UHASSELT
KNOWLEDGE IN ACTION

Maastricht University

Faculty of Sciences
School for Information Technology
Master of Statistics and Data Science

Master's thesis

Simulation study on penalized generalized estimating equations after multiple imputation for repeated binary data with a rare event

Mark Mussner

Thesis presented in fulfillment of the requirements for the degree of Master of Statistics and Data Science, specialization Biostatistics

SUPERVISOR :

dr. Anna IVANOVA

2021
2022

Transnational University Limburg is a unique collaboration of two universities in two countries: the University of Hasselt and Maastricht University.

UHASSELT
KNOWLEDGE IN ACTION

www.uhasselt.be
Universiteit Hasselt
Campus Hasselt:
Martelarenlaan 42 | 3500 Hasselt
Campus Diepenbeek:
Agoralaan Gebouw D | 3590 Diepenbeek

Maastricht University

Faculty of Sciences
School for Information Technology
Master of Statistics and Data Science

Master's thesis

Simulation study on penalized generalized estimating equations after multiple imputation for repeated binary data with a rare event

Mark Mussner

Thesis presented in fulfillment of the requirements for the degree of Master of Statistics and Data Science, specialization Biostatistics

SUPERVISOR :

dr. Anna IVANOVA

Abstract

Background: If in a study population an event is observed only rarely, then computational problems may arise in the statistical analysis. It can lead to separation, which in turn may produce infinite coefficients and standard errors (SE). The Firth penalty has been described as one solution to problems with separation. Moreover, in clinical and sociological studies, it is common for data to have a complex structure due to a longitudinal design and to entail missing values. In the present thesis, the performance of the Firth penalty in a longitudinal study design with a dichotomous response variable and multiple imputation for values missing according to the MAR mechanism is investigated.

Method: Simulation studies were performed on the basis of the longitudinal study by Sommer et al. (1983) on the influence of vitamin A deficiency on the occurrence of respiratory infections in children. The simulations considered different sample sizes and different dropout probabilities. Firth Generalized Estimating Equations (F-GEE), standard GEE, Firth-logistic regression, and logistic regression were performed before and after multiple imputation (MI) of missing values. For the multiple imputations itself, a Firth logistic regression with adjusted intercept (FLIC) and a data augmentation algorithm (DA) were used.

Results: Standard GEE did not prove to be a valid method for analyzing data with rarely occurring events; the convergence rate was low, and the coefficients as well as SE were infinite in converged analyses for many simulation runs. This was particularly true for smaller samples or larger dropout probabilities. Classical logistic regressions showed high and uninterpretable coefficients for smaller samples. In spite of this, logistic regressions predominantly showed systematically lower SE, both with and without Firth penalty; this may be because these analyses do not consider the correlated nature of the data. F-GEE achieved a lower mean squared error and higher coverage for the 95% confidence interval of coefficient estimates after MI. FLIC imputation achieved lower bias than DA with respect to the predictor, which itself coded a rarely occurring event—namely, vitamin A deficiency.

Conclusion: The results of this thesis are consistent with previous scientific literature. F-GEE with MI according to FLIC demonstrated overall reliable performance in the analysis of the datasets simulated in this study.

Key words: Firth penalty, penalized generalized estimating equations, multiple imputation, separation, rarely occurring events

Content

1. Introduction

It is not uncommon in medical research for the clinical parameter of interest to occur rarely. For example, Sommer et al. (1983) investigated whether vitamin A deficiency in children can cause respiratory infections. Vitamin A deficiency is particularly prevalent in developing countries and manifests itself, among other things, as xerophthalmia, which is characterized by symptoms such as night blindness or Bitot spots in the eyes. The prevalence of xerophthalmia is, however, declining; it varies depending on the country and is generally below 5% (Sahile et al., 2020). It has been hypothesized for vitamin A deficiency to increase the risk of contracting a respiratory infection, the prevalence of which varies seasonally, ranging from 4% to 18% on average depending on the age of the child (Wang et al., 2016). In the statistical analysis of such relationships, a scenario with rarely occurring characteristics may lead to infinitely positive or negative regression coefficients as well as infinitely large standard errors, which are thus not interpretable (Heinze & Schemper, 2002). Further examples are given here, in which rare events have occurred in clinical research. The problem was described, for example, when analyzing the influence of radiation therapy on the development of lung cancer in smokers (Heinze & Schemper, 2002). Separation also occurred in the study of the occurrence of cannulation-site complications in the use of a new minimally invasive method for cardiac surgery (Puhr et al., 2017), as well as in the study of haematological complications after implant dentistry (Geroldinger et al., 2022).

The problem arises when the group of individuals showing the rare clinical parameter—for example, an infection—is almost completely separated from the other group by a predictor or by a combination of predictors. For example, if most of the children who have a respiratory infection at the time of the examination show a deficiency of vitamin A, and most of the other healthy children do not show this deficiency, then the separation is said to be near-to-quasi-complete (Albert & Anderson, 1984). Table 1 shows numerical examples of different scenarios of separation.

Table 1

Separation Scenarios

		Respiratory infection					
		Yes	*No*	*Yes*	*No*	*Yes*	*No*
Vitamin A	*Yes*	20	0	20	0	18	2
Deficiency	*No*	0	20	8	12	8	12
		Complete Separation		Quasi-complete Separation		Near-to-quasi complete Separation	

Note: Numbers are based on Mondol and Rahman (2019), adapted to the study of Sommer et al (1983).

In the scenario of complete separation and quasi-complete separation, a finite estimate of an association measure (i.e., odds ratio) does not even exist. The scenario of near-to-

quasi-complete separation is more common in practice and leads to infinitely large or small non-interpretable coefficients of association, as previously mentioned. It can be inferred from the examples in Table 1 that the problem of separation can arise in both large and small samples (Rogers & Stoner, 2021). The problem is aggravated for small samples because, firstly, separation is more likely and, secondly, in small samples the maximum likelihood estimator (MLE) is itself subject to bias (McCullagh & Nelder, 1989). Firth (1993) proposed a method to prevent the small sample bias, and Heinze and Schemper (2002) demonstrated that Firth's technique is also a solution in the case of infinitely large (or small) estimators or non-finite coefficients due to separation. David Firth proposed multiplying the likelihood function by a penalty (hence the term Firth's penalty) and thereby removing the first-order term of the asymptotic bias of the MLE. As a penalty term he suggested the root of the determinant of the Fisher information, $I(\theta)$. The penalized likelihood can be written as:

$$L^*(\theta) = L(\theta) \cdot |I(\theta)|^{\frac{1}{2}}$$

Firth's penalty has been shown to provide valid estimates for coefficients of a logistic regression in the context of rarely occurring events (Heinze & Schemper, 2002; Puhr et al., 2017) and recently, the method has also been shown to be valid in the context of repeated measurements with rare events applying a penalized form of Generalized Estimating Equations (GEE) (Mondol & Rahman, 2019).

The study by Sommer et al. (1983) can be used as an example of a setting with repeated measurements of rarely occurring clinical parameters: The participating children were regularly examined by a physician at three-month intervals to determine whether vitamin A deficiency or respiratory infection was present. This example also reveals another problem that has not yet been scientifically investigated in the context of rarely occurring events: missing values. There are some children who did not come for further visits and thus dropped out of the study. Missing values can be ignored in GEE only under the strict and unrealistic assumption that they are missing completely at random, which is termed MCAR (Molenberghs & Verbeke, 2005). However, if the child's dropout is dependent on previous infections, then the "missingness mechanism" is called missing at random (MAR) and missing values need to be addressed separately. In this respect, statistical research has developed techniques such as multiple imputation, which allow the application of a standard GEE model. Currently, there is no scientific research on the analysis of rarely occurring events where missing values had to be imputed before.

The aim of this master's thesis is to investigate the behavior of Firth's penalty with logistic regression and GEE after multiple imputation. The data of Sommer et al. (1983) will be present throughout this work, serving as a basis for analysis.

Chapter 1 will discuss the statistical basis in detail. The data of Sommer et al. (1983) will be presented in Chapter 2, and the setting of the simulation study will be described in Chapter 3. This will be followed by the results in Chapter 4, followed by the conclusions.

2. Description of the problem

In this chapter, the statistical basis needed to study the topic of research is described. The bias of the MLE in smaller samples is addressed first, followed by Firth's proposed solution. Next, I describe how Firth's technique not only solves the problem of small sample bias but also the issue of separation in the context of logistic regression and GEE. This is followed by a section on techniques for handling missing values and an elaboration of how multiple imputation could be used when analyzing rarely occurring events.

2.1. Small sample bias of the maximum likelihood estimate

The MLE is a consistent estimator; however, it is only asymptotically unbiased. Thus, in small samples, the MLE may be biased (Heinze & Schemper, 2002; McCullagh & Nelder, 1989). The bias of the MLE, $b = E(\hat{\beta} - \beta)$ was formulated by McCullagh and Nelder (1989) as follows:

$$b = (X^T W X)^{-1} X^T \xi$$

In the case of a binomial model with a logit link function, ξ can be written as:

$$\xi_i = h_i \left(\pi_i - \frac{1}{2}\right)$$

with h_i representing the i^{th} diagonal element of the hat matrix, H:

$$H = X(X^T W X)^{-1} X^T$$

and π_i being $P(y_i = 1 \mid x_i, \theta)$,
with $i = 1 \ldots n$ subjects, θ vector of parameters $\beta = 1 \ldots r$

The matrix $W = V^{-1}$ is the variance-covariance matrix of the modeled probability for the response variable Y having the binomial variances $\pi_i \cdot (1 - \pi_i)$ on its diagonal and $X^T W X$ being the Fisher information matrix. It should be emphasized, that in logistic regression, the variance is functionally dependent on the mean, $\mu = \pi$ (Firth, 2015). A bias adjustment to the MLE is needed to improve the approximation in small samples (McCullagh & Nelder, 1989).

2.2. Firth's solution to the small sample bias of the maximum likelihood estimate

Firth (1993) proposed a technique for removing the first-order term of the asymptotic bias, $O(n^{-1})$, of the MLE and thus for reducing the bias of the MLE in small samples. He suggested multiplying the likelihood function by a vague prior distribution. He proposed using Jeffreys invariant prior distribution, which is obtained from Jeffreys invariance principle and leads to a distribution proportional to the square root of the determinant of the Fisher information (Lesaffre & Lawson, 2012).
In the context of a logistic regression, with the Fisher information being $I(\theta) = X^T WX$ and with θ being the vector of the regression coefficients, the Jeffreys invariant prior can be written as $|X^T WX|^{½}$. The prior distribution can be thought of as a penalty function that causes the regression coefficients β to shrink toward 0 which, in this context, is equivalent to a shrinkage of the predicted probabilities, π_i, toward ½—the point where the determinant is maximized (Firth, 1993). The shrinkage is as large as needed to remove the first-order term of the bias and is asymptotically negligible (Heinze & Schemper, 2002). It should be mentioned that in the case of a logistic model the expected information, $i(\theta)$, and observed Fisher information, $I(\theta)$, do not differ, since the observed Fisher information is non-random (McCullagh & Nelder, 1989). In the following, the application of Firth's penalty in logistic regression is outlined.
The likelihood function of logistic regression is:

$$L(\theta) = \prod_{i=1}^{n} \left(\frac{e^{\theta}}{1+e^{\theta}}\right)^{y_i} \left(1 - \frac{e^{\theta}}{1+e^{\theta}}\right)^{(1-y_i)}$$

θ = vector of the regression coefficients
y = binary response variable
i = subjects *1* to *n*

The application of the penalty function to the likelihood can be formulated as:

$$L^*(\theta) = L(\theta) \cdot |I(\theta)|^{\frac{1}{2}}$$

with the penalized log-likelihood being:

$$l^*(\theta) = l(\theta) + \frac{1}{2} log|I(\theta)|$$

The score function is obtained by taking the derivative of the penalized log-likelihood with respect to the parameter vector θ:

$$U(\beta_r)^* = U(\beta_r) + \frac{1}{2} trace[I(\beta_r)^{-1}\{\partial I(\beta_r)/\ \partial\beta_r\}] = 0$$

with $r = 1 \ldots p$ β parameters in θ.

It should be mentioned that the log of a determinant of a matrix is equal to the trace of the log of the matrix. In contrast, the non-penalized, classical score functions for a (univariate) logistic regression can be written as follows:

$$U(\beta_r) = \sum_{i=1}^{n} (y_i - \pi_i) x_{ir} = 0$$

with the penalized score function resulting in:

$$U(\beta_r)^* = \sum_{i=1}^{n} \left\{ y_i - \pi_i + h_i \left(\frac{1}{2} - \pi_i \right) \right\} x_{ir} = 0$$

with $\pi = \frac{1}{1+e^{-\theta}}$.

The penalty term in the penalized score equation is $-h_i \left(\pi_i - \frac{1}{2} \right) x_{ir}$, with h_i representing the i^{th} diagonal elements of the hat matrix and it corresponds to subtracting $-i(\theta)b(\theta)$ from the score equation,

$U^*(\theta) = U(\theta) - i(\theta)b(\theta)$,

with *b(θ)* being the aforementioned bias of the MLE formulated by McCullagh and Nelder (1989). In a sense, the bias in the estimate of θ can be reduced by introducing a small bias in the score function (Firth, 1993). The presented technique is preventive rather than corrective, since the likelihood is penalized, and it is advantageous here that the asymptotic covariance matrix can be used for inference as usual. As in classical logistic regression, the Newton Raphson algorithm can be applied iteratively to the penalized score equations until convergence (Heinze & Schemper, 2002):

$\beta^{(s+1)} = \beta^{(s)} + I^{-1}\left(\beta^{(s)}\right)\ U\left(\beta^{(s)}\right)^*$, with s ... iteration 1 ... k, when convergence is met.

Heinze and Schemper (2002) showed that Firth's penalty term is a solution to the problem of separation. In a simulation study and in two re-analyses of clinical data, the authors impressively demonstrated that infinitely high and low parameter estimates from a logistic regression turned into finite parameter estimates through the Firth logistic regression.
The penalized score equations can also be understood as score equations for an augmented dataset. In Bayesian statistics, priors can be thought of as adding information to the data using pseudo-observations. As Firth's penalty is equivalent to using the Jeffreys prior for the logistic regression, applying Firth's penalty can also be seen as augmenting the data. To augment the data such that it corresponds to this specific Firth penalized equation, one would need to supplement each original observation with two pseudo-observations that receive $h_i/2$ as a weight, keeping the values in the covariates unchanged. One of the two created pseudo-observations would receive a zero as a response and the other would receive a one. If this augmented dataset were analyzed with a classical logistic regression, then the results of the Firth logistic regression using the original dataset would be obtained. Since the weights are taken from the hat matrix, and the trace of this matrix gives $(p+1)$, then the augmented dataset contains $(p+1)/2$ more events compared to the original dataset, with p equal to the number of covariates (Puhr et al., 2017).

2.3. Applying Firth's penalty to GEE

In the example by Sommer et al. (1983) mentioned in the introduction, children were repeatedly examined by a physician at intervals of three months. The observations made on the same child were thus not independent of each other, but rather dependent within a particular child. However, the (Firth-) logistic regression model outlined in the previous section does not take into account these dependencies between the observations. Therefore, the model needs to be adapted. There are several possibilities for doing this, but this thesis will focus on GEE.
GEE makes it possible to assume a correlation structure for the observations within a subject as a working guess. It is assumed that the working correlation structure and its parameters α are the same across all subjects. For each subject a separate variance-covariance matrix V_i is created with the model-based variances along the diagonal and the covariances on the off-diagonal, which are calculated from the diagonal variances by means of the assumed working correlation structure:

$$V_i = A_i^{\frac{1}{2}} R_i A_i^{\frac{1}{2}}$$

with A_i, a matrix with the marginal variances $\pi_i \cdot (1 - \pi_i)$ on the main diagonal, which are fully specified by μ through the mean-variance link in binomial models and with R_i

corresponding to the marginal working correlation matrix, which is parameterized by the vector α, which is not specified by the marginal mean μ (Molenberghs & Verbeke, 2005). The individual variance-covariance matrices are then stuck together to a block-diagonal matrix. In the score equations, the model-based variance-covariance matrix whose covariances are zero, is then replaced by this modified variance-covariance matrix. Due to this modification of the score equations, GEE is a semi-parametric, non-likelihood marginal model (Liang & Zeger, 1986; Molenberghs & Verbeke, 2005). Since the parameters of the working correlation structure are usually unknown, Liang and Zeger (1986) proposed a moment-based method for their estimation built on Pearson residuals: $e_{ij} = \frac{y_{ij}-\mu_{ij}}{\sqrt{var}_{ij}}$. The residuals are calculated in each iteration of the Newton Raphson algorithm and inserted into the following formulas—that is, when assuming an exchangeable working correlation structure:

$$\hat{\alpha} = \frac{1}{N}\sum_{i=1}^{N} \frac{1}{n_i(n_i-1)} \sum_{j\neq k} e_{ij}e_{ik}$$

or when assuming a first-order autoregressive working correlation structure, AR(1):

$$\hat{\alpha} = \frac{1}{N}\sum_{i=1}^{N} \frac{1}{n_i-1} \sum_{j\leq n_i-1} e_{ij}e_{i,j+1}$$

(Liang & Zeger, 1986; Molenberghs & Verbeke, 2005)

In order to achieve asymptotically correct β coefficients, it is necessary to correctly specify the marginal mean only. The assumed working correlation structure does not have to be correct; however, a correct assumption does lead to more efficient coefficient estimators (Molenberghs & Verbeke, 2005). In this thesis, the focus is on a dichotomous response variable, such as the presence or absence of a respiratory infection, although GEE can also be used to model response variables that have a different distribution, such as a Gaussian distribution. GEE makes it possible to leave all higher-order moments unspecified. This is an advantage in the context of dichotomous outcome variables, as the properties of multivariate normality in that case are not applicable; with dichotomous data, the full correct specification of the joint distribution would be a high burden (Molenberghs & Verbeke, 2005).
In the context of GEE, Firth's proposed penalty can now be used in the same way as with the logistic regression. This is possible because the described modifications of the variance-covariance matrix occur only at the level of the score equations and the Firth penalty is a penalty of the likelihood function. Mondol and Rahman (2019) showed in their simulation study that Firth GEE (F-GEE) leads to valid parameter estimates.

The authors formulate the penalized and modified score equations as follows:

$$U_r^*(\theta,\alpha) = \sum_{i=1}^{N}\left[R_i^{-1}(y_i-\mu_i) + h_i\left(\frac{1}{2}-\mu_i\right)\right]x_{ir} = 0$$

$r = 1$ to p coefficients

The estimation of the variances and covariances of the β parameters in θ, $var(\beta_r)$, also must consider the correlated nature of the data. As only the first moment must be correctly specified, but not the second moment, the Fisher information cannot be used as an asymptotically unbiased estimate for the variance-covariance matrix of the β regression coefficients. The so-called sandwich estimator, however, makes it possible to estimate asymptotically consistent and unbiased estimators for the standard errors of the regression coefficients, even if the working covariance structure is misspecified (Mancl & DeRouen, 2001; Rogers & Stoner, 2021). The sandwich estimator was developed for the context of GEE by Liang and Zeger (1986) and was formulated as follows:

$V(\hat{\beta}) = I_0(\hat{\beta})^{-1}\ I_1(\hat{\beta})\ I_0(\hat{\beta})^{-1}$ with $I_0(\beta)$ being the Fisher information, and

$I_1(\hat{\beta}) = XV^{-1}\ var(y)\ V^{-1}X$ with $var(y) = (y_i-\mu_i)\ (y_i-\mu_i)^T$

If $I_1(\hat{\beta})$ equals $I_0(\hat{\beta})$, the two terms cancel each other out, leading to the familiar, $V(\hat{\beta}) = I_0(\hat{\beta})^{-1}$. There is one further issue with the variances of β. With small sample sizes and in samples with rarely occurring events, the sandwich estimators are biased toward zero, and in that way they underestimate the variances of the β regression coefficients. As a consequence, the resulting *p* values would be too small and thus lead to an inflated type I error rate in small samples (Morel et al., 2003; Rogers & Stoner, 2021). Various solutions have been proposed to this issue. For example, it has been suggested to use a jackknife or bootstrap technique to derive standard errors. However, with small samples and rarely occurring events (i.e. in bootstrap), zero cell counts are encountered frequently and thus do not provide a proper solution in this case (Mancl & DeRouen, 2001). Morel et al. (2003) took a different approach to solving the problem, borrowing the idea of the design effect from sampling theory. The authors transferred the idea of clusters in sampling designs to clusters in correlated data and repeated measures in longitudinal studies. They proposed inflating the model-based covariance matrix, $I_0(\hat{\beta})^{-1}$ by an estimate of the design effect, $trace\{I_0(\hat{\beta})^{-1}\ I_1(\hat{\beta})\}$, and then multiplying it by a term of order n^{-1}, which vanishes as the number of subjects increases.

This inflated variance is then added to the sandwich estimator:

$$\hat{V}(\hat{\beta}) = V(\hat{\beta}) + \hat{\delta}_n \hat{\phi}\{I_0(\hat{\beta})\}^{-1}$$

with the overdispersion parameter

$$\hat{\phi} = max\left[1, tr\left\{I_0(\hat{\beta})^{-1} I_1(\hat{\beta})\right\} / p\right] \qquad \text{and} \qquad \hat{\delta}_n = min\left(0.5, \frac{p}{n-p}\right)$$

1 = bounded below by 1 (*over*-dispersion parameter)

0.5 = arbitrarily upper bound

and p being the number of coefficients.

In the simulation study by Rogers and Stoner (2021), the method was shown to be a valid alternative to the sandwich estimator of Liang and Zeger (1986) in the context of small samples and samples with rarely occurring events. Moreover, the correction was applied in the context of F- GEE by Mondol and Rahman (2019) and Geroldinger et al. (2022).

2.4. Missing values

The analysis of longitudinal, binary data becomes more complex when there are missing measurements for individual subjects. For example, in clinical research, patients might not return for subsequent visits (dropout pattern), or they might skip individual visits in between (intermittent pattern). Although GEE can cope with the problem of missing values by simply calculating with the observed responses, these only lead to consistent coefficients if the responses are MCAR (Lipsitz et al., 2020; Rubin, 1976). A missingness mechanism is called MCAR, if missingness neither depends on (previously) observed response values nor on the unobserved measurements; missingness is, however, allowed to depend on observed covariates (Little & Rubin, 1987; Rubin, 1987). MCAR is a strong and, in practice, often unrealistic assumption. However, if missingness depends on the (previously) observed response values and is still independent of unobserved measurements conditional on the given observed data, then the mechanism is referred to as MAR. Under MAR assumption, GEE do not lead to valid results since, due to the modification of the score equations, it is a non-likelihood, semiparametric, frequentist method. As already mentioned, the variance-covariance-matrix using a working assumption for the correlation structure can be wrong; and as the variance is involved in the prediction model for the missing values in case of MAR, it needs to be correctly specified (Liang & Zeger, 1986; Lipsitz et al., 2020; Molenberghs & Verbeke, 2005).
Various approaches have been developed to adequately deal with missing values. One of these is the technique of multiple imputation (MI), and another is the approach of weighting all observations with the inverse probability that a participant would have a

missing value at a particular measurement time point (Molenberghs & Verbeke, 2005). Under MAR, both techniques, MI-GEE and Weighted-GEE (W-GEE) yield valid results. Beunckens et al. (2008), however, showed that for small- and medium-sized samples MI-GEE could achieve results with lower bias compared to W-GEE. For this reason, only the MI approach is further pursued in this thesis, as small to medium samples and samples with rarely occurring events are the focus of research.
The following equation refers to the MCAR mechanism, where R is the missing data indicator that is not dependent on *Y,* and ψ, the vector of parameters of the dropout model:

$$f(R|Y,\psi) = f(R|\psi) \qquad \text{for all } Y,\psi$$

For MAR, where R depends on Y_{obs}, the equation can be written as

$$f(R|Y,\psi) = f(R|Y_{obs},\psi) \qquad \text{for all } Y_{mis},\psi$$

with covariates *X* suppressed from the equations (Molenberghs & Verbeke, 2005).

MI was proposed by Rubin (1976) and consists of replacing the missing values with *M* values from a distribution of likely values (Rubin & Schenker, 1986). The number of imputations *M* must not be very large, and values such as *M* = 5 or *M* = 10 are often considered. In this way, *M* completed datasets are generated, retaining the structure as well as the sampling uncertainty of the original dataset (van Buuren et al., 2006). The *M* completed datasets can then be analyzed in parallel with, for example, standard GEE, or indeed Firth-GEE. The *M* results are finally combined with Rubin's rules to draw an inference (Rubin, 1976; van Buuren et al., 2006).
According to Rubin, the MI estimate of β, $\hat{\beta}^*$, is calculated as the mean of the *M* coefficients:

$$\hat{\beta}^* = \frac{\sum_{m=1}^{M} \hat{\beta}^m}{M}$$

leading to a normal based inference with

$$\left(\hat{\beta}^* - \beta\right) \sim N(0, V)$$

The MI variance estimator, *V*, consists of two parts, the within-imputation variance, *W*, and the between-imputation variance, *B*. The variance components *W* and *B* are combined as follows:

$$V = W + \left(\frac{M+1}{M}\right) B$$

with $W = \frac{\sum_{m=1}^{M} U^m}{M}$ and $U = \widehat{var}(\hat{\beta})$

with $B = \frac{\sum_{m=1}^{M} (\hat{\beta}^m - \hat{\beta}^*)(\hat{\beta}^m - \hat{\beta}^*)'}{M-1}$

(Beunckens et al., 2008; Lipsitz et al., 2020)

MI requires a model for the prediction of the missing values. One possible approach is to specify a joint distribution, which is then factorized into the marginal distribution of the observed values and into the conditional distribution of the missing values, given the observed values. From this conditional distribution, likely values can then be drawn at random to replace the missing values. Since the parameters of the predictive model are themselves unknown and need to be estimated, the uncertainty of its estimation must be taken into account when drawing values based on the prediction model. This is implemented by either first randomly drawing a parameter value from the posterior distribution of the parameter or by randomly drawing a parameter value from an approximation of a normal distribution using the specific β coefficient as the mean and its standard error, $\hat{\beta} \pm z \cdot V\hat{}(1/2)$, with $z = 1.96$. Each prediction is then drawn based on specific, randomly selected parameter values (Beunckens et al., 2008; Molenberghs & Verbeke, 2005; van Buuren et al., 2006).

