A stochastic model to analyse risk factors for emesis in multi-cycle chemotherapies

Werner Vach[1], *Andreas du Bois*[2] *and Kirstin Schulz*[1]

Summary

We propose a stochastic model to analyse risk factors for emesis in multi-cycle chemotherapies, which allows to describe the risk of a factor by a single parameter. This model is a hybrid between a random intercept model and a transition model and it is motivated by some medical background knowledge with respect to frequency and course of emesis in cancer patients. We consider maximum likelihood estimation of the parameters of the model and additionally efficient estimation of the marginal risk in the first cycle. Investigations of finite sample properties in a simulation study indicate some difficulties in the estimation procedure. A comparison with alternative regression models for binary longitudinal data reveals that in this setting these models allow to define a time-constant regression coefficient only in a somewhat arbitrary manner; however, the impact of this problem seems to be not too serious in practice.

[1]Center for Data Analysis and Model Building, University of Freiburg & Institute of Medical Biometry and Medical Informatics, University of Freiburg

[2]St. Vincentius Hospital, Frauenklinik, Karlsruhe

1 Introduction

Nausea and vomiting are among the most distressing side effects of cancer chemotherapy. They have a strong negative impact on quality of life[1] and even may cause patients to refuse further treatment. The progress in developing new antiemetic drugs, especially of new serotonin receptor (5-HT3) antagonists[2] has increased the interest in the adequate planning, conduct and analysis of antiemetic trials[3,4]. Since many chemotherapy regimen consist of multiple cycles, data from antiemetic research consists typically of several measurements of an outcome variable for each patient. This immanent problem was widely ignored for a long time by regarding each cycle of a patient as the basic unit of the analysis. This neglects both the dependencies among the measurements from different cycles of each patient as well as a possible change of the parameter of interest over time. In recent years more attention has been paid to the adequate analysis of this special type of longitudinal data, especially with respect to the persistence of the efficacy of antiemetic drugs[5–8], but also with respect to the evaluation of risk factors[6,9].

For cancer patients treated with chemotherapy optimal supportive care with respect to protection from emesis requires knowledge of the individual risk and hence of the major factors determining it. Besides the factors sex, age, performance status, chemotherapeutic regimen and previous experience with chemotherapy or emesis little is known and further research is necessary. This requires an adequate statistical framework to assess the influence of risk factors. The longitudinal character of data from antiemetic research makes this a nontrivial problem, because it is necessary to take into account several circumstances typical for this type of data. This starts with two different sources for the dependence of the outcome variables, i.e. occurence of emesis, between different cycles of one patient, which reflect two well known properties. First, there exists a strong individual predisposition of patients with respect to emesis: some patients seem to be nearly robust to the emetogeniaty of chemotherapy and suffer never or only to a very mild degree from emesis, whereas other patients react heavily right from the first cycle. Second, occurence of emesis in one cycle increases the risk of emesis in subsequent cycles. This may be due to a change of the sensitivity of the pathophysiological mechanism involved or to an increased fear of the patient. This psychological explanation is supported by the occurence of anticipatory emesis, i.e. the occurence of vomiting or nausea prior to start of the next cycle, which is especially observed in patients who suffered from severe emesis in previous cycles.

In Section 2 we present a stochastic model to describe the influence of risk factors which takes these two sources into account. Additionally we will allow that the efficacy of antiemetic treatment may change over time. Section 3 explains the necessary steps to estimate the parameters of this model by the maximum likelihood principle. Further the use of this estimation procedure is justified if we have different numbers of cycles for different patients due to drop outs, which is a usual additional complication in antiemetic research. Section 4 considers a procedure to achieve estimates comparable to those of other studies using only the first cycle of each patient. Properties in finite samples of the estimators proposed are investigated in Section 5. Some difficulties with alternative approaches are discussed in Section 6. Section 7 presents an application of the estimation procedures proposed; this is complemented in Section 8 by a sensitivity analysis with respect to some assumptions made in the development of our model. A discussion finishes the paper.

2 The stochastic model

Suppose that we have collected data from n patients undergoing several cycles of a chemotherapy. The data consists of to risk factors known prior to the start of the chemotherapy and the occurence of emesis in each cycle. We denote the occurrence of emesis in the j-th cycle of the i-th patient by the binary outcome variable Y_{ij}. The values of the risk factors of the i-th patient are summarised in the covariate vector X_i. To relate the probability of emesis in each cycle to the risk factor values we assume that there exist a latent variable L_{ij} and biological thresholds c_j such that emesis occurs if and only if the latent variable exceeds this threshold, i.e. $Y_{ij} = 1 \Leftrightarrow L_{ij} \geq c_j$. This latent variable L_{ij} is now decomposed into several components reflecting different contributions to the occurence of emesis. The decomposition reads

$$L_{ij} = L_i + \alpha' X_i + \beta Y_{ij-1} + \beta\gamma Y_{ij-2} + \beta\gamma^2 Y_{ij-3} + ... + \beta\gamma^{j-2} Y_{i1} + \epsilon_{ij} \ .$$

The first component L_i reflects the individual predisposition of the patient. The second component $\alpha' X_i$ describes the influence of the risk factors. The occurence of vomiting in one cycle increases the latent variable by the amount β in the next cycle, and the influence on subsequent cycles decrease with an exponential decay determined by the parameter γ with $\gamma \geq 0$. For example, occurence of emesis in the second cycle increases the latent variable by the amount of β in the third cycle, by $\beta\gamma$ in the fourth cycle and by $\beta\gamma^2$ in the fifth cycle. Of course we expect that $\gamma \leq 1$. The final term ϵ_{ij} describes the remaining random fluctuation of the latent variable.

If we assume that ϵ_{ij} follows a standard logistic distribution with distribution function

$\Lambda(\cdot) = 1/(1+exp(-\cdot))$ we have

$$\begin{aligned}
& P(Y_{ij}=1|X_i=x_i, L_i=l, Y_{i1}=y_{i1},...,Y_{ij-1}=y_{ij-1}) \\
= \; & P(l+\alpha' x_i + \beta \sum_{j'<j} y_{ij'}\gamma^{j-j'-1} + \epsilon_{ij} \geq c_j) \\
= \; & P(-\epsilon_{ij} \leq -c_j + l + \alpha' x_i + \beta \sum_{j'<j} y_{ij'}\gamma^{j-j'-1}) \\
= \; & \Lambda(-c_j + l + \alpha' x_i + \beta \sum_{j'<j} y_{ij'}\gamma^{j-j'-1}) =: A_{ij}(l)
\end{aligned}$$

such that the components of α can be interpreted as log-odds-ratios for the risk of emesis, adjusted for the individual predisposition and the influence of emesis in previous cycles.

To keep this model simple, we assume a further restriction of the thresholds c_j. It seems to be not adequate to assume a constant threshold, especially because the efficacy of antiemetic treatment may change over time. However, a reasonable approximation is a linear change over time, i.e.

