias	N_{conv}
with tr. effect	GLM NB	0.933	0.03	5963	0.981	-0.15	6038	0.819	-0.41	5858
	GLM Poisson	0.936	0.03	10000	0.984	-0.16	10000	0.822	-0.41	10000
	Log rank test	0.863	-	10000	0.876	-	10000	0.713	-	10000
	Cox model	0.863	0.01	10000	0.874	-0.02	10000	0.712	0.29	10000
	LR test	0.858	-0.03	10000	0.858	0.00	10000	0.668	0.17	10000
	GF model	0.880	-0.07	9449	0.847	0.00	9996	0.707	0.12	8544
	SGF model	0.934	-0.10	9995	0.983	-0.13	10000	0.820	-0.24	9992
	χ^2-test	0.325	0.07	10000	0.837	-0.13	10000	0.642	-0.74	10000

Table 25: Comparison of methods estimating the treatment-policy estimand for scenarios with study dropout of 100% of patients, with MCAR, $\phi = 2$, and $\%_{TD_{PL}} = 20\%$

		$\%_{TD_{PL}} = 20\%$								
		$\%_{TD_A} - \%_{TD_{PL}} = -20\%$			$\%_{TD_A} - \%_{TD_{PL}} = 0\%$			$\%_{TD_A} - \%_{TD_{PL}} = +20\%$		
		type I error rate [CI]	bias	N_{conv}	type I error rate [CI]	bias	N_{conv}	type I error rate [CI]	bias	N_{conv}
no tr. effect	GLM NB	0.05 [0.046;0.054]	0.08	10000	0.051 [0.047;0.055]	0.00	10000	0.049 [0.045;0.053]	-0.12	10000
	GLM Poisson	0.163 [0.156;0.17]	0.08	10000	0.16 [0.153;0.167]	0.00	10000	0.142 [0.135;0.149]	-0.12	10000
	Log rank test	0.077 [0.072;0.082]	-	10000	0.052 [0.048;0.056]	-	10000	0.299 [0.29;0.308]	-	10000
	Cox model	0.077 [0.072;0.082]	-0.07	10000	0.052 [0.048;0.056]	-0.01	10000	0.298 [0.289;0.307]	0.19	10000
	LR test	0.093 [0.087;0.099]	-0.08	10000	0.061 [0.056;0.066]	-0.01	10000	0.286 [0.277;0.295]	0.15	10000
	GF model	0.072 [0.067;0.077]	-0.07	10000	0.052 [0.048;0.056]	-0.01	10000	0.211 [0.203;0.219]	0.13	9649
	SGF model	0.05 [0.046;0.054]	-0.01	10000	0.051 [0.047;0.055]	-0.01	10000	0.049 [0.045;0.053]	0.01	10000
	χ^2-test	0.11 [0.104;0.116]	0.17	10000	0.042 [0.038;0.046]	0.00	10000	0.322 [0.313;0.331]	-0.28	10000
		power	bias	N_{conv}	power	bias	N_{conv}	power	bias	N_{conv}
with tr. effect	GLM NB	0.915	0.09	10000	0.903	-0.06	10000	0.836	-0.23	10000
	GLM Poisson	0.971	0.09	10000	0.968	-0.06	10000	0.897	-0.23	10000
	Log rank test	0.961	-	10000	0.890	-	10000	0.831	-	10000
	Cox model	0.960	-0.07	10000	0.889	-0.01	10000	0.830	0.18	10000
	LR test	0.970	-0.10	10000	0.898	-0.02	10000	0.815	0.11	10000
	GF model	0.953	-0.19	10000	0.884	-0.10	10000	0.828	0.00	9472
	SGF model	0.915	-0.01	10000	0.903	-0.05	10000	0.836	-0.13	10000

Table 26: Comparison of methods estimating the treatment-policy estimand for scenarios with study dropout of 100% of patients, with MCAR, $\phi = 2$, and $\%_{TD_{PL}} = 50\%$

		$\%_{TD_{PL}} = 50\%$								
		$\%_{TD_A} - \%_{TD_{PL}} = -20\%$			$\%_{TD_A} - \%_{TD_{PL}} = 0\%$			$\%_{TD_A} - \%_{TD_{PL}} = +20\%$		
		type I error rate [CI]	bias	N_{conv}	type I error rate [CI]	bias	N_{conv}	type I error rate [CI]	bias	N_{conv}
no tr. effect	GLM NB	0.05 [0.046;0.054]	0.12	10000	0.05 [0.046;0.054]	0.00	10000	0.047 [0.043;0.051]	-0.18	10000
	GLM Poisson	0.139 [0.132;0.146]	0.12	10000	0.15 [0.143;0.157]	0.00	10000	0.117 [0.111;0.123]	-0.18	10000
	Log rank test	0.364 [0.355;0.373]	-	10000	0.049 [0.045;0.053]	-	10000	0.513 [0.503;0.523]	-	10000
	Cox model	0.362 [0.353;0.371]	-0.02	10000	0.048 [0.044;0.052]	0.00	10000	0.511 [0.501;0.521]	0.27	9999
	LR test	0.356 [0.347;0.365]	-0.07	10000	0.051 [0.047;0.055]	0.00	10000	0.455 [0.445;0.465]	0.19	9999
	GF model	0.29 [0.281;0.299]	-0.07	9726	0.044 [0.04;0.048]	0.00	10000	0.362 [0.352;0.372]	0.16	9300
	SGF model	0.05 [0.046;0.054]	-0.03	10000	0.049 [0.045;0.053]	0.00	10000	0.047 [0.043;0.051]	-0.03	10000
	χ^2-test	0.558 [0.548;0.568]	0.13	10000	0.038 [0.034;0.042]	0.00	10000	0.599 [0.589;0.609]	-0.42	10000
		power	bias	N_{conv}	power	bias	N_{conv}	power	bias	N_{conv}
with tr. effect	GLM NB	0.834	0.03	10000	0.854	-0.16	10000	0.723	-0.42	10000
	GLM Poisson	0.906	0.03	10000	0.941	-0.16	10000	0.794	-0.42	10000
	Log rank test	0.861	-	10000	0.794	-	10000	0.734	-	10000
	Cox model	0.860	-0.01	10000	0.792	-0.01	10000	0.733	0.31	10000
	LR test	0.862	-0.08	10000	0.794	-0.02	10000	0.697	0.18	9999
	GF model	0.874	-0.15	9578	0.778	-0.05	9997	0.741	0.10	8871
	SGF model	0.835	-0.11	10000	0.854	-0.13	10000	0.724	-0.25	10000
	χ^2-test	0.376	0.04	10000	0.783	-0.11	10000	0.661	-0.64	10000

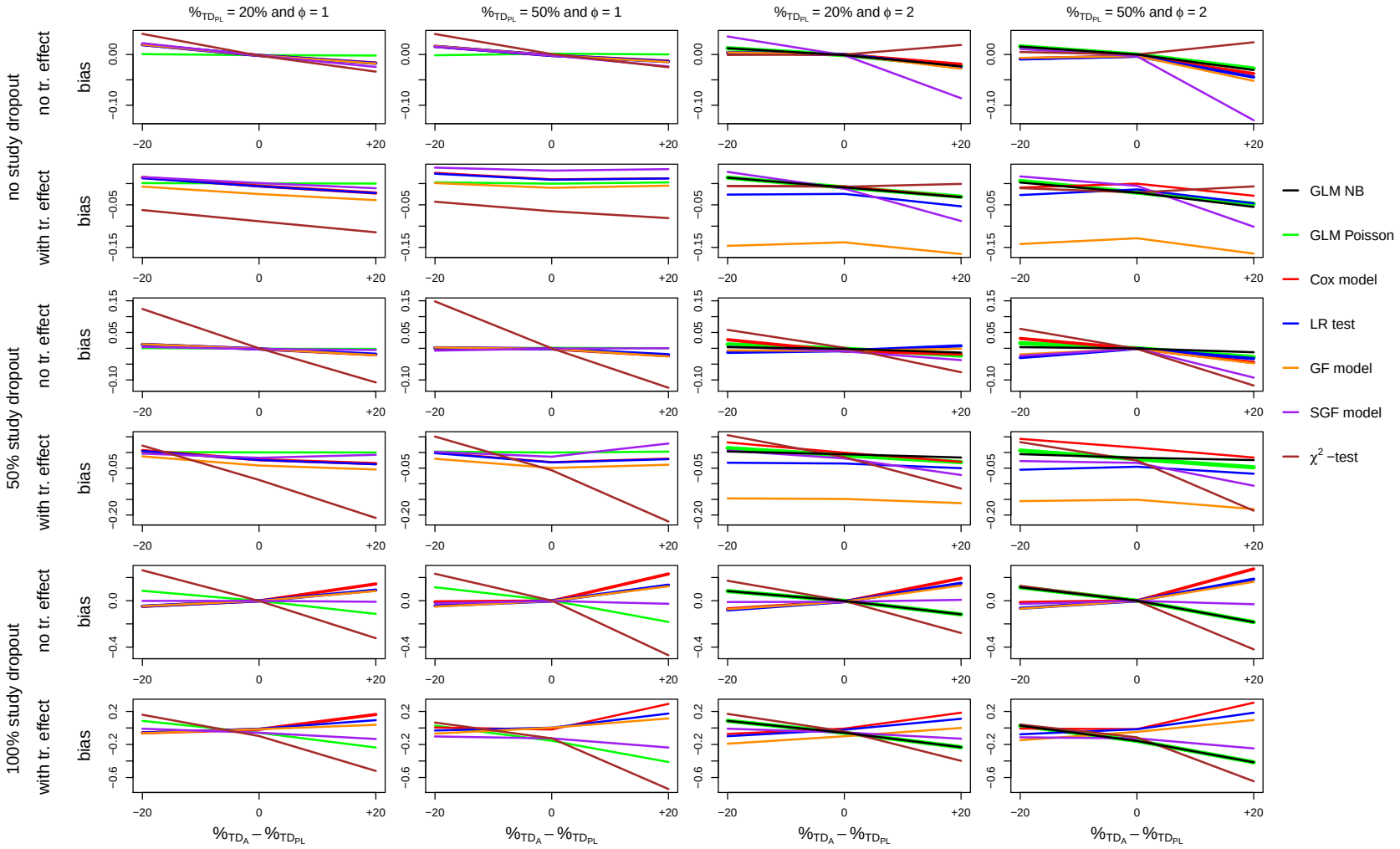

Figure 12: Bias of the Log-linear negative-binomial model (GLM NB), Log-linear Poisson model (GLM Poisson), Cox proportional hazards model (Cox model), Likelihood ratio test (LR test), Gamma frailty model (GF model), Shared gamma frailty model (SGF model) and χ^2-test for treatment-policy estimand depending on the percentage of patients with treatment discontinuation in placebo group ($\%_{TD_{PL}}$), the difference of patients with treatment discontinuation between both trial arms ($\%_{TD_A} - \%_{TD_{PL}}$) and ϕ under MCAR

7.3.2 Results for the treatment-policy estimand with treatment discontinuation mechanism MNAR

7.3.2.1 Results for the treatment-policy estimand without study dropout with treatment discontinuation mechanism MNAR

The results of the scenarios of the treatment-policy estimand without study dropout under MNAR, $\phi = 1$ and $\%_{TD_{PL}} = 27\%$ are listed in table 27 on page 103. For scenarios with $\phi = 1$ $\%_{TD_{PL}} = 60\%$, table 28 on page 104 shows the results. Furthermore, the results of the scenarios $\phi = 2$ and $\%_{TD_{PL}} = 28\%$ are collected in table 29 on page 105. Table 30 on page 106 contains the results for the scenarios with $\phi = 2$ and $\%_{TD_{PL}} = 60\%$.

In general, the results of the scenarios under MNAR are similar to those under MCAR concerning the type I error rate. The χ^2-test is a method which always provides an acceptable type I error rate, if $\%_{TD_A} - \%_{TD_{PL}} = 0\%$. Thus, the methods do not deliver a wrong test decision in more than 5% of all test decisions only in scenarios without any difference between the percentages of patients with treatment discontinuation in both trial arms. If there is such a discrepancy between both trials arms, don't matter which direction, no method respects the type I error rate. This is true for all scenarios of the treatment-policy estimand under MNAR. Additionally, it can be observed for all these scenarios that the type I error rate gets larger if more patients discontinue treatment, dropout out of the study and show individual frailties. For the scenarios under MNAR without study dropout, $\phi = 1$, $\%_{TD_{PL}} = 27\%$ and $\%_{TD_A} - \%_{TD_{PL}} = 0\%$, the Log-linear Poisson model, the Log rank test, the Cox proportional hazards model, the Gamma frailty model, the Shared gamma frailty model and the χ^2-test return an acceptable type I error rates. But not each method provides an unbiased estimate as it was the case in scenarios under MCAR. The Log-linear Poisson model possesses the largest power of 0.976 with zero bias. This is also valid under the scenario $\phi = 1$, $\%_{TD_{PL}} = 60\%$ and $\%_{TD_A} - \%_{TD_{PL}} = 0\%$. But hereby, only the Log-linear Poisson-model, the Gamma frailty model, the Shared gamma frailty model and the χ^2-test return some acceptable type I error rates, each with a bias of zero. The method with the largest power 0.769 and the smallest bias of 0.01 is again the Log-linear Poisson model. For the scenarios with $\phi = 2$, $\%_{TD_{PL}} = 28\%$ and $\%_{TD_A} - \%_{TD_{PL}} = 0\%$, the Log rank test, the Cox proportional hazards model, the Gamma frailty model and the χ^2-test show a type I error rate lower than 5%. They represent a value of 0.048 for Log rank test, a value of 0.048 for the Cox proportional hazards model, a value of 0.047 for the Gamma frailty model and a value of 0.041 for the χ^2-test. In analogy to the corresponding scenarios without overdispersion, some methods don't deliver an unbiased estimate. The Log rank test and the Cox proportional hazards model manifest the largest power of 0.850 or 0.848,

respectively. Their estimates are unbiased. In the last scenario without study dropout under MNAR, with $\phi = 2$, $\%_{TD_{PL}} = 60\%$ and $\%_{TD_A} - \%_{TD_{PL}} = 0\%$, the Log rank test, the Cox proportional hazards model, the Gamma frailty model and the χ^2-test hold the type I error rate of lower than 5%. The type I error rate of the Log rank test, of the Cox proportional hazards model and of the Gamma frailty model represents a value of 0.045 with zero bias. The type I error rate of the χ^2-test has a value of 0.039 with a bias of zero. Again, the Log rank test and the Cox proportional hazards model reveal the maximal power of 0.657 or 0.653 without any bias. As the Log rank test and the Cox proportional hazards model are actually the same procedure if only one covariate is used. The differences between the results of these two methods are caused by the implementations in the statistical software R.

In figure 13 the bias of each method is visualized for each scenario without study dropout under MNAR. These results do not differ heavily from those already described in the context of the bias description of the scenarios without study dropout under MCAR in section 7.3.1.1. However, there are some differences. In the scenarios without overdispersion, the Log-linear Poisson model shows a large negative bias if there is a negative difference between the percentages of patients with treatment discontinuation between the treatment and the placebo group. Focusing on the scenarios with $\phi = 2$, the Shared gamma frailty model underestimates the true treatment effect if there is a negative difference between the percentage of patients with treatment discontinuation in treatment and placebo arm. But in contrast to the MCAR, if this difference is positive, this model no longer reveals such a heavy bias.

Table 27: Comparison of methods estimating the tr.-policy for scenarios without study dropout, MNAR, $\phi = 1$, $\%_{TD_{PL}} = 27\%$

		$\%_{TD_{PL}} = 27\%$								
		$\%_{TD_A} - \%_{TD_{PL}} = -27\%$			$\%_{TD_A} - \%_{TD_{PL}} = 0\%$			$\%_{TD_A} - \%_{TD_{PL}} = +23\%$		
		type I error rate [CI]	bias	N_{conv}	type I error rate [CI]	bias	N_{conv}	type I error rate [CI]	bias	N_{conv}
no tr. effect	GLM NB	0.15 [0.14;0.16]	-0.09	4953	0.042 [0.036;0.048]	0.00	4314	0.137 [0.128;0.146]	-0.05	5053
	GLM Poisson	0.152 [0.145;0.159]	-0.09	10000	0.045 [0.041;0.049]	0.00	10000	0.144 [0.137;0.151]	-0.05	10000
	Log rank test	0.07 [0.065;0.075]	-	10000	0.05 [0.046;0.054]	-	10000	0.092 [0.086;0.098]	-	10000
	Cox model	0.068 [0.063;0.073]	0.02	10000	0.05 [0.046;0.054]	0.01	10000	0.091 [0.085;0.097]	-0.03	10000
	LR test	0.071 [0.066;0.076]	0.02	10000	0.052 [0.048;0.056]	0.01	10000	0.096 [0.09;0.102]	-0.03	10000
	GF model	0.066 [0.061;0.071]	0.02	10000	0.048 [0.044;0.052]	0.01	10000	0.086 [0.081;0.091]	-0.03	10000
	SGF model	0.147 [0.14;0.154]	0.06	9949	0.044 [0.04;0.048]	0.01	9999	0.139 [0.132;0.146]	0.00	9975
	χ^2-test	0.056 [0.051;0.061]	0.04	10000	0.041 [0.037;0.045]	0.00	10000	0.079 [0.074;0.084]	-0.06	10000
		$\%_{TD_A} - \%_{TD_{PL}} = -27\%$			$\%_{TD_A} - \%_{TD_{PL}} = -4\%$			$\%_{TD_A} - \%_{TD_{PL}} = +18\%$		
		power	bias	N_{conv}	power	bias	N_{conv}	power	bias	N_{conv}
with tr. effect	GLM NB	0.973	-0.09	4966	0.975	0.00	4746	0.843	-0.04	5015
	GLM Poisson	0.974	-0.09	10000	0.976	0.00	10000	0.850	-0.04	10000
	Log rank test	0.928	-	10000	0.917	-	10000	0.819	-	10000
	Cox model	0.927	0.01	10000	0.916	0.01	10000	0.817	-0.03	10000
	LR test	0.930	0.01	10000	0.918	0.00	10000	0.818	-0.03	10000
	GF model	0.923	-0.02	10000	0.912	-0.02	10000	0.808	-0.04	10000
	SGF model	0.972	0.05	10000	0.975	0.03	9999	0.847	0.02	10000
	χ^2-test	0.887	-0.07	10000	0.868	-0.08	10000	0.736	-0.13	10000