A second possibility to specify a prediction model is the direct formulation of conditional distributions of missing values conditional on the observed values without having to specify a joint distribution. In this case, a predictive model is formulated for each variable containing missing values, conditional on the observed values of all other variables in the model. This approach was proposed by van Buuren et al. (2006) and is termed Fully Conditional Specification (FCS). Also in this method, the uncertainty of the estimation of the model parameters must be taken into account as described, for example, using an approximation of the normal distribution. As an example of FCS, in a longitudinal design with a dichotomous response, a logistic regression model could be specified for each measurement time point that counts missing values, employing the other time points—or just the previous measurement time points and potentially covariates—as predictors. The FCS algorithm works iteratively until convergence by first assuming initial values for the missing values and then repeatedly going through all logistic regressions one after the other and imputing missing values under the condition of the values imputed in the previous steps. In each step, a new parameter value is first drawn from the parameter distribution to account for the variability of the parameters themselves. When convergence is reached, it is assumed that values are drawn from a stationary distribution, and these values are then kept as imputation for the missing values. This algorithm is repeated M times. The idea of modeling a joint distribution by drawing values from conditional distributions corresponds to that of Gibbs sampling (Schöning, 2013). It may now be that the joint distribution does not exist, since only the conditional distributions were specified. Simulation results showed that FCS leads to unbiased

results despite these theoretical considerations about a potentially nonexistent joint distribution. If the joint distribution exists, then FCS corresponds to a Gibbs sampling (van Buuren, 2007). Both techniques—the method via joint distribution and FCS—make use of Monte Carlo Markov Chains (MCMC) (Bartlett et al., 2015; van Buuren et al., 2006). The algorithm is briefly outlined for the first two iterations of the first imputation and is valid up to a constant of proportionality:

draw from posterior of θ | vague prior | observed given all other variables but one

$$\theta_1^{(t)} \sim f(\theta_1) \cdot f\left(y_1^{obs} | y_{-1}^{(t)}, x, \theta_1\right) \quad \text{with } y_{-1} = (y_2, \ldots y_n)$$

draw from conditional predictive distribution

$$y_1^{mis(t)} \sim f\left(y_1^{mis} | y_{-1}^{(t)}, x, \theta_1^{(t)}\right)$$

next iteration

$$\theta_2^{(t)} \sim f(\theta_2) f\left(y_2^{obs} | y_{-2}^{(t)}, x, \theta_2\right) \quad \text{with } y_{-2} = (y_1, y_3, \ldots y_n)$$

$$y_2^{mis(t)} \sim f\left(y_2^{mis} | y_{-2}^{(t)}, x, \theta_2^{(t)}\right)$$

In studies such as Sommer et al. (1983), where the presence or absence of respiratory infection is repeatedly assessed, the joint distribution is not as straightforward as in the case of normally distributed response variables. In the case of dichotomous variables, this would require the specification of all higher order associations. FCS provides a direct way of doing this in that only the logistic regression equations for each visit must be formulated. Since the focus of this thesis is on dichotomous response variables, only FCS will be considered in the following.

2.5. Firth's penalty in the context of missing values

In this section, the situation of rarely occurring events is again addressed. Particularly in the context of FCS, some points need to be considered in greater detail.
For the predictive model, in order to draw a parameter value from its distribution, in the frequentist approach, the standard error of the information matrix would be used to provide an approximation of the normal distribution from which values could be drawn. However, this approach cannot be pursued because in a scenario of separation, the standard errors of the parameters can become too large. A normal distribution approximation would be very flat and almost any value would be likely for the parameter;

a high between-imputation variance would be the consequence (White et al., 2010). At this point, it could be objected that Firth logistic regression could be used for imputation due to its ability to produce finite estimates. However, it has been shown that Firth logistic regression leads to valid parameter estimates but introduces bias into the predicted probabilities (Puhr et al., 2017; van Buuren, 2018). In the context of imputation, valid predictions are precisely what is of interest. The problem is that when applying the Firth penalty to the likelihood, the coefficients shrink toward 0, which corresponds to a shift on the probability scale in the direction of ½. Due to rarely occurring events, there are now many non-responses, for which probabilities below 0.5 are predicted. With larger numbers of smaller probabilities being modified upwards in the direction of 0.5, this results in an overestimation of the predicted proportion of events, such that it no longer holds that $\sum_{i=1}^{n} y_i = \sum_{i=1}^{n} \pi_i$. Puhr et al. (2017) proposed to re-estimate the intercept after applying the Firth logistic regression and thus to re-satisfy this equation by analogy to a calibration. The recalculation of the intercept is done by using the linear combination of the Firth coefficients without intercept as offset of a classical logistic regression. The authors have shown in simulation studies that the method leads to valid predictions. Puhr (2016) outlined the algorithm of Firth logistic regression with corrected intercept (FLIC) as follows:

1. Calculate coefficient estimates with Firth logistic regression
2. Calculate the linear predictor, η, of Firth penalized coefficient estimates without including the intercept.
3. Calculate a classic logistic regression using the linear predictor η from step 2 as offset, $\pi(\eta) = \frac{1}{1+e^{-\gamma_0-\gamma_1\eta}}$ *setting* $\gamma_1 = 1$.
4. Use γ_0 as adjusted intercept along with remaining coefficients from step 1.

FLIC will be discussed again later when the simulation study is planned.
Puhr et al. (2017) proposed a second method that achieved valid predictions in their studies. The method involves augmenting the dataset with pseudo-observations corresponding to the effect of the Jeffreys prior used as Firth's penalty and then including a covariate in the logistic model that differentiates between the augmented and the original dataset. ML estimation of the parameters then leads to valid predictions under this configuration.

3. Motivating case: Dr. Sommer's investigation of vitamin A deficiency

In the course of this thesis, the study by Sommer et al. (1983) has been repeatedly cited as an illustration. This section will describe the study in detail and use it as a motivating case for the empirical section that follows later. Subsequently, an exploratory analysis of a subset of the data from this study will be presented.

3.1. Dr. Sommer's investigation on vitamin A

Dr. Alfred Sommer is professor of ophthalmology and epidemiology at Johns Hopkins University. Dr. Sommer and his colleagues have devoted their research to studying the effect of vitamin A deficiency in children on childhood mortality in developing countries. Vitamin A deficiency leads to keratinization of the respiratory, gastrointestinal, and genitourinary tracts, increasing the risk for bacterial colonization and infection. Vitamin A deficiency is manifested by so-called Bitot spots in the ocular conjunctiva, which are the clinical appearance of xerophthalmia leading to night blindness. Contrary to severe vitamin A deficiency, a mild form of this deficiency occurs in children who are otherwise well nourished and appear healthy (Sommer et al., 1983). Dr. Sommer and his research team examined 3,593 Indonesian children up to six years of age over an 18-month period in the late 1970s. A medical team consisting of an ophthalmologist, a pediatrician, and nutritionists assessed the children at three-month intervals. The ophthalmologist diagnosed xerophthalmia using the standard diagnostic criteria of the time by means of a hand light and a magnifying lens. When severe xerophthalmia was diagnosed, children were immediately treated with vitamin A and hospitalized. Children with mild xerophthalmia were treated in the same manner as children without such symptoms. For children with a severe infection, medical care was provided and referrals were made to local physicians.

The researchers found a higher mortality even in children with mild xerophthalmia compared to those without this diagnosis. Moreover, children with mild xerophthalmia were at higher risk for respiratory infections compared with the other children.

The prevalence of xerophthalmia was 5.5% on average and respiratory infection was identified in 8.8% of all physician visits. A total of 132 child deaths were counted. The study had a profound impact on global efforts to increase vitamin A levels in children. With the help of WHO, supplements to 12 million children were distributed into 40 countries each year (Johns Hopkins Bloomberg School of Public Health, 2002).

A subset of this dataset including 275 children is publicly available and has already been used by researchers to explore statistical methods (Diggle et al., 2002; Zeger & Karim, 1991). The dataset contains up to six visits per child and the information on xerophthalmia and respiratory infection at each measurement time. In addition, there is information on gender, age, height, or the season of measurement. The available dataset does not include mortality information.

This dataset provides the opportunity to apply statistical analysis to infrequently occurring events. On the one hand is the occurrence of respiratory infection and on the other hand xerophthalmia. The aim of this thesis is to study the behavior of the penalized analyses under multiple imputation of missing data. For this purpose, however, I would like to start with a complete dataset in order to be able to perform the subsequent simulations under controlled conditions. In the following, only those children (n = 138) from the dataset who were present at the first four medical visits and had no missing values were included.

3.2. Exploratory analysis of a complete subset of Summer et al. (1983)

In this section an exploratory analysis of the subset from the data collected by Sommer et al. (1983) is presented. It should be noted at this point that the analysis of these data was done for illustrative purposes only in the context of the study of statistical methods for rarely occurring events. For precise statements on vitamin A deficiency and its effects, the reader may wish to refer to the original study.
The complete dataset with four measurement time points included 138 Indonesian children, of whom 62 (44.9%) were female.
At the start of the study, the children were on average 31 months old. The youngest child was four months, and the oldest child was 71 months (i.e., just under six years old). Due to cultural conditions on Jawa Island, it was not possible to include even younger children in the study. The proportion of children diagnosed with xerophthalmia at each measurement time point varied between 2% and 7% of the children, depending on the visit, and between 4% and 12% of the children had a respiratory infection. Boys (7%) were more likely to have symptoms of xerophthalmia than girls (2%); however, the mean age of the children was comparable between boys (30.1 months) and girls (31.5 months). Moreover, with increasing age there was an increase in the risk of xerophthalmia for both males and females (OR = 1.02). The results of the Firth-GEE model are shown in Table 2.

$$logit\,(P[\text{infection}_{ij}]) = \beta_0 + \beta_1 * \text{gender}_i + \beta_2 * \text{age}_{ij} + \beta_3 * \text{xerophthalmia}_{ij}$$

i = child 1 to 138
j = visit 1 to 4, working correlation = first-order autoregressive (AR-1)

There were no significant predictor interactions on the occurrence of respiratory infections. Children who showed symptoms of xerophthalmia were significantly more likely to have a respiratory illness. The odds of suffering from an infection increased four-fold when xerophthalmia was present. The odds of having a respiratory infection decreased over time with increasing age; specifically, the odds decreased by about 3% per month of age. There was no significant difference between boys and girls in the

probability of acquiring a respiratory infection. It was tested whether a quadratic term for age contributed to the model, however, there was no significant quadratic age effect.

Table 2
Modeling Probability of Respiratory Infection using F-GEE

	β	SE	Wald	p
(Intercept)	−2.807	0.264	112.680	<0.001
Female vs. male	−0.347	0.384	0.817	0.366
Age	−0.028	0.009	9.831	0.002
Xerophthalmia	1.418	0.557	6.483	0.011

Note: response = respiratory infection, F-GEE = Firth penalized Generalized Estimating Equations

4. Description of the methodology

The purpose of this thesis is to investigate how penalized analyses work in longitudinal study designs with binary response after imputation of missing values—that is, when applying MI-F-GEE. This research question will be answered by using a simulation study based on the study by Sommer et al. (1983). In this section, the process of data generation will first be described. Subsequently, how missing values are created for the generated dataset according to the MAR mechanism will be explained. Third, the multiple imputation of this incomplete dataset will be described and, finally, the procedure for the statistical analysis of the imputed data will be outlined. The simulation study was performed using R Statistics, version 4.1. For the purpose of replicating the results, a seed was set following Geroldinger et al. (2022), which was coupled to the index of the for-loops. Excerpts of the programmed code may be found in the appendix.

4.1. Data generation for the simulation study

For the generation of the binary response, respiratory infection, the predictors gender, age, and xerophthalmia were used with the coefficients, $\beta_0 = -2.807$, $\beta_1 = -0.347$, $\beta_2 = -0.028$, $\beta_3 = 1.418$, from the Firth-GEE model with a first-order autoregressive correlation structure, $logit\ (P[\text{infection}_{ij} = 1]) = \beta_0 + \beta_1 * \text{gender}_i + \beta_2 * \text{age}_{ij} + \beta_3 * \text{xerophthalmia}_{ij}$, $i = 1 \ldots N$, $j = 1 \ldots 4$ visits. This model was presented in the previous section. The associations between the previously described predictors were taken into account in the simulation. Xerophthalmia is a time-dependent variable, as the value could

change from visit to visit. Gender and age in months were already known at the beginning of the study. Due to the longitudinal design and the three-month intervals of the visits, the age in months has also been used to indicate the four repeated measures. In the first step, a Bernoulli random variable of length N with probability $p = 0.45$ was created as the gender variable. In the second step, a truncated normal random variable was created as the age variable, centered at 36 months. The normal distribution was truncated downward and upward to ensure that no child was younger than four months or older than six years. The mean of this centered age variable was similar to the Sommer et al. (1983) dataset. From this age variable, the time indicator was obtained for each child by adding three months from the age at the beginning of each of the four measurement points. Thus, for each child there were four values—for example 7, 10, 13, and 16 months of age at the respective measurement points. The generation of the variable for xerophthalmia was somewhat more complex as it is a time-dependent variable that was related to the other predictors age and gender. To generate this variable, and later to generate the response variable—respiratory infections—the method of Qaqish (2003) was followed. This method was used by Mondol and Rahman (2019) in creating the response variable when exploring F-GEE. Qaqish (2003) described how a correlated binary variable can be simulated by specifying only the mean structure, the correlation, and the correlation structure. The full specification of the joint distribution is not required. For the first expression of this binary variable, a Bernoulli random variable is generated, in this concrete case with the mean value given from *logit* (P[xerophthalmia$_i$ = 1]) = $\beta_0 + \beta_1 *$ gender$_i$ + $\beta_2 *$ age$_i$. The values for the further three measurement points are then generated under the condition of the already generated value(s). The assumed correlation and the correlation structure are thereby used to weight the already generated values. This conditional mean is written in Qaqish (2003) as:

$$E(Y_i|X_i = x_i) = \mu_i + \sum_{j=1}^{i-1} b_{ij}(y_j - \mu_j)$$

$i = 2, \ldots, n,$

b = weights calculated from correlation structure

For each child, this creates a vector of dependent values that have the required mean and the specified correlation magnitude and structure. In addition, Qaqish (2003) provides the check of the following two preconditions. First, it checks that the correlation matrix is positively defined, and second, depending on the predetermined correlation structure, that the covariances are within the bounds imposed by the marginal mean. Professor Qaqish has made his method available in the R package `binarySimCLF`. For the generation of the response variable, respiratory infections, at the end, an analogous procedure was followed. The method of Qaqish was used with the following model for the generation of the (conditional) mean, *logit*(P[infection$_{ij}$ = 1]) = $\beta_0 + \beta_1 *$gender$_i$ + $\beta_2 *$age$_{ij}$ + $\beta_3 *$ xero$_{ij}$. Because of the rare occurrence of both events, xerophthalmia and respiratory infection, F-GEE was calculated to obtain the coefficients estimates.

This algorithm resulted in simulated datasets that corresponded with the original dataset of Sommer et al. (1983) in its structure. The comparison of the means of 1000 simulated datasets with the original dataset is shown in Table 3. Datasets that by chance did not contain a single infection or contained only one infection, no xerophthalmia or only one xerophthalmia were rejected.

Table 3

Comparison of Means (Proportions) in Original and Simulated Datasets

	infection	**gender**	**age**	**xero**	**xero female**	**xero male**
Original	0.060	0.449	30.5	0.047	0.020	0.069
Simulation	0.059	0.452	31.7	0.049	0.024	0.070

Note: infection = respiratory infection, xero = xerophthalmia, n = 1000 simulations

4.2. Generating missing values

The algorithm described in the previous section generates complete datasets. To investigate the behavior of Firth GEE after multiple imputation, however, it is necessary to generate datasets containing missing values. To this end, a dropout pattern with MAR mechanism is considered. It was assumed that with four repeated measurement time points, children could drop out at the second, third, or fourth time point. A dropout results in both the information on respiratory infection and the information on the time-dependent variable xerophthalmia not being available for subsequent visits. Although intermittent missing patterns might also have been plausible, these are not considered in this thesis. In the scenario used for this simulation study, three different dropout patterns result, which are reproduced in Table 4.

Table 4

Three Dropout Missingness Patterns for Scenario of Sommer et al. (1983)

#	**gender**	**age**	**xero1**	**resp1**	**xero2**	**resp2**	**xero3**	**resp3**	**xero4**	**resp4**
1	1	1	1	1	1	1	1	1	**0**	**0**
2	1	1	1	1	1	1	**0**	**0**	**0**	**0**
3	1	1	1	1	**0**	**0**	**0**	**0**	**0**	**0**

Note: 1 = observed, 0 = missing, 1–4 = medical visits, # = pattern, resp = respiratory infection, xero = xerophthalmia

As described previously, the MAR mechanism consists of the probability of the dropout depending on already observed response values. It is further possible that the dropout probabilities additionally depend on covariates. In this specific case, the response variable is the dichotomous variable with rarely occurring respiratory infections. In order to produce a dataset with missing values corresponding to MAR, the dropout

probabilities must now depend on whether a child was diagnosed with a respiratory infection at a previous visit. Sommer et al. (1983) described that children who were diagnosed with an infection were referred to local physicians. It is conceivable that these children would remain on treatment there and then not return for further visits with the initial physicians of the study. This would mean that at each visit, children with an infection would be more likely to drop out than children without an infection. This setting is assumed in the simulation, and 30%, 40%, 60%, and 80% dropout probabilities are modeled for infected children. For children without an infection, the lower dropout probabilities 10% and 20% were applied. The combinations of these probabilities, shown in Table 5, were then used for the simulations. The goal was to produce an MAR mechanism that was as clear and sharp as possible and not confounded with MCAR. For this reason, first, I left out the covariates when generating dropout dependencies, and second, I omitted the 30%/20% and 40%/20% combinations for infected and uninfected children, respectively, because of the small difference. Although an 80% dropout probability is rather unrealistic in practice, I included this scenario to be able to test the performance of MI-F-GEE even under an extreme situation. Generating a clear MAR mechanism was challenging, especially due to the infrequently occurring events. The MAR mechanism was validated by modeling dropout (yes / no) as an outcome variable in a logistic regression analysis with the previous response as a predictor. It should be noted that due to the rarely occurring (previous) event, which was used as a condition for generating dropout, separation may also occur in this case. Firth-logistic regression was therefore used for the analysis. The predictor had a statistically significant effect on dropout; this was taken as evidence that the dropout mechanism was not MCAR.

Table 5

Scenarios for Simulating Datasets by Sample Size and Dropout Probabilities

	Dropout probabilities conditioned on previous infection (yes / no)					
n	**0.3/0.1**	**0.4/0.1**	**0.6/0.1**	**0.8/0.1**	**0.6/0.2**	**0.8/0.2**
50	x	x	x	x		
100	x	x	x	x		
200	x	x	x	x	x	x
500	x	x	x	x	x	x

Note: x = scenario, which was applicated; in each scenario the following analyses were performed: DL-GLM, DL-F-GLM, GEE, F-GEE, MI-GLM, MI-F-GLM, MI-GEE, MI-F-GEE, multiple imputation was conducted with FLIC and with DA technique; DL = direct likelihood.

4.3. Imputing missing values

Once datasets with missing values following MAR were available, it was possible to perform a multiple imputation in this simulation study. FCS was already introduced theoretically in this thesis as a multiple imputation technique, and this technique was used for this purpose. In FCS, for each variable containing missing values, a univariate

imputation model is required. Based on the three dropout patterns given in the previous section in Table 4, the information on respiratory infection and on xerophthalmia could be missing at visits 2 to 4. The following predictive imputation models were specified for these two variables:

response variable: respiratory infection

$$logit\,(P[\text{infection}_{\text{visit 2}}= 1]) = \beta_0 + \beta_1 * \text{infection}_{\text{visit 1}} + \beta_2 * \text{xero}_{\text{visit 1}} + \beta_3 * \text{age}$$
$$logit\,(P[\text{infection}_{\text{visit 3}}= 1]) = \beta_0 + \beta_1 * \text{infection}_{\text{visit 2}} + \beta_2 * \text{xero}_{\text{visit 2}} + \beta_3 * \text{age}$$
$$logit\,(P[\text{infection}_{\text{visit 4}}= 1]) = \beta_0 + \beta_1 * \text{infection}_{\text{visit 3}} + \beta_2 * \text{xero}_{\text{visit 3}} + \beta_3 * \text{age}$$

time-dependent predictor: xerophthalmia

$$logit\,(P[\text{xero}_{\text{visit 2}}= 1]) = \beta_0 + \beta_1 * \text{infection}_{\text{visit 1}} + \beta_2 * \text{xero}_{\text{visit 1}} + \beta_3 * \text{age}$$
$$logit\,(P[\text{xero}_{\text{visit 3}}= 1]) = \beta_0 + \beta_1 * \text{infection}_{\text{visit 2}} + \beta_2 * \text{xero}_{\text{visit 2}} + \beta_3 * \text{age}$$
$$logit\,(P[\text{xero}_{\text{visit 4}}= 1]) = \beta_0 + \beta_1 * \text{infection}_{\text{visit 3}} + \beta_2 * \text{xero}_{\text{visit 3}} + \beta_3 * \text{age}$$

Based on these imputation models, $M = 10$ imputations for each missing value were generated for each dataset. The FCS procedure is implemented in the R package `MICE` (van Buuren & Groothuis-Oudshoorn, 2011). However, due to the rare occurrences, special precautions were needed to obtain valid imputations. A classical logistic regression analysis could lead to very large standard errors and possibly impute only zeros or only ones (van Buuren, 2018). In the `MICE` package, a technique of data augmentation (White et al., 2010), which is similar in approach to the Firth correction, is used to prevent perfect predictions. For each predictor, two pseudo-observations are introduced, at $\bar{x} + SD$ and at $\bar{x} - SD$; this also for dichotomous variables. One of these pseudo-observations is assigned the response 0 and the other is assigned 1. These pseudo-observations receive a low weight of $(p+1)/(2 * p *$ levels of $Y)$. The sum of the weights is equal to $(p + 1)$, which is the number of parameters in the model. The technique of White et al. (2010) likewise removes the first-order bias of the MLE. On the other hand, it has been described earlier that a Firth penalty leads to valid parameter estimates but introduces a bias in the prediction of values. Puhr et al. (2017) proposed a Firth logistic regression with recalibrated intercept: FLIC. The method is implemented in the R package `logistf` (Heinze & Schemper, 2002) and was presented in the theoretical part of this paper. Therefore, in this thesis, imputation was performed using both techniques—data augmentation (White et al., 2010) and FLIC (Puhr et al., 2017) —and the results were compared in terms of their performance. In the `mice.impute.logreg` function in the `MICE` package, the `logistf` function was called with the option FLIC = TRUE instead of the algorithm of White et al. (2010). If by chance in the course of a multiple imputation a dataset with zero or only one event in respiratory infection or xerophthalmia was generated, then this dataset was discarded.

4.4. Analysis of datasets

Once the imputed datasets were available, the data was ready to be analyzed. The same model that was presented earlier on the exploratory analyses of the data by Sommer et al. (1983) was used for the analysis, as described below.
Four analyses were performed: Classical logistic regression (GLM), standard GEE (GEE), Firth logistic regression (F-GLM), and Firth-GEE (F-GEE).
For F-GEE, the implementation of Mondol and Rahman (2019) in the extended version of Ogden (2022) in the R package `geefirthr` was used. The extension consisted of the ability to set the overdispersion parameter to 1 and to handle clusters of size 1. For F-GLM `logistf` (Heinze & Schemper, 2002) was used, and `geeM` (McDaniel et al., 2013) was used for GEE. In all analyses, the overdispersion parameter was fixed to 1 so as not to introduce additional sources for differences between the methods.
The coefficients and standard errors of the analyses of the 10 imputations per dataset were pooled according to the rules of Rubin (Rubin, 1987) presented earlier.
Furthermore, these four analyses were also applied to the incomplete datasets before values were deleted.
The four analyses were performed on 1000 simulated datasets. The four methods were compared in terms of MSE, bias and coverage of the 95% confidence interval (CI), and standard errors of the coefficients, with

$$Bias = E(\hat{\beta}) - \beta$$

$$MSE = E\left[(\hat{\beta} - \beta)^2\right] = Bias^2 + \widehat{var}(\hat{\beta})$$

$$\widehat{var}(\hat{\beta}) = \frac{\sum_{i=1}^{s} \left(\hat{\beta}_i - \bar{\hat{\beta}}\right)^2}{s-1}, \quad \text{with } i = 1 \text{ to } s \text{ simulations}$$

The analyses were repeated for datasets of different sizes (50, 100, 200, 500 children). Non-convergence rates for each type of analysis were documented.
To compare the results with and without MI, the methods of analysis were also applied to datasets with missing values. For datasets with missing values according to the MAR mechanism, methods that are full-likelihood are valid. For such cases, the term direct likelihood (DL) was coined in the literature (Molenberghs & Verbeke, 2005). (F-)GEE is not valid for missing values that are not MCAR, as explained earlier, the performance of which is investigated in this thesis for the purpose of comparison with MI-(F)-GEE. Thus, for datasets with missing values, DL-GLM, GEE, DL-F-GLM, and F-GEE are performed.