$$c_j = c + (j-1)\Delta \; .$$

3 Maximum likelihood estimation

To estimate the parameters of the model introduced in the last section by the maximum likelihood (ML) principle, we have to consider the law of the joint observations from each patient. If T is the number of cycles of the chemotherapy considered, this law is given by

$$P(Y_{i1}=y_{i1},...,Y_{iT}=y_{iT}) = \int \prod_{j=1}^{T} A_{ij}(l)^{y_{ij}}(1-A_{ij}(l))^{1-y_{ij}} dF(l) \; .$$

where F denotes the unknown distribution of the individual predisposition of the patients in the considered population. For the above derivation we have further assumed that L_i, X_i and $\epsilon_{i1},...,\epsilon_{iT}$ are independent. To make the maximum likelihood principle feasible we need a parametric assumption for F. We assume that L_i is approximatively Gaussian distributed with variance σ^2 and – without loss of generality – expectation 0. To simplify numerical integration we further approximate the standardised distribution of F by a discrete distribution on the grid $\mathcal{L} = \{l_k = -3.0 + 0.3k | k = 0,...,20\}$ with probabilities proportional to $e^{-\frac{1}{2}l_k}$. A similar approach has been used by Neuhaus, Hauck and Kalbfleisch[10]. Given a patient with the observations $(y_{i1},...,y_{iT},x_i)$ his/her contribution to the likelihood – depending on the parameters $\alpha, \beta, \gamma, c, \Delta$ and σ – is given by

$$L(\alpha,\beta,\gamma,c,\Delta,\sigma; y_{i1},...,y_{iT},x_i) = \left[\sum_{l\in\mathcal{L}} e^{-\frac{1}{2}l^2} \prod_{j=1}^{T} B_{ij}(l)^{y_{ij}}(1-B_{ij}(l))^{1-y_{ij}}\right] / \sum_{l\in\mathcal{L}} e^{-\frac{1}{2}l^2}$$

$$\text{with } B_{ij}(l) := \Lambda(-c - (j-1)\Delta + \sigma l + \alpha' x_i + \beta \sum_{j'<j} y_{ij'} \gamma^{j-j'-1}) \ .$$

Assuming independence between different patients we arrive at the ML estimate $\hat{\theta}$ of $\theta := (\alpha, \beta, \gamma, c, \Delta, \sigma)$ by maximising $\sum_{i=1}^{n} \log L(\alpha, \beta, \gamma, c, \Delta, \sigma; y_{i1}, ..., y_{iT}, x_i)$.

So far we have assumed that all patients obtain a chemotherapy with a fixed number T of cycles, and that we have for each patient data from all cycles. This is not true for most data sets used in antiemetic research. First we have often a mixture of patients with different chemotherapies with different numbers of planned cycles. Second, even if all patients have the same number of planned cycles, there may be a substantial number of patients stopping the chemotherapy before the planned end. Third, a change of antiemetic treatment causes often stopping of recording data from subsequent cycles in antiemetic trials, although this is no good practice. Summarising, we have in practice a random number T_i of cycles for each patient.

ML estimation with replacing T by T_i in the formulas above remains valid, if the *Missing at Random* (MAR) assumption[11,12] holds, which is often called the assumption of *non-informative dropout* in the setting of longitudinal data[13]. This assumption requires that

$$P(T_i > t | Y_{i1} = y_{i1}, ..., Y_{iT} = y_{iT}, X_i = x_i) = P(T_i > t | Y_{i1} = y_{i1}, ..., Y_{it} = y_{it}, X_i = x_i) \ .$$

With respect to data from antiemetic research this assumption can be violated, if a patient obtaining an alternative rescue medication due to severe emesis in cycle j is recorded as a patient withdrawn in this cycle, i.e. the true value $Y_{ij} = 1$ is not recorded in the data. If one requires that any withdrawal due to severe emesis is only recorded as a withdrawal in the next cycle, the MAR assumption is maintained, because it allows that the event $T = j$ (or its probability) depends on the value of Y_{ij}. The MAR assumption also allows that the drop out mechanism differs among subgroups defined by covariate values. This will be important in our example where we compare different chemotherapies also differing with respect to drop out probabilities.

4 Estimation of effects not adjusted for individual predisposition

In the analyses of earlier studies there was usually no attempt to adjust effect estimates of risk factors for the latent predisposition L_i and only seldom an attempt to adjust for the effect of emesis in previous cycles[6]. Of course, we may be interested in comparing these unadjusted effect estimates with those arrived from our model. If the data from earlier studies consists of a mixture with data from different cycles, such a comparison renders difficult for reasons further discussed in Section 6. Hence let us assume for simplicity that in earlier studies always the first cycle of each patient was used. Then the resulting estimate of the effect $\tilde{\alpha}$ not adjusted for the individual predisposition is not comparable to the estimates of the adjusted effect α obtained using the model of the last section. To achieve comparability is one reason to consider in this section additional estimation of $\tilde{\alpha}$ following the estimation of α; further reasons will become obvious in Section 5.

We can always obtain an estimate of $\tilde{\alpha}$ by fitting a logistic model only to the data $(Y_{i1}, X_i)_{i=1,\ldots,n}$ of the first cycles. However, it promises to be more efficient to use the relation between $\tilde{\alpha}$ and the complete parameter vector θ to obtain an estimate of $\tilde{\alpha}$ as a function of the ML estimate $\hat{\theta}$, as this allows to make use of all data. This relation is however a little bit complicated by the fact that if our model is true then the conditional distribution of Y_{i1} given X_i need not follow exactly a linear logistic model. Consequently the true value of $\tilde{\alpha}$ can only be defined as the stochastic limit of the logistic regression estimate for the unadjusted effect, which depends on the distribution of X. Using the empirical distribution of X as a naive estimate we nevertheless result in the following intuitively convincing procedure to obtain an estimate $\hat{\tilde{\alpha}}$ based on $\hat{\theta}$. The basic idea is to generate a new data set with the same values for X_i, but with two weighted observations $Y_{i1} = 0$ and $Y_{i1} = 1$, where the weights $p_{\hat{\theta}}^{i,y_{i1}}$ are just the conditional probabilities for the outcome in the first cycle arrived by our model evaluated at the ML estimate $\hat{\theta}$. These weights can be easily computed as

$$\begin{aligned} p_{\hat{\theta}}^{i,y} &= P_\theta(Y_{i1} = y | X_i = x_i) \\ &= \left[\sum_{l \in \mathcal{L}} \mathrm{e}^{-\frac{1}{2} l^2} \Lambda(-\hat{c} + \hat{\sigma} l + \hat{\alpha}' x_i)^y (1 - \Lambda(-\hat{c} + \hat{\sigma} l + \hat{\alpha}' x_i))^{1-y} \right] / \sum_{l \in \mathcal{L}} \mathrm{e}^{-\frac{1}{2} l^2} \end{aligned}$$

This new data set is then analysed by linear logistic regression resulting in an estimate $\hat{\tilde{\alpha}}$. It can be computed by standard logistic regression; in the Appendix we derive also a procedure to estimate its variance.

5 Finite sample properties

Statistical inference based on the ML principle is justified by asymptotic results. The estimates are consistent, efficient and asymptotically Gaussian and the asymptotic variance can be estimated in a consistent manner. However, in our setting it is questionable whether the asymptotic results imply sufficient properties in finite samples, especially small bias and valid confidence intervals based on the normal approximation. There are two reasons for this: First, the asymptotic theory requires that all parameters are in the interior of the parameter space. In our setting there are two parameters with a natural inequality restriction: The scale parameter σ must be nonnegative, and the factor γ must be nonnegative, too. We cannot exclude a priori for a given data sets, that in the underlying model both parameters are different from 0, as it may happen that we have no individual predisposition or no effect of emesis in one cycle on the next but one cycle. It is plausible that this does not happen, but at least σ and γ may be rather small such that asymptotic works only for very large samples. Second, investigations of simulated data sets reveals a frequent existence of multiple roots of the maximum likelihood equation. We often observed up to three roots corresponding to three local extrema of the likelihood. The first shows a value of β near 0 and a large value for σ, such that the correlation among the outcome variables is totally explained by the individual predisposition. The second shows a value of σ near 0 and a large value of β, such that the correlation is totally explained by the increased risk due to emesis in previous cycles. The third shows moderate values of both β and σ and hence takes both explanations into account. Asymptotic theory guarantees only the existence of one consistent root (see for example Lehmann[14], chapter 6.3), hence we have the problem to choose the "correct" one. In the situation described above it may be straightforward to choose the third solution, but often the role of the different solutions is not so obvious. Additionally, we need an algorithm to compute all roots, but we usually have only tools to compute one root depending on start values.