Table 28: Comparison of methods estimating the tr.-policy for scenarios without study dropout, MNAR, $\phi = 1$, $\%_{TD_{PL}} = 60\%$

		$\%_{TD_{PL}} = 60\%$								
		$\%_{TD_A} - \%_{TD_{PL}} = -21\%$			$\%_{TD_A} - \%_{TD_{PL}} = 0\%$			$\%_{TD_A} - \%_{TD_{PL}} = +18\%$		
		type I error rate [CI]	bias	N_{conv}	type I error rate [CI]	bias	N_{conv}	type I error rate [CI]	bias	N_{conv}
no tr. effect	GLM NB	0.353 [0.34;0.366]	-0.13	4819	0.044 [0.037;0.051]	0.00	3816	0.195 [0.184;0.206]	-0.08	4904
	GLM Poisson	0.362 [0.353;0.371]	-0.13	10000	0.043 [0.039;0.047]	0.00	10000	0.202 [0.194;0.21]	-0.08	10000
	Log rank test	0.136 [0.129;0.143]	-	10000	0.052 [0.048;0.056]	-	10000	0.109 [0.103;0.115]	-	10000
	Cox model	0.134 [0.127;0.141]	-0.01	10000	0.051 [0.047;0.055]	0.00	10000	0.107 [0.101;0.113]	-0.04	10000
	LR test	0.14 [0.133;0.147]	-0.01	10000	0.054 [0.05;0.058]	0.00	10000	0.112 [0.106;0.118]	-0.04	10000
	GF model	0.13 [0.123;0.137]	-0.01	10000	0.05 [0.046;0.054]	0.00	10000	0.104 [0.098;0.11]	-0.04	10000
	SGF model	0.356 [0.347;0.365]	0.05	9977	0.042 [0.038;0.046]	0.00	9997	0.198 [0.19;0.206]	0.01	9993
	χ^2-test	0.116 [0.11;0.122]	0.01	10000	0.041 [0.037;0.045]	0.00	10000	0.088 [0.082;0.094]	-0.06	10000
		$\%_{TD_A} - \%_{TD_{PL}} = -26\%$			$\%_{TD_A} - \%_{TD_{PL}} = -6\%$			$\%_{TD_A} - \%_{TD_{PL}} = +14\%$		
		power	bias	N_{conv}	power	bias	N_{conv}	power	bias	N_{conv}
with tr. effect	GLM NB	0.667	-0.12	4380	0.760	0.01	4110	0.456	-0.07	4437
	GLM Poisson	0.672	-0.12	10000	0.769	0.01	10000	0.449	-0.07	10000
	Log rank test	0.705	-	10000	0.736	-	10000	0.541	-	10000
	Cox model	0.703	-0.01	10000	0.734	0.01	10000	0.539	-0.03	10000
	LR test	0.708	-0.01	10000	0.739	0.01	10000	0.541	-0.03	10000
	GF model	0.693	-0.03	10000	0.724	-0.01	10000	0.525	-0.05	10000
	SGF model	0.668	0.07	9994	0.767	0.06	9995	0.446	0.05	9995
	χ^2-test	0.608	-0.07	10000	0.618	-0.05	10000	0.421	-0.11	10000

Table 29: Comparison of methods estimating the tr.-policy for scenarios without study dropout, MNAR, $\phi = 2$, $\%_{TD_{PL}} = 28\%$

		$\%_{TD_{PL}} = 28\%$								
		$\%_{TD_A} - \%_{TD_{PL}} = -28\%$			$\%_{TD_A} - \%_{TD_{PL}} = 0\%$			$\%_{TD_A} - \%_{TD_{PL}} = +22\%$		
		type I error rate [CI]	bias	N_{conv}	type I error rate [CI]	bias	N_{conv}	type I error rate [CI]	bias	N_{conv}
no tr. effect	GLM NB	0.298 [0.289;0.307]	0.00	10000	0.053 [0.049;0.057]	0.00	10000	0.367 [0.358;0.376]	0.00	9940
	GLM Poisson	0.46 [0.45;0.47]	0.00	10000	0.114 [0.108;0.12]	0.00	10000	0.486 [0.476;0.496]	0.00	10000
	Log rank test	0.106 [0.1;0.112]	-	10000	0.048 [0.044;0.052]	-	10000	0.259 [0.25;0.268]	-	10000
	Cox model	0.104 [0.098;0.11]	0.01	10000	0.048 [0.044;0.052]	0.01	10000	0.257 [0.248;0.266]	-0.02	10000
	LR test	0.128 [0.121;0.135]	0.01	10000	0.059 [0.054;0.064]	0.01	10000	0.274 [0.265;0.283]	-0.02	9999
	GF model	0.097 [0.091;0.103]	0.02	10000	0.047 [0.043;0.051]	0.01	10000	0.222 [0.214;0.23]	-0.02	9997
	SGF model	0.298 [0.289;0.307]	0.15	10000	0.053 [0.049;0.057]	0.01	10000	0.362 [0.353;0.371]	-0.01	9980
	χ^2-test	0.1 [0.094;0.106]	0.00	10000	0.041 [0.037;0.045]	0.00	10000	0.273 [0.264;0.282]	0.03	10000
		$\%_{TD_A} - \%_{TD_{PL}} = -28\%$			$\%_{TD_A} - \%_{TD_{PL}} = -4\%$			$\%_{TD_A} - \%_{TD_{PL}} = +17\%$		
		power	bias	N_{conv}	power	bias	N_{conv}	power	bias	N_{conv}
with tr. effect	GLM NB	0.614	0.00	10000	0.844	0.00	10000	0.622	-0.01	9936
	GLM Poisson	0.756	0.00	10000	0.921	0.00	10000	0.730	0.00	10000
	Log rank test	0.833	-	10000	0.850	-	10000	0.755	-	10000
	Cox model	0.832	0.00	10000	0.848	0.00	10000	0.753	-0.03	10000
	LR test	0.850	-0.02	10000	0.863	-0.02	10000	0.766	-0.05	10000
	GF model	0.829	-0.14	10000	0.843	-0.13	10000	0.745	-0.16	9986
	SGF model	0.614	0.13	10000	0.844	0.02	10000	0.622	0.00	9986
	χ^2-test	0.792	-0.01	10000	0.805	-0.01	10000	0.724	0.00	10000

Table 30: Comparison of methods estimating the tr.-policy for scenarios without study dropout, MNAR, $\phi = 2$, $\%_{TD_{PL}} = 60\%$

		$\%_{TD_{PL}} = 60\%$								
		$\%_{TD_A} - \%_{TD_{PL}} = -21\%$			$\%_{TD_A} - \%_{TD_{PL}} = 0\%$			$\%_{TD_A} - \%_{TD_{PL}} = +18\%$		
		type I error rate [CI]	bias	N_{conv}	type I error rate [CI]	bias	N_{conv}	type I error rate [CI]	bias	N_{conv}
no tr. effect	GLM NB	0.600 [0.59;0.61]	0.00	9772	0.052 [0.048;0.056]	0.00	9852	0.489 [0.479;0.499]	-0.01	9482
	GLM Poisson	0.653 [0.644;0.662]	0.00	10000	0.079 [0.074;0.084]	0.00	10000	0.562 [0.552;0.572]	0.00	10000
	Log rank test	0.363 [0.354;0.372]	-	10000	0.046 [0.042;0.05]	-	10000	0.398 [0.388;0.408]	-	10000
	Cox model	0.361 [0.352;0.37]	-0.01	10000	0.045 [0.041;0.049]	0.00	10000	0.395 [0.385;0.405]	-0.03	10000
	LR test	0.385 [0.375;0.395]	0.00	9999	0.054 [0.05;0.058]	0.00	10000	0.416 [0.406;0.426]	-0.04	9999
	GF model	0.315 [0.306;0.324]	0.02	10000	0.045 [0.041;0.049]	0.00	10000	0.345 [0.336;0.354]	-0.03	9999
	SGF model	0.582 [0.572;0.592]	0.17	9952	0.052 [0.048;0.056]	0.00	9998	0.492 [0.482;0.502]	0.02	9898
	χ^2-test	0.374 [0.365;0.383]	0.02	10000	0.039 [0.035;0.043]	0.00	10000	0.407 [0.397;0.417]	0.04	10000
		$\%_{TD_A} - \%_{TD_{PL}} = -25\%$			$\%_{TD_A} - \%_{TD_{PL}} = -5\%$			$\%_{TD_A} - \%_{TD_{PL}} = +15\%$		
		power	bias	N_{conv}	power	bias	N_{conv}	power	bias	N_{conv}
with tr. effect	GLM NB	0.364	-0.01	9764	0.618	0.00	9897	0.382	-0.03	9439
	GLM Poisson	0.445	0.00	10000	0.695	0.00	10000	0.441	0.00	10000
	Log rank test	0.579	-	10000	0.657	-	10000	0.548	-	10000
	Cox model	0.576	-0.02	10000	0.653	0.00	10000	0.546	-0.04	10000
	LR test	0.589	-0.03	10000	0.675	-0.01	10000	0.552	-0.06	10000
	GF model	0.576	-0.13	9997	0.654	-0.12	10000	0.529	-0.16	9994
	SGF model	0.366	0.17	9972	0.622	0.04	10000	0.386	0.04	9922
	χ^2-test	0.546	-0.01	10000	0.564	-0.02	10000	0.520	0.00	10000

7.3.2.2 Results for the treatment-policy estimand study dropout of 50% of patients with treatment discontinuation with MNAR

The table 31 on page 109 shows the results for the scenarios under MNAR, with study dropout of 50% of the patients with treatment discontinuation, $\phi = 1$ and $\%_{TD_{PL}} = 27\%$. The results for the corresponding scenarios with $\phi = 1$ and $\%_{TD_{PL}} = 60\%$ are listed in table 32 on page 110. In table 33 on page 111, the results of the scenarios with $\phi = 2$ and $\%_{TD_{PL}} = 28\%$ are collected. The results of $\phi = 2$ and $\%_{TD_{PL}} = 60\%$ are provided in table 34 on page 112.

If $\phi = 1$, $\%_{TD_{PL}} = 27\%$ and $\%_{TD_A} - \%_{TD_{PL}} = 0\%$, the type I error rate of each interpretable method denotes a value smaller than 5%. But only the Log-linear Poisson model and the χ^2-test provide a bias of zero. The Log-linear Poisson model and the Shared gamma frailty model possess the largest power with a value of 0.980 for the Log-linear Poisson Model and a value of 0.981 for the Gamma frailty model. Both methods estimate the treatment effect without any bias. In the scenario with $\phi = 1$, $\%_{TD_{PL}} = 60\%$ and $\%_{TD_A} - \%_{TD_{PL}} = 0\%$, the type I error rate of the Cox proportional hazards model, the Gamma frailty model, the Shared gamma frailty model and of the χ^2-test are smaller than 5%. Each of these methods reveals a bias of zero. The Shared gamma frailty model returns the largest power. If $\phi = 2$, $\%_{TD_{PL}} = 28\%$ and $\%_{TD_A} - \%_{TD_{PL}} = 0\%$, the Log rank test, the Cox proportional hazards model, the Gamma frailty model and the χ^2-test reveal an acceptable type I error rate. The Log rank test and the Cox proportional hazards model manifest the largest power with a value of 0.862 and 0.860 respectively. Both methods show the same bias with a value of -0.02. The last scenario under MNAR and with a study dropout of 50%, where an acceptable type I error rate is found, has the conditions of $\phi = 2$, $\%_{TD_{PL}} = 60\%$ and $\%_{TD_A} - \%_{TD_{PL}} = 0\%$. In this context, the Log rank test, the Cox proportional hazards model, the Gamma frailty model, the Shared gamma frailty model and the χ^2-test hold the requirement of a type I error rate lower than 5%. These have a value of 0.045 for the Log rank test, and a value of 0.044 with a bias of 0.01 for the Cox proportional hazards model and the Gamma frailty model. The type I error rate of the Shared gamma frailty model is 0.05 with a bias of 0.01. The χ^2-test delivers a type I error rate of 0.035 without bias. Hereby the Log rank test possesses the maximal power of these methods with a value of 0.696. Actually, the Cox proportional hazards model should obtain the same power, but because of different implementations in the statistical software a little different value of 0.693 with a bias of -0.04 results.

The results for the bias of the methods are similar to those in the MCAR case. The results under MNAR are plotted in figure 13. In analogy to the scenarios without study

dropout, the Log-linear Poisson model clearly overestimates the true treatment effect in scenarios without overdispersion and a negative difference between the difference of patients with treatment discontinuation in both trial groups. One further parallelism to the scenarios without study dropout exists in the way that the Shared gamma frailty model underestimates the treatment-policy estimand if this difference is negative. Otherwise, if it is positive, this model reveals a positive bias.

Table 31: Comparison of methods estimating the tr.-policy for scenarios with study dropout of 50%, MNAR, $\phi = 1$, $\%_{TD_{PL}} = 27\%$

		$\%_{TD_{PL}} = 27\%$								
		$\%_{TD_A} - \%_{TD_{PL}} = -27\%$			$\%_{TD_A} - \%_{TD_{PL}} = 0\%$			$\%_{TD_A} - \%_{TD_{PL}} = +23\%$		
		type I error rate [CI]	bias	N_{conv}	type I error rate [CI]	bias	N_{conv}	type I error rate [CI]	bias	N_{conv}
no tr. effect	GLM NB	0.057 [0.051;0.063]	-0.09	5656	0.049 [0.043;0.055]	0.00	5578	0.259 [0.248;0.27]	-0.05	6586
	GLM Poisson	0.066 [0.061;0.071]	-0.09	10000	0.049 [0.045;0.053]	0.00	10000	0.206 [0.198;0.214]	-0.05	10000
	Log rank test	0.063 [0.058;0.068]	-	10000	0.049 [0.045;0.053]	-	10000	0.079 [0.074;0.084]	-	10000
	Cox model	0.062 [0.057;0.067]	0.01	10000	0.049 [0.045;0.053]	0.01	10000	0.078 [0.073;0.083]	-0.03	10000
	LR test	0.064 [0.059;0.069]	0.01	10000	0.05 [0.046;0.054]	0.01	10000	0.079 [0.074;0.084]	-0.03	10000
	GF model	0.059 [0.054;0.064]	0.01	10000	0.047 [0.043;0.051]	0.01	9999	0.074 [0.069;0.079]	-0.03	9999
	SGF model	0.097 [0.091;0.103]	0.03	9996	0.046 [0.042;0.05]	0.01	9997	0.081 [0.076;0.086]	0.02	9999
	χ^2-test	0.115 [0.109;0.121]	0.14	10000	0.037 [0.033;0.041]	0.00	10000	0.298 [0.289;0.307]	-0.16	10000
		$\%_{TD_A} - \%_{TD_{PL}} = -27\%$			$\%_{TD_A} - \%_{TD_{PL}} = -4\%$			$\%_{TD_A} - \%_{TD_{PL}} = +18\%$		
		power	bias	N_{conv}	power	bias	N_{conv}	power	bias	N_{conv}
with tr. effect	GLM NB	0.994	-0.09	5808	0.977	0.01	6056	0.775	-0.01	6428
	GLM Poisson	0.993	-0.09	10000	0.980	0.00	10000	0.800	-0.04	10000
	Log rank test	0.933	-	10000	0.930	-	10000	0.823	-	10000
	Cox model	0.932	0.00	10000	0.929	-0.02	10000	0.821	-0.04	10000
	LR test	0.934	0.00	10000	0.931	-0.02	10000	0.831	-0.05	10000
	GF model	0.930	-0.02	9989	0.926	-0.04	9990	0.825	-0.06	9990
	SGF model	0.983	0.02	10000	0.981	0.00	9999	0.789	0.02	9998
	χ^2-test	0.780	0.04	10000	0.865	-0.07	10000	0.824	-0.22	10000

Table 32: Comparison of methods estimating the tr.-policy for scenarios with study dropout of 50%, MNAR, $\phi = 1$, $\%_{TD_{PL}} = 60\%$

		$\%_{TD_{PL}} = 60\%$								
		$\%_{TD_A} - \%_{TD_{PL}} = -21\%$			$\%_{TD_A} - \%_{TD_{PL}} = 0\%$			$\%_{TD_A} - \%_{TD_{PL}} = +18\%$		
		type I error rate [CI]	bias	N_{conv}	type I error rate [CI]	bias	N_{conv}	type I error rate [CI]	bias	N_{conv}
no tr. effect	GLM NB	0.193 [0.184;0.202]	-0.14	7017	0.047 [0.042;0.052]	0.00	7035	0.416 [0.405;0.427]	-0.08	7973
	GLM Poisson	0.17 [0.163;0.177]	-0.13	10000	0.051 [0.047;0.055]	0.00	10000	0.371 [0.362;0.38]	-0.08	10000
	Log rank test	0.11 [0.104;0.116]	-	10000	0.052 [0.048;0.056]	-	10000	0.094 [0.088;0.1]	-	10000
	Cox model	0.109 [0.103;0.115]	-0.02	10000	0.05 [0.046;0.054]	0.00	10000	0.093 [0.087;0.099]	-0.04	10000
	LR test	0.114 [0.108;0.12]	-0.02	10000	0.054 [0.05;0.058]	0.00	10000	0.096 [0.09;0.102]	-0.04	10000
	GF model	0.109 [0.103;0.115]	-0.01	9998	0.048 [0.044;0.052]	0.00	9994	0.09 [0.084;0.096]	-0.04	9999
	SGF model	0.211 [0.203;0.219]	0.03	9996	0.048 [0.044;0.052]	0.00	9998	0.157 [0.15;0.164]	0.04	9997
	χ^2-test	0.487 [0.477;0.497]	0.13	10000	0.038 [0.034;0.042]	0.00	10000	0.48 [0.47;0.49]	-0.19	10000
		$\%_{TD_A} - \%_{TD_{PL}} = -26\%$			$\%_{TD_A} - \%_{TD_{PL}} = -6\%$			$\%_{TD_A} - \%_{TD_{PL}} = +14\%$		
		power	bias	N_{conv}	power	bias	N_{conv}	power	bias	N_{conv}
with tr. effect	GLM NB	0.819	-0.12	7147	0.815	0.00	7672	0.647	-0.04	7872
	GLM Poisson	0.835	-0.12	10000	0.834	0.01	10000	0.652	-0.07	10000
	Log rank test	0.716	-	10000	0.776	-	10000	0.556	-	10000
	Cox model	0.714	-0.03	10000	0.774	-0.03	10000	0.555	-0.06	10000
	LR test	0.721	-0.03	10000	0.779	-0.03	10000	0.566	-0.06	10000
	GF model	0.712	-0.05	9995	0.772	-0.05	9996	0.559	-0.08	9993
	SGF model	0.726	0.02	10000	0.827	0.01	10000	0.456	0.04	10000
	χ^2-test	0.393	0.06	10000	0.580	-0.02	10000	0.587	-0.24	10000