5. Results

This section reports the results of the analyses. The goal of this thesis was to investigate the performance of F-GEE after multiple imputation of values that were missing according to the MAR mechanism. For four sample sizes (50, 100, 200, and 500 children), the results of 1,000 converged analyses were collected and analyzed for scenarios with different dropout probabilities conditioned on previous respiratory infections. The percentage of analyses that did not converge were recorded. The evaluation is based on MSE, bias, coverage of 95% CI, and SE. Results are tabled separately for the four regression coefficients: β_0 = intercept, β_1 = gender, β_2 = age, and β_3 = xerophthalmia. The coefficient β_3 is of particular interest here because not only the response but also the predictor addresses a rarely occurring event. Because a large number of analyses were performed, not all tables and figures could be included in this section. In general, for F-GEE with incomplete datasets and MI-F-GEE after FLIC imputation, I have tabulated the results for all scenarios in this section. The results of all other analyses—DL-GLM, GEE, DL-F-GLM, MI-GLM, MI-F-GLM, MI-F-GEE (DA) —are tabulated in Appendices 1 to 20. For the graphical representations, I used a red tone for analyses with Firth penalty after MI and a blueish tone for the other analyses. Ideally, the scaling on the y-axis should always be kept constant; however, it was not possible everywhere, partly because the coefficients varied in magnitude and partly because the values became lower with increasing sample size. Since this thesis was primarily concerned with validity of parameter estimates and not with predictions, the values for the intercepts were not shown in the figures.

5.1. Mean squared error

Methods of analysis with Firth penalty after MI showed lower MSE values compared to analyses without Firth penalty or analyses of incomplete datasets. This can be seen from most of the lines in red being lower than other lines in Figure 1. Standard GEE showed infinitely high values in both cases, with incomplete and multiple imputed datasets; in those cases, values are not depicted in Figure 1. In logistic regression, with smaller datasets high but finite values were observed for incomplete datasets as well as after MI. Among the methods of analysis with a Firth penalty, F-GEE and DL-F-GLM had higher MSE values than other methods with Firth penalty. Among those, often MI-F-GEE predominantly achieved the lowest MSE values. Especially for the coefficient with the rare event, xerophthalmia, MI-F-GEE performed better after MI with the FLIC method than after MI with the DA method. This was mostly true when the proportion of dropouts was higher than 40% in children who had a previous infection. The MSE with DA imputation for the rare event was also larger compared to analyses with incomplete data in this case. MI-F-GLM performed well in terms of MSE; however, this method does not

take the intra-child correlation into account, which was simulated with 0.4. The problem with this method is highlighted later in the discussion of SE.

MSE became lower with increasing sample size, and so did the variance in MSE across the different methods of analysis. In summary, with respect to MSE there is evidence for MI-F-GEE showing a reliable performance, when using FLIC as a MI technique. In general, a low MSE indicates that estimates have little spread around the true value; thus, an estimator with small MSE is efficient. There is evidence for MI-F-GEE (FLIC) producing efficient estimators.

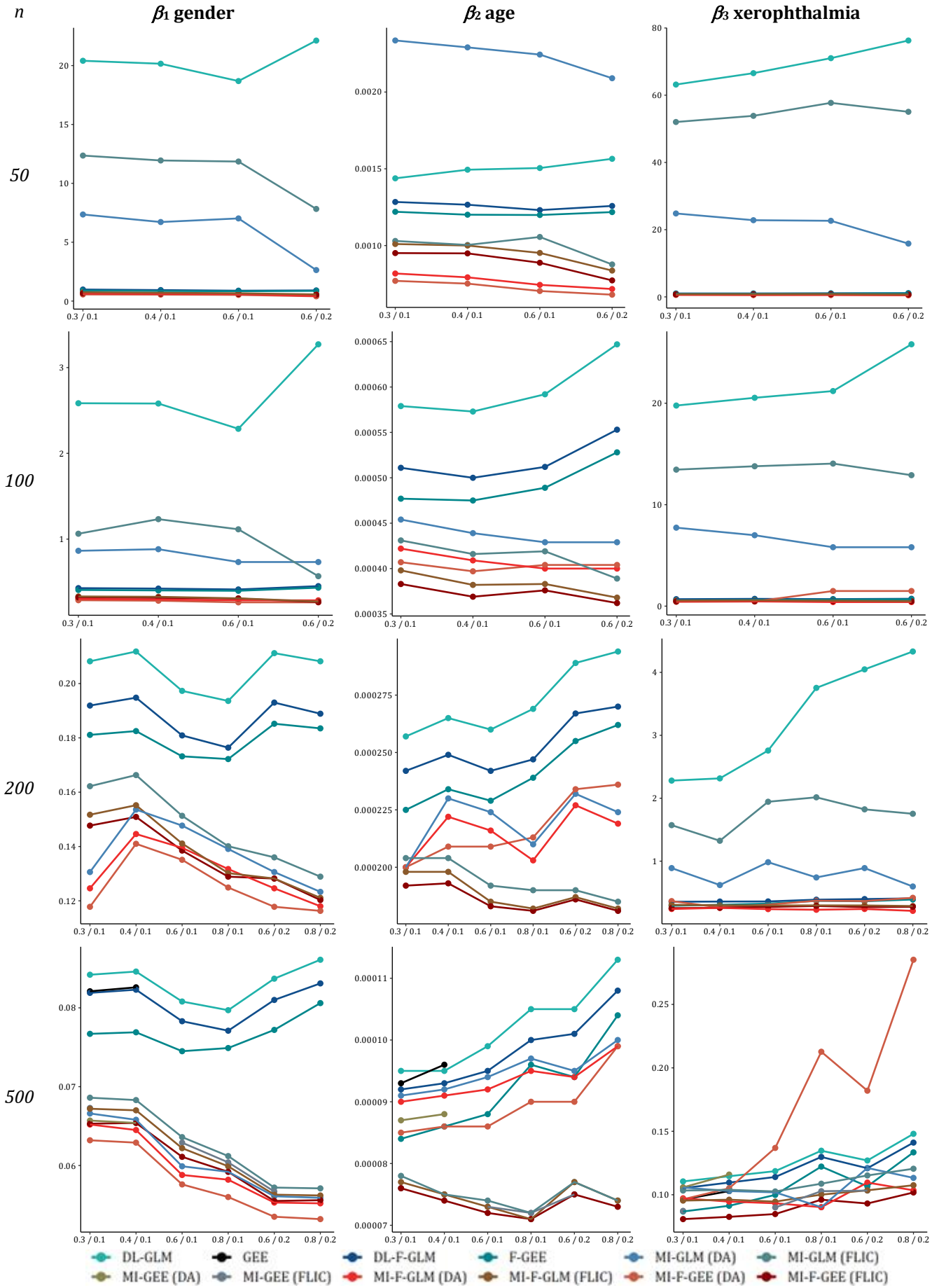
Table 6 compares the results of MI-F-GEE (FLIC) with the F-GEE method applied on incomplete datasets. As already mentioned, all other results can be found in the appendices (1 to 20). Further, Figure 1 presents MSE graphically for all analyses.

Table 6. MSE from Simulations for MI-F-GEE (FLIC) and F-GEE, 1000 Simulations.

	MI-F-GEE (FLIC)						F-GEE (incomplete dataset)					
p	**0.3/0.1**	**0.4/0.1**	**0.6/0.1**	**0.8/0.1**	**0.6/0.2**	**0.8/0.2**	**0.3/0.1**	**0.4/0.1**	**0.6/0.1**	**0.8/0.1**	**0.6/0.2**	**0.8/0.2**
n = 50												
β_0	0.3655	0.3680	0.3825		0.3551		0.4563	0.4571	0.4736		0.4698	
β_1	0.6579	0.6344	0.5978		0.5380		0.8952	0.8586	0.8370		0.8667	
β_2	0.00095	0.00095	0.00089		0.00077		0.00122	0.00120	0.00120		0.00122	
β_3	0.7277	0.7433	0.7833		0.7017		0.9133	0.9347	1.0012		1.0500	
n = 100												
β_0	0.1687	0.1801	0.1908		0.1784		0.2042	0.2198	0.2347		0.2520	
β_1	0.3204	0.3157	0.3066		0.2660		0.4074	0.4021	0.3976		0.4338	
β_2	0.00038	0.00037	0.00038		0.00036		0.00048	0.00048	0.00049		0.00053	
β_3	0.4667	0.4936	0.5036		0.4704		0.5727	0.6184	0.6456		0.6798	
n = 200												
β_0	0.0783	0.0850	0.1050	0.1368	0.0962	0.1229	0.0871	0.0928	0.1103	0.1455	0.1153	0.1480
β_1	0.1477	0.1509	0.1385	0.1289	0.1282	0.1203	0.1811	0.1825	0.1732	0.1722	0.1852	0.1835
β_2	0.00019	0.00019	0.00018	0.00018	0.00019	0.00018	0.00023	0.00023	0.00023	0.00024	0.00026	0.00026
β_3	0.2533	0.2590	0.2719	0.2879	0.2693	0.2751	0.2923	0.3063	0.3254	0.3694	0.3623	0.3909
n = 500												
β_0	0.0383	0.0462	0.0648	0.1081	0.0562	0.0973	0.0379	0.0427	0.0533	0.0937	0.0533	0.0909
β_1	0.0653	0.0654	0.0611	0.0592	0.0555	0.0556	0.0767	0.0769	0.0745	0.0749	0.0772	0.0806
β_2	0.00008	0.00007	0.00007	0.00007	0.00008	0.00007	0.00008	0.00009	0.00009	0.00010	0.00009	0.00010
β_3	0.0808	0.0826	0.0848	0.0961	0.0931	0.1018	0.0868	0.0913	0.0999	0.1222	0.1066	0.1334

Note: Results for other analyses are tabled in the attachment; β_0 = intercept, β_1 = gender, β_2 = age, β_3 = xerophthalmia; response variable = respiratory infection; *p* = probability of dropout conditional on previous infection (yes / no); *FLIC* = imputation with Firth logistic regression with adjusted intercept; *MI* = multiple imputation; *F* = Firth penalty; *MSE* = mean squared error, *n* = number of children.

Figure 1. MSE Values for Coefficients β_1, β_2, β_3 of Simulated Scenarios (1,000 Simulations each)



x-axis = probability of dropout conditioned on previous response (yes / no), *n* = number of children

5.2. Bias

The interpretation of the results of the bias of the coefficient estimates was less straightforward than the interpretation of the results of the MSE. Figure 2 plots the bias values of all analyses, with the dashed line indicating a 0 bias. Standard GEE showed an infinitely high bias in most scenarios, both for incomplete datasets and after MI. In these cases, the values are not plotted in Figure 2. Logistic regressions achieved higher bias values for the coefficients β_1 and β_3 in small samples compared to the other methods. The bias here was up to four times larger than the values of the other methods. In spite of these issues, the bias was comparable between Firth-penalized and non-penalized methods of analysis and also between analyses with incomplete datasets and those after MI. It was noticeable that, with respect to β_3 xerophthalmia, the bias in MI-F-GEE after FLIC imputation was smaller than after DA imputation at higher dropout probabilities. This corresponds to the results reported with respect to MSE. Table 7 shows the values for MI-F-GEE (FLIC) and F-GEE. In Figure 2, bias values of all analyses are shown.

Table 7. Bias from Simulations for MI-F-GEE (FLIC) and F-GEE, 1000 Simulations.

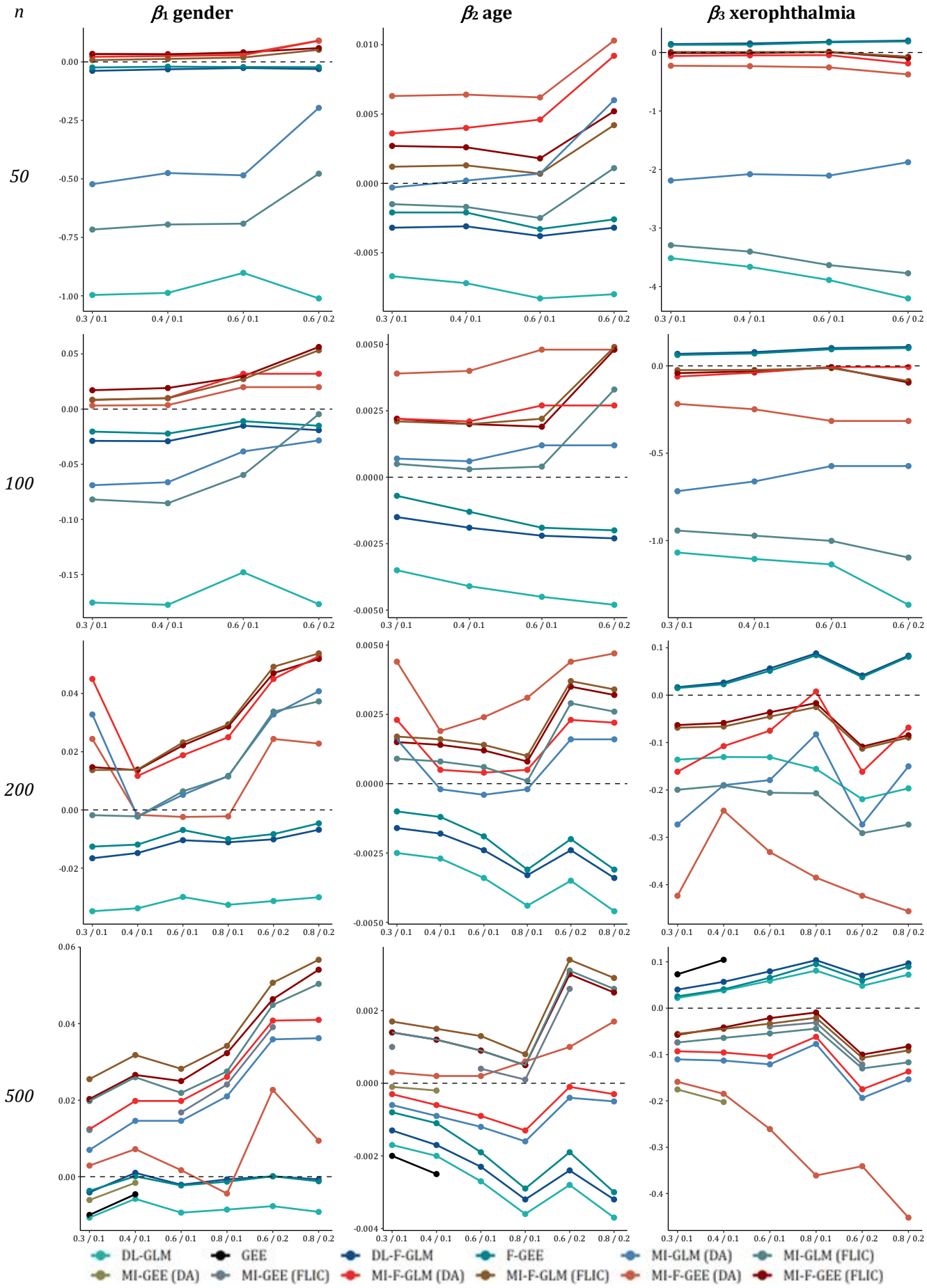
	MI-F-GEE (FLIC)						F-GEE (incomplete dataset)					
	0.3/0.1	0.4/0.1	0.6/0.1	0.8/0.1	0.6/0.2	0.8/0.2	0.3/0.1	0.4/0.1	0.6/0.1	0.8/0.1	0.6/0.2	0.8/0.2
n = 50												
β_0	−0.0906	−0.1182	−0.1845		−0.1182		−0.1414	−0.1604	−0.2271		−0.2030	
β_1	0.0331	0.0321	0.0410		0.0590		−0.0240	−0.0206	−0.0221		−0.0228	
β_2	0.0027	0.0026	0.0018		0.0052		−0.0021	−0.0021	−0.0033		−0.0026	
β_3	−0.0082	−0.0087	0.0075		−0.0932		0.1324	0.1358	0.1745		0.1906	
n = 100												
β_0	−0.0751	−0.1107	−0.1873		−0.1473		−0.0812	−0.1107	−0.1824		−0.1824	
β_1	0.0171	0.0191	0.0298		0.0563		−0.0204	−0.0222	−0.0110		−0.0151	
β_2	0.0022	0.0020	0.0019		0.0048		−0.0007	−0.0013	−0.0019		−0.0020	
β_3	−0.0415	−0.0308	−0.0084		−0.0956		0.0623	0.0710	0.0950		0.1020	
n = 200												
β_0	−0.0578	−0.0935	−0.1746	−0.2564	−0.1431	−0.2277	−0.0378	−0.0658	−0.1406	−0.2340	−0.1288	−0.2252
β_1	0.0147	0.0137	0.0222	0.0287	0.0470	0.0519	−0.0126	−0.0119	−0.0069	−0.0100	−0.0083	−0.0046
β_2	0.0015	0.0014	0.0012	0.0008	0.0035	0.0032	−0.0010	−0.0012	−0.0019	−0.0031	−0.0020	−0.0031
β_3	−0.0632	−0.0586	−0.0361	−0.0167	−0.1086	−0.0846	0.0150	0.0231	0.0514	0.0843	0.0381	0.0810
n = 500												
β_0	−0.0667	−0.1046	−0.1821	−0.2774	−0.1572	−0.2555	−0.0360	−0.0622	−0.1295	−0.2369	−0.1220	−0.2248
β_1	0.0203	0.0266	0.0250	0.0323	0.0464	0.0541	−0.0037	0.0001	−0.0023	−0.0013	0.0001	−0.0012
β_2	0.0014	0.0012	0.0009	0.0005	0.0030	0.0025	−0.0008	−0.0011	−0.0019	−0.0029	−0.0019	−0.0030
β_3	−0.0576	−0.0416	−0.0214	−0.0093	−0.1002	−0.0826	0.0256	0.0408	0.0657	0.0952	0.0595	0.0896

Note: Results for other analyses are tabled in the attachment; β_0 = intercept, β_1 = gender, β_2 = age, β_3 = xerophthalmia; response variable = respiratory infection; *p* = probability of dropout conditional on previous infection (yes / no); *FLIC* = imputation with Firth logistic regression with adjusted intercept; *MI* = multiple imputation; *F* = Firth penalty, *n* = number of children.

5.3. Coverage for the 95% confidence interval

Coverage of the 95% CI of the coefficient estimates was not calculated for standard GEE with incomplete datasets and after MI because of infinitely large SE; the results would have been meaningless. Despite that, analyses after MI obtained higher coverage of the

Figure 2. Bias for Coefficients β_1, β_2, β_3 of Simulated Scenarios (1,000 Simulations each)



x-axis = probability of dropout conditioned on previous response (yes / no), *n* = number of children
dashed line represent 0 bias

95% CI for the three coefficients β_1, β_2 and β_3 compared to analyses with incomplete data. However, it should be noted that even though the coverage values for incomplete datasets were lower, all values were around 0.90 or higher and thus satisfactory. The range of coverage values of MI-F-GEE after FLIC, for example, was 0.926 to 0.988, and for F-GEE with incomplete data it was 0.887 to 0.962. The higher coverage after FLIC imputation, however, cannot be attributed to higher SE values since these were comparably large with those in incomplete datasets. The coverage of MI-F-GEE (DA) with the β_3 coefficient dropped with increasing dropout probabilities with 500 children. However, it should also be noted here that this may not be overestimated, as coverage was still 93% after having dropped.

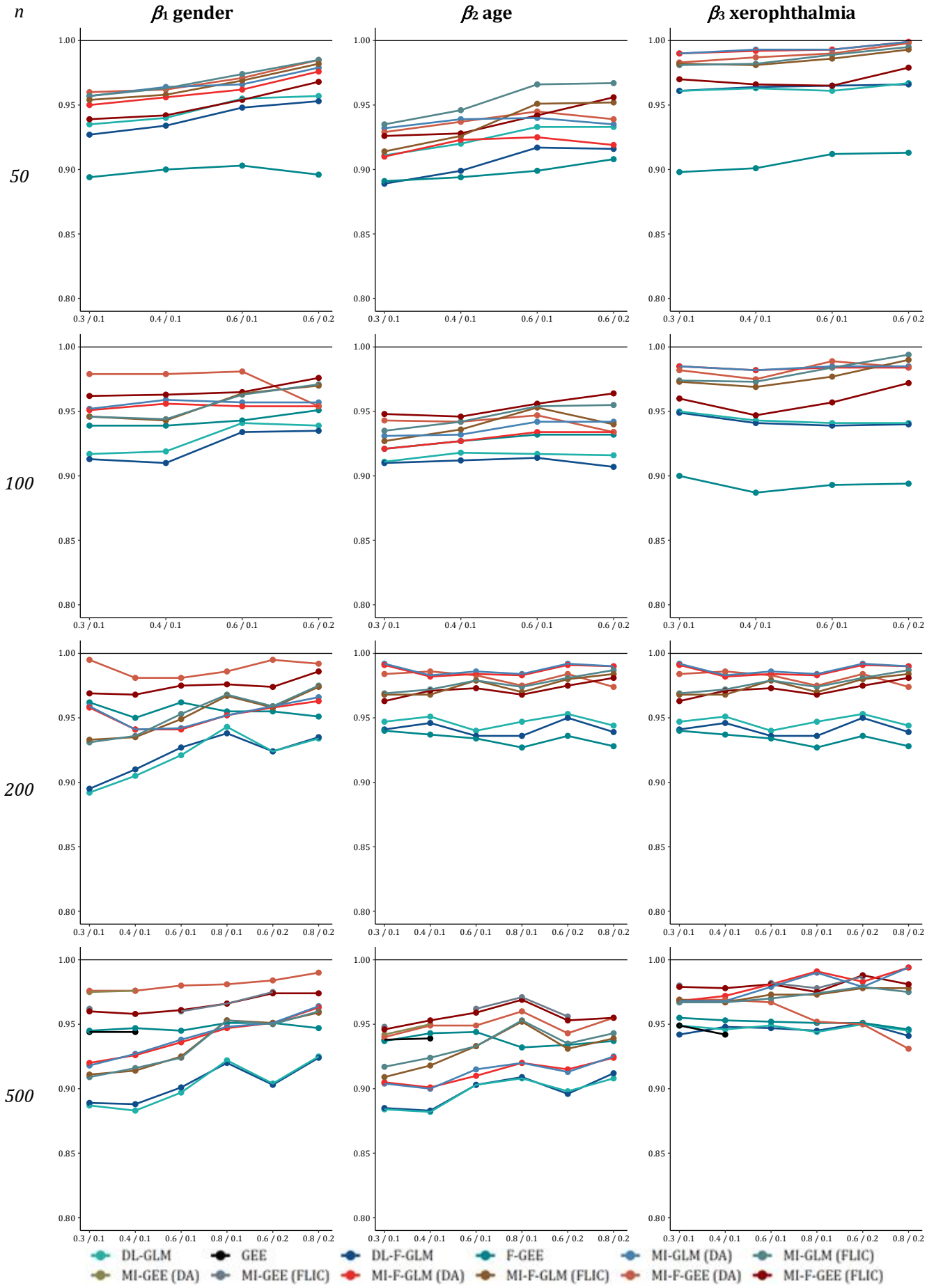
Coverage values for intercepts, on the other hand, were lower for 500 children after FLIC imputation and for incomplete data than after DA imputation. For example, with 500 children, the value for MI-F-GEE (FLIC) intercept was 0.677, for F-GEE intercept 0.745, and for MI-F-GEE (DA) intercept 0.936. However, as the focus of this work was on estimation of coefficients and not on prediction of probabilities, no particular importance was given to the intercept at this stage. If the focus had been on predictions, then it might be that the FLIC method should have been used not only for MI of missing values, but also for analyses of the imputed data itself. Here, research could be done on applying the FLIC method developed for logistic regression to (F-)GEE. The coverage values of F-GEE and MI-F-GEE after FLIC imputation are presented in Table 8; the values of all analyses for coefficients β_1, β_2 and β_3 are depicted in Figure 3.

Table 8. Coverage from Simulations for MI-F-GEE (FLIC) and F-GEE, 1000 Simulations.

	MI-F-GEE (FLIC)						F-GEE (incomplete dataset)					
	0.3/0.1	0.4/0.1	0.6/0.1	0.8/0.1	0.6/0.2	0.8/0.2	0.3/0.1	0.4/0.1	0.6/0.1	0.8/0.1	0.6/0.2	0.8/0.2
n = 50												
β_0	0.955	0.962	0.965		0.967		0.920	0.925	0.933		0.924	
β_1	0.939	0.942	0.954		0.968		0.894	0.900	0.903		0.896	
β_2	0.926	0.928	0.942		0.956		0.891	0.894	0.899		0.908	
β_3	0.970	0.966	0.965		0.979		0.898	0.901	0.912		0.913	
n = 100												
β_0	0.954	0.948	0.952		0.968		0.933	0.920	0.918		0.926	
β_1	0.962	0.963	0.965		0.976		0.939	0.939	0.943		0.951	
β_2	0.948	0.946	0.956		0.964		0.921	0.927	0.932		0.932	
β_3	0.960	0.947	0.957		0.972		0.900	0.887	0.893		0.894	
n = 200												
β_0	0.956	0.945	0.933	0.902	0.941	0.919	0.942	0.933	0.914	0.880	0.920	0.896
β_1	0.969	0.968	0.975	0.976	0.974	0.986	0.962	0.950	0.962	0.955	0.955	0.951
β_2	0.941	0.943	0.948	0.960	0.938	0.945	0.930	0.926	0.935	0.935	0.910	0.917
β_3	0.963	0.971	0.973	0.968	0.975	0.981	0.940	0.937	0.934	0.927	0.936	0.928
n = 500												
β_0	0.935	0.915	0.854	0.677	0.882	0.732	0.935	0.926	0.896	0.745	0.909	0.778
β_1	0.960	0.958	0.961	0.966	0.974	0.974	0.945	0.947	0.945	0.951	0.951	0.947
β_2	0.946	0.953	0.959	0.969	0.953	0.955	0.937	0.943	0.944	0.932	0.934	0.937
β_3	0.979	0.978	0.981	0.975	0.988	0.981	0.955	0.953	0.952	0.951	0.951	0.946

Note: Coverage of 95% CI. Results for other analyses tabled in attachment; β_0 = intercept, β_1 = gender, β_2 = age, β_3 = xerophthalmia; response variable = respiratory infection; p = probability of dropout conditional on previous infection (yes / no); *FLIC* = imputation with Firth logistic regression with adjusted intercept; *MI* = multiple imputation; *F* = Firth penalty, n = number of children.