These arguments against working of the asymptotic theory in finite samples makes investigations of finite sample properties extremely necessary. We have tried to do this by extensive simulation studies from which we will report some results. However, the conduct of such a simulation study is limited by the existence of multiple roots. Given a numerical procedure to compute a root in practical data analysis we can try several start values to check for the existence of multiple roots and choose the most convincing one, but this is difficult to implement in a definite, reproducible procedure. We try to circumvent this problem by choosing always the true parameter values as start values for our maximisation procedure. The choice of this procedure is also crucial. The Newton-Raphson procedure is not sufficient in this setting as the matrix of second derivatives is numerically nonsingular only in a small neighbourhood of the root and it may even result in an inadmissible search direction. Hence we use a Levenberg-Marquardt type algorithm trying several search directions in each step, including the gradient and the direction of a Newton-Raphson step. We further use the reparametrization $\gamma = \exp(\gamma')$ to handle the inequality $\gamma \geq 0$. No attempt was made to ensure $\sigma \geq 0$ as the likelihood behaves symmetrically with respect to σ and $-\sigma$. Only if the true value of σ was 0, we use 0.5 as a start value for σ, because otherwise we can only find a root with $\sigma = 0$, as the derivative of the likelihood with respect to σ is 0 at $\sigma = 0$.

In our simulation study we consider the case of one binary, balanced covariate. With respect to the parameter constellation we consider several choices of α, β and σ, whereas we keep the values $c = -1.0$, $\Delta = 0.25$ and $\gamma = 0.5$ fixed. We consider always the situation with $T = 6$ planned cycles and first only the situation with no drop outs (A) as well as with a drop out probability of 29% in each cycle (B) for each patient, such that we have in the average three cycles for each patient. We first consider only a sample size of $n = 100$. Table 1 summarises the results of our simulation study. We report for all considered parameter constellations and for each of the six parameters of our model the mean of the difference between the estimates and the true values (bias), the empirical standard deviations (sdv) and the coverage probabilities (cvg) of a nominal 95% confidence interval based in the normal approximation and the estimated asymptotic variance. For $\hat{\gamma}$ and $\hat{\sigma}$ we also report the fraction of estimates equal to 0. All reported results are based on 1000 repetitions in the simulation study. Confidence intervals for γ are based on an exponential transformation of the confidence intervals for γ'.

In the first part of Table 1 we keep $\beta = 1.5$ and $\sigma = 1.0$ fixed and vary α. With respect to the estimation of α we observe a slight overestimation increasing with the value of α. Coverage probabilities indicate an anti-conservative behaviour of confidence intervals for large values of α, which is more distinct in the case of no drop outs. Coverage probabilities smaller – and sometimes distinctly smaller – than the nominal level can be observed also for the other parameters. There are two possible explanations for this. First, the approximation of the distribution of $\hat{\theta}$ by a multivariate normal distribution may be poor, which is indicated by substantial fractions of zeros for $\hat{\sigma}$ and $\hat{\gamma}$. However, confidence intervals for σ and γ behave not poorer than those for other parameters, such that this explanation is not convincing. Second, the existence of multiple roots and the choice of one of them gives a source of variability not accounted for in the asymptotic variance. This explanation coincides with the observation that among all parameters the worst violation of the nominal coverage probability is found for Δ. Indeed, we observed in our examples with multiple roots that estimates of α, β, σ and Δ can differ in a comparable magnitude of 1.0 or more. However, $\hat{\Delta}$ has the smallest asymptotic and empirical variance of all estimates, hence here the choice of the root shows the largest relative increase of variability. With respect to the bias of the other parameter estimates we observe a substantial underestimation of σ and moderate bias for the other parameters.

In the second part of Table 1 we investigate the dependence of finite sample properties on the true value of σ, i.e. on the degree of variation of the individual predisposition.

We observe that the bias of $\hat{\alpha}$ is largest for $\sigma = 0$ and vanishes for large values of σ. A dependence of the bias on the true value of σ can be also observed for the other parameters. For all parameters except of σ the difference between the actual and the nominal level of the confidence intervals increases with increasing σ.

In the third part we investigate the influence of β on the finite sample properties. Omitting the results for $\beta = 0$ the standard errors for all parameters increase with β, i.e. larger values of β make estimation more difficult. Also the bias seems to be a function of β, especially if β is large there may be a substantial overestimation of α. The results for $\beta = 0$ do not fit well in these observations, here there may be other difficulties. Note that γ is not identifiable if $\beta = 0$, such that bias and coverage probabilities cannot be assessed for $\hat{\gamma}$.

In the fourth part we investigate the influence of the drop out mechanism. We consider mechanisms where the drop out probability in cycle j depends only on the value of Y_{ij-1} and X_i, i.e. we assume

$$P(T_i > t | Y_{i1} = y_{i1}, ..., Y_{it} = y_{it}, X_i = x_i) = P(T_i > t | Y_{it} = y_{it}, X_i = x_i) =: q(y_{it}, x_i) \ ,$$

such that $1 - q(y_{it}, x_i)$ are drop out probabilities. The drop out mechanism C is defined by the values $q(0, \cdot) = 0.9$ and $q(1, \cdot) = 0.5$, i.e. occurence of emesis determines the probability of drop out. The drop out mechanism D is defined by $q(\cdot, 0) = 0.9$ and $q(\cdot, 1) = 0.5$, such that drop out probabilities differ only between subpopulations defined by covariate values. The drop out mechanism E combines both sources by assuming $q(0, 0) = 0.9$, $q(0, 1) = q(1, 0) = 0.7$, and $q(1, 1) = 0.5$. We observe no great impact on the bias and coverage with respect to estimation of α, but for other parameters this seems to be different.

In the last part of Table 1 we investigate the impact of the sample size. We observe that increasing the sample size always implies a decrease of the bias. However, the change from n=100 patients to n=400 patients implies for most parameters including α an increase of the difference between nominal and actual level of the confidence intervals. Approximatively valid confidence intervals are obtained for $n = 1600$ patients in the case of no drop outs, but otherwise we observe this only for $n = 6400$ patients. This is due to the fact that the number of multiple roots and the numerical difference between multiple roots does not vanish with the same rate as the variance of the "correct" root decreases. Consequently for moderate sample sizes the choice of an "incorrect" root results often in a failure of the coverage of the confidence interval centered around this root, whereas in small samples this is often avoided due to the large estimated variance. A further hint for the necessary sample sizes to ensure validity of asymptotic results is give by the ratio of standard errors. If asymptotic works, enlarging the sample size by a factor of 4 shrinks the standard error by a factor $\frac{1}{2}$. This is never observed for the change from n=100 to n=400 patients. It can be approximatively observed for the change from n=400 to n=1600 patients in the case of no drop outs, and for the change from n=1600 to n=6400 patients in the considered case of drop outs.

So far we have considered estimation of the parameters of the model introduced in Section 2. In Section 3 we have also considered a procedure to obtain estimates for $\tilde{\alpha}$, i.e. for the effect of the risk factors not adjusted for the individual predisposition. In Table 2 we report properties of this estimator and the corresponding confidence intervals based on estimation of the asymptotic variance as outlined in the Appendix are reported. In comparison to Table 1 the bias is distinctly decreased and confidence intervals show only a very slight tendency to be anti-conservative. So the properties of this estimator are much more satisfactory than those of the ML estimator for α. We can also observe that the standard deviations of the estimates $\hat{\tilde{\alpha}}$ are much smaller than those of $\hat{\alpha}$, but this statement needs some care, as $\tilde{\alpha}$ is usually smaller than α. However, simple comparisons reveal that also the coefficients of variation are smaller for $\hat{\tilde{\alpha}}$ than for $\hat{\alpha}$. This suggests that significance tests based on $\hat{\tilde{\alpha}}$ are more powerful than significance tests based on $\hat{\alpha}$. This is corroborated in Table 3, where we compare the power of the tests to reject the (equivalent) null hypotheses $\alpha = 0$ or $\tilde{\alpha} = 0$, respectively, if the 95% confidence intervals do not include 0. To explain this gain in power one has to note that $\hat{\alpha}$ and $\hat{\sigma}$ are highly correlated; in our simulation studies we always observed a Pearson correlation of at least 0.45 for all constellations with $\alpha = 1.0$. Now considering the dependence of $\tilde{\alpha}$ on α and σ, $\tilde{\alpha}$ decreases if σ increases, such that due to the correlation in the average large values of $\hat{\alpha}$ are shrunken to a higher degree than small values of $\hat{\alpha}$ by the procedure of Section 3, which overall decreases the variability. This also explains to some degree the better behaviour of $\hat{\tilde{\alpha}}$ compared to $\hat{\alpha}$ with respect to bias and coverage, as the multiple roots show usually a correlation between $\hat{\alpha}$ and $\hat{\sigma}$, too, such that different multiple roots result in similar values of $\tilde{\alpha}$. We have included in Table 3 also the power of a test based on $\hat{\alpha}_1$, which denotes the estimate based on logistic regression using only the data from the first cycle. For large values of β and σ this test can be even more powerful than the test based on $\hat{\alpha}$.