Table 33: Comparison of methods estimating the tr.-policy for scenarios with study dropout of 50%, MNAR, $\phi = 2$, $\%_{TD_{PL}} = 28\%$

		$\%_{TD_{PL}} = 28\%$								
		$\%_{TD_A} - \%_{TD_{PL}} = -28\%$			$\%_{TD_A} - \%_{TD_{PL}} = 0\%$			$\%_{TD_A} - \%_{TD_{PL}} = +22\%$		
		type I error rate [CI]	bias	N_{conv}	type I error rate [CI]	bias	N_{conv}	type I error rate [CI]	bias	N_{conv}
no tr. effect	GLM NB	0.17 [0.163;0.177]	-0.02	10000	0.054 [0.05;0.058]	0.00	10000	0.296 [0.287;0.305]	0.00	9998
	GLM Poisson	0.333 [0.324;0.342]	0.00	10000	0.118 [0.112;0.124]	0.00	10000	0.421 [0.411;0.431]	0.00	10000
	Log rank test	0.082 [0.077;0.087]	-	10000	0.047 [0.043;0.051]	-	10000	0.226 [0.218;0.234]	-	10000
	Cox model	0.081 [0.076;0.086]	0.00	10000	0.046 [0.042;0.05]	0.00	10000	0.224 [0.216;0.232]	0.03	10000
	LR test	0.093 [0.087;0.099]	-0.01	10000	0.057 [0.052;0.062]	0.00	10000	0.229 [0.221;0.237]	0.04	9999
	GF model	0.077 [0.072;0.082]	0.01	9998	0.046 [0.042;0.05]	0.00	10000	0.199 [0.191;0.207]	0.03	10000
	SGF model	0.204 [0.196;0.212]	0.10	10000	0.051 [0.047;0.055]	-0.00	10000	0.279 [0.27;0.288]	0.06	10000
	χ^2-test	0.112 [0.106;0.118]	0.08	10000	0.042 [0.038;0.046]	0.00	10000	0.265 [0.256;0.274]	-0.11	10000
		$\%_{TD_A} - \%_{TD_{PL}} = -28\%$			$\%_{TD_A} - \%_{TD_{PL}} = -4\%$			$\%_{TD_A} - \%_{TD_{PL}} = +17\%$		
		power	bias	N_{conv}	power	bias	N_{conv}	power	bias	N_{conv}
with tr. effect	GLM NB	0.761	-0.03	10000	0.848	0.00	10000	0.708	0.00	9994
	GLM Poisson	0.843	0.00	10000	0.929	0.00	10000	0.786	0.00	10000
	Log rank test	0.853	-	10000	0.862	-	10000	0.758	-	10000
	Cox model	0.852	-0.01	10000	0.860	-0.02	10000	0.756	0.01	10000
	LR test	0.875	-0.04	10000	0.878	-0.04	10000	0.770	0.00	10000
	GF model	0.845	-0.15	10000	0.856	-0.16	10000	0.752	-0.11	10000
	SGF model	0.698	0.08	10000	0.856	-0.01	10000	0.659	0.05	10000
	χ^2-test	0.720	0.07	10000	0.805	-0.01	10000	0.692	-0.12	10000

Table 34: Comparison of methods estimating the tr.-policy for scenarios with study dropout of 50%, MNAR, $\phi = 2$, $\%_{TD_{PL}} = 60\%$

		$\%_{TD_{PL}} = 60\%$								
		$\%_{TD_A} - \%_{TD_{PL}} = -21\%$			$\%_{TD_A} - \%_{TD_{PL}} = 0\%$			$\%_{TD_A} - \%_{TD_{PL}} = +18\%$		
		type I error rate [CI]	bias	N_{conv}	type I error rate [CI]	bias	N_{conv}	type I error rate [CI]	bias	N_{conv}
no tr. effect	GLM NB	0.431 [0.421;0.441]	-0.03	9989	0.056 [0.051;0.061]	0.00	9994	0.471 [0.461;0.481]	0.00	9983
	GLM Poisson	0.551 [0.541;0.561]	0.00	10000	0.087 [0.081;0.093]	0.00	10000	0.559 [0.549;0.569]	0.00	10000
	Log rank test	0.282 [0.273;0.291]	-	10000	0.045 [0.041;0.049]	-	10000	0.349 [0.34;0.358]	-	10000
	Cox model	0.279 [0.27;0.288]	0.03	10000	0.044 [0.04;0.048]	0.01	10000	0.347 [0.338;0.356]	0.03	10000
	LR test	0.278 [0.269;0.287]	0.03	9999	0.053 [0.049;0.057]	0.01	10000	0.344 [0.335;0.353]	0.04	9999
	GF model	0.245 [0.237;0.253]	0.05	10000	0.044 [0.04;0.048]	0.01	10000	0.307 [0.298;0.316]	0.04	10000
	SGF model	0.485 [0.475;0.495]	0.16	10000	0.05 [0.046;0.054]	0.01	10000	0.44 [0.43;0.45]	0.11	10000
	χ^2-test	0.408 [0.398;0.418]	0.07	10000	0.035 [0.031;0.039]	0.00	10000	0.409 [0.399;0.419]	-0.15	10000
		$\%_{TD_A} - \%_{TD_{PL}} = -25\%$			$\%_{TD_A} - \%_{TD_{PL}} = -5\%$			$\%_{TD_A} - \%_{TD_{PL}} = +15\%$		
		power	bias	N_{conv}	power	bias	N_{conv}	power	bias	N_{conv}
with tr. effect	GLM NB	0.540	-0.04	9983	0.663	-0.01	9992	0.525	0.00	9957
	GLM Poisson	0.573	0.00	10000	0.752	0.00	10000	0.563	0.00	10000
	Log rank test	0.598	-	10000	0.696	-	10000	0.552	-	10000
	Cox model	0.596	0.01	10000	0.693	-0.04	10000	0.550	-0.01	10000
	LR test	0.630	-0.01	10000	0.720	-0.05	10000	0.562	-0.02	10000
	GF model	0.597	-0.10	9999	0.692	-0.16	9998	0.541	-0.12	9999
	SGF model	0.396	0.15	10000	0.678	-0.02	10000	0.400	0.08	10000
	χ^2-test	0.400	0.06	10000	0.544	-0.01	10000	0.455	-0.18	10000

7.3.2.3 Results for the treatment-policy estimand with a study dropout of 100% of patients with treatment discontinuation with MNAR

Table 35 on page 114 contains the results of the scenarios with a study dropout of 100% of the patients with treatment discontinuation and MNAR, $\phi = 1$ and $\%_{TD_{PL}} = 27\%$. The results for the according scenario with $\%_{TD_{PL}} = 60\%$ are provided in table 36 on page 115. If $\phi = 2$ and $\%_{TD_{PL}} = 28\%$, the table 37 on page 116 shows the relevant results. The results of the scenario with $\phi = 2$ and $\%_{TD_{PL}} = 60\%$ are listed in table 38 on page 117.

In analogy to the scenarios with a study dropout of 100% under MCAR, the calculated type I error rates and power are the same as those of the hypothetical estimand under MNAR. But the results of the bias calculations are different.

Finally, the results of the bias calculations of each method in the scenarios with a study dropout of 100% of the patients with treatment discontinuation are plotted in figure 13. Hereby, it can be notified that these results are similar to those of the according scenarios with a study dropout of 100% under MCAR. Furthermore, it attracts the attention that the bias of the Log-linear negative-binomial model and the one of the Log-linear Poisson model is not identical any more, if $\phi = 2$.

Table 35: Comparison of methods estimating the tr.-policy for scenarios with study dropout of 100%, MNAR, $\phi = 1$, $\%_{TD_{PL}} = 27\%$

		$\%_{TD_{PL}} = 27\%$								
		$\%_{TD_A} - \%_{TD_{PL}} = -27\%$			$\%_{TD_A} - \%_{TD_{PL}} = 0\%$			$\%_{TD_A} - \%_{TD_{PL}} = +23\%$		
		type I error rate [CI]	bias	N_{conv}	type I error rate [CI]	bias	N_{conv}	type I error rate [CI]	bias	N_{conv}
no tr. effect	GLM NB	0.089 [0.081;0.097]	0.04	5035	0.037 [0.031;0.043]	0.00	4349	0.08 [0.072;0.088]	-0.17	5000
	GLM Poisson	0.098 [0.092;0.104]	0.04	10000	0.044 [0.04;0.048]	0.00	10000	0.085 [0.08;0.09]	-0.18	10000
	Log rank test	0.055 [0.051;0.059]	-	10000	0.049 [0.045;0.053]	-	10000	0.391 [0.381;0.401]	-	10000
	Cox model	0.054 [0.05;0.058]	-0.07	10000	0.048 [0.044;0.052]	0.01	10000	0.391 [0.381;0.401]	0.18	10000
	LR test	0.051 [0.047;0.055]	-0.07	10000	0.046 [0.042;0.05]	0.01	10000	0.3 [0.291;0.309]	0.11	10000
	GF model	0.048 [0.044;0.052]	-0.07	10000	0.043 [0.039;0.047]	0.01	10000	0.212 [0.204;0.22]	0.10	9642
	SGF model	0.093 [0.087;0.099]	0.03	9995	0.043 [0.039;0.047]	0.01	9999	0.081 [0.076;0.086]	0.02	9996
	χ^2-test	0.329 [0.32;0.338]	0.31	10000	0.039 [0.035;0.043]	0.00	10000	0.507 [0.497;0.517]	-0.45	10000
		$\%_{TD_A} - \%_{TD_{PL}} = -27\%$			$\%_{TD_A} - \%_{TD_{PL}} = -4\%$			$\%_{TD_A} - \%_{TD_{PL}} = +18\%$		
		power	bias	N_{conv}	power	bias	N_{conv}	power	bias	N_{conv}
with tr. effect	GLM NB	0.983	0.04	5015	0.987	-0.04	4446	0.869	-0.25	4995
	GLM Poisson	0.980	0.04	10000	0.988	-0.04	10000	0.867	-0.27	10000
	Log rank test	0.982	-	10000	0.940	-	10000	0.815	-	10000
	Cox model	0.982	-0.08	10000	0.939	-0.02	10000	0.814	0.17	10000
	LR test	0.981	-0.08	10000	0.937	-0.01	10000	0.780	0.09	10000
	GF model	0.980	-0.09	9999	0.932	-0.01	9999	0.785	0.03	9048
	SGF model	0.980	0.02	9999	0.988	-0.04	10000	0.867	-0.11	9997
	χ^2-test	0.529	0.21	10000	0.864	-0.06	10000	0.718	-0.55	10000

Table 36: Comparison of methods estimating the tr.-policy for scenarios with study dropout of 100%, MNAR, $\phi = 1$, $\%_{TD_{PL}} = 60\%$

		$\%_{TD_{PL}} = 60\%$								
		$\%_{TD_A} - \%_{TD_{PL}} = -21\%$			$\%_{TD_A} - \%_{TD_{PL}} = 0\%$			$\%_{TD_A} - \%_{TD_{PL}} = +18\%$		
		type I error rate [CI]	bias	N_{conv}	type I error rate [CI]	bias	N_{conv}	type I error rate [CI]	bias	N_{conv}
no tr. effect	GLM NB	0.247 [0.234;0.26]	0.12	4467	0.037 [0.031;0.043]	0.00	3465	0.193 [0.181;0.205]	-0.19	4318
	GLM Poisson	0.252 [0.243;0.261]	0.09	10000	0.041 [0.037;0.045]	0.00	10000	0.204 [0.196;0.212]	-0.24	10000
	Log rank test	0.327 [0.318;0.336]	-	10000	0.053 [0.049;0.057]	-	10000	0.622 [0.612;0.632]	-	10000
	Cox model	0.325 [0.316;0.334]	-0.02	10000	0.053 [0.049;0.057]	0.00	10000	0.621 [0.611;0.631]	0.26	10000
	LR test	0.227 [0.219;0.235]	-0.07	10000	0.041 [0.037;0.045]	0.00	10000	0.451 [0.441;0.461]	0.14	10000
	GF model	0.191 [0.183;0.199]	-0.08	9693	0.038 [0.034;0.042]	0.00	10000	0.379 [0.369;0.389]	0.13	9244
	SGF model	0.244 [0.236;0.252]	0.05	9991	0.041 [0.037;0.045]	0.00	9997	0.199 [0.191;0.207]	0.05	9983
	χ^2-test	0.858 [0.851;0.865]	0.20	10000	0.043 [0.039;0.047]	0.00	10000	0.782 [0.774;0.79]	-0.63	10000
		$\%_{TD_A} - \%_{TD_{PL}} = -26\%$			$\%_{TD_A} - \%_{TD_{PL}} = -6\%$			$\%_{TD_A} - \%_{TD_{PL}} = +14\%$		
		power	bias	N_{conv}	power	bias	N_{conv}	power	bias	N_{conv}
with tr. effect	GLM NB	0.737	0.04	4138	0.925	-0.11	3292	0.576	-0.39	4066
	GLM Poisson	0.721	0.02	10000	0.921	-0.10	10000	0.558	-0.41	10000
	Log rank test	0.854	-	10000	0.839	-	10000	0.724	-	10000
	Cox model	0.853	-0.04	10000	0.837	-0.04	10000	0.723	0.26	10000
	LR test	0.836	-0.09	10000	0.811	-0.02	10000	0.641	0.13	10000
	GF model	0.874	-0.13	9283	0.801	-0.02	9992	0.718	0.07	8303
	SGF model	0.719	-0.04	9991	0.921	-0.12	9997	0.557	-0.19	9988
	χ^2-test	0.282	0.11	10000	0.630	-0.06	10000	0.615	-0.78	10000

Table 37: Comparison of methods estimating the tr.-policy for scenarios with study dropout of 100%, MNAR, $\phi = 2$, $\%_{TD_{PL}} = 28\%$

		$\%_{TD_{PL}} = 28\%$								
		$\%_{TD_A} - \%_{TD_{PL}} = -28\%$			$\%_{TD_A} - \%_{TD_{PL}} = 0\%$			$\%_{TD_A} - \%_{TD_{PL}} = +22\%$		
		type I error rate [CI]	bias	N_{conv}	type I error rate [CI]	bias	N_{conv}	type I error rate [CI]	bias	N_{conv}
no tr. effect	GLM NB	0.122 [0.116;0.128]	0.10	10000	0.055 [0.051;0.059]	0.00	10000	0.113 [0.107;0.119]	-0.14	9999
	GLM Poisson	0.315 [0.306;0.324]	0.13	10000	0.121 [0.115;0.127]	0.00	10000	0.265 [0.256;0.274]	-0.11	10000
	Log rank test	0.073 [0.068;0.078]	-	10000	0.045 [0.041;0.049]	-	10000	0.375 [0.366;0.384]	-	10000
	Cox model	0.072 [0.067;0.077]	-0.08	10000	0.045 [0.041;0.049]	0.00	10000	0.374 [0.365;0.383]	0.29	9999
	LR test	0.086 [0.081;0.091]	-0.09	10000	0.054 [0.05;0.058]	0.00	10000	0.362 [0.353;0.371]	0.23	9999
	GF model	0.067 [0.062;0.072]	-0.08	10000	0.045 [0.041;0.049]	0.00	10000	0.271 [0.262;0.28]	0.21	9517
	SGF model	0.12 [0.114;0.126]	0.07	10000	0.052 [0.048;0.056]	0.00	10000	0.11 [0.104;0.116]	0.11	10000
	χ^2-test	0.229 [0.221;0.237]	0.22	10000	0.042 [0.038;0.046]	0.00	10000	0.434 [0.424;0.444]	-0.39	10000
		$\%_{TD_A} - \%_{TD_{PL}} = -28\%$			$\%_{TD_A} - \%_{TD_{PL}} = -4\%$			$\%_{TD_A} - \%_{TD_{PL}} = +17\%$		
		power	bias	N_{conv}	power	bias	N_{conv}	power	bias	N_{conv}
with tr. effect	GLM NB	0.751	0.09	10000	0.886	-0.06	10000	0.658	-0.23	10000
	GLM Poisson	0.840	0.12	10000	0.946	-0.04	10000	0.720	-0.19	10000
	Log rank test	0.940	-	10000	0.871	-	10000	0.799	-	10000
	Cox model	0.939	-0.10	10000	0.869	-0.02	10000	0.798	0.27	10000
	LR test	0.951	-0.12	10000	0.882	-0.04	10000	0.776	0.18	10000
	GF model	0.931	-0.21	10000	0.865	-0.11	10000	0.792	0.07	9399
	SGF model	0.748	0.05	10000	0.883	-0.07	10000	0.655	-0.04	10000
	χ^2-test	0.514	0.21	10000	0.810	-0.02	10000	0.702	-0.43	10000

Table 38: Comparison of methods estimating the tr.-policy for scenarios with study dropout of 100%, MNAR, $\phi = 2$, $\%_{TD_{PL}} = 60\%$