Figure 3. Coverage for Coefficients β_1, β_2, β_3 of Simulated Scenarios (1,000 Simulations each)



x-axis = probability of dropout conditioned on previous response (yes / no), *n* = number of children
solid line represents 100% coverage

5.4. Standard Error

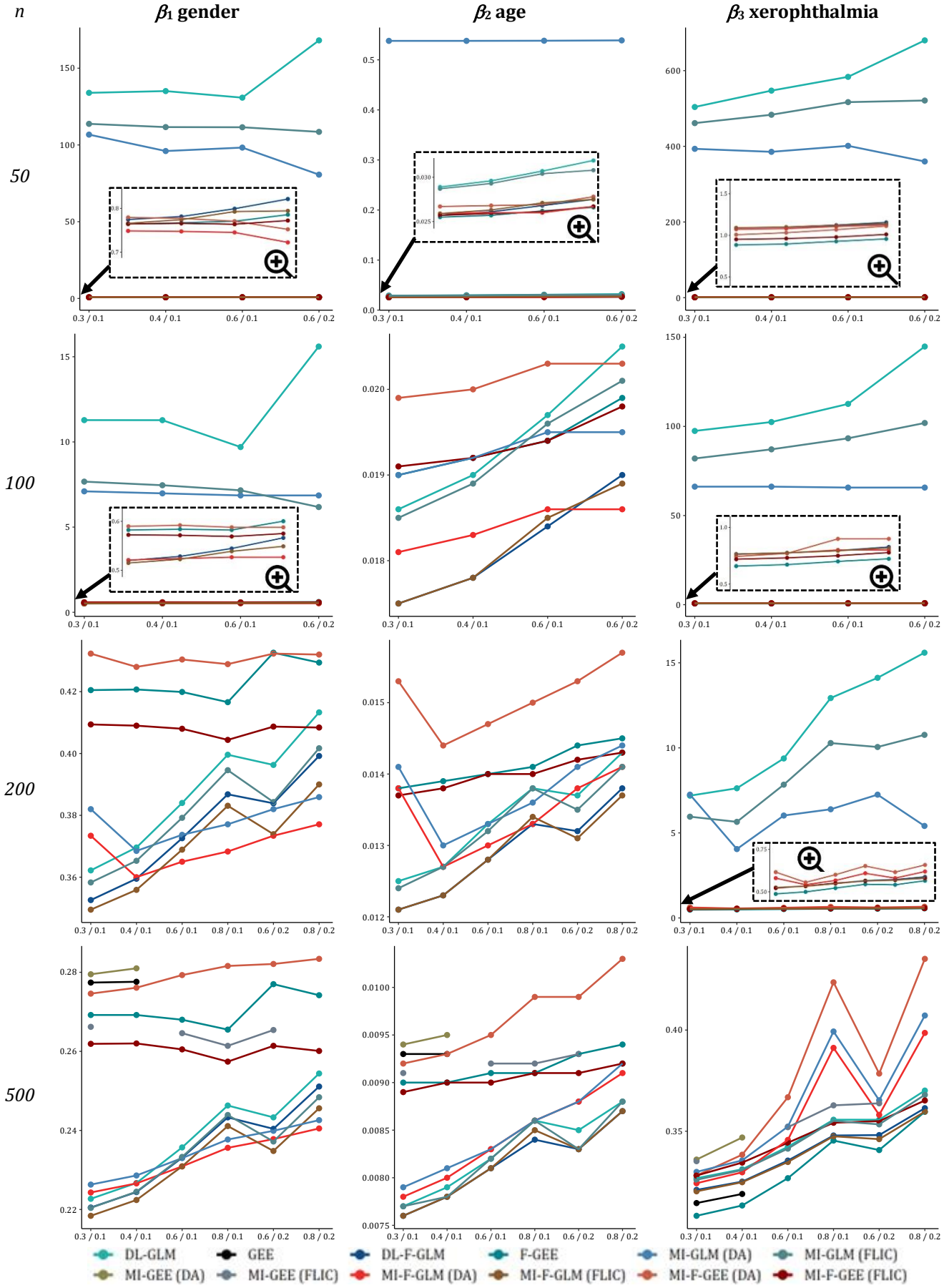
In terms of SE, severe issues were observed with logistic regression for incomplete data as well as after MI when no Firth penalty was applied. The values were not infinitely high, but they were larger than the values of the other analyses by a factor of 10 to 100. These values are thus uninterpretable. This problem occurred mainly with smaller samples, with β_3 even up to 200 children. In most analyses, SE of standard GEE and MI-GEE analyses were infinitely high; those values are thus not depicted in Figure 4. Apart from these problems, SE were smaller when intra-child correlations—simulated with 0.4—were not considered. This was the case in variants of logistic regressions (e.g., DL-GLM, DL-F-GLM, MI-GLM). However, with SE it is not like with bias or MSE that low values are better. It may also be the other way around, that low SE underestimate the true variance and thus lead to inflation of the type I error rate when testing null hypotheses. For this reason, SE achieved with methods that consider the correlated nature of the data are more trustworthy. Generally, SE of MI-F-GEE after FLIC were lower than MI-F-GEE after DA. If the analyses with complete data are used as a reference (Appendix 21)—that is, those generated complete datasets before having deleted values according to MAR mechanism—then one sees that the SE obtained with F-GEE at coefficients β_0, β_1 and β_2 correspond to those of MI-F-GEE after FLIC imputation and not to those after DA. For β_3, the SE were slightly higher in all cases compared to the analyses of complete data. It should be recalled here that β_3 describes the association between the predictor with rarely occurring events and the response with rarely occurring events. SE became smaller

Table 9. SE from Simulations for MI-F-GEE (FLIC) and F-GEE, 1000 Simulations.

	MI-F-GEE (FLIC)						F-GEE (incomplete dataset)					
	0.3/0.1	0.4/0.1	0.6/0.1	0.8/0.1	0.6/0.2	0.8/0.2	0.3/0.1	0.4/0.1	0.6/0.1	0.8/0.1	0.6/0.2	0.8/0.2
n = 50												
β_0	0.5331	0.5382	0.5407		0.5512		0.5314	0.5355	0.5416		0.5518	
β_1	0.7646	0.7659	0.7639		0.7723		0.7644	0.7671	0.7706		0.7857	
β_2	0.0257	0.0259	0.0261		0.0267		0.0255	0.0257	0.0262		0.0266	
β_3	0.9509	0.9606	0.9788		1.0109		0.8837	0.8955	0.9274		0.9563	
n = 100												
β_0	0.3862	0.3887	0.3902		0.3991		0.3906	0.3925	0.3932		0.4070	
β_1	0.5723	0.5715	0.5690		0.5749		0.5822	0.5838	0.5822		0.6001	
β_2	0.0191	0.0192	0.0194		0.0198		0.0190	0.0192	0.0194		0.0199	
β_3	0.7189	0.7312	0.7500		0.7776		0.6582	0.6716	0.6999		0.7223	
n = 200												
β_0	0.2735	0.2743	0.2761	0.2761	0.2797	0.2818	0.2788	0.2789	0.2801	0.2803	0.2896	0.2898
β_1	0.4094	0.4090	0.4080	0.4044	0.4087	0.4084	0.4205	0.4207	0.4199	0.4166	0.4326	0.4294
β_2	0.0137	0.0138	0.0140	0.0140	0.0142	0.0143	0.0138	0.0139	0.0140	0.0141	0.0144	0.0145
β_3	0.5218	0.5345	0.5509	0.5633	0.5701	0.5801	0.4877	0.4998	0.5220	0.5446	0.5414	0.5652
n = 500												
β_0	0.1737	0.1742	0.1746	0.1748	0.1768	0.1780	0.1773	0.1773	0.1780	0.1782	0.1844	0.1844
β_1	0.2619	0.2620	0.2605	0.2574	0.2614	0.2601	0.2692	0.2692	0.2680	0.2655	0.2770	0.2742
β_2	0.0089	0.0090	0.0090	0.0091	0.0091	0.0092	0.0090	0.0090	0.0091	0.0091	0.0093	0.0094
β_3	0.3281	0.3345	0.3443	0.3542	0.3550	0.3652	0.3082	0.3133	0.3268	0.3454	0.3407	0.3596

Note: Results for other analyses in the attachment; β_0 = intercept, β_1 = gender, β_2 = age, β_3 = xerophthalmia; response variable = respiratory infection; p = probability of dropout conditional on previous infection (yes / no); *FLIC* = imputation with Firth logistic regression with adjusted intercept; *MI* = multiple imputation; *F* = Firth penalty, *n* = number of children.

Figure 4. SE for Coefficients β_1, β_2, β_3 of Simulated Scenarios (1,000 Simulations each)



x-axis = probability of dropout conditioned on previous response (yes / no), *n* = number of children

with increasing sample size; this is simply because SE is a function of *n*. On the other hand, SE tended to become larger with increasing dropout probability conditioned on previous respiratory infections. This was less clearly the case for MI-F-GEE (FLIC), where SE were more constant across scenarios (Table 9, Figure 4).

5.5. Failure of convergence

Standard GEE with incomplete data and after MI had not only infinite coefficients as well as SE in converged analyses, but also a very high proportion of analyses, which did not converge. This problem mainly affected samples with 50, 100, or 200 children. For 50 children, for example, up to 76% of the analyses did not converge. Across all methods of analysis, the rates of not converged analyses were lowest when analyzing incomplete data, followed by analyses after FLIC imputation. The percentage of not converged analyses after DA imputation was almost always higher than FLIC. Analyses with F-GLM, either with incomplete or after MI, reached the lowest non-convergence rates. Even with 50 children only, there were 0% or, in the worst cases, 0.2% non-convergence rates. GLM was slightly above this with values between 0.5% and 6.1%, followed by F-GEE (with and without MI). With MI-F-GEE, a distinction should be made between MI-F-GEE after FLIC imputation (range: 2.3%–3.7%) and after DA imputation (range: 6.2%–14.4%). In terms of coverage, FLIC imputation seems to work more reliably than DA imputation for rarely occurring events. In large samples of 500 children, there were almost no problems with convergence for all methods of analysis. Even when including GEE, more than 95% of the analyses converged. It was also noticeable that the proportion of non-converged analyses increased with the increasing dropout probability of children. The percentages of analyses that did not converge are tabled in Appendices 1 to 20 in the last line of each table.

6. Discussion

In this thesis, a longitudinal study with four measurement time points was simulated with a dichotomous response variable. The basis for the simulation was the study by Sommer et al. (1983) on the influence of vitamin A deficiency on the occurrence of respiratory infections in children. For the analysis of these data, GEE would be a suitable marginal model in the case of absence of missingness. The aim of this work was to investigate the performance of GEE in the context of rarely occurring events in the response variable and in one of the predictors. In the theoretical section, it was discussed and hypothesized that GEE is not suitable because rarely occurring events can generate separation and result in infinite coefficients and SE. As a solution to this problem, the Firth penalty was considered. The behavior of GEE and the application of the Firth penalty was also investigated under the condition of missing data. It was simulated that the data were missing after the MAR mechanism. It was hypothesized that F-GEE after multiple imputation would be a valid method for analyzing the data. Due to the fact that MI of missing values is a problem of rarely occurring events itself, the Firth penalty was also investigated in the context of imputations. It had to be considered that models with Firth penalty produce valid estimates, but not valid predictions. Since valid predictions are necessary for imputations, it was hypothesized that the application of the FLIC method, which recalibrates the intercept after applying the Firth penalty, would solve the problem. After presenting the results in the previous chapter, this section will discuss these results and put them in perspective with the theory.

The results showed that standard GEE was not suitable for analyzing the simulated data with rarely occurring events. First, the proportion of analyses that did not converge was high; and second, for those analyses that did converge, there were infinite estimates for coefficients and for SE. Standard GEE was unsuitable both for analyzing datasets with missing values and for datasets with imputed values. The problems with this method of analysis were more prevalent with small samples, but also with large samples when the proportion of missing children increased. These results were expected based on theoretical assumptions about separation for rarely occurring events (Heinze & Schemper, 2002) and are consistent with analyses by Mondol and Rahman (2019) and Geroldinger et al. (2022). The solution proposed in the theory—namely the application of a Firth penalty to GEE, F-GEE—resulted in estimates for coefficients and SE that were neither infinite, nor finite large and not interpretable. This replicated the findings of Mondol and Rahman (2019). The proportion of analyses that did not converge was also much lower for F-GEE compared to GEE. However, the convergence rate was not at 100%, which again is consistent with the finding of Geroldinger et al. (2022) that, contrary to Mondol and Rahman (2019), some F-GEE analyses did not converge.

Classical logistic regression analyses with and without Firth-penalty achieved the highest convergence rates. Without Firth penalty, there were finite but large and uninterpretable estimates for coefficients and SE in small samples. With Firth penalty, this was not the case. However, despite problems with interpretability of SE, variants of logistic regression analyses (DL-GLM, DL-F-GLM, MI-GLM, MI-F-GLM) showed systematically lower SE compared to the associated variants of GEE (GEE, F-GEE, MI-MI-GEE, F-GEE).

This could be because the logistic regressions do not account for the correlated nature of the data. It is well known from the literature that SE can be underestimated if the dependence structure of the data is ignored (Molenberghs & Verbeke, 2005). In the present simulations, an intra-child correlation of 0.4 was used. Thus, the results are consistent with theoretical considerations, and logistic regressions were not found to be suitable for the analysis of longitudinal data, even when bias and MSE values for F-GLM were on the lower side.

The question then arose whether F-GEE yields better results after MI than when analyzing incomplete datasets. Here, however, the analysis of the imputation technique itself must be discussed beforehand. Both imputation methods, FLIC and DA, produced higher coverage of the 95% CI of the coefficient estimates compared to coverage rates with incomplete data. For FLIC, this was not due to higher SE, because its SE were comparable in size with those for incomplete analyses. DA analyses had slightly larger SE than FLIC and it was noticeable that the bias of the coefficient estimates for β_3, xerophthalmia, was much larger for DA than FLIC. For larger samples, the MSE became larger for DA than for FLIC as dropout rates increased. Overall, the proportion of analyses that converged were higher for FLIC than for DA. For the coefficient estimates of the other predictors, bias and MSE of DA were partially lower than for FLIC and, partially comparable in magnitude. In summary, the FLIC method performed more reliably with respect to the predictor with rarely occurring events and performed comparably with respect to the other parameters. These results confirm the research of Puhr et al (2017) that recalibration of the intercept leads to valid predictions, which were imputations in this case.

MI-F-GEE had lower MSE values compared to F-GEE. These results suggest that MI-F-GEE produces the most efficient estimators. MI-F-GEE after FLIC imputation also had mostly lower bias than F-GEE for small samples in coefficients β_2 and β_3. For 200 and 500 children, with DA, the bias was partially smaller and partially the values of both methods of analysis were comparable.

The counterargument of White et al. (2010) that multiple imputation with a Firth penalty is more time-consuming and thus computationally unrealistic can be questioned here. For 500 children, imputing with FLIC and running all analyses on 1,000 simulated datasets on a regular Windows 10 machine took about 48 hours. When imputation was done using the algorithm of White et al. (2010), DA, the time was reduced by three hours. However, in practice, 1,000 imputations are not performed and then analyzed using a wide variety of techniques. Therefore, it would be an additional effort of a matter of seconds, which probably no one would hesitate to do in order to obtain more valid results. The present analyses have shown that a Firth penalty is indeed feasible in the context of the MICE package. In summary, MI-F-GEE performed well in both smaller and larger samples and also in both lower and higher dropout rates.

There are also some limitations that need to be addressed. Due to the rarely occurring event in the response variable, it was a challenge to generate missing values according to the MAR mechanism. Even if 80% of children with a previous infection drop out, only 4% of children are missing when the incidence is 5%. Conversely, with 80% dropout rate in uninfected children, almost all children would be dropped out. The goal of this thesis was to explicitly study the behavior of F-GEE under MAR. Based on the information in the

original study, it was assumed that children with a previous infection were more likely to be absent. A higher dropout probability among non-infected children was not simulated. The dropout mechanism was also not made dependent on covariates because otherwise the line between MAR and MCAR would not have been sharp due to the rarely occurring event. Because of limited resources of time, the intra-child correlation was kept constant at 0.4; the effect of different intra-child correlations was not investigated. Simulations with fewer than 50 children were not performed because the proportion of analyses that did not converge increased markedly.
The generated datasets represented near-to-quasi-complete separation scenarios. However, due to the Monte Carlo process, datasets with quasi-complete separation were also included. No distinction was made between the two scenarios in the analyses, which was analogous to the study by Mondol and Rahman (2019). One suggestion for future studies would be to investigate this further.

7. Ethical considerations

At this point I would like to mention a few ethical considerations concerning this topic. My master's thesis was dedicated to statistical methods for the analysis of rarely occurring events. An event that occurs rarely can have severe consequences for the affected person and for society at large. For example, the event could concern a serious illness or the uncovering of sexual abuse, to name just two examples. Scientific research into these cases could help to reduce, end or prevent the suffering of those affected and the suffering within a society. A severe illness may also cause individual visits to be missed by the patients. The present thesis has investigated solutions as to how such topics can be researched even under these more difficult circumstances.
The computational problems that occur in the context of separation have in the past often been inadequately addressed. For example, the predictor that caused separation was simply omitted in the absence of a solution, or the rare event was not evaluated with inferential statistics (Heinze & Schemper, 2002; Puhr et al., 2017). This is a wood-chipper approach that leaves available information unused and potentially undercuts new insights. The Firth penalty presented in this paper allows inferential statistical analysis of rare events. For a method to be used in medical research, for example, the implementation must be low threshold. The best technique is of little use if the application is so complex that no one knows how to use it. Firth-penalty is a technique that is easy to apply. In the most common statistical programs—such as SAS, R, Stata and SPSS—there are options for using Firth logistic regression, either directly or as a macro. In SAS, for example, the option "FIRTH" can be used in PROC LOGISTIC, in R the same syntax can be used for logistf() as for the familiar lm() function. For F-GEE there is a package in R, in the other programs it requires still relevant software developments. The Firth penalty showed promising results in past studies, and this was also the case in this thesis. Accordingly, researchers should continue to develop implementations of the Firth penalty in statistical software, especially for F-GEE.

References

Albert, A., & Anderson, J. A. (1984). On the Existence of Maximum Likelihood Estimates in Logistic Regression Models. *Biometrika*, *71*. https://doi.org/10.1093/biomet/71.1.1

Bartlett, J. W., Seaman, S. R., White, I. R., & Carpenter, J. R. (2015). Multiple imputation of covariates by fully conditional specification: Accommodating the substantive model. *Statistical Methods in Medical Research*, *24*(4), 462–487. https://doi.org/10.1177/0962280214521348

Beunckens, C., Sotto, C., & Molenberghs, G. (2008). A simulation study comparing weighted estimating equations with multiple imputation based estimating equations for longitudinal binary data. *Comput. Stat. Data Anal.* https://doi.org/10.1016/j.csda.2007.04.020

Diggle, P., Heagerty, P., Liang, K.-Y., & Zeger, S. (2002). The Analysis of Longitudinal Data. In *Journal of the American Statistical Association* (Bd. 90). https://doi.org/10.2307/2291352

Firth, D. (1993). Bias Reduction of Maximum Likelihood Estimates. *Biometrika*, *80*(1), 27–38. https://doi.org/10.2307/2336755

Firth, D. (2015). Recent Developments in Quasi-Likelihood Methods. *Bulletin of the International Statistical Institute*, *55*, 341–358.

Geroldinger, A., Blagus, R., Ogden, H., & Heinze, G. (2022). *An Investigation of Penalization and Data Augmentation to Improve Convergence of Generalized Estimating Equations for Clustered Binary Outcomes* [Preprint]. In Review. https://doi.org/10.21203/rs.3.rs-1369776/v1

Heinze, G., & Schemper, M. (2002). A solution to the problem of separation in logistic regression. *Statistics in Medicine*, *21*(16), 2409–2419. https://doi.org/10.1002/sim.1047

Johns Hopkins Bloomberg School of Public Health. (2002). *The Story of Dean Sommer and Vitamin A*. https://publichealth.jhu.edu/2003/sommer-vita

Lesaffre, E., & Lawson, A. B. (2012). *Bayesian Biostatistics*. John Wiley & Sons.

Liang, K.-Y., & Zeger, S. L. (1986). Longitudinal data analysis using generalized linear models. *Biometrika*, *73*(1), 13–22. https://doi.org/10.1093/biomet/73.1.13

Lipsitz, S., Fitzmaurice, G., & Weiss, R. (2020). Using Multiple Imputation with GEE with Non-monotone Missing Longitudinal Binary Outcomes. *Psychometrika*, *85*. https://doi.org/10.1007/s11336-020-09729-y

Little, R. J. A., & Rubin, D. B. (1987). *Statistical Analysis with Missing Data*. Wiley.

Mancl, L. A., & DeRouen, T. A. (2001). A covariance estimator for GEE with improved small-sample properties. *Biometrics*, *57*(1), 126–134. https://doi.org/10.1111/j.0006-341x.2001.00126.x

McCullagh, P., & Nelder, J. A. (1989). *Generalized Linear Models*. Springer US. https://doi.org/10.1007/978-1-4899-3242-6

McDaniel, L., S., Henderson, N., C., & Rathouz, P., J. (2013). Fast Pure R Implementation of GEE: Application of the Matrix Package. *The R Journal*, *5*(1), 181. https://doi.org/10.32614/RJ-2013-017

Molenberghs, G., & Verbeke, G. (2005). *Models for Discrete Longitudinal Data*. https://doi.org/10.1007/0-387-28980-1

Mondol, M. H., & Rahman, M. S. (2019). Bias-reduced and separation-proof GEE with small or sparse longitudinal binary data. *Statistics in Medicine*, *38*(14), 2544–2560. https://doi.org/10.1002/sim.8126

Morel, J. G., Bokossa, M., & Neerchal, N. (2003). *Small Sample Correction for the Variance of GEE Estimators*. https://doi.org/10.1002/BIMJ.200390021

Ogden, H. (2022). *A modified version of the package at https://github.com/mhmondol/geefirthr for fitting Firth-type GEE for correlated binary data with separation or near-to-separation.* https://github.com/heogden/geefirthr/

Puhr, R., Heinze, G., Nold, M., Lusa, L., & Geroldinger, A. (2017). Firth's logistic regression with rare events: Accurate effect estimates and predictions? *Statistics in Medicine*, *36*(14), 2302–2317. https://doi.org/10.1002/sim.7273

Qaqish, B. (2003). A family of multivariate binary distributions for simulating correlated binary variables. *Biometrika*, *90*, 455–463. https://doi.org/10.1093/biomet/90.2.455

Rogers, P., & Stoner, J. (2021). Assessment of a Modified Sandwich Estimator for Generalized Estimating Equations with Application to Opioid Poisoning in MIMIC-IV ICU Patients. *Stats*, *4*(3), 650–664. https://doi.org/10.3390/stats4030039

Rubin, D. B. (1976). Inference and Missing Data. *Biometrika*, *63*(3), 581–592. https://doi.org/10.2307/2335739

Rubin, D. B. (1987). *Multiple Imputation for Nonresponse in Surveys*. John Wiley & Sons Inc.