6 Comparison with other approaches

In the literature there exist many other approaches to develop regression models for longitudinal binary outcome data (see for example Zeger and Liang[15] or Diggle, Liang and Zeger[16]). These approaches differ with respect to the interpretation of the regression parameters. In this section we want to sketch some properties and difficulties with these alternative approaches if applied to data from antiemetic research. Especially we try to explain why parameters of the alternative approaches tend to change over time such that – contrary to our model – a unique parameter describing the effect of a risk factor can only be defined in a somewhat arbitrary manner. We further try to explain why the parameters of the alternative approaches tend to be smaller or larger than α; of course this is no argument in favour of any approach and reflects mainly the different interpretation. We also discuss whether the alternative approaches can be used to perform a significance test for the null hypothesis of "no effect"of a risk factor. This requires first that the null hypothesis of no effect in our model is equivalent to an appropriate null hypothesis in the alternative approach, and second that the distribution of an appropriate test statistic can be well approximated under the null hypothesis.

We will illustrate the time dependency of the regression parameters of the alternative approaches by some of their numerical values in Table 4. Precisely, the values are the stochastic limits of the parameter estimates if we fit the alternative models to data according to our model. For the latter we consider again the situation with one binary, balanced covariate; we consider several choices of α, β, and σ and keep $c = 0$, $\Delta = 0$ and $\gamma = 0.5$ fixed. As a consequence of this time dependency the stochastic limits of the artificially defined unique parameters depend on the number of planned cycles and the drop out mechanism, as they are weighted averages of the time-varying parameters. Hence we report in Table 5 also these limits for eight combinations of two different choices of the planned number of cycles, two different drop out probabilities and two different parameter constellations.

We start our considerations with marginal models (see for example Chapter 8 of Diggle, Liang and Zeger[16]), which consider the observable effect of X_i on Y_{ij} separately in each cycle. The regression parameter α_j^M refers hence to a logistic model for the conditional distribution of Y_{ij} given X_i for each cycle, i.e. the assumed model is

$$P(Y_{ij} = 1|X_i = x_i) = \Lambda(\alpha_{0j}^M + (\alpha_j^M)' x_i) \ .$$

Estimates for α_j^M can be obtained by fitting logistic models to the data from each cycle, i.e. a cross-sectional analysis. In the first cycle α_1^M tends to be smaller than α. This is due to the lack of an adjustment for the effect of L_i, which is independent of X_i by assumption. From investigations of the bias due to omitted covariates[17,18] it is known that in logistic models this results in an underestimation of α. Further, the bias increases with the influence of the omitted covariate, hence the difference between α and α_1^M increases with σ, cf. also Zeger, Liang and Albert[19] and Neuhaus, Kalbfleisch and Hauck[20]. If

$\alpha > 0$ in the following cycles the values of α_j^M increase over time. Formally, this is due to the continuing omission of "covariates" $Y_{i1}, ..., Y_{ij-1}$, which are correlated with X_i and have also a positive effect on Y_{ij}. The latter is due to $\beta > 0$, and the amount of increase of α_j^M over time depends on β (cf. Table 4). Conceptually, the assumption that emesis in one cycle increases the probability of emesis in the next cycle implies that the observable association between X_i and Y_{ij} increases over time, because patients identifiable as high risk patients by their covariate values have a high risk to vomit in the first cycle and hence many of them have as increased risk in the second cycle, whereas only few of the patients identifiable as low risk patients by their covariate values vomit in the first cycle and suffer from an increased risk in the second cycle.

Neglecting the time dependency of α_j^M we can artificially assume $\alpha_j^M \equiv \alpha^M$. Then estimates can be obtained by fitting logistic regression models simultaneously to all cycles[21] or using the GEE approach[22,23]. For a specific parameter constellation we observe in the first row of the right part of Table 5 a variation of α^M between 0.94 and 1.08 due to differences in the number of planned courses and drop out mechanisms.

Biased estimates due to drop outs are an additional difficulty with marginal models. Using the estimation procedures mentioned so far to analyse all data of the observed cycles the resulting estimates for α_j^M refer to the conditional distribution of Y_{ij} given X_i and the fact that the j-th cycle is observed. The MAR assumption is not sufficient to guarantee coincidence of this conditional distribution with the original one. The bias of $\hat{\alpha}_j^M$ depends on the quantities $\tilde{q}_j(y, x) := P(T_i \geq j | Y_{ij} = y, X_i = x)$. We can exclude a bias if a decomposition $\tilde{q}_j(y, x) = \tilde{q}_j(y)\tilde{q}_j(x)$ is possible[24,25]. If the conditional drop out probabilities $q(y, x)$ defined in the last section satisfy $q(y, x) = q(x)$ this decomposition holds. However, even in the simple situation of drop out probabilities depending only on the previous outcome, i.e. if $q(y, x) = q(y)$ holds, a decomposition of $\tilde{q}$ need not be possible and $\hat{\alpha}_j^M$ is biased. For example for the drop out mechanism C we find for the six regression parameters of the marginal model in the constellation of the fourth line of Table 4 the stochastic limits 0.833, 0.932, 0.994, 1.029, 1.043 and 1.041 slightly smaller than the true ones. The bias can be larger and of the opposite direction, for example for the drop out mechanism E we observe the stochastic limits 0.833, 0.973, 1.080, 1.162, 1.228 and 1.289. Such a bias can be only avoided by using more complicated estimation procedures[26].

The null hypothesis of no effect in our model, i.e. $\alpha = 0$ is equivalent to the null hypothesis $\alpha^M = 0$ under the assumption that the model of Section 2 is correct. The estimation procedures mentioned above suggest to use a Wald Test to reject the null hypothesis. The approximation by a χ^2-distribution is asymptotically valid in the case of no drop outs, as this relies only on the correct specification of the marginal conditional expectations and variances. The power of this test depends on the estimation procedure and the working correlation matrix chosen, which renders a comparison difficult, especially as we have no software to specify a working correlation matrix according to the model of Section 2. In our comparison in Table 5 we consider the GEE approach of Liang and Zeger[22] with three different working correlation matrices, corresponding to the assumptions of independence, of an exchangeable correlation and of a correlation structure according to an AR(1) process. We further use the assumption $\alpha_{0j}^M = \alpha_0^M + \Delta^M(j-1)$ with respect to the intercepts in fitting the marginal models. We observe that the power of a marginal model with a working correlation matrix according to an AR(1) process is of similar magnitude as the power of a test best on $\hat{\hat{\alpha}}$. Even if σ is large and hence we have the largest deviation between the true and the working correlation matrix, we observe only some loss of power in the design with drop outs. In contrast, the use of an working correlation matrix corresponding to an exchangable correlation structure can imply a substantial loss; the latter is especialy large, if β is large, i.e. if the true and the working correlation matrix differ substantial. Note that in the case of the drop out mechanism E the stochastic limit of $\hat{\alpha}^M$ is equal to 0.040 and not 0, but this difference is not large enough to show a substantial difference between nominal and true level of the test based on fitting a marginal model using the GEE approach.