		$\%_{TD_{PL}} = 60\%$								
		$\%_{TD_A} - \%_{TD_{PL}} = -21\%$			$\%_{TD_A} - \%_{TD_{PL}} = 0\%$			$\%_{TD_A} - \%_{TD_{PL}} = +18\%$		
		type I error rate [CI]	bias	N_{conv}	type I error rate [CI]	bias	N_{conv}	type I error rate [CI]	bias	N_{conv}
no tr. effect	GLM NB	0.35 [0.341;0.359]	0.18	9966	0.058 [0.053;0.063]	0.00	9970	0.3 [0.291;0.309]	-0.13	9914
	GLM Poisson	0.604 [0.594;0.614]	0.23	10000	0.086 [0.081;0.091]	0.00	10000	0.491 [0.481;0.501]	-0.09	10000
	Log rank test	0.396 [0.386;0.406]	-	10000	0.049 [0.045;0.053]	-	10000	0.618 [0.608;0.628]	-	10000
	Cox model	0.394 [0.384;0.404]	0.06	9999	0.048 [0.044;0.052]	0.01	10000	0.617 [0.607;0.627]	0.41	9999
	LR test	0.361 [0.352;0.37]	0.00	9999	0.048 [0.044;0.052]	0.01	10000	0.553 [0.543;0.563]	0.30	9999
	GF model	0.308 [0.299;0.317]	-0.00	9600	0.043 [0.039;0.047]	0.01	9999	0.475 [0.465;0.485]	0.29	9238
	SGF model	0.34 [0.331;0.349]	0.18	10000	0.054 [0.05;0.058]	0.01	10000	0.294 [0.285;0.303]	0.20	10000
	χ^2-test	0.768 [0.76;0.776]	0.12	10000	0.038 [0.034;0.042]	0.00	10000	0.696 [0.687;0.705]	-0.54	10000
		$\%_{TD_A} - \%_{TD_{PL}} = -25\%$			$\%_{TD_A} - \%_{TD_{PL}} = -5\%$			$\%_{TD_A} - \%_{TD_{PL}} = +15\%$		
		power	bias	N_{conv}	power	bias	N_{conv}	power	bias	N_{conv}
with tr. effect	GLM NB	0.353	0.11	9977	0.792	-0.15	9971	0.286	-0.32	9934
	GLM Poisson	0.365	0.17	10000	0.842	-0.13	10000	0.284	-0.26	10000
	Log rank test	0.811	-	10000	0.741	-	10000	0.706	-	10000
	Cox model	0.809	0.02	10000	0.738	-0.04	10000	0.704	0.38	10000
	LR test	0.804	-0.06	10000	0.738	-0.04	10000	0.654	0.24	10000
	GF model	0.819	-0.12	9537	0.719	-0.07	9999	0.707	0.18	8894
	SGF model	0.343	0.06	10000	0.782	-0.15	10000	0.278	-0.07	9999
	χ^2-test	0.260	0.09	10000	0.623	-0.09	10000	0.573	-0.65	10000

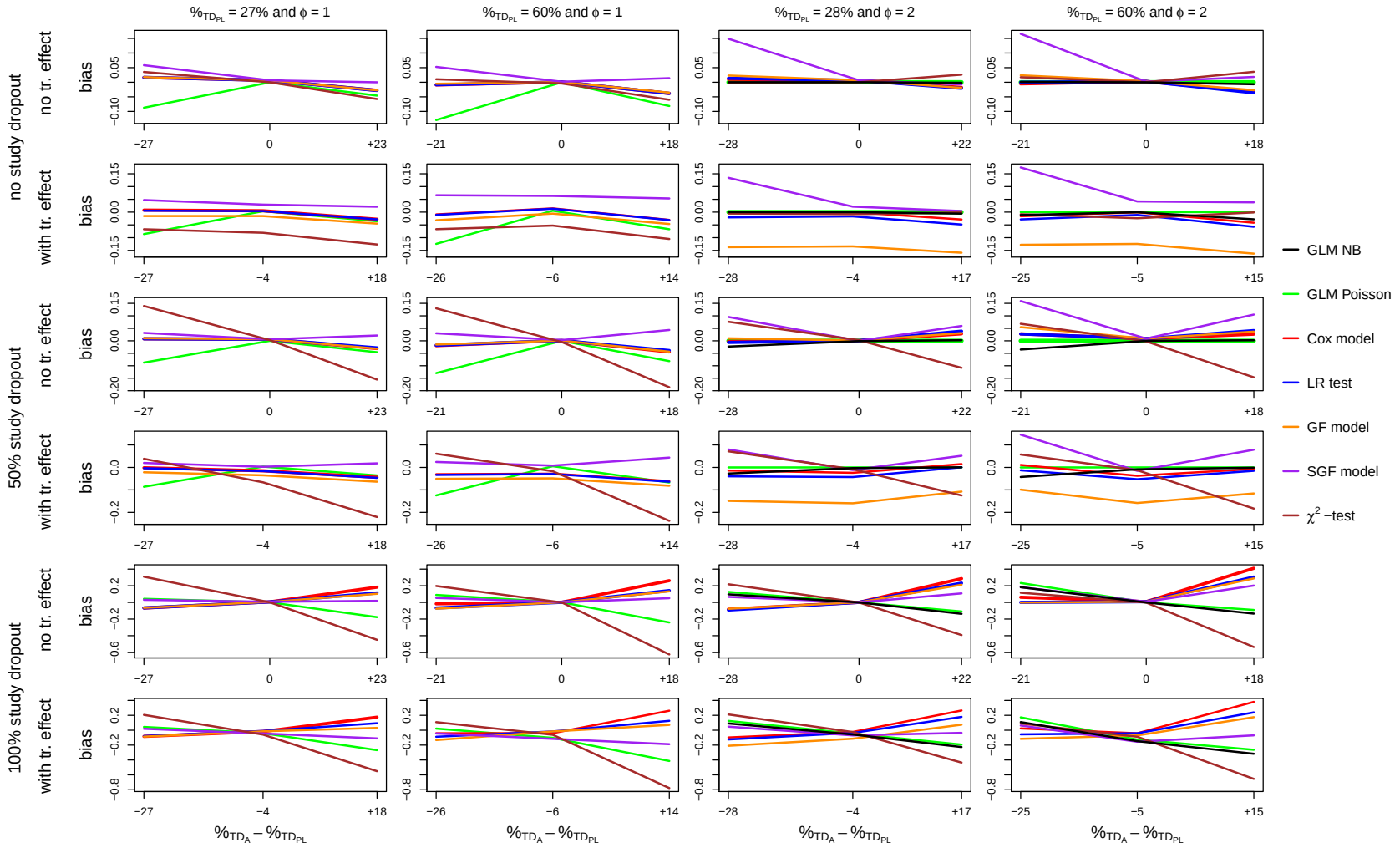

Figure 13: Bias of the Log-linear negative-binomial model (GLM NB), Log-linear Poisson model (GLM Poisson), Cox proportional hazards model (Cox model), Likelihood ratio test (LR test), Gamma frailty model (GF model), Shared gamma frailty model (SGF model) and χ^2-test for treatment-policy estimand depending on the percentage of patients with treatment discontinuation in placebo group ($\%_{TD_{PL}}$), the difference of patients with treatment discontinuation between both trial arms ($\%_{TD_A} - \%_{TD_{PL}}$) and ϕ under MNAR

8 Discussion

The discussion of the results previously presented is structured in two parts. In the first part, the impact of the overdispersion, the different percentages of treatment discontinuation, study dropout and the two mechanisms of treatment discontinuation on the performance of the methods can be generalized from the results presented in section 7. In the second part, advice will be given for the correct evaluation of confirmatory clinical trials in asthma under these different conditions.

In the following first part of the discussion, the impact of the different conditions which is examined in this thesis can be generalized for both hypothetical and treatment-policy estimand. In scenarios with overdispersion, the type I error rate is greatly increased for both the hypothetical and the treatment-policy estimand. Especially, this can be observed in the context of the Log-linear Poisson model, the Log rank test, the Cox proportional hazards model and the Likelihood ratio test. These are the methods which do not consider individual frailties for their evaluation. The type I error rate of the remaining models also rises with the presence of the overdispersion, but the increase is not as high compared to these previous methods. Interestingly, the χ^2-test is not as much influenced as those other methods which do not take into account overdispersion. The power of all methods will shrink, if overdispersion occurs. Whereas the bias is not much changed. To create the link to real confirmatory clinical trials, individual frailties and therefore different severities of the disease are present. Thus, scenarios based on overdispersion capture reality best.

The percentage of patients with treatment discontinuation has the strongest impact on the type I error rate, power and bias, if there is a difference between both trial arms. The type I error rate and bias increase most, if this difference between both groups is positive. Hence, the percentage of treatment discontinuation is larger in treatment group than in placebo group. The count data models, the Shared gamma frailty model and the χ^2-test overestimate the true treatment effect in most of these scenarios. Whereas, the Cox proportional hazards model, the Likelihood ratio test and the Gamma frailty model tend to provide a positive bias. If the difference between both trial arms is smaller than zero, usually the direction of the bias of the methods is reversed. The effect of the difference between the treatment discontinuation in both trail arms on the growing type I error rate and bias is intensified with a larger percentage of patients with treatment discontinuation in the placebo group. The power will decrease the more patients discontinue treatment even if they are followed up afterwards like in the case of treatment-policy estimand. To yield reliable estimates in a real confirmatory clinical trial, the goal should be to keep the percentages of patients with treatment discontinuation as similar as possible in both

trial arms. Of course, the fewer patients stop their treatment the fewer problems will arise during the statistical evaluation.

Study dropout affects the bias. The more patients stop their participation in a clinical trial, the more biased the estimated treatment effects will be. The type I error rate of the time-to-first-event-methods will decrease, if 50% of the patients drop out of the study. For the remaining methods, the type I error rate grows. But if 100% of the patients with treatment discontinuation finish the participation in the study, the type I error rate of all methods except the count data methods increases greatly. In these scenarios with a study dropout of 100% of the patients with treatment discontinuation, the type I error rate is higher than the type I error rate in the scenarios without study dropout. These larger type I error rates are especially visible for the methods evaluating only the time to the first event and the χ^2-test in the scenarios with $\%_{TD_A} - \%_{TD_{PL}} = +20\%$. The type I error rates of the count data methods in scenarios with a study dropout of 100% of the patients with study treatment discontinuation are even smaller than those in scenarios without study dropout. The type I error rates of scenarios without a difference between the percentages of patients with treatment discontinuation in both groups hardly change if the percentage of study dropout varies.

The last influencing factor which was examined in this thesis is the mechanism of treatment discontinuation. It has an impact on the type I error rate, power and bias. Under MNAR, the type I error rate increases and the power shrinks the most, if the count data models, the Shared gamma frailty model and the χ^2-test are used. These methods except the χ^2-test evaluate the total time of the study and encounter the MNAR mechanism the most. The remaining methods just need the information until the first event. If this happens before the potential treatment discontinuation, these models are not contacted to this mechanism. The rise of the type I error rates is reinforced the higher the percentage of patients with treatment discontinuation in the placebo group is. This effect is even strengthened in the context of the estimation of the hypothetical estimand. Furthermore, the MNAR mechanism increases the bias and sometimes turns the direction of the bias of the count data models and the Shared gamma frailty model in scenarios with $\phi = 2$.

In the following second part of the discussion, advice is given for the correct evaluation of a confirmatory clinical trial in asthma. This second part structured in estimation of the hypothetical estimand and in the estimation of the treatment-policy estimand. In general, it must be recognized that the more realistic the scenarios are, the fewer methods are at our disposal. Neither for the hypothetical nor for the treatment-policy estimand, there is no method respecting the type I error rate of lower than 5%, if scenarios are

simulated with overdispersion and with a difference between the percentages of patients with treatment discontinuation in both trial arms under MNAR.

In the context of the scenarios without overdispersion, the results of the Log-linear negative-binomial model must be treated with caution. These limitations must be respected for the evaluation of both the hypothetical estimand and the treatment-policy estimand. In these scenarios the number of converged models is too small to draw valid conclusions. Therefore, this method does not play a role in the proposed strategies of evaluation of scenarios without overdispersion. However, the Log-linear negative-binomial model returns the type I error rates, power and bias, which are similar to those of the Log-linear Poisson model. Although the number of converged models of the Log-linear negative-binomial model is quite small.

Firstly, the results of the hypothetical estimand will be classified. This can be divided according to the scenarios without or with overdispersion and according to the mechanism of treatment discontinuation. If the scenarios are simulated with $\%_{TD_A} - \%_{TD_{PL}} = 0\%$, the χ^2-test always returns an acceptable type I error rate. But unfortunately, its power is lower and its bias is higher compared to the other models respecting the type I error rate. Thus, the χ^2-test does not possess a good capability to recognize new effective asthma drugs, but rather classifies them as ineffective. In general, if no overdispersion occurs, the Shared gamma frailty model is the method of choice. Its advantage consists in a smaller bias. Consequently, for the analysis of more or less unrealistic scenarios with $\phi = 1$, the Shared gamma frailty model is superior to the Log-linear Poisson model. The methods considering the time until the first event deliver a smaller power than the Log-linear Poisson model and the Shared gamma frailty model. Therefore, it is not recommended to use these methods, if the Shared gamma frailty model or the Log-linear Poisson model return a type I error rate of lower than 5%. The Shared gamma frailty model and the Log-linear Poisson model recognize an ineffective drug even if there is a difference between the percentages of patients with treatment discontinuation in both trials arms under MCAR. Especially in this context, the bias of the Shared gamma frailty model is still smaller than the one of the Log-linear Poisson model. As the Shared gamma frailty model considers the time between all succeeding events, it detects ineffective drugs with a smaller bias despite a different proportion of patients in treatment and placebo group.

However, if there is a difference between the proportion of treatment discontinuation in both trial arms under MNAR in scenarios without overdispersion, even the Shared gamma frailty model and the Log-linear Poisson model manifest problems to provide an acceptable type I error rate. In the scenario with $\phi = 1$, $\%_{TD_{PL}} = 27\%$ and $\%_{TD_A} - \%_{TD_{PL}} = -27\%$ under MNAR, the Gamma frailty model is the unique method to respect

the type I error rate of lower or equal to 5%. In all other scenarios under MNAR with a difference of treatment discontinuation between both trial arms, no method is appropriate for the evaluation. Especially, if more patients discontinue treatment in the active group than in the placebo group, the methods seem to detect a treatment effect where actually none exists. This can be reasoned by the fact that less information about the effect of the new drug is available and the information which is collected is overrated.

A problem of evaluation rises if overdispersion plays a role. But this is a more realistic condition than a trial with equal frailties amongst all patients. Under MCAR, it is recommended to use the Shared gamma frailty model or the Log-linear negative-binomial model for the evaluation of scenarios without a difference between the proportions of patients with treatment discontinuation between both trial arms. These two methods reveal the largest power and unbiased estimates by respecting the type I error rate. Although the Shared gamma frailty model and the Log-linear negative-binomial model provide an type I error rate larger than 5% in the scenario with $\phi = 2$, $\%_{TD_{PL}} = 20\%$ and $\%_{TD_A} - \%_{TD_{PL}} = 0\%$, an evaluation with the help of these methods is advised. The 95%-confidence interval includes an acceptable type I error rate as its lower bound. Furthermore, both methods return unbiased estimates. This represents a great advantage compared to the χ^2-test which is the single method respecting the type I error rate. The Shared gamma frailty model provides a smaller bias compared to the one of the Log-linear negative-binomial model, if there is a large difference between the percentages of patients with treatment discontinuation in both trial arms. Consequently, an evaluation with the help of the Shared gamma frailty model in scenarios with a discrepancy in the proportion of treatment discontinuation is proposed.

In the scenarios with $\phi = 2$ with MNAR, no appropriate method was found to analyze scenarios with a difference between the percentages of treatment discontinuation between both trials arms. In this context no method respects the type I error rate. But for scenarios with $\%_{TD_A} - \%_{TD_{PL}} = 0\%$, advice can be given. The Log rank test, the Cox proportional hazards model or the Gamma frailty model are recommended. The Log rank test and the Cox proportional hazards model deliver a higher power than the Gamma frailty model by revealing a similar absolute value of the bias. Especially if $\%_{TD_{PL}}$ increases up to 60%, the Log rank test and the Cox proportional hazards model manifest a larger power. Hence, these two methods can even be preferred to the Gamma frailty model.

Secondly, the results of the treatment-policy estimand are discussed. In principal, those methods which can be proposed for the evaluation of the hypothetical estimand are also optimal for the analysis of the treatment-policy estimand. In analogy to the hypothet-

ical estimand, the suggestions for the evaluation of the treatment-policy estimand are structured according to the scenarios with or without overdispersion and according to the treatment discontinuation mechanism.

A piece of advice for the estimation of the treatment-policy estimand mainly can be given if there is no difference of treatment discontinuation between both trial arms. The χ^2-test always respects the type I error rate if $\%_{TD_A} - \%_{TD_{PL}} = 0\%$. But its power is lower than that of the other methods by providing a higher bias. This was already the case in the context of the hypothetical estimand. In the scenarios without overdispersion under MCAR, the Shared gamma frailty model and the Log-linear Poisson model again reveal their superiority over the remaining methods. The perfect but rather unreal situation represents a clinical trial without study dropout. In such a scenario with $\phi = 1$, $\%_{TD_A} - \%_{TD_{PL}} = 0\%$, the Shared gamma frailty method always manifests an acceptable type I error rate by providing a larger power without any bias in the estimation of the treatment effect. If $\%_{TD_{PL}}$ rises, the Log-linear Poisson model can respect the type I error rate. This reveals the Shared gamma frailty model as the optimal method in this rather unrealistic scenario without any overdispersion. If $\phi = 1$, $\%_{TD_{PL}} = 20\%$ and $\%_{TD_A} - \%_{TD_{PL}} = -20\%$, the χ^2-test can be used for the analysis. Now, one problem in the context of the evaluation of the treatment-policy estimand comes to the fore. The actual idea of the treatment-policy estimand is only captured by methods which respect the whole period of time until the end of the study. In this way, the effect of the combination of several medications can be analyzed. The time-to-first-event methods and the χ^2-test wouldn't respect the effect of the rescue medication, if it was taken after the first event. This aspect is picked up later. If study dropout rises, the Shared gamma frailty model is superior to the χ^2-test in scenarios with a negative difference between the percentage of patients with treatment discontinuation in both trial arms.