Rubin, D. B., & Schenker, N. (1986). *Multiple Imputation for Interval Estimation from Simple Random Samples with Ignorable Nonresponse*. https://doi.org/10.1080/01621459.1986.10478280

Sahile, Z., Yilma, D., Tezera, R., Bezu, T., Haileselassie, W., Seifu, B., & Haidar, J. (2020). Prevalence of Vitamin A Deficiency among Preschool Children in Ethiopia: A Systematic Review and Meta-Analysis. *BioMed Research International*, *2020*, 1–12. https://doi.org/10.1155/2020/8032894

Schöning, N. (2013). *Modellbasierte Imputationsverfahren im Vergleich*. https://www.gsi.uni-muenchen.de/lehreinheiten/le_policy_analysis/dokumente/impu1.pdf

Sommer, A., Tarwotjo, I., Hussaini, G., & Susanto, D. (1983). Increased mortality in children with mild vitamin A deficiency. *Lancet (London, England)*, *2*(8350), 585–588. https://doi.org/10.1016/s0140-6736(83)90677-3

van Buuren, S. (2007). Multiple imputation of discrete and continuous data by fully conditional specification. *Statistical Methods in Medical Research*, *16*(3), 219–242. https://doi.org/10.1177/0962280206074463

van Buuren, S. (2018). *Flexible Imputation of Missing Data* (2. Aufl.). Chapman and Hall/CRC. https://doi.org/10.1201/9780429492259

van Buuren, S., Brand, J., Groothuis-Oudshoorn, C., & Rubin, D. B. (2006). Fully Conditional Specification in Multivariate Imputation. *Journal of Statistical Computation and Simulation*, *76*. https://doi.org/10.1080/10629360600810434

van Buuren, S., & Groothuis-Oudshoorn, K. (2011). mice: Multivariate Imputation by Chained Equations in R. *Journal of Statistical Software*, *45*, 1–67. https://doi.org/10.18637/jss.v045.i03

Wang, H., Zheng, Y., Deng, J., Wang, W., Liu, P., Yang, F., & Jiang, H. (2016). Prevalence of respiratory viruses among children hospitalized from respiratory infections in Shenzhen, China. *Virology Journal*, *13*(1), 39. https://doi.org/10.1186/s12985-016-0493-7

White, I. R., Daniel, R., & Royston, P. (2010). Avoiding bias due to perfect prediction in multiple imputation of incomplete categorical variables. *Computational Statistics & Data Analysis*, *54*(10), 2267–2275. https://doi.org/10.1016/j.csda.2010.04.005

Zeger, S. L., & Karim, M. R. (1991). Generalized Linear Models with Random Effects; a Gibbs Sampling Approach. *Journal of the American Statistical Association*, *86*(413), 79–86. https://doi.org/10.1080/01621459.1991.10475006

Appendices

Appendix 1. *Simulation with N = 50 Children, Dropout Probabilities 0.3 vs. 0.1 Conditional on Respiratory Infection at Previous Visit (yes vs. no) respectively.*

	Analysis of incomplete datasets				MI following DA imputation (White et al. 2010)				MI following FLIC imputation (Puhr et al. 2017)			
	DL-GLM	GEE	DL-F-GLM	F—GEE	MI—GLM	MI—GEE	MI—F—GLM	MI—F—GEE	MI—GLM	MI—GEE	MI—F—GLM	MI—F—GEE
Coefficients												
β_0	−3.2101	*	−3.0056	−2.9484	−2.9356	*	−2.7639	−2.6936	−3.1168	*	−2.9550	−2.8976
β_1	−1.3438	*	−0.3853	−0.3711	−0.8699	*	−0.3256	−0.3125	−1.0640	*	−0.3398	−0.3141
β_2	−0.0351	*	−0.0306	−0.0306	−0.0287	*	−0.0248	−0.0221	−0.0299	*	−0.0272	−0.0257
β_3	−2.0978	*	1.5625	1.5500	−0.7707	*	1.3597	1.1920	−1.8766	*	1.4283	1.4095
SE												
β_0	5.8378	*	0.5164	0.5314	19.7392	*	0.5265	0.5670	4.1348	*	0.5170	0.5331
β_1	133.9532	*	0.7747	0.7644	106.7334	*	0.7487	0.7794	113.7417	*	0.7659	0.7646
β_2	0.0289	*	0.0257	0.0255	0.5379	*	0.0259	0.0267	0.0287	*	0.0259	0.0257
β_3	503.9182	*	1.0904	0.8837	393.3481	*	1.0729	1.0073	461.3192	*	1.0900	0.9509
Bias												
β_0	−0.4031	*	−0.1986	−0.1414	−0.1287	*	0.0431	0.1134	−0.3098	*	−0.1481	−0.0906
β_1	−0.9966	*	−0.0382	−0.0240	−0.5228	*	0.0216	0.0347	−0.7168	*	0.0074	0.0331
β_2	−0.0067	*	−0.0032	−0.0021	−0.0003	*	0.0036	0.0063	−0.0015	*	0.0012	0.0027
β_3	−3.5154	*	0.1449	0.1324	−2.1883	*	−0.0580	−0.2256	−3.2942	*	0.0107	−0.0082
MSE												
β_0	1.3658	*	0.4715	0.4563	1.2410	*	0.2549	0.2639	0.6996	*	0.3994	0.3655
β_1	20.4067	*	0.9809	0.8952	7.3541	*	0.6457	0.5618	12.3567	*	0.7493	0.6579
β_2	0.001438	*	0.001284	0.001220	0.002336	*	0.000818	0.000770	0.00103	*	0.00101	0.000951
β_3	63.1700	*	1.0203	0.9133	24.8166	*	0.6113	0.6113	52.0265	*	0.8117	0.7277
Coverage												
β_0	0.941	‡	0.935	0.920	0.944	‡	0.926	0.952	0.959	‡	0.949	0.955
β_1	0.935	‡	0.927	0.894	0.957	‡	0.950	0.960	0.957	‡	0.954	0.939
β_2	0.911	‡	0.889	0.891	0.932	‡	0.910	0.929	0.935	‡	0.914	0.926
β_3	0.961	‡	0.961	0.898	0.990	‡	0.990	0.983	0.981	‡	0.982	0.970
NC	0.7%	9.9%	0%	0.7%	3.8%	70.9%	0%	8.6%	4.3%	66.8%	0%	3.7%

Note: Results based on $n = 1{,}000$ converged simulation analyses using coefficients β_0 (intercept) = −2.8070; β_1 (gender) = −0.3472; β_2 (age) = −0.0284; β_3 (xerophthalmia) = 1.4176 and a first−order autoregressive correlation structure with $\alpha = 0.4$. *MI* = multiple imputation; *DA* = data augmentation; *FLIC* = Firth logistic regression with intercept correction; *DL* = direct likelihood; *GLM* = logistic regression; *GEE* = generalized estimating equations; *F* = Firth; *SE* = Standard Error; *MSE* = Mean Squared Error; *NC* = percentage of non−convergence; * value larger $1x10^{10}$ or smaller $-1x10^{10}$; ‡ not calculated due to uninterpretable coefficients and SE.

Appendix 2. *Simulation with N = 50 Children, Dropout Probabilities 0.4 vs. 0.1 Conditional on Respiratory Infection at Previous Visit (yes vs. no) respectively.*

	Analysis of incomplete datasets				MI following DA imputation (White et al. 2010)				MI following FLIC imputation (Puhr et al. 2017)			
	DL-GLM	GEE	DL-F-GLM	F—GEE	MI—GLM	MI—GEE	MI—F—GLM	MI—F—GEE	MI—GLM	MI—GEE	MI—F—GLM	MI—F—GEE
Coefficients												
β_0	−3.2452	*	−3.0302	−2.9674	−2.9209	*	−2.7574	−2.6840	−3.1478	*	−2.9811	−2.9252
β_1	−1.3349	*	−0.3789	−0.3677	−0.8224	*	−0.3220	−0.3136	−1.0425	*	−0.3343	−0.3150
β_2	−0.0356	*	−0.0315	−0.0305	−0.0283	*	−0.0244	−0.0220	−0.0301	*	−0.0271	−0.0258
β_3	2.2462	*	1.5738	1.5534	−0.6617	*	1.3689	1.1851	−1.9854	*	1.4260	1.4090
SE												
β_0	5.8496	*	0.5236	0.5355	19.1572	*	0.5348	0.5768	4.1462	*	0.5252	0.5382
β_1	135.0976	*	0.7814	0.7671	96.0382	*	0.7475	0.7773	111.6302	*	0.7743	0.7659
β_2	0.0296	*	0.0261	0.0257	0.5379	*	0.0260	0.0268	0.0293	*	0.0263	0.0259
β_3	547.1840	*	1.0998	0.8955	385.4460	*	1.0807	1.0308	483.4106	*	1.0978	0.9606
Bias												
β_0	−0.4382	*	−0.2232	−0.1604	−0.1140	*	0.0496	0.1230	−0.3408	*	−0.1741	−0.1182
β_1	−0.9877	*	−0.0317	−0.0206	−0.4752	*	0.0252	0.0336	−0.6953	*	0.0129	0.0321
β_2	−0.0072	*	−0.0031	−0.0021	0.0002	*	0.0040	0.0064	−0.0017	*	0.0013	0.0026
β_3	−3.6638	*	0.1561	0.1358	−2.0793	*	−0.0488	−0.2325	−3.4031	*	0.0084	−0.0087
MSE												
β_0	1.3920	*	0.4753	0.4571	1.2227	*	0.2504	0.2663	0.7219	*	0.4083	0.3680
β_1	20.1563	*	0.9337	0.8586	6.7131	*	0.6145	0.5490	11.9414	*	0.7058	0.6344
β_2	0.001494	*	0.001266	0.001201	0.002291	*	0.000793	0.000752	0.001004	*	0.001000	0.000949
β_3	66.5580	*	1.0254	0.9347	22.8156	*	0.5632	0.5822	53.8589	*	0.8121	0.7433
Coverage												
β_0	0.948	‡	0.943	0.925	0.948	‡	0.927	0.947	0.964	‡	0.958	0.962
β_1	0.940	‡	0.934	0.900	0.964	‡	0.956	0.962	0.963	‡	0.958	0.942
β_2	0.920	‡	0.899	0.894	0.939	‡	0.923	0.937	0.946	‡	0.926	0.928
β_3	0.963	‡	0.964	0.901	0.993	‡	0.992	0.987	0.982	‡	0.981	0.966
NC	0.5%	9.9%	0%	0.5%	3.8%	72.9%	0%	10.0%	4.6%	68.9%	0%	3.1%

Note: Results based on $n = 1{,}000$ converged simulation analyses using coefficients β_0 (intercept) = −2.8070; β_1 (gender) = −0.3472; β_2 (age) = −0.0284; β_3 (xerophthalmia) = 1.4176 and a first−order autoregressive correlation structure with $\alpha = 0.4$. *MI* = multiple imputation; *DA* = data augmentation; *FLIC* = Firth logistic regression with intercept correction; *DL* = direct likelihood; *GLM* = logistic regression; *GEE* = generalized estimating equations; *F* = Firth; *SE* = Standard Error; *MSE* = Mean Squared Error; *NC* = percentage of non−convergence; * value larger $1x10^{10}$ or smaller $-1x10^{10}$; ‡ not calculated due to uninterpretable coefficients and SE.

Appendix 3. *Simulation with N = 50 Children, Dropout Probabilities 0.6 vs. 0.1 Conditional on Respiratory Infection at Previous Visit (yes vs. no) respectively.*

	Analysis of incomplete datasets				MI following DA imputation (White et al. 2010)				MI following FLIC imputation (Puhr et al. 2017)			
	DL-GLM	GEE	DL-F-GLM	F—GEE	MI—GLM	MI—GEE	MI—F—GLM	MI—F—GEE	MI—GLM	MI—GEE	MI—F—GLM	MI—F—GEE
Coefficients												
β_0	−3.3194	*	−3.0819	−3.0341	−2.9134	*	−2.7440	−2.6821	−3.2139	*	−3.0389	−2.9915
β_1	−1.2491	*	−0.3732	−0.3693	−0.8322	*	−0.3183	−0.3128	−1.0387	*	−0.3282	−0.3061
β_2	−0.0367	*	−0.0322	−0.0318	−0.0277	*	−0.0238	−0.0222	−0.0309	*	−0.0277	−0.0266
β_3	−2.4713	*	1.6017	1.5921	−0.6865	*	1.3738	1.1639	−2.2162	*	1.4309	1.4251
SE												
β_0	5.8984	*	0.5387	0.5416	19.1877	*	0.5590	0.5992	2.9624	*	0.5408	0.5407
β_1	130.8061	*	0.7992	0.7706	98.2800	*	0.7452	0.7707	111.5247	*	0.7926	0.7639
β_2	0.0307	*	0.0268	0.0262	0.5382	*	0.0260	0.0269	0.0304	*	0.0271	0.0261
β_3	583.5836	*	1.1229	0.9274	401.3080	*	1.1069	1.0676	516.7869	*	1.1176	0.9788
Bias												
β_0	−0.5124	*	−0.2749	−0.2271	−0.1064	*	0.0629	0.1248	−0.4069	*	−0.2319	−0.1845
β_1	−0.9020	*	−0.0261	−0.0221	−0.4851	*	0.0288	0.0344	−0.6916	*	0.0189	0.0410
β_2	−0.0083	*	−0.0038	−0.0033	0.0007	*	0.0046	0.0062	−0.0025	*	0.0007	0.0018
β_3	−3.8889	*	0.1841	0.1745	−2.1042	*	−0.0438	−0.2538	−3.6338	*	0.0132	0.0075
MSE												
β_0	1.4656	*	0.4877	0.4736	1.2134	*	0.2479	0.2585	0.7381	*	0.4300	0.3825
β_1	18.6846	*	0.8826	0.8370	7.0199	*	0.5934	0.5250	11.8509	*	0.6656	0.5978
β_2	0.001505	*	0.001231	0.001199	0.002244	*	0.000744	0.000704	0.001056	*	0.000952	0.000888
β_3	71.0184	*	1.0589	1.0012	22.6385	*	0.5664	0.6044	57.7227	*	0.8129	0.7833
Coverage												
β_0	0.958	‡	0.960	0.933	0.958	‡	0.940	0.956	0.976	‡	0.968	0.965
β_1	0.955	‡	0.948	0.903	0.966	‡	0.962	0.971	0.974	‡	0.969	0.954
β_2	0.933	‡	0.917	0.899	0.940	‡	0.925	0.945	0.966	‡	0.951	0.942
β_3	0.961	‡	0.965	0.912	0.993	‡	0.993	0.990	0.989	‡	0.986	0.965
NC	0.3%	10.0%	0%	0.3%	3.6%	76.4%	0%	11.4%	5.2%	72.0%	0.1%	2.3%

Note: Results based on $n = 1{,}000$ converged simulation analyses using coefficients β_0 (intercept) = −2.8070; β_1 (gender) = −0.3472; β_2 (age) = −0.0284; β_3 (xerophthalmia) = 1.4176 and a first−order autoregressive correlation structure with $\alpha = 0.4$. *MI* = multiple imputation; *DA* = data augmentation; *FLIC* = Firth logistic regression with intercept correction; *DL* = direct likelihood; *GLM* = logistic regression; *GEE* = generalized estimating equations; *F* = Firth; *SE* = Standard Error; *MSE* = Mean Squared Error; *NC* = percentage of non−convergence; * value larger $1x10^{10}$ or smaller $-1x10^{10}$; ‡ not calculated due to uninterpretable coefficients and SE.

Appendix 4. *Simulation with N = 50 Children, Dropout Probabilities 0.6 vs. 0.2 Conditional on Respiratory Infection at Previous Visit (yes vs. no) respectively.*

	Analysis of incomplete datasets				MI following DA imputation (White et al. 2010)				MI following FLIC imputation (Puhr et al. 2017)			
	DL-GLM	GEE	DL-F-GLM	F—GEE	MI—GLM	MI—GEE	MI—F—GLM	MI—F—GEE	MI—GLM	MI—GEE	MI—F—GLM	MI—F—GEE
Coefficients												
β_0	−3.3383	*	−3.0488	−3.0100	−2.7383	*	−2.5820	−2.5338	−3.1630	*	−2.9737	−2.9252
β_1	−1.3584	*	−0.3777	−0.3700	−0.5437	*	−0.2569	−0.2561	−0.8252	*	−0.2955	−0.2882
β_2	−0.0364	*	−0.0316	−0.0310	−0.0224	*	−0.0192	−0.0181	−0.0273	*	−0.0242	−0.0232
β_3	−2.7855	*	1.6222	1.6083	−0.4570	*	1.2318	1.0434	−2.3557	*	1.3475	1.3244
SE												
β_0	14.3113	*	0.5537	0.5518	20.2178	*	0.6009	0.6353	8.2258	*	0.5537	0.5512
β_1	168.1259	*	0.8213	0.7857	80.6602	*	0.7225	0.7523	108.5289	*	0.7943	0.7723
β_2	0.0319	*	0.0275	0.0266	0.5389	*	0.0267	0.0278	0.0308	*	0.0275	0.0267
β_3	679.9824	*	1.1558	0.9563	359.9850	*	1.1261	1.1114	521.0655	*	1.1387	1.0109
Bias												
β_0	−0.5313	*	−0.2418	−0.2030	0.0687	*	0.2249	0.2732	−0.3560	*	−0.1667	−0.1182
β_1	−1.0112	*	−0.0306	−0.0228	−0.1966	*	0.0902	0.0911	−0.4780	*	0.0517	0.0590
β_2	−0.0080	*	−0.0032	−0.0026	0.0060	*	0.0092	0.0103	0.0011	*	0.0042	0.0052
β_3	−4.2031	*	0.2046	0.1906	−1.8746	*	−0.1859	−0.3743	−3.7733	*	−0.0702	−0.0932
MSE												
β_0	2.2565	*	0.4787	0.4698	1.1653	*	0.2623	0.2963	1.1024	*	0.3935	0.3551
β_1	22.1225	*	0.9055	0.8667	2.6346	*	0.4312	0.4044	7.8246	*	0.5692	0.5380
β_2	0.001565	*	0.001258	0.001218	0.002090	*	0.000717	0.000680	0.000877	*	0.000837	0.000773
β_3	76.2732	*	1.1055	1.0500	15.8561	*	0.4704	0.5864	55.0616	*	0.7205	0.7017
Coverage												
β_0	0.968	‡	0.954	0.924	0.948	‡	0.923	0.939	0.975	‡	0.964	0.967
β_1	0.957	‡	0.953	0.896	0.979	‡	0.976	0.985	0.985	‡	0.982	0.968
β_2	0.933	‡	0.916	0.908	0.935	‡	0.919	0.939	0.967	‡	0.952	0.956
β_3	0.967	‡	0.966	0.913	0.999	‡	0.999	0.998	0.995	‡	0.993	0.979
NC	0.6%	10.0%	0.1%	0.6%	3.4%	73.4%	0.1%	14.4%	6.1%	75.2%	0.2%	2.6%

Note: Results based on $n = 1{,}000$ converged simulation analyses using coefficients β_0 (intercept) = −2.8070; β_1 (gender) = −0.3472; β_2 (age) = −0.0284; β_3 (xerophthalmia) = 1.4176 and a first−order autoregressive correlation structure with $\alpha = 0.4$. *MI* = multiple imputation; *DA* = data augmentation; *FLIC* = Firth logistic regression with intercept correction; *DL* = direct likelihood; *GLM* = logistic regression; *GEE* = generalized estimating equations; *F* = Firth; *SE* = Standard Error; *MSE* = Mean Squared Error; *NC* = percentage of non−convergence; * value larger $1x10^{10}$ or smaller $-1x10^{10}$; ‡ not calculated due to uninterpretable coefficients and SE.

Appendix 5. *Simulation with N = 100 Children, Dropout Probabilities 0.3 vs. 0.1 Conditional on Respiratory Infection at Previous Visit (yes vs. no) respectively.*

	Analysis of incomplete datasets				MI following DA imputation (White et al. 2010)				MI following FLIC imputation (Puhr et al. 2017)			
	DL-GLM	GEE	DL-F-GLM	F—GEE	MI—GLM	MI—GEE	MI—F—GLM	MI—F—GEE	MI—GLM	MI—GEE	MI—F—GLM	MI—F—GEE
Coefficients												
β_0	−3.0354	*	−2.9459	−2.8882	−2.8559	*	−2.7888	−2.7337	−2.9852	*	−2.9096	−2.8821
β_1	−0.5226	*	−0.3760	−0.3675	−0.4162	*	−0.3389	−0.3439	−0.4290	*	−0.3387	−0.3300
β_2	−0.0319	*	−0.0299	−0.0291	−0.0277	*	−0.0262	−0.0245	−0.0279	*	−0.0263	−0.0262
β_3	0.3492	*	1.4864	1.4799	0.7004	*	1.3564	1.2000	0.4750	*	1.3934	1.3762
SE												
β_0	0.3622	*	0.3436	0.3906	0.3705	*	0.3539	0.4041	0.3596	*	0.3425	0.3862
β_1	11.2814	*	0.5200	0.5822	7.0998	*	0.5210	0.5897	7.6696	*	0.5149	0.5723
β_2	0.0186	*	0.0175	0.0190	0.0190	*	0.0181	0.0199	0.0185	*	0.0175	0.0191
β_3	97.4576	*	0.7647	0.6582	66.2025	*	0.7622	0.7432	82.0032	*	0.7626	0.7189
Bias												
β_0	−0.2285	*	−0.1389	−0.0812	−0.0489	*	0.0182	0.0733	−0.1782	*	−0.1026	−0.0751
β_1	−0.1754	*	−0.0288	−0.0204	−0.0690	*	0.0082	0.0032	−0.0819	*	0.0085	0.0171
β_2	−0.0035	*	−0.0015	−0.0007	0.0007	*	0.0022	0.0039	0.0005	*	0.0021	0.0022
β_3	−1.0684	*	0.0688	0.0623	−0.7172	*	−0.0612	−0.2176	−0.9427	*	−0.0242	−0.0415
MSE												
β_0	0.2851	*	0.2169	0.2042	0.1807	*	0.1571	0.1579	0.2230	*	0.1762	0.1687
β_1	2.5841	*	0.4287	0.4074	0.8639	*	0.3029	0.2877	1.0628	*	0.3318	0.3204
β_2	0.000579	*	0.000511	0.000477	0.000454	*	0.000422	0.000407	0.000431	*	0.000398	0.000383
β_3	19.7709	*	0.6856	0.5727	7.7330	*	0.4581	0.4241	13.4538	*	0.5537	0.4667
Coverage												
β_0	0.882	‡	0.891	0.933	0.928	‡	0.915	0.951	0.915	‡	0.916	0.954
β_1	0.917	‡	0.913	0.939	0.952	‡	0.951	0.979	0.946	‡	0.946	0.962
β_2	0.911	‡	0.910	0.921	0.931	‡	0.921	0.943	0.935	‡	0.927	0.948
β_3	0.950	‡	0.949	0.900	0.985	‡	0.985	0.982	0.974	‡	0.973	0.960
NC	0.3%	7.6%	0%	0.3%	0%	37.6%	0%	5.7%	0%	34.4%	0%	3.2%

Note: Results based on $n = 1{,}000$ converged simulation analyses using coefficients β_0 (intercept) = −2.8070; β_1 (gender) = −0.3472; β_2 (age) = −0.0284; β_3 (xerophthalmia) = 1.4176 and a first—order autoregressive correlation structure with $\alpha = 0.4$. *MI* = multiple imputation; *DA* = data augmentation; *FLIC* = Firth logistic regression with intercept correction; *DL* = direct likelihood; *GLM* = logistic regression; *GEE* = generalized estimating equations; *F* = Firth; *SE* = Standard Error; *MSE* = Mean Squared Error; *NC* = percentage of non—convergence; * value larger $1x10^{10}$ or smaller $-1x10^{10}$; ‡ not calculated due to uninterpretable coefficients and SE

Appendix 6. *Simulation with N = 100 Children, Dropout Probabilities 0.4 vs. 0.1 Conditional on Respiratory Infection at Previous Visit (yes vs. no) respectively.*

	Analysis of incomplete datasets				MI following DA imputation (White et al. 2010)				MI following FLIC imputation (Puhr et al. 2017)			
	DL-GLM	GEE	DL-F-GLM	F—GEE	MI—GLM	MI—GEE	MI—F—GLM	MI—F—GEE	MI—GLM	MI—GEE	MI—F—GLM	MI—F—GEE
Coefficients												
β_0	−3.0714	*	−2.9786	−2.9176	−2.8515	*	−2.7845	−2.7254	−3.0205	*	−2.9422	−2.9177
β_1	−0.5246	*	−0.3762	−0.3693	−0.4135	*	−0.3372	−0.3436	−0.4325	*	−0.3373	−0.3280
β_2	−0.0325	*	−0.0303	−0.0297	−0.0278	*	−0.0263	−0.0244	−0.0281	*	−0.0264	−0.0264
β_3	0.3123	*	1.4962	1.4886	0.7562	*	1.3792	1.1690	0.4459	*	1.3944	1.3868
SE												
β_0	0.3699	*	0.3501	0.3925	0.3794	*	0.3622	0.4118	0.3685	*	0.3502	0.3887
β_1	11.2775	*	0.5284	0.5838	6.9813	*	0.5241	0.5919	7.4576	*	0.5226	0.5715
β_2	0.0190	*	0.0178	0.0192	0.0192	*	0.0183	0.0200	0.0189	*	0.0178	0.0192
β_3	102.4559	*	0.7756	0.6716	66.2149	*	0.7744	0.7702	87.1336	*	0.7744	0.7312
Bias												
β_0	−0.2644	*	−0.1716	−0.1107	−0.0445	*	0.0225	0.0816	−0.2135	*	−0.1352	−0.1107
β_1	−0.1774	*	−0.0291	−0.0222	−0.0663	*	0.0100	0.0036	−0.0853	*	0.0098	0.0191
β_2	−0.0041	*	−0.0019	−0.0013	0.0006	*	0.0021	0.0040	0.0003	*	0.0020	0.0020
β_3	−1.1053	*	0.0786	0.0710	−0.6614	*	−0.0385	−0.2486	−0.9718	*	−0.0232	−0.0308
MSE												
β_0	0.3121	*	0.2330	0.2198	0.1894	*	0.1654	0.1694	0.2437	*	0.1891	0.1801
β_1	2.5809	*	0.4233	0.4021	0.8824	*	0.2963	0.2820	1.2327	*	0.3275	0.3157
β_2	0.000573	*	0.000500	0.000475	0.000439	*	0.000409	0.000397	0.000416	*	0.000382	0.000369
β_3	20.5359	*	0.7163	0.6184	6.9889	*	0.4631	0.4529	13.7877	*	0.5725	0.4936
Coverage												
β_0	0.873	‡	0.894	0.920	0.924	‡	0.919	0.945	0.920	‡	0.933	0.948
β_1	0.919	‡	0.910	0.939	0.959	‡	0.956	0.979	0.944	‡	0.943	0.963
β_2	0.918	‡	0.912	0.927	0.932	‡	0.927	0.942	0.942	‡	0.936	0.946
β_3	0.943	‡	0.941	0.887	0.982	‡	0.982	0.975	0.973	‡	0.969	0.947
NC	0.2%	8.1%	0%	0.2%	0%	40.2%	0%	6.2%	0%	35.2%	0%	2.3%

Note: Results based on $n = 1{,}000$ converged simulation analyses using coefficients β_0 (intercept) = −2.8070; β_1 (gender) = −0.3472; β_2 (age) = −0.0284; β_3 (xerophthalmia) = 1.4176 and a first−order autoregressive correlation structure with $\alpha = 0.4$. *MI* = multiple imputation; *DA* = data augmentation; *FLIC* = Firth logistic regression with intercept correction; *DL* = direct likelihood; *GLM* = logistic regression; *GEE* = generalized estimating equations; *F* = Firth; *SE* = Standard Error; *MSE* = Mean Squared Error; *NC* = percentage of non−convergence; * value larger $1x10^{10}$ or smaller $-1x10^{10}$; ‡ not calculated due to uninterpretable coefficients and *SE*.