It is worth mentioning that earlier studies investigating risk factors and regarding each cycle of a patient as a single unit also tried to estimate the regression parameter of a marginal model. In many of these studies it was not required to observe each patient right from the first cycle, resulting in a rather arbitrary mixture of cycles. For a specific parameter constellation the parameters estimated in these studies can be regarded as some arbitrary averages of the values shown in one line of the first part of Table 4.

Second we discuss transition models (see for example chapter 10 of Diggle, Liang and Zeger[16] or Bonney[27]), considering the effect of the covariate X_i on the outcome Y_{ij} adjusted for the effect of the outcome in previous cycles. The latter can be modeled in different manners; to keep the comparison simple we consider here only the analogue to our model. Hence the parameters α_j^T refer to a logistic model for the conditional distribution of Y_{ij} given X_i and $Y_{i1}, ..., Y_{ij-1}$ assuming

$$P(Y_{ij} = 1 | X_i = x_i, Y_{i1} = y_{i1}, ..., Y_{ij-1} = y_{ij-1}) = \Lambda(\alpha_{0j}^T + (\alpha_j^T)' x_i + \beta \sum_{j'<j} y_{ij'} \gamma^{j-j'-1}) \ .$$

Parameter estimates can be easily obtained by fitting this model to appropriately arranged data.

The parameters α_j^T are always smaller than α, because in each cycle we omit the "covariate" L_i in comparison to our model (cf. the considerations for α_1^M above and note that $\alpha_1^T = \alpha_1^M$.) Although this mistake is the same in each cycle, α_j^T decrease over time. This is due to the fact that the correlation among the outcome variables due to the existence of L_i with $Var(L_i) > 0$ is partially transferred to $\hat{\beta}$. For example the stochastic limits of $\hat{\beta}$ in the seven lines of the middle part of Table 4 are 1.45, 1.46, 1.47, 0.57, 2.42, 1.16 and 2.02, respectively. This overestimation of β implies an increasing over-adjustment as with increasing time more and more "covariates" $Y_{i1}, ..., Y_{ij-1}$ enter the model, and consequently α_j^T decrease over time. Also, the amount of the change from one cycle to the next increases with σ, as this implies increasing over-adjustment. If one uses an alternative transition model, the stochastic limits can be different. Replacing $\beta \sum_{j'<j} y_{ij'}\gamma^{j-j'-1}$ by $\sum_{j'<j} \beta_{jj'}y_{ij'}$ as done in a paper by the Italian Group of Antiemetic Research[6], α_j^T still decrease over time. However, using a simple Markov model only including the term $\beta_j y_{ij-1}$ we will observe an increase of α_j^T after the second cycle.

Assuming artificially a constant parameter $\alpha_j^T \equiv \alpha^T$ we can observe in the left part of Table 4 a variation of the stochastic limit of $\hat{\alpha}_T$ between 0.68 and 0.77 due to differences in the number of planned courses and drop out mechanisms. As above the null hypothesis $\alpha = 0$ is equivalent to $\alpha^T = 0$. However, using the ML principle for statistical inference, the usual approximation for the distribution of a Wald-statistic or the likelihood ratio test need not be valid, as the model is misspecified under the null hypothesis, because the additional dependence among the outcome variables due to L_i is neglected. Valid inference can be obtained using robust variance estimates in analogue to the GEE approach for marginal models (cf. Diggle, Liang and Zeger[16], Section 10.2).

A further drawback of transitional models concerns the interpretation of a change of the intercept over time. In a transition model such a change is mainly due to the omission of the individual predisposition, implying a decrease of the risk of emesis over time for subjects without emesis in previous cycles.

Third we consider random intercept models, a special case of random effect models (see for example Chapter 9 of Diggle, Liang and Zeger[16] or Laird and Ware[28]). In a random intercept model the effects of the risk factors are adjusted for an individual predisposition which is regarded as the only source of correlation between the outcome variables. Hence the parameters α_j^R are based on the assumption

$$P(Y_{ij} = 1 | L_i = l, X_i = x_i) = \Lambda(c_j + l + (\alpha_j^R)' x_i) \ .$$

Fitting such models by the ML principle requires assumptions on the distribution of L_i; we consider here only the analogue to our model. Although for $j = 1$ this model coincides with our model, α_1^R is much larger than α. This is due the fact that estimates of σ based on the random intercept model are larger than the true σ of our model, because the effect of β is partially transfered to σ, as both explain the correlation among outcome variables. This explanation is corroborated by an increase of the difference between α_1^R and α with increasing β as observable in the lower part of Table 4. The parameters α_j^R increase over time as we omit the necessary adjustment for $Y_{i1}, ..., Y_{ij-1}$, which are positively correlated with X_i (if $\alpha > 0$), and the number of omitted terms increase over time. However, as the terms added in each cycle become relatively small over time, the amount of increase decreases, and for one constellation we observe a decrease from cycle 5 to cycle 6.

Assuming artificially a constant parameter $\alpha_j^R \equiv \alpha^R$ we can observe in the right part of Table 5 a variation of the stochastic limit of $\hat{\alpha}_R$ between 1.894 and 2.030 due to differences in the number of planned courses and drop out mechanisms. As above the null hypothesis $\alpha = 0$ is equivalent to $\alpha^R = 0$, but again asymptotic results cannot be used to construct a significance test due to a misspecification under the null hypothesis.

Finally we take a look on simple binomial models relating the number of cycles with emesis to X_i. Here the parameter α^B refers to the assumption

$$P(\sum_{j=1}^{T_i} Y_{ij} = k | X_i = x_i, T_i = t_i) = \binom{t_i}{k} \Lambda(\alpha_0^B + \alpha^{B\prime} x_i)^k (1 - \Lambda(\alpha_0^B + (\alpha^B)' x_i))^{t_i - k} \ .$$

Such models can be easily fitted using logistic regression for count data. The interpretation of α^B is similar to that of α^M; α^B compares the probability of an event in dependence on the value of x on the logit scale, and - as in truth these probabilities are time dependent - it actually estimates some average of these probabilities. This similarity can be observed in Table 5, where the stochastic limits of α^M and α^B are of similar magnitude. The null hypothesis $\alpha = 0$ is equivalent to the null hypothesis $\alpha^B = 0$, at least in the case of no drop outs. Assessing the distribution of $\hat{\alpha}^B$ under the null hypothesis requires some care; first we have to allow for overdispersion to take into account the correlation among the outcome variables; second it is necessary to allow a change of α_0^B in dependence of the value of T_i in the case of drop outs. For the latter we replace α_0^B by $\alpha_0^B + (t_i - 1)\Delta^B$; for the estimation of an over-dispersion parameter we use the method of Williams[29] as provided by PROC LOGISTIC of SAS, version 6.11. This way we achieve a reasonable approximation under the null hypothesis (see Table 3). In the case of no drop outs the power using the binomial model is comparable or even slightly larger in comparison to a test based on $\hat{\hat{\alpha}}$. In the considered case of drop outs the power is slightly smaller. However, as observable for the drop out mechanism E, the binomial model may fail to define a valid significance test in the case of drop outs. The parallels to marginal models shown above illustrate that binomial models suffer from the same source of bias, and – at least in the one case considered here – suffer more. For the drop out mechanism E

and $\alpha = 0$ (second but last line of Table 3) the stochastic limit of $\hat{\alpha}^B$ is equal to 0.112, resulting in a substantial violation of the nominal level.