Under MNAR, $\phi = 1$, $\%_{TD_{PL}} = 27\%$ and $\%_{TD_A} - \%_{TD_{PL}} = 0\%$ or $\%_{TD_{PL}} = 60\%$ and $\%_{TD_A} - \%_{TD_{PL}} = 0\%$, the performance of the Log-linear Poisson model and the Shared gamma frailty model can be compared as both deliver a type I error rate lower than 5% with similar power and bias. In scenarios without any study dropout, the Log-linear Poisson model might be considered superior to the Shared gamma frailty model as it provides a little larger power and a smaller bias. If the study dropout rises, both models are equivalent. No method respects the type I error rate in scenarios with a large difference between the proportion of patients with treatment discontinuation under MNAR. Therefore, no optimal method of evaluation is found in these scenarios.

The scenarios with overdispersion under MCAR, reveal the same problem that already appeared in the context of the hypothetical estimand. If $\%_{TD_{PL}} = 27\%$ and $\%_{TD_A} -$

$\%_{TD_{PL}} = 0\%$, only the χ^2-test respects the type I error rate. This fact doesn't change with higher study dropout. With a higher $\%_{TD_{PL}}$, the Log rank test and the Cox proportional hazards model additionally respect the type I error rate by providing the largest power and the smallest bias. But these methods aren't the optimal choice to capture the treatment-policy estimand. Referring to the 95%-confidence intervals of the estimated type I error rates, the Log-linear negative-binomial model and the Shared gamma frailty model represent an alternative. The lower bounds of their intervals include an acceptable type I error rate in most scenarios without a difference between the percentages of patients with treatment discontinuation in both trial arms. Hence, if the Log-linear negative-binomial model and the Shared gamma frailty model respected the type I error rate, they would be the optimal methods of evaluation. The Cox proportional hazards model and the Log rank test deliver reliably a type I error rate lower than 5% with the largest power and smallest bias. This fact justifies the examination of these methods in the context of the treatment-policy estimand, although they don't represent the perfect methods to evaluate the treatment-policy estimand.

The evaluation of the scenarios under MNAR with overdispersion reveals the same problems as under MCAR. Valid statements about the recommendation of a method of analysis can only be issued, if the difference of patients with treatment discontinuation between both trials arms is as small as possible. The Shared gamma frailty model and the Log-linear negative-binomial model as the optimal methods are recommended, if their lower bounds of the 95%-confidence intervals cover an acceptable type I error rate. In contrast to the scenarios under MCAR, the Shared gamma frailty model is superior to the Log-linear negative-binomial model, if the percentage of patients who drop out of the study rises. The power of the Log-linear negative-binomial model is smaller than that of the Shared gamma frailty model, in scenarios with higher study dropout. If 100% of the patients with treatment discontinuation stop their participation in the study, the estimation of the treatment-policy estimand is not possible. The combined effects of the examined drug with the rescue medication is not captured. In this case, the hypothetical estimand should be preferred.

It can be inferred that if extremely high study dropout rates occur, the clinical outcome should be measured in form of the average hazard ratio. The estimation of the treatment effect under the treatment-policy estimand with a study dropout of 100% of those with treatment discontinuation in form of the average hazard ratio by the Shared gamma frailty model is less biased than the ratio of the number of events in both trial groups in the context of the first endpoint. This is also true for the hypothetical estimand where each treatment discontinuation equals a study dropout. Hence, the estimation of the

hypothetical estimand always yields a high percentage of study dropout if treatment dropout occurs. In general, those methods which consider the whole period of the study for their analysis are those models which implement the idea of the treatment-policy estimand. Therefore, both count data models and the Shared gamma frailty model are able to include the number of events or the time between the events after a change of medication respectively. If they respect the type I error rate, these are the methods which are the chosen methods for the evaluation of both estimands in most situations. The remaining time-to-first-events methods basically just imply the estimation of the hypothetical estimand as they only consider the time to the first asthma exacerbation, no matter if the conditions of medication change after the first event. Surely, if the first event does not occur before treatment discontinuation, they are capable of involving the new event rates into their analysis.

9 Conclusions

This thesis performs a simulation study to examine several analysis methods for recurrent event data in a confirmatory clinical trial in asthma. The comparison of the models is investigated for the hypothetical and treatment-policy estimand. The goal of this thesis is to compare the chosen methods with respect to their type I error rate, power and bias. With these parameters, advice was given for a correct evaluation of recurrent event data under different conditions. The 'ICH E9 addendum on estimands and sensitivity analysis in clinical trials to the guideline on statistical principles in clinical trials' (2017, p. 5) proposes to address which estimand will be measured in the context of a confirmatory clinical trial. Hence, this thesis covers a field of research that is of interest for both the pharmaceutical companies and the regulatory agencies.

The record of recurrent event data leads to three different endpoints in this thesis. The first endpoint considers the total number of events until the end of the study or the study dropout respectively. The summary statistics which include the treatment effect of the first endpoint is the rate ratio. The second endpoint examines the time until the first asthma exacerbation or the time between all succeeding events. The treatment effect is measured with the average hazard ratio. The occurrence of at least one event is accounted for with the third endpoint. Its treatment effect is the odds ratio. For each of these endpoints, different methods are used. The Log-linear Poisson model and the Log-linear negative-binomial model evaluate the information of the first endpoint. For the analysis of the second endpoint, the Log rank test, the Cox proportional hazards model, the Likelihood ratio test, the Gamma frailty model and the Shared gamma frailty model are used. The χ^2-test serves for the analysis of the third endpoint.

The influence of four different factors is examined with the help of the simulations. Hereby, the results of the hypothetical and treatment-policy estimand reveal the same problems. Firstly, overdispersion can either occur or not. If it occurs, the type I error rate of each method will increase and the power will shrink. However, the bias of the methods doesn't change a lot. The results of the Log-linear negative-binomial model must be treated with caution in scenarios without overdispersion. In these scenarios, at least 40% of the Log-linear negative-binomial models don't converge. Secondly, the influence of the percentage of patients with treatment discontinuation is revealed. In the context of the hypothetical estimand, a treatment discontinuation leads to missing data. Whereas for the treatment-policy estimand, patients can be followed up. In this simulation study, treatment discontinuation is performed in the placebo group either with a lower or higher percentage. The proportion of patients with treatment discontinuation in the treatment group can either be smaller, equal or larger than in the placebo group.

Treatment discontinuation will have the strongest impact if the percentage of patients in the treatment group is larger than in the placebo group. The type I error rate and bias increase, the power shrinks. Valid statements about the recommendation of an optimal analysis of clinical trials can only be made if the treatment discontinuation is equal in both groups. Hence, a balance of treatment discontinuation in both trial arms must be the goal for real clinical trials. The third factor of influence is the study dropout which plays a role in the context of the analysis of the treatment-policy estimand. Hereby, patients stop their participation in the clinical trial after treatment discontinuation if they are followed up. With higher study dropout, the estimates are more biased and the power decreases. The type I error rates of scenarios without a difference in the percentage of patients with treatment discontinuation are hardly affected by the study dropout. In scenarios with a difference between the proportions of treatment discontinuation, there is a change of the type I error rates. If 50% of the patients with treatment discontinuation stop their participation in the trial, the type I error rates of the time-to-first-event methods decrease. Those of the remaining methods rise. If 100% of the patients with treatment discontinuation drop out of the study, the type I error rates increase greatly. This is especially visible for the methods evaluating only the time to the first event and the χ^2-test. Fourthly, the influence of the mechanism of treatment discontinuation is examined, which can either appear missing completely at random or missing not at random. Under MNAR, the type I error rate and the bias increase and the power shrinks the most for the count data models, the Shared gamma frailty model and the χ^2-test.

The more real the simulated scenarios are, the fewer methods provide an acceptable type I error rate. Furthermore, if there is a larger difference between the proportions of patients with treatment discontinuation in both trials arms, very high type I error rates are returned. If more patients in treatment group change to rescue medication than in the placebo group, this effect is even stronger. In general, the Shared gamma frailty model or one of the count data models is the method of choice. In scenarios with overdispersion, the Log rank test and the Cox proportional hazards model often are the only methods which return an acceptable type I error rate. However, for the evaluation of the treatment-policy estimand, the Shared gamma frailty model is recommended as providing the smallest bias in scenarios with a high study dropout. This model is proposed, although its estimated type I error rate is not lower or equal to 5%. But the lower bounds of their 95%-confidence intervals of the estimated type I error rates cover an acceptable value. To capture the idea of the treatment-policy estimand of analyzing the combination of new drug and rescue medication, the adequate method has to evaluate the whole period of time. This emphasizes the use of the Shared gamma frailty model in the context of scenarios with overdispersion for the treatment-policy estimand.

Further research could deal with a simulation study that includes the rescue medication with the same treatment effect as the new drug, but with stronger side effects. To recognize these side effects in the analysis, an endpoint must be defined which considers the occurrence of side effects. This endpoint can be based the record of the number of events or the time between all succeeding events. This thesis just covers the disease of asthma. The analysis of further diseases with recurrent events, e. g. multiple sclerosis or epilepsy, is more difficult. Their events often appear temporally close to each other. Therefore, it is difficult to recognize and record the single events.

References

Akacha, M. and Benda, N. (2010). The impact of dropouts on the analysis of dose-finding studies with recurrent event data. *Statistics in medicine*, 29 (15): 1635–1646. DOI: 10.1002/sim.3930.

Akacha, M., Bretz, F., Ohlssen, D., Rosenkranz, G. and Schmidli, H. (2017a). Estimands and their role in clinical trials. *Statistics in Biopharmaceutical Research*, 9 (3): 268–271. DOI: 10.1080/19466315.2017.1302358.

Akacha, M., Bretz, F. and Ruberg, S. (2017b). Estimands in clinical trials–broadening the perspective. *Statistics in medicine*, 36 (1): 5–19. DOI: 10.1002/sim.7033.

Bamberg, G., Baur, F. and Krapp, M. (2011). *Statistik*. 16th Edition. München: Oldenbourg. ISBN: 978-3-486-70258-3.

Benda, N., Branson, M., Maurer, W. and Friede, T. (2010). Aspects of modernizing drug development using clinical scenario planning and evaluation. *Drug Information Journal*, 44 (3): 299–315. DOI: 10.1177/009286151004400312.

Box-Steffensmeier, J. M. and De Boef, S. (2006). Repeated events survival models: the conditional frailty model. *Statistics in medicine*, 25 (20): 3518–3533. DOI: 10.1002/sim.2434.

Brueckner, M. (2016). R-Package 'AHR': Estimation and Testing of Average Hazard Ratios. URL: https://CRAN.R-project.org/package=AHR.

Carpenter, J., Pocock, S. and Johan Lamm, C. (2002). Coping with missing data in clinical trials: A model-based approach applied to asthma trials. *Statistics in medicine*, 21 (8): 1043–1066. URL: 10.1002/sim.1065.

Committee for Medicinal Products for Human Use (CHMP) et al. (2010). Guideline on missing data in confirmatory clinical trials. *London: European Medicines Agency*. URL: https://www.google.com/url?sa=t&rct=j&q=&esrc=s&source=web&cd=1&cad=rja&uact=8&ved=0ahUKEwj7k5Xu0dLaAhWDVSwKHY7UCn4QFggvMAA&url=http%3A%2F%2Fwww.ema.europa.eu%2Fdocs%2Fen_GB%2Fdocument_library%2FScientific_guideline%2F2010%2F09%2FWC500096793.pdf&usg=AOvVaw34NbJFsgmK8tmcYdeC9g-X.

Committee for Medicinal Products for Human Use (CHMP) et al. (2015). Guideline on the clinical investigation of medicinal products for the treatment of asthma. Tech. rep. CHMP/EWP/2922/01 Rev. 1. URL: http://www.%20ema.%20europa.%20eu/docs/en%5C_GB/document%5C_library/Scientific%5C_guideline/2015/12/WC500198877.%20pdf.

Committee for Medicinal Products for Human Use (CHMP) et al. (2017). ICH E9 (R1) addendum on estimands and sensitivity analysis in clinical trials to the guideline on statistical principles for clinical trials, Step 2b. Tech. rep. EMA/CHMP/ICH/436221/2017 Rev. 1. URL: http://www.ema.europa.eu/docs/en_GB/document_library/Scientific_guideline/2017/08/WC500233916.pdf.

Cook, R. J. and Lawless, J. F. (2002). Analysis of repeated events. *Statistical Methods in Medical Research*, 11 (2): 141–166. DOI: 10.1191/0962280202sm278ra.

Cook, R. J. and Lawless, J. F. (2007). *The statistical analysis of recurrent events*. New York: Springer Science & Business Media. DOI: 10.1007/978-0-387-69810-6.

Cook, R. J., Bergeron, P.-J., Boher, J.-M. and Liu, Y. (2009). Two-stage design of clinical trials involving recurrent events. *Statistics in medicine*, 28 (21): 2617–2638. DOI: 10.1002/sim.3645.

Duchateau, L. and Janssen, P. (2007). *The frailty model*. New York: Springer Science & Business Media. DOI: 10.1007/978-0-387-72835-3.

Fahrmeir, L., Kneib, T., Lang, S. and Marx, B. (2013). *Regression: models, methods and applications*. Berlin, Heidelberg: Springer Science & Business Media. DOI: 10.1007/978-3-642-34333-9.

Friede, T., Schmidli, H., et al. (2010). Blinded sample size reestimation with negative binomial counts in superiority and non-inferiority trials. *Methods of Information in Medicine*, 49 (6): 618–624. DOI: 10.3414/ME09-02-0060.

Furman, E. (2007). On the convolution of the negative binomial random variables. *Statistics & probability letters*, 77 (2): 169–172. DOI: 10.1016/j.spl.2006.06.007.

Hardin, J. W. and Hilbe, J. M. (2012). *Generalized linear models and extensions*. Third Edition. Lakeaway Drive: Stata press. ISBN: 978-1597181051.

Held, L. and Sabanés Bové, D. (2014). *Applied Statistical Inference: Likelihood and Bayes*. Berlin, Heidelberg: Springer-Verlag. DOI: 10.1007/978-3-642-37887-4.

Hernán, M. A., Hernández-Diaz, S. and Robins, J. M. (2013). Randomized trials analyzed as observational studies. *Annals of internal medicine*, 159 (8): 560–562. DOI: 10.7326/0003-4819-159-8-201310150-00709.

Hilbe, J. M. (2011). *Negative binomial regression*. Second Edition. New York: Cambridge University Press. DOI: 10.1111/j.1751-5823.2011.00159_4.x.

Ibrahim, J. G., Chen, M.-H., Lipsitz, S. R. and Herring, A. H. (2005). Missing-data methods for generalized linear models: A comparative review. *Journal of the American Statistical Association*, 100 (469): 332–346. DOI: 10.1198/016214504000001844.

Jahn-Eimermacher, A. (2008). Comparison of the Andersen–Gill model with Poisson and negative binomial regression on recurrent event data. *Computational Statistics & Data Analysis*, 52 (11): 4989–4997. DOI: 10.1016/j.csda.2008.04.009.

Johnson, N. L., Kotz, S. and Balakrishnan, N. (1994). *Continuous univariate distributions. Volume 1*. Second Edition. Hoboken: John Wiley & Sons. ISBN: 978-0-471-58495-7.

Johnson, N. L., Kotz, S. and Kemp, A. W. (1993). *Univariate discrete distributions*. Second Edition. Hoboken: John Wiley & Sons. ISBN: 978-0-471-54897-3.

Kalbfleisch, J. D. and Prentice, R. L. (1981). Estimation of the average hazard ratio. *Biometrika*, 68 (1): 105–112. DOI: 10.1093/biomet/68.1.105.

Keene, O. N., Jones, M. R., Lane, P. W. and Anderson, J. (2007). Analysis of exacerbation rates in asthma and chronic obstructive pulmonary disease: example from the TRISTAN study. *Pharmaceutical statistics*, 6 (2): 89–97. DOI: 10.1002/pst.250.

Keene, O. N., Roger, J. H., Hartley, B. F. and Kenward, M. G. (2014). Missing data sensitivity analysis for recurrent event data using controlled imputation. *Pharmaceutical statistics*, 13 (4): 258–264. DOI: 10.1002/pst.1624.

Kenward, M. G. (2013). The handling of missing data in clinical trials. *Clinical Investigation*, 3 (3): 241–250. DOI: 10.4155/CLI.13.7.

Klein, J. P. and Moeschberger, M. L. (1997). *Survival analysis: techniques for censored and truncated data*. New York et al.: Springer Science & Business Media. DOI: 10.1007/b97377.

Kneib, T. and Hambuckers, J. (winter term 2017/2018). Statistical Inference: Likelihood & Bayes. *Handouts for lecture 'Statistical Inference: Likelihood & Bayes', Georg-August-University Göttingen.*

Lawless, J. F. (1995). The analysis of recurrent events for multiple subjects. *Applied Statistics*, 44 (4): 487–498. DOI: 10.2307/2986140.

Lawless, J. F. (1987). Regression methods for Poisson process data. *Journal of the American Statistical Association*, 82 (399): 808–815. DOI: 10.1080/01621459.1987.10478502.

Lee, E. T. and Wang, J. (2013). *Statistical methods for survival data analysis.* New Jersey: John Wiley & Sons. DOI: 10.1002/0471458546.

Leuchs, A.-K., Brandt, A., Zinserling, J. and Benda, N. (2017). Disentangling estimands and the intention-to-treat principle. *Pharmaceutical statistics*, 16 (1): 12–19. DOI: 10.1002/pst.1791.

Leuchs, A.-K., Zinserling, J., Brandt, A., Wirtz, D. and Benda, N. (2015). Choosing appropriate estimands in clinical trials. *Therapeutic Innovation & Regulatory Science*, 49 (4): 584–592. DOI: 10.1177/2168479014567317.