Appendix 7. *Simulation with N = 100 Children, Dropout Probabilities 0.6 vs. 0.1 Conditional on Respiratory Infection at Previous Visit (yes vs. no) respectively.*

	Analysis of incomplete datasets				MI following DA imputation (White et al. 2010)				MI following FLIC imputation (Puhr et al. 2017)			
	DL-GLM	GEE	DL-F-GLM	F—GEE	MI—GLM	MI—GEE	MI—F—GLM	MI—F—GEE	MI—GLM	MI—GEE	MI—F—GLM	MI—F—GEE
Coefficients												
β_0	−3.1447	*	−3.0453	−2.9894	−2.8517	*	−2.7853	−2.7206	−3.0961	*	−3.0130	−2.9943
β_1	−0.4951	*	−0.3623	−0.3582	−0.3856	*	−0.3152	−0.3273	−0.4068	*	−0.3199	−0.3173
β_2	−0.0329	*	−0.0306	−0.0303	−0.0272	*	−0.0258	−0.0237	−0.0280	*	−0.0262	−0.0265
β_3	0.2812	*	1.5196	1.5126	0.8442	*	1.4112	1.1023	0.4160	*	1.4047	1.4093
SE												
β_0	0.3854	*	0.3633	0.3932	0.3996	*	0.3811	0.4266	0.3844	*	0.3640	0.3902
β_1	9.7005	*	0.5445	0.5822	6.8592	*	0.5267	0.5877	7.1591	*	0.5388	0.5690
β_2	0.0197	*	0.0184	0.0194	0.0195	*	0.0186	0.0203	0.0196	*	0.0185	0.0194
β_3	112.6163	*	0.7953	0.6999	65.6889	*	0.8002	0.8984	93.3291	*	0.7932	0.7500
Bias												
β_0	−0.3377	*	−0.2383	−0.1824	−0.0447	*	0.0217	0.0864	−0.2892	*	−0.2060	−0.1873
β_1	−0.1479	*	−0.0151	−0.0110	−0.0384	*	0.0320	0.0199	−0.0597	*	0.0272	0.0298
β_2	−0.0045	*	−0.0022	−0.0019	0.0012	*	0.0027	0.0048	0.0004	*	0.0022	0.0019
β_3	−1.1364	*	0.1019	0.0950	−0.5734	*	−0.0065	−0.3153	−1.0016	*	−0.0130	−0.0084
MSE												
β_0	0.3444	*	0.2475	0.2347	0.1879	*	0.1638	0.1756	0.2667	*	0.1982	0.1908
β_1	2.2862	*	0.4132	0.3976	0.7329	*	0.2863	0.2629	1.1144	*	0.3098	0.3066
β_2	0.000592	*	0.000512	0.000489	0.000429	*	0.000400	0.000404	0.000419	*	0.000383	0.000376
β_3	21.1994	*	0.6914	0.6456	5.8061	*	0.4001	1.4902	14.0497	*	0.5319	0.5036
Coverage												
β_0	0.881	‡	0.904	0.918	0.931	‡	0.927	0.951	0.934	‡	0.944	0.952
β_1	0.941	‡	0.934	0.943	0.957	‡	0.954	0.981	0.963	‡	0.964	0.965
β_2	0.917	‡	0.914	0.932	0.942	‡	0.934	0.947	0.954	‡	0.953	0.956
β_3	0.941	‡	0.939	0.893	0.985	‡	0.984	0.989	0.984	‡	0.977	0.957
NC	0.1%	8.5%	0%	0.1%	0.1%	48%	0%	8%	0.1%	38%	0%	1.9%

Note: Results based on $n = 1{,}000$ converged simulation analyses using coefficients β_0 (intercept) = −2.8070; β_1 (gender) = −0.3472; β_2 (age) = −0.0284; β_3 (xerophthalmia) = 1.4176 and a first−order autoregressive correlation structure with $\alpha = 0.4$. *MI* = multiple imputation; *DA* = data augmentation; *FLIC* = Firth logistic regression with intercept correction; *DL* = direct likelihood; *GLM* = logistic regression; *GEE* = generalized estimating equations; *F* = Firth; *SE* = Standard Error; *MSE* = Mean Squared Error; *NC* = percentage of non−convergence; * value larger $1x10^{10}$ or smaller $-1x10^{10}$; ‡ not calculated due to uninterpretable coefficients and *SE*.

Appendix 8. *Simulation with N = 100 Children, Dropout Probabilities 0.6 vs. 0.2 Conditional on Respiratory Infection at Previous Visit (yes vs. no) respectively.*

	Analysis of incomplete datasets				MI following DA imputation (White et al. 2010)				MI following FLIC imputation (Puhr et al. 2017)			
	DL-GLM	GEE	DL-F-GLM	F—GEE	MI—GLM	MI—GEE	MI—F—GLM	MI—F—GEE	MI—GLM	MI—GEE	MI—F—GLM	MI—F—GEE
Coefficients												
β_0	−3.1598	*	−3.0366	−2.9894	−2.8517	*	−2.7853	−2.7206	−3.0551	*	−2.9723	−2.9543
β_1	−0.5240	*	−0.3662	−0.3623	−0.3856	*	−0.3152	−0.3273	−0.3517	*	−0.2938	−0.2909
β_2	−0.0332	*	−0.0307	−0.0304	−0.0272	*	−0.0258	−0.0237	−0.0251	*	−0.0235	−0.0236
β_3	0.0502	*	1.5258	1.5196	0.8442	*	1.4112	1.1023	0.3205	*	1.3309	1.3221
SE												
β_0	1.7403	*	0.3782	0.4070	0.3996	*	0.3811	0.4266	0.9501	*	0.3774	0.3991
β_1	15.6008	*	0.5662	0.6001	6.8592	*	0.5267	0.5877	6.1775	*	0.5489	0.5749
β_2	0.0205	*	0.0190	0.0199	0.0195	*	0.0186	0.0203	0.0201	*	0.0189	0.0198
β_3	144.8085	*	0.8271	0.7223	65.6889	*	0.8002	0.8984	101.9256	*	0.8190	0.7776
Bias												
β_0	−0.3528	*	−0.2297	−0.1824	−0.0447	*	0.0217	0.0864	−0.2481	*	−0.1653	−0.1473
β_1	−0.1768	*	−0.0191	−0.0151	−0.0284	*	0.0320	0.0199	−0.0046	*	0.0533	0.0563
β_2	−0.0048	*	−0.0023	−0.0020	0.0012	*	0.0027	0.0048	0.0033	*	0.0049	0.0048
β_3	−1.3675	*	0.1081	0.1020	−0.5734	*	−0.0065	−0.3153	−1.0972	*	−0.0867	−0.0956
MSE												
β_0	0.6833	*	0.2643	0.2520	0.1879	*	0.1638	0.1756	0.2649	*	0.1842	0.1784
β_1	3.2713	*	0.4519	0.4338	0.7329	*	0.2863	0.2629	0.5678	*	0.2723	0.2660
β_2	0.000647	*	0.000553	0.000528	0.000429	*	0.000400	0.000404	0.000389	*	0.000368	0.000362
β_3	25.8095	*	0.7251	0.6798	5.8061	*	0.4001	1.4902	12.9071	*	0.4932	0.4704
Coverage												
β_0	0.891	‡	0.918	0.926	0.931	‡	0.927	0.927	0.953	‡	0.956	0.968
β_1	0.939	‡	0.935	0.951	0.957	‡	0.954	0.954	0.971	‡	0.970	0.976
β_2	0.916	‡	0.907	0.932	0.942	‡	0.934	0.934	0.955	‡	0.940	0.964
β_3	0.941	‡	0.940	0.894	0.985	‡	0.984	0.984	0.994	‡	0.990	0.972
NC	0.2%	9.0%	0%	0.2%	0.1%	47.6%	0%	8.5%	0.1%	37.2%	0%	1.4%

Note: Results based on $n = 1{,}000$ converged simulation analyses using coefficients β_0 (intercept) = −2.8070; β_1 (gender) = −0.3472; β_2 (age) = −0.0284; β_3 (xerophthalmia) = 1.4176 and a first−order autoregressive correlation structure with $\alpha = 0.4$. *MI* = multiple imputation; *DA* = data augmentation; *FLIC* = Firth logistic regression with intercept correction; *GLM* = logistic regression; *GEE* = generalized estimating equations; *F* = Firth; *SE* = Standard Error; *MSE* = Mean Squared Error; *NC* = percentage of non−convergence; * value larger $1x10^{10}$ or smaller $-1x10^{10}$; ‡ not calculated due to uninterpretable coefficients and *SE*.

Appendix 9. *Simulation with N = 200 Children, Dropout Probabilities 0.3 vs. 0.1 Conditional on Respiratory Infection at Previous Visit (yes vs. no) respectively.*

	Analysis of incomplete datasets				MI following DA imputation (White et al. 2010)				MI following FLIC imputation (Puhr et al. 2017)			
	DL-GLM	GEE	DL-F-GLM	F—GEE	MI—GLM	MI—GEE	MI—F—GLM	MI—F—GEE	MI—GLM	MI—GEE	MI—F—GLM	MI—F—GEE
Coefficients												
β_0	−2.9371	*	−2.8977	−2.8448	−2.7598	*	−2.7299	−2.6754	−2.9145	*	−2.8799	−2.8648
β_1	−0.3820	*	−0.3637	−0.3598	−0.3144	*	−0.3022	−0.3227	−0.3490	*	−0.3335	−0.3324
β_2	−0.0309	*	−0.0300	−0.0294	−0.0268	*	−0.0261	−0.0241	−0.0275	*	−0.0268	−0.0269
β_3	1.2815	*	1.4344	1.4326	1.1446	*	1.2559	0.9941	1.2179	*	1.3490	1.3544
SE												
β_0	0.2406	*	0.2352	0.2788	0.2805	*	0.2746	0.3122	0.2387	*	0.2337	0.2735
β_1	0.3622	*	0.3526	0.4205	0.3820	*	0.3734	0.4323	0.3583	*	0.3495	0.4094
β_2	0.0125	*	0.0121	0.0138	0.0141	*	0.0138	0.0153	0.0124	*	0.0121	0.0137
β_3	7.1898	*	0.5251	0.4877	7.2492	*	0.5805	0.6159	5.9533	*	0.5251	0.5218
Bias												
β_0	−0.1301	*	−0.0908	−0.0378	0.0471	*	0.0771	0.1316	−0.1076	*	−0.0729	−0.0578
β_1	−0.0348	*	−0.0166	−0.0126	0.0328	*	0.0450	0.0244	−0.0018	*	0.0137	0.0147
β_2	−0.0025	*	−0.0016	−0.0010	0.0016	*	0.0023	0.0044	0.0009	*	0.0017	0.0015
β_3	−0.1361	*	0.0168	0.0150	−0.2730	*	−0.1617	−0.4235	−0.1997	*	−0.0687	−0.0632
MSE												
β_0	0.1095	*	0.0948	0.0871	0.0868	*	0.0861	0.0968	0.0925	*	0.0815	0.0783
β_1	0.2082	*	0.1919	0.1811	0.1306	*	0.1246	0.1178	0.1622	*	0.1517	0.1477
β_2	0.000257	*	0.000242	0.000225	0.000232	*	0.000227	0.000234	0.000204	*	0.000198	0.000192
β_3	2.2802	*	0.3554	0.2923	0.8900	*	0.2402	0.3654	1.5724	*	0.3064	0.2533
Coverage												
β_0	0.872	‡	0.893	0.942	0.922	‡	0.913	0.928	0.912	‡	0.915	0.956
β_1	0.892	‡	0.895	0.962	0.959	‡	0.958	0.995	0.931	‡	0.933	0.969
β_2	0.892	‡	0.889	0.930	0.920	‡	0.916	0.931	0.914	‡	0.907	0.941
β_3	0.947	‡	0.941	0.940	0.992	‡	0.991	0.984	0.969	‡	0.968	0.963
NC	0%	2.3%	0%	0%	0%	15.6%	0%	3.1%	0%	6.4%	0%	0.1%

Note: Results based on $n = 1{,}000$ converged simulation analyses using coefficients β_0 (intercept) = −2.8070; β_1 (gender) = −0.3472; β_2 (age) = −0.0284; β_3 (xerophthalmia) = 1.4176 and a first—order autoregressive correlation structure with $\alpha = 0.4$. *MI* = multiple imputation; *DA* = data augmentation; *FLIC* = Firth logistic regression with intercept correction; *DL* = direct likelihood; *GLM* = logistic regression; *GEE* = generalized estimating equations; *F* = Firth; *SE* = Standard Error; *MSE* = Mean Squared Error; *NC* = percentage of non—convergence; * value larger $1\text{x}10^{10}$ or smaller $-1\text{x}10^{10}$; ‡ not calculated due to uninterpretable coefficients and *SE*.

Appendix 10. *Simulation with N = 200 Children, Dropout Probabilities 0.4 vs. 0.1 Conditional on Respiratory Infection at Previous Visit (yes vs. no) respectively.*

	Analysis of incomplete datasets				MI following DA imputation (White et al. 2010)				MI following FLIC imputation (Puhr et al. 2017)			
	DL-GLM	GEE	DL-F-GLM	F—GEE	MI—GLM	MI—GEE	MI—F—GLM	MI—F—GEE	MI—GLM	MI—GEE	MI—F—GLM	MI—F—GEE
Coefficients												
β_0	−2.9757	*	−2.9346	−2.8728	−2.8057	*	−2.7744	−2.7330	−2.9516	*	−2.9157	−2.9005
β_1	−0.3809	*	−0.3620	−0.3590	−0.3494	*	−0.3355	−0.3489	−0.3494	*	−0.3333	−0.3334
β_2	−0.0311	*	−0.0302	−0.0296	−0.0286	*	−0.0279	−0.0266	−0.0276	*	−0.0268	−0.0270
β_3	1.2870	*	1.4441	1.4407	1.2278	*	1.3101	1.1738	1.2269	*	1.3514	1.3591
SE												
β_0	0.2458	*	0.2401	0.2789	0.2534	*	0.2483	0.2911	0.2442	*	0.2389	0.2743
β_1	0.3696	*	0.3595	0.4207	0.3685	*	0.3600	0.4280	0.3653	*	0.3559	0.4090
β_2	0.0127	*	0.0123	0.0139	0.0130	*	0.0127	0.0144	0.0127	*	0.0123	0.0138
β_3	7.6237	*	0.5339	0.4998	4.0496	*	0.5419	0.5558	5.6494	*	0.5326	0.5345
Bias												
β_0	−0.1687	*	−0.1276	−0.0658	0.0012	*	0.0326	0.0740	−0.1446	*	−0.1088	−0.0935
β_1	−0.0338	*	−0.0148	−0.0119	−0.0023	*	0.0117	−0.0017	−0.0022	*	0.0139	0.0137
β_2	−0.0027	*	−0.0018	−0.0012	−0.0002	*	0.0005	0.0019	0.0008	*	0.0016	0.0014
β_3	−0.1306	*	0.0265	0.0231	−0.1899	*	−0.1076	−0.2438	−0.1908	*	−0.0662	−0.0586
MSE												
β_0	0.1227	*	0.1041	0.0928	0.0815	*	0.0781	0.0802	0.1030	*	0.0889	0.0850
β_1	0.2118	*	0.1948	0.1825	0.1536	*	0.1446	0.1410	0.1663	*	0.1552	0.1509
β_2	0.000265	*	0.000249	0.000234	0.000230	*	0.000222	0.000209	0.000204	*	0.000198	0.000193
β_3	2.3153	*	0.3580	0.3063	0.6206	*	0.2574	0.2643	1.3243	*	0.2992	0.2590
Coverage												
β_0	0.866	‡	0.882	0.933	0.922	‡	0.911	0.947	0.893	‡	0.913	0.945
β_1	0.905	‡	0.910	0.950	0.941	‡	0.941	0.981	0.936	‡	0.935	0.968
β_2	0.891	‡	0.888	0.926	0.910	‡	0.904	0.945	0.917	‡	0.907	0.943
β_3	0.951	‡	0.946	0.937	0.983	‡	0.982	0.986	0.972	‡	0.968	0.971
NC	0%	0.3%	0%	0%	0%	9.1%	0%	1.7%	0%	7.4%	0%	0.4%

Note: Results based on $n = 1{,}000$ converged simulation analyses using coefficients β_0 (intercept) = −2.8070; β_1 (gender) = −0.3472; β_2 (age) = −0.0284; β_3 (xerophthalmia) = 1.4176 and a first−order autoregressive correlation structure with $\alpha = 0.4$. *MI* = multiple imputation; *DA* = data augmentation; *FLIC* = Firth logistic regression with intercept correction; *DL* = direct likelihood; *GLM* = logistic regression; *GEE* = generalized estimating equations; *F* = Firth; *SE* = Standard Error; *MSE* = Mean Squared Error; *NC* = percentage of non−convergence; * value larger $1x10^{10}$ or smaller $-1x10^{10}$; ‡ not calculated due to uninterpretable coefficients and *SE*.

Appendix 11. *Simulation with N = 200 Children, Dropout Probabilities 0.6 vs. 0.1 Conditional on Respiratory Infection at Previous Visit (yes vs. no) respectively.*

	Analysis of incomplete datasets				MI following DA imputation (White et al. 2010)				MI following FLIC imputation (Puhr et al. 2017)			
	DL-GLM	GEE	DL-F-GLM	F—GEE	MI—GLM	MI—GEE	MI—F—GLM	MI—F—GEE	MI—GLM	MI—GEE	MI—F—GLM	MI—F—GEE
Coefficients												
β_0	−3.0551	*	−3.0094	−2.9476	−2.8081	*	−2.7766	−2.7195	−3.0329	*	−2.9942	−2.9816
β_1	−0.3771	*	−0.3575	−0.3540	−0.3420	*	−0.3284	−0.3496	−0.3408	*	−0.3239	−0.3250
β_2	−0.0318	*	−0.0308	−0.0303	−0.0288	*	−0.0280	−0.0260	−0.0278	*	−0.0270	−0.0272
β_3	1.2866	*	1.4741	1.4690	1.2383	*	1.3427	1.0865	1.2117	*	1.3723	1.3816
SE												
β_0	0.2567	*	0.2501	0.2801	0.2710	*	0.2652	0.3065	0.2555	*	0.2495	0.2761
β_1	0.3840	*	0.3726	0.4199	0.3737	*	0.3650	0.4304	0.3792	*	0.3689	0.4080
β_2	0.0133	*	0.0128	0.0140	0.0133	*	0.0130	0.0147	0.0132	*	0.0128	0.0140
β_3	9.3766	*	0.5492	0.5220	6.0217	*	0.5668	0.6007	7.8365	*	0.5482	0.5509
Bias												
β_0	−0.2481	*	−0.2025	−0.1406	−0.0011	*	0.0304	0.0875	−0.2259	*	−0.1872	−0.1746
β_1	−0.0299	*	−0.0104	−0.0069	0.0052	*	0.0188	−0.0024	0.0064	*	0.0232	0.0222
β_2	−0.0034	*	−0.0024	−0.0019	−0.0004	*	0.0004	0.0024	0.0006	*	0.0014	0.0012
β_3	−0.1310	*	0.0565	0.0514	−0.1793	*	−0.0750	−0.3312	−0.2060	*	−0.0453	−0.0361
MSE												
β_0	0.1540	*	0.1264	0.1103	0.0856	*	0.0819	0.0883	0.1307	*	0.1095	0.1050
β_1	0.1973	*	0.1809	0.1732	0.1477	*	0.1395	0.1351	0.1513	*	0.1411	0.1385
β_2	0.000260	*	0.000242	0.000229	0.000224	*	0.000216	0.000209	0.000192	*	0.000185	0.000183
β_3	2.7577	*	0.3600	0.3254	0.9836	*	0.2380	0.3096	1.9440	*	0.2958	0.2719
Coverage												
β_0	0.852	‡	0.879	0.914	0.925	‡	0.918	0.936	0.889	‡	0.907	0.933
β_1	0.921	‡	0.927	0.962	0.942	‡	0.941	0.981	0.953	‡	0.949	0.975
β_2	0.906	‡	0.905	0.935	0.921	‡	0.916	0.944	0.928	‡	0.924	0.948
β_3	0.940	‡	0.936	0.934	0.986	‡	0.984	0.983	0.979	‡	0.979	0.973
NC	0.1%	3.4%	0%	0.1%	0%	13.6%	0%	2.3%	0%	9.6%	0%	0.1%

Note: Results based on $n = 1{,}000$ converged simulation analyses using coefficients β_0 (intercept) = −2.8070; β_1 (gender) = −0.3472; β_2 (age) = −0.0284; β_3 (xerophthalmia) = 1.4176 and a first−order autoregressive correlation structure with $\alpha = 0.4$. *MI* = multiple imputation; *DA* = data augmentation; *FLIC* = Firth logistic regression with intercept correction; *DL* = direct likelihood; *GLM* = logistic regression; *GEE* = generalized estimating equations; *F* = Firth; *SE* = Standard Error; *MSE* = Mean Squared Error; *NC* = percentage of non−convergence; * value larger $1x10^{10}$ or smaller $-1x10^{10}$; ‡ not calculated due to uninterpretable coefficients and *SE*.

Appendix 12. *Simulation with N = 200 Children, Dropout Probabilities 0.8 vs. 0.1 Conditional on Respiratory Infection at Previous Visit (yes vs. no) respectively.*

	Analysis of incomplete datasets				MI following DA imputation (White et al. 2010)				MI following FLIC imputation (Puhr et al. 2017)			
	DL-GLM	GEE	DL-F-GLM	F—GEE	MI—GLM	MI—GEE	MI—F—GLM	MI—F—GEE	MI—GLM	MI—GEE	MI—F—GLM	MI—F—GEE
Coefficients												
β_0	−3.1335	*	−3.0843	−3.0409	−2.8124	*	−2.7807	−2.7153	−3.1143	*	−3.0726	−3.0639
β_1	−0.3798	*	−0.3582	−0.3571	−0.3355	*	−0.3221	−0.3493	−0.3357	*	−0.3178	−0.3182
β_2	−0.0328	*	−0.0317	−0.0315	−0.0286	*	−0.0279	−0.0253	−0.0283	*	−0.0274	−0.0276
β_3	1.2619	*	1.5058	1.5019	1.3351	*	1.4254	1.0325	1.2105	*	1.3924	1.4009
SE												
β_0	0.2682	*	0.2609	0.2803	0.2945	*	0.2880	0.3276	0.2671	*	0.2603	0.2761
β_1	0.3996	*	0.3868	0.4166	0.3771	*	0.3683	0.4289	0.3946	*	0.3831	0.4044
β_2	0.0138	*	0.0133	0.0141	0.0136	*	0.0133	0.0150	0.0138	*	0.0134	0.0140
β_3	12.9279	*	0.5661	0.5446	6.3891	*	0.6091	0.6521	10.2813	*	0.5643	0.5633
Bias												
β_0	−0.3265	*	−0.2774	−0.2340	−0.0055	*	0.0263	0.0916	−0.3073	*	−0.2657	−0.2564
β_1	−0.0326	*	−0.0111	−0.0100	0.0117	*	0.0250	−0.0022	0.0115	*	0.0294	0.0287
β_2	−0.0044	*	−0.0033	−0.0031	−0.0002	*	0.0005	0.0031	0.0001	*	0.0010	0.0008
β_3	−0.1557	*	0.0882	0.0843	−0.0826	*	0.0077	−0.3852	−0.2072	*	−0.0252	−0.0167
MSE												
β_0	0.1975	*	0.1607	0.1455	0.1050	*	0.0998	0.1202	0.1703	*	0.1410	0.1368
β_1	0.1936	*	0.1764	0.1722	0.1391	*	0.1317	0.1249	0.1401	*	0.1302	0.1289
β_2	0.000269	*	0.000247	0.000239	0.000210	*	0.000203	0.000213	0.000190	*	0.000182	0.000181
β_3	3.7533	*	0.3879	0.3694	0.7413	*	0.2278	0.3717	2.0146	*	0.2972	0.2879
Coverage												
β_0	0.819	‡	0.852	0.880	0.927	‡	0.924	0.924	0.856	‡	0.889	0.902
β_1	0.943	‡	0.938	0.955	0.952	‡	0.952	0.986	0.968	‡	0.967	0.976
β_2	0.921	‡	0.916	0.935	0.933	‡	0.931	0.952	0.946	‡	0.941	0.960
β_3	0.947	‡	0.936	0.927	0.984	‡	0.983	0.975	0.974	‡	0.970	0.968
NC	0%	0.4%	0%	0%	0%	31.6%	0%	8.6%	0%	13.2%	0%	0%

Note: Results based on $n = 1{,}000$ converged simulation analyses using coefficients β_0 (intercept) = −2.8070; β_1 (gender) = −0.3472; β_2 (age) = −0.0284; β_3 (xerophthalmia) = 1.4176 and a first−order autoregressive correlation structure with $\alpha = 0.4$. *MI* = multiple imputation; *DA* = data augmentation; *FLIC* = Firth logistic regression with intercept correction; *DL* = direct likelihood; *GLM* = logistic regression; *GEE* = generalized estimating equations; *F* = Firth; *SE* = Standard Error; *MSE* = Mean Squared Error; *NC* = percentage of non−convergence; * value larger $1x10^{10}$ or smaller $-1x10^{10}$; ‡ not calculated due to uninterpretable coefficients and *SE*.