7 Example

In our example we consider data collected in an open study during the development of the antiemetic drug granisetron and published so far only in part[30]. We consider patients from five subgroups treated with different chemotherapies. Except of patients treated with cisplatin all patients obtain only the chemotherapeutic drugs mentioned in Table 6 and these drugs were administered in a single dose. All cisplatin patients have in common to be treated with a cisplatin dose of at least 50 mg/m^2 on the first day of each cycle, but they received a variety of additional drugs, partially also on subsequent days. We divide the cisplatin patients in a male and female subgroup; all other groups include only female patients. The exact dose of the administered drugs was not recorded in the data, however it is known that the dose of carboplatin was always at least 300 mg/m^2 and the dose of cyclophosphamid was at least 500 mg/m^2. The outcome variable considered in each cycle is the occurence of vomiting within 8 days after start of chemotherapy. In Table 6 we report the percentage of patients vomiting in the first cycle to give a first impression of the emetic potential of the regimens. All patients received in each cycle for the protection against emesis granisetron in a dose of 3 mg prior to chemotherapy. In the case of emesis further treatment with granisetron or other antiemetic drugs was allowed. Note that in this example there was a strict rule in the study protocol to report all data including the outcome also for those cycles where a change of antiemetic treatment occurs after start of chemotherapy or due to emesis. Only data from subsequent cycles were not recorded and appeared as drop outs.

In the first part of the example we compare the male and female cisplatin patients. The first risk factor considered is sex, and the second is age, coded by two dummy variables for age groups with the youngest patients as the reference category. The estimates for the effect of sex confirm the well known increased risk for women (Table 7). To establish significance it is necessary to rely on the more powerful inference based on $\tilde{\alpha}$. Age shows an increased risk for elder patients, however, this fails to be significant. In the second part we compare the five female patient groups with respect to the influence of the regime and additionally of age. For the comparison of the regimens we choose CMF as the reference category. Distinct differences between the regimens can be revealed. CMF and the combination of Carboplatin with Cyclophosphamid shows the lowest emetogenic potential, whereas EC and FEC are associated with an increased risk. Patients with cisplatin show the highest risk. The last result coincides with the literature regarding cisplatin as a high emetogenic drug contrary to the other drugs considered here. A little bit surprising is the low risk of patients treated with Carboplatin and Cyclophosphamid, as both drugs are regarded as moderate emetogenic. However, the interpretation of these results is hampered by insufficient knowledge which patients are naive to chemotherapy, as this is an important confounder both related to the choice of therapy and the risk of emesis. The effect of age appears here to be not monotone and far away from being significant.

A further look at the other parameters reveals some interesting differences and coincidences. First we should mention that in both parts we have tried different start values but were unable to find alternative roots. Nevertheless, in the first part of the example $\hat{\beta}$ is large and $\hat{\sigma}$ is small contrary to the second part. However, there may be a conceptual explanation for this. First, it is well known that cisplatin patients suffer not only from frequent, but also often from severe emesis. Severe emesis may have a stronger impact on the occurence of emesis in the following cycles than mild emesis. Hence the true value of β may indeed be smaller in the second part than in the first, as in the second part we mainly consider patients receiving moderate emetogenic drugs. Second, combining patients from different chemotherapies may increase the heterogeneity with respect to other important, unobserved risk factors which constitute a part of the latent individual predisposition L_i. Consequently, the true value of σ may be indeed larger for patients in the second part. The values of γ are rather small in both parts of the example, suggesting that the influence of emesis in one cycle on the next but one and following cycles is not very substantial. The estimates of Δ are positive in both parts, but nearly 0 in the second part. Interpretation of the first result as an indicator for a gain in efficacy of the antiemetic drug should be taken with care; this will be further discussed in the final section.

8 Sensitivity analyses

Any model trying to reduce a complex data generating mechanism to a few parameter values can be criticised due to its simplifying assumptions. This simplification can be justified two-fold; first, it is necessary to obtain interpretable results and second it is necessary to ensure estimability from finite samples. Nevertheless, one should try to check whether one of the simplifying assumptions has a substantial impact on the results. Such a sensitivity analyses is reported in this section; all numerical results are summarised in Table 8.

First, we weaken the assumption of a linear trend in the change of the thresholds by allowing a separate threshold for each cycle. This means that in the likelihood $B_{ij}(l)$ is replaced by $\Lambda(-c_j + \sigma l + \alpha' x_i + \beta \sum_{j'<j} y_{ij'}\gamma^{j-j'-1})$. The results are shown in row a) of Table 8. The estimated c_j suggest no need to allow for complex changes of the thresholds, and the effect estimates of the risk factors are only slightly changed compared to the original analysis. Second, we allow that the effect of emesis in one cycle on emesis in subsequent cycles may vary over time by allowing cycle-specific values of β. Here $B_{ij}(l)$ is replaced by $\Lambda(-c-(j-1)\Delta + \sigma l + \alpha' x_i + \beta_j \sum_{j'<j} y_{ij'}\gamma^{j-j'-1}$. The results of row b) shows little evidence for a change of β over time and the impact on the estimates of α is small, again.

Third, we consider the assumption that L_i is approximatively Gaussian in the population. This assumption is indeed often highly questionable, because other alternatives are of similar plausibility. For example, if there exists an important, unobserved, dichotomous risk factor, L_i is expected to have a bimodal distribution. In a first step we consider the alternative that L_i follows approximatively a mixture of two Gaussian distributions. This model we fit based on the likelihood

$$L(\alpha, \beta, \gamma, c, \Delta, p_1, p_2, \sigma_1, \sigma_2, \mu_1, \mu_2; y_{i1}, ..., y_{iT_i}, x_i)$$

$$= \left[\sum_{l\in\mathcal{L}} \mathrm{e}^{-\frac{1}{2}l^2} \sum_{k=1}^{2} p_k \prod_{j=1}^{T_i} B_{ijk}(l)^{y_{ij}}(1-B_{ijk}(l))^{1-y_{ij}} \right] / \sum_{l\in\mathcal{L}} \mathrm{e}^{-\frac{1}{2}l^2}$$

$$\text{with } B_{ijk}(l) := \Lambda(-c-(j-1)\Delta + \sigma_k l + \mu_k + \beta \sum_{j'<j} y_{ij'}\gamma^{j-j'-1})$$

and the restrictions $\mu_1 = 0$ and $p_2 = 1 - p_1$. In the first part of the example this analysis (row c1) suggests a two-point distribution for L_i with a distance of 2.69 between the two mass points on the logit scale. All risk estimates are increased by factors between 1.6 and 2.0. In the second part this analysis suggest a distribution of L_i with additional mass close to the mean compared to a Gaussian distribution. The risk effect estimates are comparable to those of the original analysis except of a slight change for $\hat{\alpha}(CC)$. In a second step we consider a three point lattice distribution on $\{-\sigma, 0, \sigma\}$ with arbitrary probabilities, such that the likelihood reads

$$L(\alpha, \beta, \gamma, c, \Delta, p_{-1}, p_0, p_1, \sigma; y_{i1}, ..., y_{iT_i}, x_i)$$

$$= \sum_{l\in\{-1,0,1\}} p_l \prod_{j=1}^{T_i} B_{ij}(l)^{y_{ij}}(1-B_{ij}(l))^{1-y_{ij}}$$

with the restriction $p_{-1} + p_0 + p_1 = 1$. In the first part of the example the results (row c2) coincide with those of the first step. In the second part the distribution of L_i is again distinctly peaked close to the mean, and the risk estimates are very similar to the first step.

Forth we consider a simplification somewhat hidden in the formulation of Section 2, but with some importance for the specific situation of this paper. We assumed in Section 2, that the variance of the remaining error terms ϵ_{ij} are constant over time. This need not be true, as this term reflects also the influence of unobserved risk factors varying over time, and these risk factors may be more or less important over time. Now let us assume that ϵ_{ij} are still independent and from a logistic distribution but with varying scale factors σ_j such that $P(\epsilon_{ij} \leq z) = \Lambda(z/\sigma_j)$. Then

$$\begin{aligned} & P(Y_{ij} = 1 | X_i = x_i, L_i = l, Y_{i1} = y_{i1}, ..., Y_{ij-1} = y_{ij-1}) \\ = \; & P(-\epsilon_{ij} \leq -c - (j-1)\Delta + l + \alpha' X_i + \beta \sum_{j'<j} y_{ij'} \gamma^{j-j'-1}) \\ = \; & \Lambda(-c/\sigma_j - (j-1)\Delta/\sigma_j + l/\sigma_j + \alpha' X_i/\sigma_j + \beta \sum_{j'<j} y_{ij'} \gamma^{j-j'-1}/\sigma_j) \end{aligned}$$

Hence an analysis under the erroneous assumption of a constant variance results in estimates for α estimating some average of the values α/σ_j and hence loosing easy interpretability. Fitting a model with varying variances with the arbitrary restriction $\sigma_1 = 1$ (row d) we observe in the first part slight evidence for an increase of the variance over time. The resulting risk estimates are similar to the original one; however, they can be interpreted now only as log odds ratios for the effect in the first cycle adjusted for the individual predisposition. In the second part we found no evidence for a change of the variance over time, and the risk effect estimates are similar to the original ones.