Leuchs, A.-K., Zinserling, J., Schlosser-Weber, G., Berres, M., Neuhäuser, M. and Benda, N. (2014). Estimation of the treatment effect in the presence of non-compliance and missing data. *Statistics in medicine*, 33 (2): 193–208. DOI: 10.1002/sim.5924.

Little, R. and Kang, S. (2015). Intention-to-treat analysis with treatment discontinuation and missing data in clinical trials. *Statistics in Medicine*, 34 (16): 2381–2390. DOI: 10.1002/sim.6352.

Machin, D., Cheung, Y. B. and Parmar, M. (2006). *Survival analysis: a practical approach.* Second Edition. Hoboken et al.: John Wiley & Sons. DOI: 10.1002/0470034572.

Mallinckrodt, C. H. (2013). *Preventing and treating missing data in longitudinal clinical trials: a practical guide.* New York: Cambridge University Press. DOI: 10.1017/CBO9781139381666.

Mallinckrodt, C., Molenberghs, G. and Rathmann, S. (2017). Choosing estimands in clinical trials with missing data. *Pharmaceutical statistics*, 16 (1): 29–36. DOI: 10.1002/pst.1765.

McCullagh, P. and Nelder, J. A. (1987). *Generalized Linear Models*. London et al.: Chapman and Hall. DOI: 10.1002/bimj.4710290217.

Mehrotra, D. V., Hemmings, R. J., Russek-Cohen, E. and Group, I. E. E. W. (2016). Seeking harmony: estimands and sensitivity analyses for confirmatory clinical trials. *Clinical Trials*, 13 (4): 456–458. DOI: 10.1177/1740774516633115.

Metcalfe, C. and Thompson, S. G. (2006). The importance of varying the event generation process in simulation studies of statistical methods for recurrent events. *Statistics in medicine*, 25 (1): 165–179. DOI: 10.1002/sim.2310.

Munda, M., Rotolo, F. and Legrand, C. (2017). Package 'parfm' Version 2.7.5. URL: https://CRAN.R-project.org/package=parfm.

Munda, M., Rotolo, F., Legrand, C., et al. (2012). Parfm: parametric frailty models in R. *Journal of Statistical Software*, 51 (11): 1–20. DOI: 10.18637/jss.v051.i11.

National Research Council et al. (2010). *The prevention and treatment of missing data in clinical trials*. Washington D. C.: National Academies Press. DOI: 10.17226/12955.

Nicholas, R., Straube, S., Schmidli, H., Schneider, S. and Friede, T. (2011). Trends in annualized relapse rates in relapsing–remitting multiple sclerosis and consequences for clinical trial design. *Multiple Sclerosis Journal*, 17 (10): 1211–1217. DOI: 10.1177/1352458511406309.

Omori, Y. and Johnson, R. A. (1993). The influence of random effects on the unconditional hazard rate and survival functions. *Biometrika*, 80 (4): 910–914. DOI: 10.1093/biomet/80.4.910.

Pavord, I. D., Korn, S., Howarth, P., Bleecker, E. R., Buhl, R., Keene, O. N., Ortega, H. and Chanez, P. (2012). Mepolizumab for severe eosinophilic asthma (DREAM): a multicentre, double-blind, placebo-controlled trial. *The Lancet*, 380 (9842): 651–659. DOI: 10.1056/NEJMoa1708208.

Permutt, T. (2016). A taxonomy of estimands for regulatory clinical trials with discontinuations. *Statistics in medicine*, 35 (17): 2865–2875. DOI: doi.org/10.1002/sim.6841.

Permutt, T. (2017). Comments on 'Estimands in clinical trials–broadening the perspective'. *Statistics in medicine*, 36 (1): 20–21. DOI: 10.1002/sim.7160.

Phillips, A., Abellan-Andres, J., Soren, A., Bretz, F., Fletcher, C., France, L., Garrett, A., Harris, R., Kjaer, M., Keene, O., et al. (2016). Estimands: discussion points from the PSI estimands and sensitivity expert group. *Pharmaceutical statistics*, 16 (1): 6–11. DOI: 10.1002/pst.1745.

R Core Team (2017). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. Vienna, Austria. URL: http://www.R-project.org/.

Ripley, B., Venables, B., Bates, D. M., Hornik, K., Gebhardt, A., Firth, D. and Ripley, M. B. (2017). Package 'MASS' Version 7.3-47. URL: https://CRAN.R-project.org/package=MASS.

Rodríguez, G. (2005). Unobserved heterogeneity. *Handouts for 'Pop509: Survival Cerca con Google*. URL: http://data.princeton.edu/eco572/unobservedHeterogeneity.pdf.

RStudio Team (2017). *RStudio: Integrated Development Environment for R*. RStudio, Inc. Boston, MA. URL: http://www.rstudio.com/.

Schemper, M., Wakounig, S. and Heinze, G. (2009). The estimation of average hazard ratios by weighted Cox regression. *Statistics in medicine*, 28 (19): 2473–2489. DOI: 10.1002/sim.3623.

Schneider, S., Schmidli, H. and Friede, T. (2013). Blinded sample size re-estimation for recurrent event data with time trends. *Statistics in medicine*, 32 (30): 5448–5457. DOI: 10.1002/sim.5977.

Siri, P., Henninger, E. and Sormani, M. P. (2012). A parametric model fitting time to first event for overdispersed data: application to time to relapse in multiple sclerosis. *Lifetime data analysis*, 18 (2): 139–156. DOI: 10.1007/s10985-011-9207-z.

Sölkner, J. (1996). Frailty models in survival analysis. *Journal of computing and information technology - CIT*, 4 (4): 235–239.

Therneau, T. M. and Grambsch, P. M. (2000). The Cox model. *Modeling survival data: extending the Cox model*. New York: Springer, pp. 39–77. DOI: 10.1007/978-1-4757-3294-8.

Therneau, T. M. and Lumley, T. (2017). Package 'survival' Version 2.41-3. URL: https://CRAN.R-project.org/package=survival.

Wan, F., Small, D., Bekelman, J. E. and Mitra, N. (2015). Bias in estimating the causal hazard ratio when using two-stage instrumental variable methods. *Statistics in medicine*, 34 (14): 2235–2265. DOI: 10.1002/sim.6470.

Wienke, A. (2010). *Frailty models in survival analysis*. Boca Raton: Chapman & Hall. DOI: 10.1201/9781420073911.

Ziegler, A., Lange, S. and Bender, R. (2007). Überlebenszeitanalyse: Der Log-Rang-Test. *DMW-Deutsche Medizinische Wochenschrift*, 132 (S 01): e39–e41. DOI: 10.1055/s-2007-959040.

Zucchini, W., Schlegel, A., Nenadic, O. and Sperlich, S. (2009). *Statistik für Bachelor-und Masterstudenten: eine Einführung für Wirtschafts-und Sozialwissenschaftler*. Berlin, Heidelberg: Springer-Verlag. DOI: 10.1007/978-3-540-88987-8.

A Percentages for scenarios

A.1 Percentages for scenarios of hypothetical estimand

Table 39: Percentages for scenarios of hypothetical estimand

No.	N	$\%_{TD_A}$	$\%_{TD_{PL}}$	$\%_{E_{1_A}}$	$\%_{E_{1_{PL}}}$	TD mechanism	ϕ
1	330	0.00	0.20	0.50	0.64	MCAR	1
2	330	0.20	0.20	0.45	0.64	MCAR	1
3	330	0.40	0.20	0.40	0.64	MCAR	1
4	330	0.30	0.50	0.43	0.54	MCAR	1
5	330	0.50	0.50	0.37	0.54	MCAR	1
6	330	0.70	0.50	0.31	0.54	MCAR	1
7	330	0.00	0.20	0.70	0.64	MCAR	1
8	330	0.20	0.20	0.64	0.64	MCAR	1
9	330	0.40	0.20	0.57	0.64	MCAR	1
10	330	0.30	0.50	0.61	0.54	MCAR	1
11	330	0.50	0.50	0.54	0.54	MCAR	1
12	330	0.70	0.50	0.45	0.54	MCAR	1
13	330	0.00	0.27	0.50	0.62	MNAR	1
14	330	0.23	0.27	0.44	0.62	MNAR	1
15	330	0.45	0.27	0.39	0.62	MNAR	1
16	330	0.34	0.60	0.42	0.48	MNAR	1
17	330	0.54	0.60	0.35	0.48	MNAR	1
18	330	0.74	0.60	0.29	0.48	MNAR	1
19	330	0.00	0.27	0.70	0.62	MNAR	1
20	330	0.27	0.27	0.62	0.61	MNAR	1
21	330	0.50	0.27	0.54	0.61	MNAR	1
22	330	0.39	0.60	0.58	0.48	MNAR	1
23	330	0.60	0.60	0.48	0.48	MNAR	1
24	330	0.78	0.60	0.40	0.48	MNAR	1
25	330	0.00	0.20	0.38	0.52	MCAR	2
26	330	0.20	0.20	0.35	0.52	MCAR	2
27	330	0.40	0.20	0.31	0.52	MCAR	2
28	330	0.30	0.50	0.33	0.44	MCAR	2
29	330	0.50	0.50	0.29	0.44	MCAR	2
30	330	0.70	0.50	0.25	0.44	MCAR	2
31	330	0.00	0.20	0.57	0.52	MCAR	2

Table 39: Percentages for scenarios of hypothetical estimand

No.	**N**	$\%_{\boldsymbol{TD_A}}$	$\%_{\boldsymbol{TD_{PL}}}$	$\%_{\boldsymbol{E_{1_A}}}$	$\%_{\boldsymbol{E_{1_{PL}}}}$	TD mechanism	ϕ
32	330	0.20	0.20	0.52	0.52	MCAR	2
33	330	0.40	0.20	0.47	0.52	MCAR	2
34	330	0.30	0.50	0.49	0.44	MCAR	2
35	330	0.50	0.50	0.44	0.44	MCAR	2
36	330	0.70	0.50	0.38	0.44	MCAR	2
37	330	0.00	0.28	0.38	0.49	MNAR	2
38	330	0.24	0.28	0.34	0.49	MNAR	2
39	330	0.45	0.28	0.30	0.49	MNAR	2
40	330	0.35	0.60	0.32	0.38	MNAR	2
41	330	0.55	0.60	0.26	0.39	MNAR	2
42	330	0.75	0.60	0.23	0.38	MNAR	2
43	330	0.00	0.28	0.57	0.49	MNAR	2
44	330	0.28	0.28	0.49	0.49	MNAR	2
45	330	0.50	0.28	0.44	0.49	MNAR	2
46	330	0.39	0.60	0.47	0.38	MNAR	2
47	330	0.60	0.60	0.39	0.39	MNAR	2
48	330	0.78	0.60	0.33	0.38	MNAR	2

A.2 Percentages for scenarios of treatment-policy estimand

Table 40: Percentages for scenarios of treatment-policy estimand

No.	N	$\%_{TD_A}$	$\%_{TD_{PL}}$	$\%_{SD_A}$	$\%_{SD_{PL}}$	$\%_{E_{1_A}}$	$\%_{E_{1_{PL}}}$	$\%_{E_{2_A}}$	$\%_{E_{2_{PL}}}$	$\%_{E_{total_A}}$	$\%_{E_{total_{PL}}}$	TD mechanism	ϕ
1	330	0.00	0.20	0.00	0.00	0.50	0.64	0.00	0.07	0.50	0.69	MCAR	1
2	330	0.20	0.20	0.00	0.00	0.45	0.64	0.07	0.07	0.51	0.69	MCAR	1
3	330	0.40	0.20	0.00	0.00	0.40	0.64	0.15	0.07	0.52	0.69	MCAR	1
4	330	0.30	0.50	0.00	0.00	0.43	0.54	0.11	0.19	0.52	0.67	MCAR	1
5	330	0.50	0.50	0.00	0.00	0.37	0.54	0.19	0.19	0.53	0.67	MCAR	1
6	330	0.70	0.50	0.00	0.00	0.31	0.54	0.28	0.19	0.54	0.67	MCAR	1
7	330	0.00	0.20	0.00	0.00	0.70	0.64	0.00	0.07	0.70	0.69	MCAR	1
8	330	0.20	0.20	0.00	0.00	0.64	0.64	0.07	0.07	0.69	0.69	MCAR	1
9	330	0.40	0.20	0.00	0.00	0.57	0.64	0.15	0.07	0.68	0.69	MCAR	1
10	330	0.30	0.50	0.00	0.00	0.61	0.54	0.11	0.19	0.68	0.67	MCAR	1
11	330	0.50	0.50	0.00	0.00	0.54	0.54	0.19	0.19	0.67	0.67	MCAR	1
12	330	0.70	0.50	0.00	0.00	0.45	0.54	0.28	0.19	0.66	0.67	MCAR	1
13	330	0.00	0.20	0.00	0.10	0.50	0.64	0.00	0.07	0.50	0.69	MCAR	1
14	330	0.20	0.20	0.10	0.10	0.45	0.64	0.07	0.07	0.51	0.69	MCAR	1
15	330	0.40	0.20	0.21	0.10	0.40	0.64	0.15	0.07	0.52	0.69	MCAR	1
16	330	0.30	0.50	0.16	0.27	0.43	0.54	0.11	0.19	0.52	0.67	MCAR	1
17	330	0.50	0.50	0.26	0.27	0.37	0.54	0.19	0.19	0.53	0.67	MCAR	1
18	330	0.70	0.50	0.40	0.27	0.31	0.54	0.28	0.19	0.54	0.67	MCAR	1
19	330	0.00	0.20	0.00	0.10	0.70	0.64	0.00	0.07	0.70	0.69	MCAR	1
20	330	0.20	0.20	0.10	0.10	0.64	0.64	0.07	0.07	0.69	0.69	MCAR	1

Table 40: Percentages for scenarios of treatment-policy estimand

No.	N	$\%_{TD_A}$	$\%_{TD_{PL}}$	$\%_{SD_A}$	$\%_{SD_{PL}}$	$\%_{E_{1_A}}$	$\%_{E_{1_{PL}}}$	$\%_{E_{2_A}}$	$\%_{E_{2_{PL}}}$	$\%_{E_{total_A}}$	$\%_{E_{total_{PL}}}$	TD mechanism	ϕ
21	330	0.40	0.20	0.21	0.10	0.57	0.64	0.15	0.07	0.68	0.69	MCAR	1
22	330	0.30	0.50	0.15	0.27	0.61	0.54	0.11	0.19	0.68	0.67	MCAR	1
23	330	0.50	0.50	0.27	0.26	0.54	0.54	0.19	0.19	0.67	0.67	MCAR	1
24	330	0.70	0.50	0.39	0.26	0.45	0.54	0.28	0.19	0.66	0.67	MCAR	1
25	330	0.00	0.20	1.00	1.00	0.50	0.64	0.00	0.07	0.50	0.69	MCAR	1
26	330	0.20	0.20	1.00	1.00	0.45	0.64	0.07	0.07	0.51	0.69	MCAR	1
27	330	0.40	0.20	1.00	1.00	0.40	0.64	0.15	0.07	0.52	0.69	MCAR	1
28	330	0.30	0.50	1.00	1.00	0.43	0.54	0.11	0.19	0.52	0.67	MCAR	1
29	330	0.50	0.50	1.00	1.00	0.37	0.54	0.19	0.19	0.53	0.67	MCAR	1
30	330	0.70	0.50	1.00	1.00	0.31	0.54	0.28	0.19	0.54	0.67	MCAR	1
31	330	0.00	0.20	1.00	1.00	0.70	0.64	0.00	0.07	0.70	0.69	MCAR	1
32	330	0.20	0.20	1.00	1.00	0.64	0.64	0.07	0.07	0.69	0.69	MCAR	1
33	330	0.40	0.20	1.00	1.00	0.57	0.64	0.15	0.07	0.68	0.69	MCAR	1
34	330	0.30	0.50	1.00	1.00	0.61	0.54	0.11	0.19	0.68	0.67	MCAR	1
35	330	0.50	0.50	1.00	1.00	0.54	0.54	0.19	0.19	0.67	0.67	MCAR	1
36	330	0.70	0.50	1.00	1.00	0.45	0.54	0.28	0.19	0.66	0.67	MCAR	1
37	330	0.00	0.27	0.00	0.00	0.50	0.62	0.00	0.10	0.50	0.68	MNAR	1
38	330	0.23	0.27	0.00	0.00	0.44	0.62	0.08	0.10	0.51	0.68	MNAR	1
39	330	0.45	0.27	0.00	0.00	0.39	0.62	0.17	0.10	0.53	0.68	MNAR	1
40	330	0.34	0.60	0.00	0.00	0.42	0.48	0.13	0.24	0.52	0.65	MNAR	1
41	330	0.54	0.60	0.00	0.00	0.35	0.48	0.21	0.24	0.52	0.65	MNAR	1