Appendix 13. *Simulation with N = 200 Children, Dropout Probabilities 0.6 vs. 0.2 Conditional on Respiratory Infection at Previous Visit (yes vs. no) respectively.*

	Analysis of incomplete datasets				MI following DA imputation (White et al. 2010)				MI following FLIC imputation (Puhr et al. 2017)			
	DL-GLM	GEE	DL-F-GLM	F—GEE	MI—GLM	MI—GEE	MI—F—GLM	MI—F—GEE	MI—GLM	MI—GEE	MI—F—GLM	MI—F—GEE
Coefficients												
β_0	−3.0377	*	−2.9896	−2.9358	−2.7598	*	−2.7299	−2.6754	−2.9976	*	−2.9603	−2.9501
β_1	−0.3785	*	−0.3573	−0.3555	−0.3144	*	−0.3022	−0.3227	−0.3134	*	−0.2979	−0.3002
β_2	−0.0319	*	−0.0308	−0.0304	−0.0268	*	−0.0261	−0.0241	−0.0255	*	−0.0247	−0.0249
β_3	1.1980	*	1.4592	1.4557	1.1446	*	1.2559	0.9941	1.1264	*	1.3046	1.3090
SE												
β_0	0.2657	*	0.2585	0.2896	0.2805	*	0.2746	0.3122	0.2625	*	0.2563	0.2797
β_1	0.3963	*	0.3839	0.4326	0.3820	*	0.3734	0.4323	0.3843	*	0.3739	0.4087
β_2	0.0137	*	0.0132	0.0144	0.0141	*	0.0138	0.0153	0.0135	*	0.0131	0.0142
β_3	14.1176	*	0.5724	0.5414	7.2492	*	0.5805	0.6159	10.0507	*	0.5683	0.5701
Bias												
β_0	−0.2307	*	−0.1826	−0.1288	0.0471	*	0.0771	0.1316	−0.1906	*	−0.1533	−0.1431
β_1	−0.0313	*	−0.0101	−0.0083	0.0328	*	0.0450	0.0244	0.0338	*	0.0492	0.0470
β_2	−0.0035	*	−0.0024	−0.0020	0.0016	*	0.0023	0.0044	0.0029	*	0.0037	0.0035
β_3	−0.2196	*	0.0415	0.0381	−0.2730	*	−0.1617	−0.4235	−0.2912	*	−0.1130	−0.1086
MSE												
β_0	0.1563	*	0.1285	0.1153	0.0868	*	0.0861	0.0968	0.1174	*	0.0993	0.0962
β_1	0.2112	*	0.1930	0.1852	0.1306	*	0.1246	0.1178	0.1360	*	0.1282	0.1282
β_2	0.000289	*	0.000267	0.000255	0.000232	*	0.000227	0.000234	0.000190	*	0.000187	0.000186
β_3	4.0476	*	0.3976	0.3623	0.8900	*	0.2402	0.3654	1.8236	*	0.2964	0.2693
Coverage												
β_0	0.856	‡	0.880	0.920	0.922	‡	0.913	0.928	0.907	‡	0.926	0.941
β_1	0.924	‡	0.924	0.955	0.959	‡	0.958	0.995	0.959	‡	0.958	0.974
β_2	0.890	‡	0.885	0.910	0.920	‡	0.916	0.931	0.927	‡	0.922	0.938
β_3	0.953	‡	0.950	0.936	0.992	‡	0.991	0.984	0.981	‡	0.980	0.975
NC	0%	2.9%	0%	0%	0%	15.6%	0%	3.1%	0%	9.1%	0%	0.5%

Note: Results based on $n = 1{,}000$ converged simulation analyses using coefficients β_0 (intercept) = −2.8070; β_1 (gender) = −0.3472; β_2 (age) = −0.0284; β_3 (xerophthalmia) = 1.4176 and a first—order autoregressive correlation structure with $\alpha = 0.4$. *MI* = multiple imputation; *DA* = data augmentation; *FLIC* = Firth logistic regression with intercept correction; *DL* = direct likelihood; *GLM* = logistic regression; *GEE* = generalized estimating equations; *F* = Firth; *SE* = Standard Error; *MSE* = Mean Squared Error; *NC* = percentage of non—convergence; * value larger $1x10^{10}$ or smaller $-1x10^{10}$; ‡ not calculated due to uninterpretable coefficients and *SE*.

Appendix 14. *Simulation with N = 200 Children, Dropout Probabilities 0.8 vs. 0.2 Conditional on Respiratory Infection at Previous Visit (yes vs. no) respectively.*

	Analysis of incomplete datasets				MI following DA imputation (White et al. 2010)				MI following FLIC imputation (Puhr et al. 2017)			
	DL-GLM	GEE	DL-F-GLM	F—GEE	MI—GLM	MI—GEE	MI—F—GLM	MI—F—GEE	MI—GLM	MI—GEE	MI—F—GLM	MI—F—GEE
Coefficients												
β_0	−3.1224	*	−3.0700	−3.0316	−2.7699	*	−2.7395	−2.6751	−3.0832	*	−3.0429	−3.0347
β_1	−0.3771	*	−0.3540	−0.3516	−0.3063	*	−0.2943	−0.3243	−0.3099	*	−0.2934	−0.2953
β_2	−0.0330	*	−0.0318	−0.0315	−0.0268	*	−0.0262	−0.0237	−0.0258	*	−0.0250	−0.0252
β_3	1.2208	*	1.5010	1.4983	1.2674	*	1.3492	0.9616	1.1444	*	1.3282	1.3330
SE												
β_0	0.2782	*	0.2700	0.2898	0.3065	*	0.2998	0.3350	0.2756	*	0.2685	0.2818
β_1	0.4133	*	0.3992	0.4294	0.3859	*	0.3771	0.4320	0.4017	*	0.3900	0.4084
β_2	0.0143	*	0.0138	0.0145	0.0144	*	0.0141	0.0157	0.0141	*	0.0137	0.0143
β_3	15.5949	*	0.5882	0.5652	5.4097	*	0.6196	0.6583	10.7643	*	0.5827	0.5801
Bias												
β_0	−0.3154	*	−0.2630	−0.2252	0.0370	*	0.0675	0.1319	−0.2762	*	−0.2359	−0.2277
β_1	−0.0300	*	−0.0068	−0.0046	0.0408	*	0.0528	0.0228	0.0373	*	0.0538	0.0519
β_2	−0.0046	*	−0.0034	−0.0031	0.0016	*	0.0022	0.0047	0.0026	*	0.0034	0.0032
β_3	−0.1968	*	0.0833	0.0810	−0.1502	*	−0.0685	−0.4561	−0.2732	*	−0.0895	−0.0846
MSE												
β_0	0.1986	*	0.1604	0.1480	0.1040	*	0.1017	0.1263	0.1526	*	0.1265	0.1229
β_1	0.2082	*	0.1889	0.1835	0.1233	*	0.1180	0.1163	0.1289	*	0.1211	0.1203
β_2	0.000294	*	0.000270	0.000262	0.000224	*	0.000219	0.000236	0.000185	*	0.000182	0.000181
β_3	4.3316	*	0.4090	0.3909	0.5975	*	0.2101	0.4178	1.7531	*	0.2865	0.2751
Coverage												
β_0	0.835	‡	0.871	0.896	0.925	‡	0.916	0.920	0.894	‡	0.913	0.919
β_1	0.934	‡	0.935	0.951	0.966	‡	0.963	0.992	0.975	‡	0.974	0.986
β_2	0.916	‡	0.910	0.917	0.927	‡	0.922	0.947	0.941	‡	0.936	0.945
β_3	0.944	‡	0.939	0.928	0.990	‡	0.990	0.974	0.987	‡	0.984	0.981
NC	0.1%	4.6%	0%	0.1%	0%	27.6%	0%	8.3%	0%	12.1%	0%	0.1%

Note: Results based on $n = 1{,}000$ converged simulation analyses using coefficients β_0 (intercept) = −2.8070; β_1 (gender) = −0.3472; β_2 (age) = −0.0284; β_3 (xerophthalmia) = 1.4176 and a first−order autoregressive correlation structure with $\alpha = 0.4$. *MI* = multiple imputation; *DA* = data augmentation; *FLIC* = Firth logistic regression with intercept correction; *DL* = direct likelihood; *GLM* = logistic regression; *GEE* = generalized estimating equations; *F* = Firth; *SE* = Standard Error; *MSE* = Mean Squared Error; *NC* = percentage of non−convergence; * value larger $1x10^{10}$ or smaller $-1x10^{10}$; ‡ not calculated due to uninterpretable coefficients and *SE*.

Appendix 15. *Simulation with N = 500 Children, Dropout Probabilities 0.3 vs. 0.1 Conditional on Respiratory Infection at Previous Visit (yes vs. no) respectively.*

	Analysis of incomplete datasets				MI following DA imputation (White et al. 2010)				MI following FLIC imputation (Puhr et al. 2017)			
	DL-GLM	GEE	DL-F-GLM	F—GEE	MI—GLM	MI—GEE	MI—F—GLM	MI—F—GEE	MI—GLM	MI—GEE	MI—F—GLM	MI—F—GEE
Coefficients												
β_0	−2.9121	−2.9134	−2.8970	−2.8430	−2.8090	−2.7998	−2.7965	−2.7778	−2.8985	−2.8940	−2.8850	−2.8736
β_1	−0.3579	−0.3572	−0.3512	−0.3509	−0.3402	−0.3532	−0.3348	−0.3443	−0.3274	−0.3350	−0.3217	−0.3269
β_2	−0.0301	−0.0305	−0.0297	−0.0292	−0.0290	−0.0285	−0.0287	−0.0281	−0.0270	−0.0274	−0.0267	−0.0270
β_3	1.4397	1.4907	1.4577	1.4433	1.3071	1.2422	1.3246	1.2587	1.3436	1.2428	1.3624	1.3600
SE												
β_0	0.1486	0.1836	0.1473	0.1773	0.1528	0.1853	0.1516	0.1827	0.1475	0.1762	0.1464	0.1737
β_1	0.2227	0.2774	0.2205	0.2692	0.2263	0.2795	0.2243	0.2746	0.2204	0.2662	0.2184	0.2619
β_2	0.0077	0.0093	0.0076	0.0090	0.0079	0.0094	0.0078	0.0092	0.0077	0.0091	0.0076	0.0089
β_3	0.3268	0.3145	0.3210	0.3082	0.3299	0.3360	0.3243	0.3281	0.3260	0.3352	0.3203	0.3281
Bias												
β_0	−0.1052	−0.1065	−0.0900	−0.0360	−0.0020	0.0071	0.0105	0.0292	−0.0915	−0.0871	−0.0780	−0.0667
β_1	−0.0107	−0.0100	−0.0041	−0.0037	0.0070	−0.0061	0.0124	0.0029	0.0198	0.0122	0.0255	0.0203
β_2	−0.0017	−0.0020	−0.0013	−0.0008	−0.0006	−0.0001	−0.0003	0.0003	0.0014	0.0010	0.0017	0.0014
β_3	0.0221	0.0731	0.0401	0.0256	−0.1105	−0.1754	−0.0930	−0.1589	−0.0740	−0.0738	−0.0553	−0.0576
MSE												
β_0	0.0495	0.0492	0.0456	0.0379	0.0359	0.0350	0.0352	0.0348	0.0438	0.0424	0.0407	0.0383
β_1	0.0842	0.0821	0.0819	0.0767	0.0666	0.0657	0.0652	0.0632	0.0686	0.0673	0.0672	0.0653
β_2	0.000095	0.000093	0.000092	0.000084	0.000091	0.000087	0.000090	0.000085	0.000078	0.000076	0.000077	0.000076
β_3	0.1105	0.0963	0.1052	0.0868	0.1057	0.1058	0.0972	0.0964	0.1033	0.0873	0.0954	0.0808
Coverage												
β_0	0.845	0.926	0.864	0.935	0.878	0.943	0.872	0.931	0.868	0.930	0.885	0.935
β_1	0.887	0.944	0.889	0.945	0.918	0.975	0.920	0.976	0.909	0.962	0.911	0.960
β_2	0.884	0.938	0.885	0.937	0.904	0.942	0.905	0.940	0.917	0.948	0.909	0.946
β_3	0.949	0.949	0.942	0.955	0.968	0.967	0.968	0.969	0.967	0.980	0.969	0.979
NC	0%	0%	0%	0%	0%	0.2%	0%	0%	0%	0%	0%	0%

Note: Results based on $n = 1{,}000$ converged simulation analyses using coefficients β_0 (intercept) = −2.8070; β_1 (gender) = −0.3472; β_2 (age) = −0.0284; β_3 (xerophthalmia) = 1.4176 and a first−order autoregressive correlation structure with $\alpha = 0.4$. *MI* = multiple imputation; *DA* = data augmentation; *FLIC* = Firth logistic regression with intercept correction; *DL* = direct likelihood; *GLM* = logistic regression; *GEE* = generalized estimating equations; *F* = Firth; *SE* = Standard Error; *MSE* = Mean Squared Error; *NC* = percentage of non−convergence; * value larger $1x10^{10}$ or smaller $-1x10^{10}$; ‡ not calculated due to uninterpretable coefficients and *SE*.

Appendix 16. *Simulation with N = 500 Children, Dropout Probabilities 0.4 vs. 0.1 Conditional on Respiratory Infection at Previous Visit (yes vs. no) respectively.*

	Analysis of incomplete datasets				MI following DA imputation (White et al. 2010)				MI following FLIC imputation (Puhr et al. 2017)			
	DL-GLM	GEE	DL-F-GLM	F—GEE	MI—GLM	MI—GEE	MI—F—GLM	MI—F—GEE	MI—GLM	MI—GEE	MI—F—GLM	MI—F—GEE
Coefficients												
β_0	−2.9504	−2.9505	−2.9346	−2.8692	−2.8106	−2.7977	−2.7981	−2.7755	−2.9371	*	−2.9232	−2.9116
β_1	−0.3529	−0.3517	−0.3461	−0.3471	−0.3326	−0.3488	−0.3274	−0.3399	−0.3212	*	−0.3154	−0.3206
β_2	−0.0305	−0.0309	−0.0301	−0.0295	−0.0293	−0.0287	−0.0290	−0.0282	−0.0272	*	−0.0269	−0.0272
β_3	1.4559	1.5221	1.4745	1.4584	1.3045	1.2152	1.3219	1.2329	1.3536	*	1.3729	1.3760
SE												
β_0	0.1518	0.1841	0.1504	0.1773	0.1574	0.1880	0.1561	0.1856	0.1509	*	0.1496	0.1742
β_1	0.2267	0.2776	0.2244	0.2692	0.2286	0.2810	0.2266	0.2761	0.2245	*	0.2224	0.2620
β_2	0.0079	0.0093	0.0078	0.0090	0.0081	0.0095	0.0080	0.0093	0.0078	*	0.0078	0.0090
β_3	0.3312	0.3190	0.3251	0.3133	0.3356	0.3469	0.3297	0.3384	0.3308	*	0.3248	0.3345
Bias												
β_0	−0.1434	−0.1435	−0.1276	−0.0622	−0.0036	0.0093	0.0088	0.0315	−0.1301	*	−0.1162	−0.1046
β_1	−0.0058	−0.0046	0.0010	0.0001	0.0146	−0.0016	0.0198	0.0072	0.0260	*	0.0318	0.0266
β_2	−0.0020	−0.0025	−0.0017	−0.0011	−0.0009	−0.0002	−0.0006	0.0002	0.0012	*	0.0015	0.0012
β_3	0.0383	0.1044	0.0568	0.0408	−0.1131	−0.2024	−0.0957	−0.1847	−0.0641	*	−0.0447	−0.0416
MSE												
β_0	0.0607	0.0598	0.0553	0.0427	0.0375	0.0368	0.0367	0.0367	0.0536	*	0.0493	0.0462
β_1	0.0846	0.0826	0.0823	0.0769	0.0658	0.0654	0.0645	0.0629	0.0683	*	0.0670	0.0654
β_2	0.000095	0.000096	0.000093	0.000086	0.000092	0.000088	0.000091	0.000086	0.000075	*	0.000075	0.000074
β_3	0.1146	0.1029	0.1096	0.0913	0.1029	0.1159	0.0946	0.1051	0.1039	*	0.0961	0.0826
Coverage												
β_0	0.808	0.888	0.828	0.926	0.895	0.939	0.889	0.935	0.829	‡	0.850	0.915
β_1	0.883	0.944	0.888	0.947	0.927	0.976	0.926	0.976	0.916	‡	0.914	0.958
β_2	0.882	0.939	0.883	0.943	0.900	0.950	0.901	0.949	0.924	‡	0.918	0.953
β_3	0.946	0.942	0.948	0.953	0.968	0.968	0.972	0.969	0.967	‡	0.967	0.978
NC	0%	0.1%	0%	0%	0%	0.3%	0%	0.2%	0%	0.1%	0%	0%

Note: Results based on $n = 1{,}000$ converged simulation analyses using coefficients β_0 (intercept) = −2.8070; β_1 (gender) = −0.3472; β_2 (age) = −0.0284; β_3 (xerophthalmia) = 1.4176 and a first−order autoregressive correlation structure with $\alpha = 0.4$. *MI* = multiple imputation; *DA* = data augmentation; *FLIC* = Firth logistic regression with intercept correction; *DL* = direct likelihood; *GLM* = logistic regression; *GEE* = generalized estimating equations; *F* = Firth; *SE* = Standard Error; *MSE* = Mean Squared Error; *NC* = percentage of non−convergence; * value larger $1x10^{10}$ or smaller $-1x10^{10}$; ‡ not calculated due to uninterpretable coefficients and *SE*.

Appendix 17. *Simulation with N = 500 Children, Dropout Probabilities 0.6 vs. 0.1 Conditional on Respiratory Infection at Previous Visit (yes vs. no) respectively.*

	Analysis of incomplete datasets				MI following DA imputation (White et al. 2010)				MI following FLIC imputation (Puhr et al. 2017)			
	DL-GLM	GEE	DL-F-GLM	F—GEE	MI—GLM	MI—GEE	MI—F—GLM	MI—F—GEE	MI—GLM	MI—GEE	MI—F—GLM	MI—F—GEE
Coefficients												
β_0	−3.0247	*	−3.0075	−2.9365	−2.8035	*	−2.7911	−2.7612	−3.0141	−3.0082	−2.9991	−2.9891
β_1	−0.3566	*	−0.3492	−0.3495	−0.3326	*	−0.3274	−0.3454	−0.3252	−0.3303	−0.3189	−0.3222
β_2	−0.0311	*	−0.0397	−0.0303	−0.0296	*	−0.0293	−0.0282	−0.0275	−0.0280	−0.0271	−0.0276
β_3	1.4769	*	1.4971	1.4833	1.2964	*	1.3137	1.1568	1.3631	1.3779	1.3842	1.3962
SE												
β_0	0.1582	*	0.1567	0.1780	0.1693	*	0.1679	0.1958	0.1571	0.1770	0.1557	0.1746
β_1	0.2357	*	0.2331	0.2680	0.2330	*	0.2309	0.2793	0.2333	0.2646	0.2309	0.2605
β_2	0.0082	*	0.0081	0.0091	0.0083	*	0.0083	0.0095	0.0082	0.0092	0.0081	0.0090
β_3	0.3423	*	0.3355	0.3268	0.3523	*	0.3457	0.3668	0.3414	0.3520	0.3348	0.3443
Bias												
β_0	−0.2177	*	−0.2005	−0.1295	0.0035	*	0.0159	0.0458	−0.2071	−0.2013	−0.1921	−0.1821
β_1	−0.0094	*	−0.0021	−0.0023	0.0146	*	0.0198	0.0017	0.0219	0.0168	0.0282	0.0250
β_2	−0.0027	*	−0.0023	−0.0019	−0.0012	*	−0.0009	0.0002	0.0009	0.0004	0.0013	0.0009
β_3	0.0592	*	0.0795	0.0657	−0.1212	*	−0.1040	−0.2608	−0.0545	−0.0398	−0.0334	−0.0214
MSE												
β_0	0.0845	*	0.0763	0.0533	0.0371	*	0.0365	0.0381	0.0757	0.0729	0.0689	0.0648
β_1	0.0808	*	0.0783	0.0745	0.0599	*	0.0588	0.0576	0.0636	0.0629	0.0622	0.0611
β_2	0.000099	*	0.000095	0.000088	0.000094	*	0.000092	0.000086	0.000074	0.000073	0.000073	0.000072
β_3	0.1186	*	0.1140	0.0999	0.1019	*	0.0933	0.1369	0.1026	0.0900	0.0946	0.0848
Coverage												
β_0	0.715	‡	0.748	0.896	0.912	‡	0.912	0.946	0.749	0.822	0.771	0.854
β_1	0.897	‡	0.901	0.945	0.938	‡	0.936	0.980	0.924	0.960	0.925	0.961
β_2	0.903	‡	0.903	0.944	0.915	‡	0.910	0.949	0.933	0.962	0.933	0.959
β_3	0.949	‡	0.947	0.952	0.979	‡	0.981	0.967	0.970	0.982	0.973	0.981
NC	0%	0.3%	0%	0%	0%	1%	0%	0.5%	0%	0.4%	0%	0%

Note: Results based on $n = 1{,}000$ converged simulation analyses using coefficients β_0 (intercept) = −2.8070; β_1 (gender) = −0.3472; β_2 (age) = −0.0284; β_3 (xerophthalmia) = 1.4176 and a first−order autoregressive correlation structure with $\alpha = 0.4$. *MI* = multiple imputation; *DA* = data augmentation; *FLIC* = Firth logistic regression with intercept correction; *DL* = direct likelihood; *GLM* = logistic regression; *GEE* = generalized estimating equations; *F* = Firth; *SE* = Standard Error; *MSE* = Mean Squared Error; *NC* = percentage of non−convergence; * value larger $1x10^{10}$ or smaller $-1x10^{10}$; ‡ not calculated due to uninterpretable coefficients and *SE*.

***Appendix 18**. Simulation with N = 500 Children, Dropout Probabilities 0.8 vs. 0.1 Conditional on Respiratory Infection at Previous Visit (yes vs. no) respectively.*

	Analysis of incomplete datasets				MI following DA imputation (White et al. 2010)				MI following FLIC imputation (Puhr et al. 2017)			
	DL-GLM	GEE	DL-F-GLM	F—GEE	MI—GLM	MI—GEE	MI—F—GLM	MI—F—GEE	MI—GLM	MI—GEE	MI—F—GLM	MI—F—GEE
Coefficients												
β_0	−3.1130	*	−3.0939	−3.0439	−2.8032	*	−2.7908	−2.7432	−3.1078	−3.1032	−3.0914	−3.0844
β_1	−0.3558	*	−0.3478	−0.3484	−0.3261	*	−0.3210	−0.3515	−0.3197	−0.3230	−0.3129	−0.3148
β_2	−0.0320	*	−0.0316	−0.0314	−0.0300	*	−0.0297	−0.0278	−0.0280	−0.0283	−0.0276	−0.0279
β_3	1.4987	*	1.5212	1.5129	1.3405	*	1.3558	1.0566	1.3735	1.3865	1.3969	1.4083
SE												
β_0	0.1661	*	0.1644	0.1782	0.1975	*	0.1957	0.2207	0.1655	0.1771	0.1639	0.1748
β_1	0.2463	*	0.2433	0.2655	0.2377	*	0.2356	0.2816	0.2439	0.2614	0.2411	0.2574
β_2	0.0086	*	0.0084	0.0091	0.0086	*	0.0086	0.0099	0.0086	0.0092	0.0085	0.0091
β_3	0.3557	*	0.3479	0.3454	0.3993	*	0.3912	0.4236	0.3551	0.3628	0.3475	0.3542
Bias												
β_0	−0.3060	*	−0.2869	−0.2369	0.0038	*	0.0162	0.0638	−0.3008	−0.2962	−0.2844	−0.2774
β_1	−0.0086	*	−0.0007	−0.0013	0.0210	*	0.0261	−0.0044	0.0275	0.0241	0.0342	0.0323
β_2	−0.0036	*	−0.0032	−0.0029	−0.0016	*	−0.0013	0.0006	0.0005	0.0001	0.0008	0.0005
β_3	0.0811	*	0.1035	0.0952	−0.0772	*	−0.0618	−0.3610	−0.0441	−0.0311	−0.0207	−0.0093
MSE												
β_0	0.1306	*	0.1182	0.0937	0.0450	*	0.0442	0.0487	0.1226	0.1196	0.1121	0.1081
β_1	0.0797	*	0.0771	0.0749	0.0592	*	0.0582	0.0560	0.0612	0.0604	0.0599	0.0592
β_2	0.000105	*	0.000100	0.000096	0.000097	*	0.000095	0.000090	0.000072	0.000072	0.000071	0.000071
β_3	0.1347	*	0.1298	0.1222	0.0902	*	0.0902	0.2127	0.1087	0.1028	0.1001	0.0961
Coverage												
β_0	0.562	‡	0.605	0.745	0.912	‡	0.916	0.936	0.580	0.641	0.612	0.677
β_1	0.922	‡	0.920	0.951	0.948	‡	0.947	0.981	0.953	0.966	0.953	0.966
β_2	0.908	‡	0.909	0.932	0.920	‡	0.920	0.960	0.953	0.971	0.952	0.969
β_3	0.944	‡	0.945	0.951	0.990	‡	0.991	0.952	0.974	0.978	0.973	0.975
NC	0%	0.5%	0%	0%	0%	4.9%	0%	2.1%	0%	0.5%	0%	0%

Note: Results based on $n = 1{,}000$ converged simulation analyses using coefficients β_0 (intercept) = −2.8070; β_1 (gender) = −0.3472; β_2 (age) = −0.0284; β_3 (xerophthalmia) = 1.4176 and a first−order autoregressive correlation structure with $\alpha = 0.4$. *MI* = multiple imputation; *DA* = data augmentation; *FLIC* = Firth logistic regression with intercept correction; *DL* = direct likelihood; *GLM* = logistic regression; *GEE* = generalized estimating equations; *F* = Firth; *SE* = Standard Error; *MSE* = Mean Squared Error; *NC* = percentage of non−convergence; * value larger $1x10^{10}$ or smaller $-1x10^{10}$; ‡ not calculated due to uninterpretable coefficients and *SE*.