9 Discussion

We have proposed a stochastic model to analyse risk factors for emesis in multi-cycle chemotherapies. The model is a hybrid between a transitional model and a random intercept model, and this way it takes into account the two major sources for dependencies among the outcome variables: an individually varying predisposition and a change in the biological or psychological mechanism caused by an event. This adequateness with respect to the medical background is contrasted by poor statistical properties of the maximum likelihood estimates due to the existence of multiple roots. However, we can partially succeed in overcoming this drawback by considering a function of the ML estimates which gives an estimate for the marginal risk in the first cycle. Considering this estimate allows also a much more powerful assessing of the significance of the influence of a risk factor.

In view of these problems and their solution one may prefer to use directly a marginal model to assess the influence of risk factors. Our considerations reveal three – however minor – arguments against the use of marginal models in practice. First in marginal models there exists no natural unique parameter to describe the influence of a covariate, as the regression parameters of a marginal model change over time. Artificially defining a time-constant regression parameter implies that its value depend on the drop out mechanism. However, the possible differences are not very large, such that in the comparison of results from different studies other reasons are of greater importance to explain differences in risk estimates. Second, drop out mechanisms plausible in antiemetic research can lead to biased estimates if one of the common estimation procedures[21−23] are used. Avoiding this bias is however possible using estimation procedures as proposed by Robins, Rotnitzky and Zhao[26]. Power comparisons also reveal no advantage in favour of marginal models compared to tests based on the derived estimate for the marginal risk based on our model; however, marginal models prove sometimes to be efficient only with an appropriate working correlation matrix.

Neglecting these difficulties there remains only one major argument for using our model in practice: In the long run estimates adjusted for individual predisposition may be preferable to marginal risks. This may be especially true if we want to compare results from different studies, and due to different selection procedures we have to conjecture, that the variance of the individual predisposition is different among the studies. Then we can expect nevertheless similar results using our model, but have to expect different results using marginal models. Further pros and cons have been discussed by Zeger, Liang and Albert[19]. From this point of view there remains some interest beyond marginal models. With respect to our model there remains some need for better estimation procedures; for example we may expect that Bayes estimates and confidence regions based on marginal posterior distributions may suffer less from the multimodal shape of the multivariate likelihood. As we have observed in the sensitivity analysis also some impact of the assumption on the distribution of the individual predisposition on the risk estimates, there is also an interest in adapting approaches avoiding this assumption[31,32].

Even if the use of our model for practical statistical inference is hampered by some difficulties, its adequateness with respect to the medical background gives it some value. So it can be used as a framework to analyse properties and limitations of procedures for statistical inference. For example we could show that transitional models sometimes used to analyse the persistence of risk factors[6] will tend to indicate a decreasing risk, and that simple binomial models based on counting the cycles with emesis for each patient can be adequate, but only in the case of no drop outs. Another example would be procedures to assess a change of efficacy of an antiemetic drug over time. Although our model provides a simple parameter Δ to describe such a change, it simultaneously gives some insights into the difficulties to estimate it: Assuming that we have no data on risk factors and only two cycles for each patient, we have no chance to estimate Δ, because we have only four different values to be observable and hence only three degrees of freedom, but the four parameters β, σ, c and Δ to be estimated. This suggests that also in the presence of data from more than two cycles estimates of Δ may be sensitive to the simplifying assumptions chosen to describe the dependence of the outcome variables and allowing identification of Δ.

Finally we should comment on our choice to assume a priori that the parameter α is constant over time. One may argue that this is as artificial as in a marginal model. However, we believe, that in our model we have considered the major sources for the development of a patient over time for most applications, so that our assumption is much more natural than that of a marginal model, where time dependence is a consequence of the source of the correlation among the outcome variables. There may exist settings where the assumption of a constant α is questionable, for example the change of efficacy of an antiemetic drug may be different for different chemotherapies. Formally this would imply a change of α over time, however conceptually this must be regarded as an interaction between the risk factors and the course of the thresholds, and not as a real change of the influence of the risk factor. We have also hesitated to formulate our model for time-varying covariates, although this is mathematically no problem. However, time-varying covariates introduce a further development over time, and it is not straightforward in which sense the influence of a risk factors should be defined in this case.

Acknowledgement

The authors want to thank H.J. Staab, SmithKline Beecham Pharma, Munich, for the kind permission to use the data set.

Appendix: Asymptotic distribution of $\hat{\tilde{\alpha}}$

Given data from n patients $\hat{\tilde{\alpha}}_n$ is the solution of the equation

$$0 = S_n(\tilde{\alpha}, \hat{\theta}_n) = \sum_{i=1}^{n} \frac{\partial}{\partial \tilde{\alpha}} S(\tilde{\alpha}, \hat{\theta}_n, X_i)$$

where $\hat{\theta}_n$ denotes the ML estimate of θ and

$$S(\tilde{\alpha}, \theta, x) = \log \Lambda(\tilde{\alpha}_1 + \tilde{\alpha}'x) P_\theta(Y_1 = 1|X = x) + \log(1 - \Lambda(\tilde{\alpha}_1 + \tilde{\alpha}'x)) P_\theta(Y_1 = 0|X = x)$$

and hence

$$\frac{\partial}{\partial \tilde{\alpha}} S(\tilde{\alpha}, \theta, x) = \left[\frac{\partial}{\partial \tilde{\alpha}}(\tilde{\alpha}_1 + \tilde{\alpha}' x)\right] \left[P_\theta(Y_1 = 1 | X = x) - \Lambda(\tilde{\alpha}_1 + \tilde{\alpha}' x)\right] \ .$$

Let us denote by $\tilde{\alpha}_0$ and θ_o the true values of $\tilde{\alpha}$ and θ and define

$$I_{\tilde{\alpha}\tilde{\alpha}} = E\left[-\frac{\partial}{\partial \tilde{\alpha}}\frac{\partial}{\partial \tilde{\alpha}} S(\tilde{\alpha}, \theta, X)|_{\tilde{\alpha}_0, \theta_0}\right] \text{ and } I_{\tilde{\alpha}\theta} = E\left[-\frac{\partial}{\partial \tilde{\alpha}}\frac{\partial}{\partial \theta} S(\tilde{\alpha}, \theta, X)|_{\tilde{\alpha}_0, \theta_0}\right] \ .$$

A Taylor expansion of $S_n(\tilde{\alpha}, \hat{\theta}_n)$ around $\tilde{\alpha}_0$ yields

$$I_{\tilde{\alpha}\tilde{\alpha}} \sqrt{n}(\hat{\tilde{\alpha}}_n - \tilde{\alpha}_0) = \frac{1}{\sqrt{n}} S_n(\tilde{\alpha}_0, \hat{\theta}_n) + o_p(1)$$

and a second expansion around θ_0 yields

$$I_{\tilde{\alpha}\tilde{\alpha}} \sqrt{n}(\hat{\tilde{\alpha}}_n - \tilde{\alpha}_0) = \frac{1}{\sqrt{n}} S_n(\tilde{\alpha}_0, \theta_0) - I_{\tilde{\alpha}\theta} \sqrt{n}(\hat{\theta}_n - \theta_0) + o_p(1) \ .$$