Table 40: Percentages for scenarios of treatment-policy estimand

No.	N	$\%_{TD_A}$	$\%_{TD_{PL}}$	$\%_{SD_A}$	$\%_{SD_{PL}}$	$\%_{E_{1_A}}$	$\%_{E_{1_{PL}}}$	$\%_{E_{2_A}}$	$\%_{E_{2_{PL}}}$	$\%_{E_{total_A}}$	$\%_{E_{total_{PL}}}$	TD mechanism	ϕ
42	330	0.74	0.60	0.00	0.00	0.29	0.48	0.31	0.24	0.55	0.65	MNAR	1
43	330	0.00	0.27	0.00	0.00	0.70	0.62	0.00	0.10	0.70	0.68	MNAR	1
44	330	0.27	0.27	0.00	0.00	0.62	0.61	0.10	0.10	0.68	0.68	MNAR	1
45	330	0.50	0.27	0.00	0.00	0.54	0.61	0.19	0.10	0.67	0.68	MNAR	1
46	330	0.39	0.60	0.00	0.00	0.58	0.48	0.15	0.24	0.68	0.65	MNAR	1
47	330	0.60	0.60	0.00	0.00	0.48	0.48	0.24	0.24	0.65	0.65	MNAR	1
48	330	0.78	0.60	0.00	0.00	0.40	0.48	0.33	0.24	0.65	0.65	MNAR	1
49	330	0.00	0.27	0.00	0.14	0.50	0.62	0.00	0.10	0.50	0.68	MNAR	1
50	330	0.23	0.27	0.12	0.14	0.44	0.62	0.08	0.10	0.51	0.68	MNAR	1
51	330	0.45	0.27	0.24	0.14	0.39	0.62	0.17	0.10	0.53	0.68	MNAR	1
52	330	0.34	0.60	0.18	0.34	0.42	0.48	0.13	0.24	0.52	0.65	MNAR	1
53	330	0.54	0.60	0.29	0.34	0.35	0.48	0.21	0.24	0.52	0.65	MNAR	1
54	330	0.74	0.60	0.43	0.34	0.29	0.48	0.31	0.24	0.55	0.65	MNAR	1
55	330	0.00	0.27	0.00	0.14	0.70	0.62	0.00	0.10	0.70	0.68	MNAR	1
56	330	0.27	0.27	0.14	0.14	0.62	0.61	0.10	0.10	0.68	0.68	MNAR	1
57	330	0.50	0.27	0.26	0.14	0.54	0.61	0.19	0.10	0.67	0.68	MNAR	1
58	330	0.39	0.60	0.20	0.34	0.58	0.48	0.15	0.24	0.68	0.65	MNAR	1
59	330	0.60	0.60	0.34	0.33	0.48	0.48	0.24	0.24	0.65	0.65	MNAR	1
60	330	0.78	0.60	0.45	0.33	0.40	0.48	0.33	0.24	0.65	0.65	MNAR	1
61	330	0.00	0.27	1.00	1.00	0.50	0.62	0.00	0.10	0.50	0.68	MNAR	1
62	330	0.23	0.27	1.00	1.00	0.44	0.62	0.08	0.10	0.51	0.68	MNAR	1

Table 40: Percentages for scenarios of treatment-policy estimand

No.	**N**	$\%_{TD_A}$	$\%_{TD_{PL}}$	$\%_{SD_A}$	$\%_{SD_{PL}}$	$\%_{E_{1_A}}$	$\%_{E_{1_{PL}}}$	$\%_{E_{2_A}}$	$\%_{E_{2_{PL}}}$	$\%_{E_{total_A}}$	$\%_{E_{total_{PL}}}$	TD mechanism	ϕ
63	330	0.45	0.27	1.00	1.00	0.39	0.62	0.17	0.10	0.53	0.68	MNAR	1
64	330	0.34	0.60	1.00	1.00	0.42	0.48	0.13	0.24	0.52	0.65	MNAR	1
65	330	0.54	0.60	1.00	1.00	0.35	0.48	0.21	0.24	0.52	0.65	MNAR	1
66	330	0.74	0.60	1.00	1.00	0.29	0.48	0.31	0.24	0.55	0.65	MNAR	1
67	330	0.00	0.27	1.00	1.00	0.70	0.62	0.00	0.10	0.70	0.68	MNAR	1
68	330	0.27	0.27	1.00	1.00	0.62	0.61	0.10	0.10	0.68	0.68	MNAR	1
69	330	0.50	0.27	1.00	1.00	0.54	0.61	0.19	0.10	0.67	0.68	MNAR	1
70	330	0.39	0.60	1.00	1.00	0.58	0.48	0.15	0.24	0.68	0.65	MNAR	1
71	330	0.60	0.60	1.00	1.00	0.48	0.48	0.24	0.24	0.65	0.65	MNAR	1
72	330	0.78	0.60	1.00	1.00	0.40	0.48	0.33	0.24	0.65	0.65	MNAR	1
73	330	0.00	0.20	0.00	0.00	0.38	0.52	0.00	0.06	0.38	0.56	MCAR	2
74	330	0.20	0.20	0.00	0.00	0.35	0.52	0.06	0.06	0.40	0.56	MCAR	2
75	330	0.40	0.20	0.00	0.00	0.31	0.52	0.12	0.06	0.42	0.56	MCAR	2
76	330	0.30	0.50	0.00	0.00	0.33	0.44	0.09	0.15	0.41	0.56	MCAR	2
77	330	0.50	0.50	0.00	0.00	0.29	0.44	0.16	0.16	0.42	0.56	MCAR	2
78	330	0.70	0.50	0.00	0.00	0.25	0.44	0.23	0.16	0.44	0.56	MCAR	2
79	330	0.00	0.20	0.00	0.00	0.57	0.52	0.00	0.06	0.57	0.56	MCAR	2
80	330	0.20	0.20	0.00	0.00	0.52	0.52	0.06	0.06	0.56	0.56	MCAR	2
81	330	0.40	0.20	0.00	0.00	0.47	0.52	0.12	0.06	0.56	0.56	MCAR	2
82	330	0.30	0.50	0.00	0.00	0.49	0.44	0.09	0.16	0.56	0.56	MCAR	2
83	330	0.50	0.50	0.00	0.00	0.44	0.44	0.16	0.16	0.56	0.56	MCAR	2

Table 40: Percentages for scenarios of treatment-policy estimand

No.	N	$\%_{TD_A}$	$\%_{TD_{PL}}$	$\%_{SD_A}$	$\%_{SD_{PL}}$	$\%_{E_{1_A}}$	$\%_{E_{1_{PL}}}$	$\%_{E_{2_A}}$	$\%_{E_{2_{PL}}}$	$\%_{E_{total_A}}$	$\%_{E_{total_{PL}}}$	TD mechanism	ϕ
84	330	0.70	0.50	0.00	0.00	0.38	0.44	0.23	0.15	0.55	0.56	MCAR	2
85	330	0.00	0.20	0.00	0.10	0.38	0.52	0.00	0.06	0.38	0.56	MCAR	2
86	330	0.20	0.20	0.10	0.10	0.35	0.52	0.06	0.06	0.40	0.56	MCAR	2
87	330	0.40	0.20	0.21	0.10	0.31	0.52	0.12	0.06	0.42	0.56	MCAR	2
88	330	0.30	0.50	0.16	0.27	0.33	0.44	0.09	0.15	0.41	0.56	MCAR	2
89	330	0.50	0.50	0.26	0.27	0.29	0.44	0.16	0.16	0.42	0.56	MCAR	2
90	330	0.70	0.50	0.40	0.27	0.25	0.44	0.23	0.16	0.44	0.56	MCAR	2
91	330	0.00	0.20	0.00	0.10	0.57	0.52	0.00	0.06	0.57	0.56	MCAR	2
92	330	0.20	0.20	0.10	0.10	0.52	0.52	0.06	0.06	0.56	0.56	MCAR	2
93	330	0.40	0.20	0.21	0.10	0.47	0.52	0.12	0.06	0.56	0.56	MCAR	2
94	330	0.30	0.50	0.15	0.27	0.49	0.44	0.09	0.16	0.56	0.56	MCAR	2
95	330	0.50	0.50	0.27	0.26	0.44	0.44	0.16	0.16	0.56	0.56	MCAR	2
96	330	0.70	0.50	0.39	0.26	0.38	0.44	0.23	0.15	0.55	0.56	MCAR	2
97	330	0.00	0.20	1.00	1.00	0.38	0.52	0.00	0.06	0.38	0.56	MCAR	2
98	330	0.20	0.20	1.00	1.00	0.35	0.52	0.06	0.06	0.40	0.56	MCAR	2
99	330	0.40	0.20	1.00	1.00	0.31	0.52	0.12	0.06	0.42	0.56	MCAR	2
100	330	0.30	0.50	1.00	1.00	0.33	0.44	0.09	0.15	0.41	0.56	MCAR	2
101	330	0.50	0.50	1.00	1.00	0.29	0.44	0.16	0.16	0.42	0.56	MCAR	2
102	330	0.70	0.50	1.00	1.00	0.25	0.44	0.23	0.16	0.44	0.56	MCAR	2
103	330	0.00	0.20	1.00	1.00	0.57	0.52	0.00	0.06	0.57	0.56	MCAR	2
104	330	0.20	0.20	1.00	1.00	0.52	0.52	0.06	0.06	0.56	0.56	MCAR	2

Table 40: Percentages for scenarios of treatment-policy estimand

No.	N	$\%_{TD_A}$	$\%_{TD_{PL}}$	$\%_{SD_A}$	$\%_{SD_{PL}}$	$\%_{E_{1_A}}$	$\%_{E_{1_{PL}}}$	$\%_{E_{2_A}}$	$\%_{E_{2_{PL}}}$	$\%_{E_{total_A}}$	$\%_{E_{total_{PL}}}$	TD mechanism	ϕ
105	330	0.40	0.20	1.00	1.00	0.47	0.52	0.12	0.06	0.56	0.56	MCAR	2
106	330	0.30	0.50	1.00	1.00	0.49	0.44	0.09	0.16	0.56	0.56	MCAR	2
107	330	0.50	0.50	1.00	1.00	0.44	0.44	0.16	0.16	0.56	0.56	MCAR	2
108	330	0.70	0.50	1.00	1.00	0.38	0.44	0.23	0.15	0.55	0.56	MCAR	2
109	330	0.00	0.28	0.00	0.00	0.38	0.49	0.00	0.09	0.38	0.55	MNAR	2
110	330	0.24	0.28	0.00	0.00	0.34	0.49	0.07	0.09	0.39	0.55	MNAR	2
111	330	0.45	0.28	0.00	0.00	0.30	0.49	0.14	0.09	0.42	0.55	MNAR	2
112	330	0.35	0.60	0.00	0.00	0.32	0.38	0.11	0.20	0.41	0.53	MNAR	2
113	330	0.55	0.60	0.00	0.00	0.26	0.39	0.18	0.20	0.41	0.53	MNAR	2
114	330	0.75	0.60	0.00	0.00	0.23	0.38	0.25	0.20	0.45	0.52	MNAR	2
115	330	0.00	0.28	0.00	0.00	0.57	0.49	0.00	0.09	0.57	0.55	MNAR	2
116	330	0.28	0.28	0.00	0.00	0.49	0.49	0.09	0.09	0.55	0.55	MNAR	2
117	330	0.50	0.28	0.00	0.00	0.44	0.49	0.16	0.09	0.56	0.55	MNAR	2
118	330	0.39	0.60	0.00	0.00	0.47	0.38	0.13	0.20	0.56	0.53	MNAR	2
119	330	0.60	0.60	0.00	0.00	0.39	0.39	0.20	0.20	0.53	0.53	MNAR	2
120	330	0.78	0.60	0.00	0.00	0.33	0.38	0.27	0.20	0.55	0.53	MNAR	2
121	330	0.00	0.28	0.00	0.15	0.38	0.49	0.00	0.09	0.38	0.55	MNAR	2
122	330	0.24	0.28	0.13	0.15	0.34	0.49	0.07	0.09	0.39	0.55	MNAR	2
123	330	0.45	0.28	0.24	0.15	0.30	0.49	0.14	0.09	0.42	0.55	MNAR	2
124	330	0.35	0.60	0.18	0.34	0.32	0.38	0.11	0.20	0.41	0.53	MNAR	2
125	330	0.55	0.60	0.30	0.34	0.26	0.39	0.18	0.20	0.41	0.53	MNAR	2

Table 40: Percentages for scenarios of treatment-policy estimand

No.	N	$\%_{TD_A}$	$\%_{TD_{PL}}$	$\%_{SD_A}$	$\%_{SD_{PL}}$	$\%_{E_{1_A}}$	$\%_{E_{1_{PL}}}$	$\%_{E_{2_A}}$	$\%_{E_{2_{PL}}}$	$\%_{E_{total_A}}$	$\%_{E_{total_{PL}}}$	TD mechanism	ϕ
126	330	0.75	0.60	0.44	0.34	0.23	0.38	0.25	0.20	0.45	0.52	MNAR	2
127	330	0.00	0.28	0.00	0.15	0.57	0.49	0.00	0.09	0.57	0.55	MNAR	2
128	330	0.28	0.28	0.15	0.15	0.49	0.49	0.09	0.09	0.55	0.55	MNAR	2
129	330	0.50	0.28	0.27	0.15	0.44	0.49	0.16	0.09	0.56	0.55	MNAR	2
130	330	0.39	0.60	0.20	0.34	0.47	0.38	0.13	0.20	0.56	0.53	MNAR	2
131	330	0.60	0.60	0.34	0.33	0.39	0.39	0.20	0.20	0.53	0.53	MNAR	2
132	330	0.78	0.60	0.46	0.33	0.33	0.38	0.27	0.20	0.55	0.53	MNAR	2
133	330	0.00	0.28	1.00	1.00	0.38	0.49	0.00	0.09	0.38	0.55	MNAR	2
134	330	0.24	0.28	1.00	1.00	0.34	0.49	0.07	0.09	0.39	0.55	MNAR	2
135	330	0.45	0.28	1.00	1.00	0.30	0.49	0.14	0.09	0.42	0.55	MNAR	2
136	330	0.35	0.60	1.00	1.00	0.32	0.38	0.11	0.20	0.41	0.53	MNAR	2
137	330	0.55	0.60	1.00	1.00	0.26	0.39	0.18	0.20	0.41	0.53	MNAR	2
138	330	0.75	0.60	1.00	1.00	0.23	0.38	0.25	0.20	0.45	0.52	MNAR	2
139	330	0.00	0.28	1.00	1.00	0.57	0.49	0.00	0.09	0.57	0.55	MNAR	2
140	330	0.28	0.28	1.00	1.00	0.49	0.49	0.09	0.09	0.55	0.55	MNAR	2
141	330	0.50	0.28	1.00	1.00	0.44	0.49	0.16	0.09	0.56	0.55	MNAR	2
142	330	0.39	0.60	1.00	1.00	0.47	0.38	0.13	0.20	0.56	0.53	MNAR	2
143	330	0.60	0.60	1.00	1.00	0.39	0.39	0.20	0.20	0.53	0.53	MNAR	2
144	330	0.78	0.60	1.00	1.00	0.33	0.38	0.27	0.20	0.55	0.53	MNAR	2

B R-code

B.1 Functions for the simulation of scenarios under MCAR

```
make.asthmatrial.longitudinal.MCAR <- function(N, dropout.tr = NULL,
                                               dropout.pl = NULL,
                                               event1.tr = NULL,
                                               event1.pl = NULL,
                                               event2 = NULL,
                                               overdispersion = NULL){

  # Generate N Cases
  id <- 1:N

  # 1. Grouping Variable
  group <- c(rep(1, floor(N/2)), rep(0, ceiling(N/2)))
  group <- sample(group)

  n.tr <- sum(group == 1)
  n.pl <- sum(group == 0)


  # 2. Time to treatment-discontinuation

  # Two Szenarios for Time to treatment-discontinuation:
  # 2.1. Time to treatment-discontinuation *different* for both groups

  if(dropout.tr != dropout.pl){

    dolambda.tr = (-1)*log(1 - dropout.tr)  # calculate lambda
    dolambda.pl = (-1)*log(1 - dropout.pl)

    time.dropout <- rep(NA, N)
    time.dropout[group == 1] <- ifelse(dolambda.tr != 0,
                                       rexp(n.tr, rate = dolambda.tr), 999)
    time.dropout[group == 0] <- rexp(n.pl, rate = dolambda.pl)


    # 2.2. Time to treatment-discontinuation *equal* for both groups
  } else {

    dolambda <- (-1)*log(1 - dropout.pl)
    time.dropout <- rexp(N, rate = dolambda)
  }