Appendix 19. *Simulation with N = 500 Children, Dropout Probabilities 0.6 vs. 0.2 Conditional on Respiratory Infection at Previous Visit (yes vs. no) respectively.*

	Analysis of incomplete datasets				MI following DA imputation (White et al. 2010)				MI following FLIC imputation (Puhr et al. 2017)			
	DL-GLM	GEE	DL-F-GLM	F—GEE	MI—GLM	MI—GEE	MI—F—GLM	MI—F—GEE	MI—GLM	MI—GEE	MI—F—GLM	MI—F—GEE
Coefficients												
β_0	−3.0087	*	−2.9902	−2.9290	−2.7863	*	−2.7742	−2.7433	−2.9875	−2.9830	−2.9729	−2.9642
β_1	−0.3549	*	−0.3471	−0.3473	−0.3113	*	−0.3064	−0.3245	−0.3023	−0.3081	−0.2964	−0.3008
β_2	−0.0312	*	−0.0308	−0.0303	−0.0288	*	−0.0285	−0.0274	−0.0253	−0.0258	−0.0250	−0.0254
β_3	1.4659	*	1.4879	1.4771	1.2240	*	1.2427	1.0768	1.2876	1.2962	1.3105	1.3174
SE												
β_0	0.1638	*	0.1621	0.1844	0.1766	*	0.1751	0.2007	0.1617	0.1791	0.1602	0.1768
β_1	0.2433	*	0.2404	0.2770	0.2399	*	0.2378	0.2821	0.2372	0.2654	0.2348	0.2614
β_2	0.0085	*	0.0083	0.0093	0.0088	*	0.0088	0.0099	0.0083	0.0093	0.0083	0.0091
β_3	0.3557	*	0.3481	0.3407	0.3653	*	0.3581	0.3785	0.3534	0.3638	0.3461	0.3550
Bias												
β_0	−0.2017	*	−0.1832	−0.1220	0.0206	*	0.0328	0.0636	−0.1805	−0.1760	−0.1659	−0.1572
β_1	−0.0077	*	0.0001	0.0001	0.0359	*	0.0408	0.0227	0.0449	0.0391	0.0507	0.0464
β_2	−0.0028	*	−0.0024	−0.0019	−0.0004	*	−0.0001	0.0010	0.0031	0.0026	0.0034	0.0030
β_3	0.0483	*	0.0702	0.0595	−0.1936	*	−0.1749	−0.3409	−0.1300	−0.1214	−0.1071	−0.1002
MSE												
β_0	0.0801	*	0.0718	0.0533	0.0376	*	0.0375	0.0399	0.0652	0.0634	0.0593	0.0562
β_1	0.0837	*	0.0810	0.0772	0.0561	*	0.0553	0.0535	0.0572	0.0567	0.0563	0.0555
β_2	0.000105	*	0.000101	0.000094	0.000095	*	0.000094	0.000090	0.000077	0.000075	0.000077	0.000075
β_3	0.1270	*	0.1209	0.1066	0.1211	*	0.1096	0.1820	0.1152	0.1033	0.1035	0.0931
Coverage												
β_0	0.759	‡	0.790	0.909	0.928	‡	0.928	0.944	0.813	0.861	0.825	0.882
β_1	0.904	‡	0.903	0.951	0.951	‡	0.951	0.984	0.950	0.975	0.951	0.974
β_2	0.898	‡	0.896	0.934	0.913	‡	0.915	0.943	0.935	0.956	0.931	0.953
β_3	0.950	‡	0.951	0.951	0.979	‡	0.983	0.950	0.979	0.987	0.978	0.988
NC	0%	0.1%	0%	0%	0%	0.9%	0%	0.5%	0%	0%	0%	0%

Note: Results based on $n = 1{,}000$ converged simulation analyses using coefficients β_0 (intercept) = −2.8070; β_1 (gender) = −0.3472; β_2 (age) = −0.0284; β_3 (xerophthalmia) = 1.4176 and a first—order autoregressive correlation structure with $\alpha = 0.4$. *MI* = multiple imputation; *DA* = data augmentation; *FLIC* = Firth logistic regression with intercept correction; *DL* = direct likelihood; *GLM* = logistic regression; *GEE* = generalized estimating equations; *F* = Firth; *SE* = Standard Error; *MSE* = Mean Squared Error; *NC* = percentage of non—convergence; * value larger $1x10^{10}$ or smaller $-1x10^{10}$; ‡ not calculated due to uninterpretable coefficients and *SE*.

Appendix 20. *Simulation with N = 500 Children, Dropout Probabilities 0.8 vs. 0.2 Conditional on Respiratory Infection at Previous Visit (yes vs. no) respectively.*

	Analysis of incomplete datasets				MI following DA imputation (White et al. 2010)				MI following FLIC imputation (Puhr et al. 2017)			
	DL-GLM	GEE	DL-F-GLM	F—GEE	MI—GLM	MI—GEE	MI—F—GLM	MI—F—GEE	MI—GLM	MI—GEE	MI—F—GLM	MI—F—GEE
Coefficients												
β_0	−3.0969	*	−3.0765	−3.0317	−2.7783	*	−2.7662	−2.7160	−3.0847	*	−3.0686	−3.0625
β_1	−0.3563	*	−0.3479	−0.3483	−0.3110	*	−0.3062	−0.3378	−0.2968	*	−0.2905	−0.2931
β_2	−0.0321	*	−0.0317	−0.0314	−0.0290	*	−0.0287	−0.0267	−0.0259	*	−0.0255	−0.0259
β_3	1.4899	*	1.5144	1.5073	1.2640	*	1.2809	0.9657	1.3008	*	1.3264	1.3350
SE												
β_0	0.1720	*	0.1701	0.1844	0.2043	*	0.2024	0.2265	0.1704	*	0.1687	0.1780
β_1	0.2544	*	0.2511	0.2742	0.2426	*	0.2405	0.2834	0.2484	*	0.2456	0.2601
β_2	0.0088	*	0.0087	0.0094	0.0092	*	0.0091	0.0103	0.0088	*	0.0087	0.0092
β_3	0.3701	*	0.3614	0.3596	0.4072	*	0.3986	0.4352	0.3680	*	0.3597	0.3652
Bias												
β_0	−0.2899	*	−0.2695	−0.2248	0.0287	*	0.0408	0.0910	−0.2777	*	−0.2616	−0.2555
β_1	−0.0092	*	−0.0007	−0.0012	0.0362	*	0.0410	0.0094	0.0504	*	0.0567	0.0541
β_2	−0.0037	*	−0.0032	−0.0030	−0.0005	*	−0.0003	0.0017	0.0026	*	0.0029	0.0025
β_3	0.0723	*	0.0967	0.0896	−0.1536	*	−0.1367	−0.4519	−0.1168	*	−0.0912	−0.0826
MSE												
β_0	0.1242	*	0.1115	0.0909	0.0458	*	0.0456	0.0547	0.1101	*	0.1005	0.0973
β_1	0.0861	*	0.0831	0.0806	0.0559	*	0.0552	0.0532	0.0571	*	0.0562	0.0556
β_2	0.000113	*	0.000108	0.000104	0.000100	*	0.000099	0.000099	0.000074	*	0.000074	0.000073
β_3	0.1480	*	0.1411	0.1334	0.1133	*	0.1035	0.2852	0.1205	*	0.1075	0.1018
Coverage												
β_0	0.629	‡	0.657	0.778	0.935	‡	0.930	0.931	0.653	‡	0.682	0.732
β_1	0.925	‡	0.924	0.947	0.964	‡	0.963	0.990	0.960	‡	0.959	0.974
β_2	0.908	‡	0.912	0.937	0.925	‡	0.924	0.955	0.943	‡	0.939	0.955
β_3	0.945	‡	0.941	0.946	0.994	‡	0.994	0.931	0.975	‡	0.978	0.981
NC	0%	0.3%	0%	0%	0%	0.5%	0%	0.2%	0%	0.5%	0%	0%

Note: Results based on $n = 1{,}000$ converged simulation analyses using coefficients β_0 (intercept) = −2.8070; β_1 (gender) = −0.3472; β_2 (age) = −0.0284; β_3 (xerophthalmia) = 1.4176 and a first—order autoregressive correlation structure with $\alpha = 0.4$. *MI* = multiple imputation; *DA* = data augmentation; *FLIC* = Firth logistic regression with intercept correction; *DL* = direct likelihood; *GLM* = logistic regression; *GEE* = generalized estimating equations; *F* = Firth; *SE* = Standard Error; *MSE* = Mean Squared Error; *NC* = percentage of non—convergence; * value larger $1x10^{10}$ or smaller $-1x10^{10}$; ‡ not calculated due to uninterpretable coefficients and *SE*.

Appendix 21. *Simulation with N = 50, 100, 200 and 500 Children Without any Dropout (Complete Datasets).*

	N=50				N=100				N=200				N=500			
	GLM	GEE	F—GLM	F—GEE	GLM	GEE	F—GLM	F—GEE	GLM	GEE	F—GLM	F—GEE	GLM	GEE	F—GLM	F—GEE
Coefficients																
β_0	−3.0892	*	−2.9174	−2.8452	−2.9389	*	−2.8628	−2.8289	−2.8491	*	−2.8162	−2.7999	−2.8299	−2.8303	−2.8172	−2.8077
β_1	−1.2140	*	−0.3900	−0.3670	−0.4792	*	−0.3716	−0.3565	−0.3874	*	−0.3711	−0.3630	−0.3553	−0.3582	−0.3496	−0.3492
β_2	−0.0341	*	−0.0300	−0.0283	−0.0312	*	−0.0291	−0.0286	−0.0298	*	−0.0291	−0.0286	−0.0292	−0.0291	−0.0289	−0.0286
β_3	−1.8076	*	1.4862	1.4610	0.6361	*	1.4465	1.4454	1.2585	*	1.3875	1.4025	1.4181	1.4193	1.4332	1.4301
SE																
β_0	2.4520	*	0.4689	0.5088	0.3287	*	0.3142	0.3778	0.2197	*	0.2156	0.2697	0.1361	0.1743	0.1351	0.1717
β_1	102.8803	*	0.7169	0.7427	7.3653	*	0.4786	0.5661	0.3337	*	0.3262	0.4084	0.2051	0.2667	0.2034	0.2620
β_2	0.0264	*	0.0236	0.0242	0.0170	*	0.0161	0.0184	0.0115	*	0.0112	0.0134	0.0071	0.0089	0.0070	0.0087
β_3	403.5412	*	1.0262	0.8096	62.3698	*	0.7097	0.6116	5.2479	*	0.4890	0.4517	0.3023	0.2913	0.2977	0.2865
Bias																
β_0	−0.2822	*	−0.1104	−0.0382	−0.1319	*	−0.0558	−0.0219	−0.0421	*	−0.0092	0.0071	−0.0229	−0.0234	−0.0102	−0.0007
β_1	−0.8669	*	−0.0428	−0.0198	−0.1320	*	−0.0244	−0.0093	−0.0403	*	−0.0240	−0.0158	−0.0081	−0.0110	−0.0025	−0.0020
β_2	−0.0057	*	−0.0016	0.0001	−0.0028	*	−0.0007	−0.0002	−0.0014	*	−0.0006	−0.0002	−0.0008	−0.0007	−0.0005	−0.0002
β_3	−3.2253	*	0.0686	0.0434	−0.7816	*	0.0289	−0.0278	−0.1592	*	−0.0301	0.0152	0.0005	0.0017	0.0155	0.0125
MSE																
β_0	0.8528	*	0.3659	0.3433	0.2224	*	0.1794	0.1719	0.0855	*	0.0791	0.0767	0.0343	0.0339	0.0332	0.0323
β_1	16.9726	*	0.9085	0.7803	1.9638	*	0.4016	0.3649	0.1953	*	0.1824	0.1721	0.0762	0.0746	0.0744	0.0716
β_2	0.0013	*	0.0010	0.0009	0.000516	*	0.000461	0.000434	0.000229	*	0.000218	0.000206	0.000083	0.000081	0.000082	0.000079
β_3	56.7331	*	0.9511	0.7771	14.1693	*	0.6284	0.4974	1.9146	*	0.3166	0.2465	0.0958	0.0762	0.0913	0.0726
Coverage																
β_0	0.923	‡	0.912	0.919	9.871	‡	0.873	0.942	0.875	‡	0.873	0.945	0.865	0.943	0.869	0.941
β_1	0.928	‡	0.919	0.906	0.891	‡	0.891	0.943	0.880	‡	0.877	0.956	0.861	0.951	0.861	0.951
β_2	0.887	‡	0.873	0.905	0.879	‡	0.876	0.923	0.873	‡	0.873	0.931	0.873	0.943	0.869	0.938
β_3	0.960	‡	0.960	0.897	0.952	‡	0.948	0.896	0.942	‡	0.940	0.941	0.953	0.960	0.952	0.962
NC	1.2%	9.8%	0%	1.2%	1.1%	6.7%	0%	1.1%	0.1%	1.6%	0%	0.1%	0%	0%	0%	0%

Note: Results based on 1,000 converged simulation analyses using coefficients β_0 (intercept) = −2.8070; β_1 (gender) = −0.3472; β_2 (age) = −0.0284; β_3 (xerophthalmia) = 1.4176 and a first−order autoregressive correlation structure with $\alpha = 0.4$. *GLM* = logistic regression; *GEE* = generalized estimating equations; *F* = Firth; *SE* = Standard Error; *MSE* = Mean Squared Error; *NC* = percentage of non−convergence; * value larger $1x10^{10}$ or smaller $-1x10^{10}$; ‡ not calculated due to uninterpretable coefficients and *SE*.

R-Code

Generation of simulated datasets (in analogy to Mondol & Rahman, 2019)

```
#Algorithm by Qaqish (2003) for generating correlated binary values
#install_github("cran/binarySimCLF")
library(binarySimCLF)
#Mondol & Rahman (2019)
#install_github("heogden/geefirthr")
library(geefirthr)
#Firth-logistic regression including FLIC
library(logistf)
#Generalized estimating equations (GEE)
library(geeM)
#data from Sommer et al. (1983) are contained in gamlss
library(gamlss)
#multiple imputation
library(mice)
#wide to long and long to wide
library(reshape)
#to generate age variable
library(truncnorm)
#seed
seed <- 1234
set.seed(seed)


#Firth GEE full model using data from Sommer et al. (1983) within gamlss
#time = binary response variable identifying the presence of respiratory infection
#xero = binary response variable identifying the presence of xerophthalmia
#female = gender factor with levels 0 is male 1 is female
#age = the age in months (centered around 36)
result <- geefirth(time ~ female + age + xero, id=id,
                   corstr="ar1",
                   data=respInf_complete,
                   est_dispersion = FALSE)

#beta coefficients collected for simulation
betaCoef <- c(result$coefficients[1][1,],
              result$coefficients[1][2,],
              result$coefficients[1][3,],
              result$coefficients[1][4,])

#Firth GEE modeling xerophthalmia for generating xerophthalmia values in simulations
xero <- geefirth(xero ~ female + age, id=id,
                   data=respInf_complete,
                   corstr = "ar1",
                   est_dispersion = FALSE)
coefXero <- c(xero$coefficients[1][1,],
              xero$coefficients[1][2,],
              xero$coefficients[1][3,])

#parameters for simulated datasets
#number of clusters
nc <- 50  # 100, 200, 500
#intra-child correlation for response variable respiratory infection
rho <- 0.4
#overall mean proportion to generate binary predictor X = gender according to Sommer et al.
p <- 0.45
#number of repeated measurement time points
cl.size <- 4
#number of simulations
nSim <- 1000
```

```
#the following algorithm was adopted from Mondol and Rahman (2019)
#data generation
generateDataset <- function(beta, nc, cl.size, p,  rho, coefXero) {
  #generate repeated measurement points
  cl.size <- round(runif(nc,  cl.size[1], cl.size[length(cl.size)]))
  #sum of individual subjects
  N <- sum(cl.size)
  j <- cl.size
  #generate binary predictor X = gender
  X1i <- rep(rbinom(nc, 1, p), times=cl.size)
  #generate age variable according to Sommer et al. (age centered at 36 month),
  #truncated to ensure no values below 4 month and beyond 6 years
  age <- rtruncnorm(nc, mean = -12, sd = 19, a = -32, b = 35)
  age <- rep(round(age), times = cl.size)
  #define age as "repeated measurement-point"
  if(length(unique(cl.size)) == 1)  tij <- c(sapply(j, function(x) 3*(1:x)))
  if(length(unique(cl.size)) > 1)  tij <- unlist(sapply(j, function(x) 3*(1:x)))
  time <- tij
  age <- tij + age
  #intra-child correlation for predictor xerophthalmia
  rho_xero <- 0.3
  #Process of generating correlated values for predictor xerophthalmia
  #design matrix X
  intercept_xero <- rep(1, N)
  dat_xero <- cbind(intercept_xero, X1i, age)
  #linear predictor
  a_xero <- exp(apply(dat_xero, 1, function(r_xero) r_xero %*% coefXero))
  #modeled probability
  pi_xero <- a_xero/(1+a_xero)
  #set child id
  id <- rep(1:nc, times=j)
  #build data frame
  dd_xero <- data.frame(id, pi_xero)
  d.pi_xero <- split(dd_xero, id)
  #create one modeled (ar-1) correlation matrix for each child
  R_xero <- lapply(cl.size, function(x_xero) ar1(x_xero, rho_xero))
  #transform correlation matrix of each child into var-cov-matrix using the generated
  estimated proportion pi
  #diagonal = pi*(1-pi), offdiagonal according to ar1 correlation matrix = model based
  V_xero <- list()
  for(i in 1:nc){
    V_xero[[i]] <- cor2var(R_xero[[i]], d.pi_xero[[i]]$pi_xero)
  }
  #Checks CLF compatibility
  B_xero <- lapply(V_xero, function(Vi_xero)  allReg(Vi_xero));
  clf.compat_xero <- NULL
  for(i in 1:nc) {
    clf.compat_xero[i] <- blrchk(d.pi_xero[[i]][,2], V_xero[[i]])
  }
  #generate correlated xerophthalmia values (Qaqish)
  xero <- list()
  for(i in 1:nc) {
    if(clf.compat_xero[i]) {
      xero[[i]] = mbsclf(1, d.pi_xero[[i]][,2], B_xero[[i]])$y
    }
    if(clf.compat_xero[i]==FALSE) {
      xero[[i]] = rep(NA, cl.size[i])
    }
  }
  X2i <- unlist(xero)
```

```
  #Process of generating correlated response values Y
  #design matrix X
  intercept <- rep(1, N)
  dat <- cbind(intercept, X1i, obstime=age, X2i)
  #linear predictor
  a <- exp(apply(dat, 1, function(r) r %*% beta))
  #modeled probability
  pi <- a/(1+a)
  #set child id
  id <- rep(1:nc, times=j)
  #build data frame
  dd <- data.frame(id, pi)
  d.pi <- split(dd, id)
  #create one modeled ar1 correlation matrix for each child
  R <- lapply(cl.size, function(x) ar1(x, rho))
  #transform correlation matrix of each child into var-cov-matrix using the generated
  estimated proportion pi
  #diagonal = pi*(1-pi), offdiagonal according to ar1 correlation matrix = model based
  V <- list()
  for(i in 1:nc){
    V[[i]] <- cor2var(R[[i]], d.pi[[i]]$pi)
  }
  #Checks CLF compatibility
  B <- lapply(V, function(Vi)  allReg(Vi));
  clf.compat <- NULL
  for(i in 1:nc) {
    clf.compat[i] <- blrchk(d.pi[[i]][,2], V[[i]])
  }
  #generate correlated y values (Qaqish)
  y <- list()
  for(i in 1:nc) {
    if(clf.compat[i]) {
      y[[i]] = mbsclf(1, d.pi[[i]][,2], B[[i]])$y
    }
    if(clf.compat[i]==FALSE) {
      y[[i]] = rep(NA, cl.size[i])
    }
  }
  yij <- unlist(y)
  data <- data.frame(id=id, y= yij, x1=X1i, x2 = X2i, obstime=age, time)
  return(data)
}
```

Amputing values according to MAR mechanism

```
#number of imputations for each missing value
m <- 10

proportion <- prop[r]

for(k in 1:nSim) {
  #couple seed with index to enable easier reproducibility of results
  set.seed(seed * 2 * k)
  #get simulated dataset
  generated_data <- get(dataset_names_list[k, n])

  tryCatch({
    #change to wide format
    d <- longToWide(generated_data)
    #ampute values according to MAR
    #conditioning on previous infections
    for(i in 1:length(d[,1])) {
      if(d$y.3[i] == 1) {
        if(runif(1, 0, 1) < 0.60) {
          d$x2.6[i] <- NA
          d$y.6[i] <- NA
          d$x2.9[i] <- NA
          d$y.9[i] <- NA
          d$x2.12[i] <- NA
          d$y.12[i] <- NA
        }
      }
      else {
        if(d$y.6[i] == 1) {
          if(runif(1, 0, 1) < 0.60) {
            d$x2.9[i] <- NA
            d$y.9[i] <- NA
            d$x2.12[i] <- NA
            d$y.12[i] <- NA
          }
        }
        else {
          if(d$y.9[i] == 1) {
            if(runif(1, 0, 1) < 0.60) {
              d$x2.12[i] <- NA
              d$y.12[i] <- NA
            }
          }
        }
      }

      if(!is.na(d$y.3[i]) & d$y.3[i] == 0) {
        if(runif(1, 0, 1) < 0.20) {
          d$x2.6[i] <- NA
          d$y.6[i] <- NA
          d$x2.9[i] <- NA
          d$y.9[i] <- NA
          d$x2.12[i] <- NA
          d$y.12[i] <- NA
        }
      }
      else {
        if(!is.na(d$y.6[i]) & d$y.6[i] == 0) {
          if(runif(1, 0, 1) < 0.20) {
            d$x2.9[i] <- NA
            d$y.9[i] <- NA
            d$x2.12[i] <- NA
```

```
            d$y.12[i] <- NA
          }
        }
        else {
          if(!is.na(d$y.9[i]) & d$y.9[i] == 0) {
            if(runif(1, 0, 1) < 0.20) {
              d$x2.12[i] <- NA
              d$y.12[i] <- NA
            }
          }
        }
      }
    }

    #conditional specifications for FCS algorithn
    formula <- c(x2.6 ~ y.3 + x2.3  + obstime.3,
                 x2.9 ~ y.6 + x2.6  + obstime.6,
                 x2.12 ~ y.9 +x2.9 + obstime.9,
                 y.6  ~ y.3 + x2.3  + obstime.3,
                 y.9  ~ y.6 + x2.6  + obstime.6,
                 y.12 ~ y.9 + x2.9 + obstime.9)

    #multiple imputations of missing outcome values
    impute <- mice(d,
                   method = "logreg",
                   formulas = formula,
                   m = m, printFlag = F)

    #save dataset
    #[ ... ]

  }, error = function(e) {
    print(e)
  })
}
```

Method of analyses for each of m=10 imputed dataset or with incomplete data

```
#logistic regression
resultGLM <- glm(y ~ x1 + obstime + x2,
                   family=binomial(link="logit"),
                   data=d)

#standard GEE
d2 <- d[order(d$id),]
resultGEE <- geem(y ~ x1 + obstime + x2,
                   id=id,
                   data=d2, corstr = "ar1",
                   family = binomial(link = "logit"),
                   scale.fix = TRUE)

#Firth-Logistic regression
resultFirthGLM <- logistf(y ~ x1 + obstime + x2,
                   family=binomial(link="logit"),
                   data=d,
                   pl = FALSE)

#Firth-GEE
resultFirthGEE <- geefirth(y ~ x1 + obstime + x2,
                   id=id,
                   data=d,
                   corstr = "ar1",
                   est_dispersion = FALSE)
```

Function called by MICE logreg, for FLIC imputation

```
#Function which is called within mice.impute.logreg in MICE package instead of DA algorithm
callMethodOutside <- function(y, ry, x, wy = NULL) {
  ry <- unlist(ry)
  y <- unlist(y)
  y <- y[ry]

  if(length(unique(as.numeric(y)))==1) {
    vec <- y
  }
  else {
    x <- x[ry, , drop=FALSE]
    frame <- data.frame(y,x)
    #Firth Logistic regression with FLIC option from logistf package
    fit <- logistf(y ~ x, flic = TRUE, data=frame)
    beta <- coef(fit)
    #draw random value for the parameter from approximation of normal distribution
    rv <- t(chol((summary(fit)$var)))
    beta.star <- beta + rv %*% rnorm(ncol(rv))
    #draw imputations
    x <- cbind(intercept = 1,x)
    p <- 1 / (1 + exp(-(as.matrix(x) %*% beta.star)))
    vec <- (runif(nrow(p)) <= p)
    for(i in 1:length(vec)) {
      vec[i] <- ifelse(vec[i] == TRUE, 1, 0)
    }
    if(is.factor(y)) {
      vec <- as.factor(vec)
    }
  }
  vec
}
```