```

```
43
44  # 999 = no treatment-discontinuation
45  time.dropout[time.dropout > 1] = 999
46  time.observed <-  ifelse(time.dropout == 999, 1, time.dropout)
47
48
49  # 3. Time between events until treatment-discontinuation
50  #     Lambda *with* treatment effect
51  if(event1.tr != event1.pl) {
52
53    evlambda1.tr <- (-1)*log(1 - event1.tr)
54    evlambda1.pl <- (-1)*log(1 - event1.pl)
55
56    evlambda1.all.patients <- rep(0, N)
57    evlambda1.all.patients[group == 1] <- evlambda1.tr
58    evlambda1.all.patients[group == 0] <- evlambda1.pl
59
60
61  } else {
62    # Lambda *without* treatment effect
63    evlambda1 <- log(1 - event1.pl)*(-1)
64    evlambda1.all.patients <- rep(evlambda1, N)
65  }
66
67  # 3.1. Event rate individual for every person
68  if(overdispersion != 1){
69    # gamma-distributed event rates
70    # 3.1.1. *with* treatment effect
71    if(event1.tr != event1.pl) {
72
73      evlambda1.id <- rep(0, N)
74      beta1 <- 1/(overdispersion - 1)
75      alpha1.tr <- evlambda1.tr * beta1
76      alpha1.pl <- evlambda1.pl * beta1
77      evlambda1.id[group == 1]
78        <- rgamma(n.tr, shape = alpha1.tr, scale=1/beta1)
79      evlambda1.id[group == 0]
80        <- rgamma(n.pl, shape = alpha1.pl, scale = 1/beta1)
81    } else {
82
83      # 3.2.2. *without* treatment effect
84      beta1 <- 1/(overdispersion - 1)
85      alpha1 <- evlambda1 * beta1
86      evlambda1.id <- rgamma(N, shape = alpha1, scale = 1/beta1)
87    }
```

```
88
89      # Time between events for 3.1.
90      time.between.events <- vector("list", N)
91      number.event <- c()
92
93      for (i in 1:N) {
94        events <- c()
95        time.total <- 0
96
97        while(time.total <= time.observed[i]) {
98          events <- c(events, rexp(1, rate = evlambda1.id[i]))
99          time.total <- sum(events)
100       }
101
102       time.between.events[[i]] <- events
103       number.event<-c(number.event, length(time.between.events[[i]])-1)
104     }
105
106
107   } else {
108     # 3.2. Event rate *similar* for each individual
109     # Time to first events and further events for scenario *with*
          treatment effect and scenario *without* treatment effect
110
111     time.between.events <- vector("list", N)
112     number.event <- c()
113
114     for (i in 1:N) {
115       events <- c()
116       time.total <- 0
117
118       while(time.total <= time.observed[i]) {
119         events <- c(events, rexp(1, rate = evlambda1.all.patients[i]))
120         time.total <- sum(events)
121       }
122
123       time.between.events[[i]] <- events
124       number.event<-c(number.event, length(time.between.events[[i]])-1)
125     }
126
127   }
128
129
130
131
```

```
132 ##############################################################
133 # Simulation of events after treatment-discontinuation #
134 ##############################################################
135
136 # 4. Time between events after treatment-discontinuation
137
138 evlambda2 <- log(1 - event2)*(-1)
139 evlambda2.all.patients <- rep(evlambda2, N)
140
141
142
143 # 4.1. Event rate individual for every person
144 if(overdispersion != 1){
145   # gamma-distributed event rates
146
147   beta2 <- 1/(overdispersion - 1)
148   alpha2 <- evlambda2 * beta2
149   evlambda2.id <- rgamma(N, shape = alpha2, scale = 1/beta2)
150
151
152   # Time between events for 4.1.
153   number.event.total <- rep(NA, N)
154   time.between.events.total <- time.between.events
155
156   for (i in 1:N) {
157     # only for those whose treatment-discontinuation is before trial
          end
158     if(time.observed[i] < 1) {
159
160       # time to first event after treatment-discontinuation = time
            from dropout to first event + time from last event to
            treatment-discontinuation
161       events <- c(head(unlist(time.between.events[[i]]), -1),
162                   rexp(1, evlambda2.id)
163                   + (time.observed[i]
164                   - sum(head(unlist(time.between.events[[i]]), -1))))
165       time.total <- sum(events)
166
167       while(time.total <= 1) {
168         events <- c(events, rexp(1, rate = evlambda2.id[i]))
169         time.total <- sum(events)
170       }
171
172       time.between.events.total[[i]] <- events
173       number.event.total[i]<-length(time.between.events.total[[i]])-1
```

```
174         }
175
176         else{
177           number.event.total[i]<-length(time.between.events.total[[i]])-1
178         }
179       }
180     } else {
181
182       # 4.2. Event rate *similar* for each individual
183       # Time to first events and further events for scenario *with*
            treatment effect and scenario *without* treatment effect
184
185       number.event.total <- rep(NA, N)
186       time.between.events.total <- time.between.events
187
188       for (i in 1:N) {
189         # only for those whose treatment-discontinuation is before trial
              end
190         if(time.observed[i] < 1) {
191
192           # time to first event after treatment-discontinuation = time
                from dropout to first event + time from last event to
                treatment-discontinuation
193           events <- c(head(unlist(time.between.events[[i]]), -1),
194                       rexp(1, evlambda2.all.patients)
195                       + (time.observed[i]
196                       - sum(head(unlist(time.between.events[[i]]), -1))))
197           time.total <- sum(events)
198
199           while(time.total <= 1) {
200             events <- c(events, rexp(1, rate =evlambda2.all.patients[i]))
201             time.total <- sum(events)
202           }
203
204           time.between.events.total[[i]] <- events
205           number.event.total[i]<-length(time.between.events.total[[i]])-1
206         }
207         else{
208           number.event.total[i]<-length(time.between.events.total[[i]])-1
209         }
210       }
211     }
212
213
214
```

```
215   # 5. Build final dataset
216   id.final <- c()
217   group.final <- c()
218   time.dropout.final <- c()
219   time.between.events.final <- c()
220   time.between.events.total.final <- c()
221   occurrence.event.final <- c()
222   number.event.final <- c()
223   number.event.total.final <-  c()
224
225   for (i in 1:N) {
226     id.final <- c(id.final,
227       rep(id[i], length(time.between.events.total[[i]])))
228     group.final <- c(group.final,
229       rep(group[i], length(time.between.events.total[[i]])))
230     time.dropout.final <- c(time.dropout.final,
231       rep(time.observed[i], length(time.between.events.total[[i]])))
232     number.event.final <- c(number.event.final,
233       rep(number.event[i],
234       length(time.between.events.total[[i]])))
235     number.event.total.final <- c(number.event.total.final,
236       rep(number.event.total[i],
237       length(time.between.events.total[[i]])))
238     occurrence.event.final <- c(occurrence.event.final,
239       rep(1, length(time.between.events.total[[i]]) - 1), 0)
240     time.between.events.total[[i]]
241         [[length(time.between.events.total[[i]])]]
242       <- 1 - sum(head(unlist(time.between.events.total[[i]]), -1))
243     time.between.events.total.final
244       <- c(time.between.events.total.final,
245        unlist(time.between.events.total[[i]]))
246   }
247
248   asthmatrial <- data.frame(id.final, group.final,
249     time.between.events.total.final,
250     occurrence.event.final, time.dropout.final,
251     number.event.final, number.event.total.final)
252   colnames(asthmatrial) <- c("id", "group", "time.between.events",
253     "occurrence.event", "time.dropout",
254     "number.event", "number.event.total")
255
256   return(asthmatrial)
257 }
```

B.2 Functions for the simulation of scenarios under MNAR

```
make.asthmatrial.longitudinal.MNAR <- function(N, dropout.tr = NULL,
                                               dropout.pl = NULL,
                                               event1.tr = NULL,
                                               event1.pl = NULL,
                                               event2 = NULL,
                                               overdispersion = 1) {

  # Generate N Cases
  id <- 1:N

  # 1. Grouping Variable
  group <- c(rep(1, floor(N/2)), rep(0, ceiling(N/2)))
  group <- sample(group)

  n.tr <- sum(group == 1)
  n.pl <- sum(group == 0)


  # 2. Simulation of total number of events to calculate time for
      treatment-discontinuation in dependence of number of events

  #   Lambda *with* treatment effect
  if(event1.pl != event1.tr) {

    evlambda1.tr <- (-1)*log(1 - event1.tr)
    evlambda1.pl <- (-1)*log(1 - event1.pl)

    evlambda1.all.patients <- rep(0, N)
    evlambda1.all.patients[group == 1] <- evlambda1.tr
    evlambda1.all.patients[group == 0] <- evlambda1.pl


  } else {
    # Lambda *without* treatment effect
    evlambda1 <- log(1 - event1.pl)*(-1)
    evlambda1.all.patients <- rep(evlambda1, N)
  }


  # 2.1. Event rate individual for every person
  if(overdispersion != 1) {
    # gamma-distributed event rates
    # 2.1.1. *with* treatment effect

```

```
44      if(event1.tr != event1.pl) {
45
46        evlambda1.id <- rep(0, N)
47        beta1 <- 1/(overdispersion - 1)
48        alpha1.tr <- evlambda1.tr * beta1
49        alpha1.pl <- evlambda1.pl * beta1
50        evlambda1.id[group == 1] <- rgamma(n.tr, shape = alpha1.tr,
51                                           scale = 1/beta1)
52        evlambda1.id[group == 0] <- rgamma(n.pl, shape = alpha1.pl,
53                                           scale = 1/beta1)
54
55      } else {
56        # 2.1.2. *without* treatment effect
57        beta1 <- 1/(overdispersion - 1)
58        alpha1 <- evlambda1 * beta1
59        evlambda1.id <- rgamma(N, shape = alpha1, scale = 1/beta1)
60
61      }
62
63      # Time between events for 2.1.1 and 2.1.2
64      time.between.events <- vector("list", N)
65      number.event <- c()
66
67      for (i in 1:N) {
68        events <- c()
69        time.total <- 0
70
71        while(time.total <= 1) {
72          events <- c(events, rexp(1, rate = evlambda1.id[i]))
73          time.total <- sum(events)
74        }
75
76        time.between.events[[i]] <- events
77        number.event <- c(number.event,
78                          length(time.between.events[[i]])-1)
79      }
80
81
82    } else {
83      # 2.2. Event rate *similar* for each individual
84      # Time to first events and further events for scenario *with*
          treatment effect and scenario *without* treatment effect
85
86      time.between.events <- vector("list", N)
87      number.event <- c()
```

```
88
89      for (i in 1:N) {
90        events <- c()
91        time.total <- 0
92
93        while(time.total <= 1) {
94          events <- c(events, rexp(1, rate = evlambda1.all.patients[i]))
95          time.total <- sum(events)
96        }
97
98        time.between.events[[i]] <- events
99        number.event <- c(number.event,
100                          length(time.between.events[[i]]) - 1)
101     }
102
103   }
104
105   ##############################################
106   # simulation process for MNAR starts now #
107   ##############################################
108
109   # 3. Time to treatment-discontinuation depending on number of events
110
111   # Two Szenarios for Time to treatment-discontinuation:
112   # 3.1. Time to treatment-discontinuation *different* for both groups
113
114   if(dropout.tr != dropout.pl) {
115
116     dolambda.tr = (-1)*log(1 - dropout.tr)  # calculate lambda
117     dolambda.pl = (-1)*log(1 - dropout.pl)
118
119     dolambda.all.patients.mnar <- rep(0, N)
120     dolambda.all.patients.mnar[group == 1]
121         <- ifelse(number.event[group == 1] != 0,
122             dolambda.tr * number.event[group == 1], dolambda.tr)
123     dolambda.all.patients.mnar[group == 0]
124         <- ifelse(number.event[group == 0] != 0,
125             dolambda.pl * number.event[group == 0], dolambda.pl)
126
127     time.dropout <- rep(NA, N)
128     time.dropout[group == 1] <- ifelse(dolambda.tr != 0,
129       unlist(lapply(dolambda.all.patients.mnar[group == 1], rexp, n =
            1)),
130                                        999)
```

```
131       time.dropout[group == 0] <- unlist(lapply(dolambda.all.patients.
              mnar[group == 0],
132                                                 rexp, n = 1))
133
134     } else {
135
136       # 3.2. Time to treatment-discontinuation *equal* for both groups
137       dolambda <- (-1)*log(1 - dropout.pl)
138
139       dolambda.all.patients.mnar <- rep(0, N)
140       dolambda.all.patients.mnar <- ifelse (number.event != 0,
141                                             dolambda * number.event, dolambda
                                                  )
142
143       time.dropout <- unlist(lapply(dolambda.all.patients.mnar, rexp, n =
              1))
144
145
146     }
147
148     # 999 = no dropout
149     time.dropout[time.dropout > 1] = 999
150     time.observed <-  ifelse(time.dropout == 999, 1, time.dropout)
151
152
153     # 4. Number of events until treatment-discontinuation
154     number.event <- c()
155
156     for(i in 1:N) {
157       j <- 1
158       time.total <- 0
159       events <- c()
160
161       while (time.total <= time.observed[i]) {
162         time.total <- time.total + time.between.events[[i]][[j]]
163         events <- c(events, time.between.events[[i]][[j]])
164         j <- j + 1
165       }
166
167       time.between.events[[i]] <- events
168       number.event <- c(number.event, j - 2)
169
170     }
171
172
```

```
173 ##############################################################
174 # Simulation of events after treatment-discontinuation #
175 ##############################################################
176
177 # 5. Time between events after treatment-discontinuation
178 evlambda2 <- log(1 - event2)*(-1)
179 evlambda2.all.patients <- rep(evlambda2, N)
180
181
182 # 5.1. Event rate individual for every person
183 if(overdispersion != 1) {
184   # gamma-distributed event rates
185
186   beta2 <- 1/(overdispersion - 1)
187   alpha2 <- evlambda2 * beta2
188   evlambda2.id <- rgamma(N, shape = alpha2, scale = 1/beta2)
189
190
191   # Time between events for 5.1.
192   number.event.total <- rep(NA, N)
193   time.between.events.total <- time.between.events
194
195   for (i in 1:N) {
196     # only for those whose treatment-discontinuation is before trial
          end
197     if(time.observed[i] < 1) {
198
199       # time to first event after treatment-discontinuation = time
            from dropout to first event + time from last event to
            treatment-discontinuation
200       events <- c(head(unlist(time.between.events[[i]]), -1),
201                   rexp(1, evlambda2.id) + (time.observed[i]
202                   - sum(head(unlist(time.between.events[[i]]), -1))))
203       time.total <- sum(events)
204
205       while(time.total <= 1) {
206         events <- c(events, rexp(1, rate = evlambda2.id[i]))
207         time.total <- sum(events)
208       }
209
210       time.between.events.total[[i]] <- events
211       number.event.total[i]<-length(time.between.events.total[[i]])-1
212     }
213
214     else{
```

```
215           number.event.total[i]<-length(time.between.events.total[[i]])-1
216         }
217       }
218     } else {
219
220       # 5.2. Event rate *similar* for each individual
221       # Time to first events and further events for scenario *with*
              treatment effect and scenario *without* treatment effect
222
223       number.event.total <- rep(NA, N)
224       time.between.events.total <- time.between.events
225
226       for (i in 1:N) {
227         # only for those whose treatment-discontinuation is before trial
              end
228         if(time.observed[i] < 1) {
229
230           # time to first event after treatment-discontinuation = time
                from dropout to first event + time from last event to
                treatment-discontinuation
231           events <- c(head(unlist(time.between.events[[i]]), -1),
232                       rexp(1, evlambda2.all.patients) + (time.observed[i]
233                       - sum(head(unlist(time.between.events[[i]]), -1))))
234           time.total <- sum(events)
235
236           while(time.total <= 1) {
237             events <- c(events, rexp(1, rate = evlambda2.all.patients[i])
                  )
238             time.total <- sum(events)
239           }
240
241           time.between.events.total[[i]] <- events
242           number.event.total[i]<-length(time.between.events.total[[i]])-1
243         }
244
245         else{
246           number.event.total[i]<-length(time.between.events.total[[i]])-1
247         }
248       }
249     }
250
251
252     # 6. Build final dataset
253     id.final <- c()
254     group.final <- c()
```

```
255   time.dropout.final <- c()
256   time.between.events.final <- c()
257   time.between.events.total.final <- c()
258   occurrence.event.final <- c()
259   number.event.final <- c()
260   number.event.total.final <-  c()
261
262   for (i in 1:N) {
263     id.final <- c(id.final,
264       rep(id[i], length(time.between.events.total[[i]])))
265     group.final <- c(group.final,
266       rep(group[i], length(time.between.events.total[[i]])))
267     time.dropout.final <- c(time.dropout.final,
268       rep(time.observed[i], length(time.between.events.total[[i]])))
269     number.event.final <- c(number.event.final,
270       rep(number.event[i],
271       length(time.between.events.total[[i]])))
272     number.event.total.final <- c(number.event.total.final,
273       rep(number.event.total[i],
274       length(time.between.events.total[[i]])))
275     occurrence.event.final <- c(occurrence.event.final,
276       rep(1, length(time.between.events.total[[i]]) - 1), 0)
277     time.between.events.total[[i]]
278       [[length(time.between.events.total[[i]])]]
279         <- 1 - sum(head(unlist(time.between.events.total[[i]]), -1))
280     time.between.events.total.final
281       <- c(time.between.events.total.final,
282       unlist(time.between.events.total[[i]]))
283   }
284
285   asthmatrial <- data.frame(id.final, group.final,
286     time.between.events.total.final, occurrence.event.final, time.
          dropout.final,
287     number.event.final, number.event.total.final)
288   colnames(asthmatrial) <- c("id", "group", "time.between.events",
289     "occurrence.event", "time.dropout",
290     "number.event", "number.event.total")
291
292   return(asthmatrial)
293 }
```

B.3 Order of execution of the R-code

The R-files provided on the accompanying CD should be executed in the following order:

1. Simulation of scenarios
2. Evaluation of the hypothetical estimand with the Shared gamma frailty model
3. Evluation of the hypothetical estimand with the remaining methods
4. Evaluation of the treatment-policy estimand with the Shared gamma frailty model
5. Evaluation of the treatment-policy estimand with the remaining methods
6. Simulation and evaluation of the scenarios with study dropout of 50% for treatment-policy estimand
7. Calculation of the true treatment effects
8. Bias of hypothetical estimand
9. Bias of treatment-policy estimand without study dropout
10. Bias of the treatment-policy estimand with study dropout of 50%
11. Bias of the treatment-policy estimand with study dropout of 100%

For the reproduction of the graphic, please keep in mind the following order:

1. Preparation of the samples of the hypothetical estimand
2. Preparation of the samples of the treatment-policy estimand
3. Now, all files containing the code of the graphics and tables can be executed.

Verification of examination registration in FlexNow

Name: Ms Maria Elisabeth Stark
Matriculation No.: 21556770

Semester: WS17/18
Degree Course: Angewandte Statistik (Master of Science)
Module: Masterarbeit
Exam: Masterarbeit
Lecturer: PD Dr. rer. nat. Norbert Benda

Declaration

I hereby declare that I have produced this work independently and without outside assistance, and have used only the sources and tools stated.

I have clearly identified the sources of any sections from other works that I have quoted or given in essence.

I have complied with the guidelines on good academic practice at the University of Göttingen.

If a digital version has been submitted, it is identical to the written one.

I am aware that failure to comply with these principles will result in the examination being graded “nicht bestanden”, i.e. failed.

Göttingen, 20th April 2018 Maria Elisabeth Stark

Verification of examination registration in FlexNow

Name: Ms Maria Elisabeth Stark
Matriculation No.: 21556770

Semester: WS17/18
Degree Course: Angewandte Statistik (Master of Science)
Module: Masterarbeit
Exam: Masterarbeit
Lecturer: Prof. Dr. Tim Friede

Declaration

I hereby declare that I have produced this work independently and without outside assistance, and have used only the sources and tools stated.

I have clearly identified the sources of any sections from other works that I have quoted or given in essence.

I have complied with the guidelines on good academic practice at the University of Göttingen.

If a digital version has been submitted, it is identical to the written one.

I am aware that failure to comply with these principles will result in the examination being graded “nicht bestanden”, i.e. failed.

Göttingen, 20th April 2018 Maria Elisabeth Stark

Eidesstattliche Erklärung

Ich versichere, dass ich die Arbeit selbständig und ohne Benutzung anderer als der angegebenen Hilfsmittel angefertigt habe. Alle Stellen, die wörtlich oder sinngemäß aus Veröffentlichungen oder anderen Quellen entnommen sind, sind als solche kenntlich gemacht. Die schriftliche und elektronische Form der Arbeit stimmen überein. Ich stimme der Überprüfung der Arbeit durch eine Plagiatssoftware zu.

Göttingen, den 4.05.2018 Maria Elisabeth Stark