By the central limit theorem $\frac{1}{\sqrt{n}} S_n(\tilde{\alpha}_0, \theta_0)$ is asymptotically Gaussian and $\sqrt{n}(\hat{\theta}_n - \theta_0)$ is asymptotically Gaussian by the standard theory for ML estimates. The asymptotic efficiency of $\hat{\theta}_n$ and the fact that $\frac{1}{\sqrt{n}} S_n(\tilde{\alpha}_0, \theta_0)$ has asymptotically mean 0 implies that $\frac{1}{\sqrt{n}} S_n(\tilde{\alpha}_0, \theta_0)$ and $\sqrt{n}(\hat{\theta}_n - \theta_0)$ are asymptotically independent[33,34]. This asymptotic independence can be also seen, if we replace $\sqrt{n}(\hat{\theta}_n - \theta_0)$ by the usual linear transformation of the corresponding score function and if we use the fact, that this score function has expectation 0 given X, which implies a correlation of 0 between this score function and $S(\tilde{\alpha}_0, \theta_0, X)$. Hence we arrive at

$$\sqrt{n}(\hat{\tilde{\alpha}}_n - \tilde{\alpha}_0) \rightharpoonup N(0, I_{\tilde{\alpha}\tilde{\alpha}}^{-1} \mathit{Var}(\tfrac{\partial}{\partial \tilde{\alpha}} S(\tilde{\alpha}, \theta, X)|_{\tilde{\alpha}_0, \theta_0}) I_{\tilde{\alpha}\tilde{\alpha}}^{-1} + I_{\tilde{\alpha}\tilde{\alpha}}^{-1} I_{\tilde{\alpha}\theta} \Sigma_{\hat{\theta}} I_{\tilde{\alpha}\theta}' I_{\tilde{\alpha}\tilde{\alpha}}^{-1})$$

where $\Sigma_{\hat{\theta}}$ denotes the asymptotic variance of $\sqrt{n}(\hat{\theta}_n - \theta_0)$. The different components of the asymptotic variance can now be estimated in the following manner. First $\mathit{Var}(\frac{\partial}{\partial \tilde{\alpha}} S(\tilde{\alpha}, \theta, X)|_{\tilde{\alpha}_0, \theta_0})$ can be estimated by the empirical estimate $\frac{1}{n}\sum_{i=1}^n (\frac{\partial}{\partial \tilde{\alpha}} S(\tilde{\alpha}, \theta, X)|_{\hat{\tilde{\alpha}}_n, \hat{\theta}_n})(\frac{\partial}{\partial \tilde{\alpha}} S(\tilde{\alpha}, \theta, X)|_{\hat{\tilde{\alpha}}_n, \hat{\theta}_n})'$. Note that $\mathit{Var}(\frac{\partial}{\partial \tilde{\alpha}} S(\tilde{\alpha}, \theta, X)|_{\tilde{\alpha}_0, \theta_0}) = 0$ in the case of a saturated model, especially in the case of one binary covariate as considered in our simulation. In this case $S_n(\tilde{\alpha}, \theta)$ behaves asymptotically like a non-stochastic function and the expression of the asymptotic variances reduces to that of Benichou and Gail[35]. We also observed in our example with non-saturated models that the contribution to the variance of the first term was always negligible small. Second $I_{\tilde{\alpha}\tilde{\alpha}}$ can be estimated by $\frac{1}{n}\sum_{i=1}^n -\frac{\partial}{\partial \tilde{\alpha}}\frac{\partial}{\partial \tilde{\alpha}} S(\tilde{\alpha}, \theta, X_i)|_{\hat{\tilde{\alpha}}_n, \hat{\theta}_n}$. This quantity is automatically computed (and inverted) if we use a standard package for logistic regression to solve the estimating equation given above. Hence the computed estimate for the asymptotic variance of $\hat{\tilde{\alpha}}_n$ provides an estimate for $I_{\tilde{\alpha}\tilde{\alpha}}^{-1}$. Finally $I_{\tilde{\alpha}\theta}$ is estimated by $\frac{1}{n}\sum_{i=1}^n -\frac{\partial}{\partial \tilde{\alpha}}\frac{\partial}{\partial \theta} S(\tilde{\alpha}, \theta, X_i)|_{\hat{\tilde{\alpha}}_n, \hat{\theta}_n}$.

References:

1. Coates, A., Abraham, S., Kaye, S.B., Sowerbutts T., Frewin, C., Fox, R.M., Tattesall, M.H.N. 'On the receiving end – patient perception of the side effects of cancer chemotherapy', *European Journal of Cancer and Clinical Oncology,* **19**, 203-208 (1983).

2. Carmichael, J., Cantwell, B.M., Edwards, C.M., Rapeport, W.G., Harris, A.L. 'The serotonin type 3 receptor antagonist BRL 43694 and nausea and vomiting induced by cisplatin', *British Medical Journal,* **297**, 110-111 (1988).

3. Gralla, R.J., Clark, R.A., Kris, M.G., Tyson, L.B. 'Methodology in anti-emetic trials', *European Journal of Cancer,* **27**, Supplement 1:5-8 (1991).

4. Tonato, M., Roila, F., del Favero, A., Ballatori, E. 'Methodology of trials with antiemetics', *Supportive Care in Cancer,* **4**, 281-286 (1996).

5. Italian Group for Antiemetic Research ' Difference in persistence of efficacy of two antiemetic regimens on acute emesis during cisplatin chemotherapy', *Journal of Clinical Oncology,* **11**, 2396-2404 (1993).

6. Italian Group for Antiemetic Research 'Persistence of efficacy of three antiemetic regimens and prognostic factors in patients undergoing moderately emetogenic chemotherapies', *Journal of Clinical Oncology,* **13**, 2417-2426 (1995).

7. Chevalier-Evain, V., Bonneterre, J., Adenis, A., Pion, J.M., Hecquet, B. 'Loss of efficiency of ondansetron-dexamethasone during successive courses in female patients receiving high-dose cisplatin', *Bulletin du Cancer,* **81**, 219-222 (1994).

8. de Wit, R., Schmitz, P.I.M., Verweij, J., de Boer-Dennert, M., de Mulder, P.H.M., Planting, A.S.T., van der Burg, M.E.L., Stoter, G. 'Analysis of cumulative probabilities shows that the efficacy of $5HT_3$ antagonist prophylaxis is not maintained', *Journal of Clinical Oncology,* **14**, 644-651 (1996).

9. Italian Group for Antiemetic Research ' Cisplatin-induced delayed emesis: Pattern and prognostic factors during three subsequent cycles', *Annals of Oncology,* **5**, 585-589 (1994).

10. Neuhaus, J.M., Hauck, W.W., and Kalbfleisch, J.D. ' The effects of mixture distribution misspecification when fitting mixed-effects logistic models', *Biometrika,* **79**, 755-762 (1992).

11. Rubin, D.B. 'Inference and missing data', *Biometrika,* **63**, 581-592 (1976).

12. Little, R.J.A. and Rubin, D.B. *Statistical Analysis with Missing Data*, Wiley, New York, 1987.

13. Diggle, P., Kenward, M.G. 'Informative drop-out in longitudinal data analysis', *Applied Statistics,* **43**, 49-93 (1994).

14. Lehmann, E.L. *Theory of point estimation*, Wiley, New York, 1983.

15. Zeger, S.L. and Liang, K.-Y. 'An overview of methods for the analysis of longitudinal data', *Statistics in Medicine,* **11**, 1825-1839 (1992).

16. Diggle, P.J., Liang, K.-Y., Zeger, S.L. *Analysis of Longitudinal Data*, Oxford Science Publications, Oxford, 1994.

17. Gail, M.H., Wieand, S., Piantadosi, S. 'Biased estimates of treatment effect in randomized experiments with nonlinear regressions and omitted covariates', *Biometrika,* **71**, 431-444 (1984).

18. Neuhaus, J.M. and Jewell, N.P. 'A geometric approach to assess bias due to omitted covariates in generalized linear models', *Biometrika,* **80**, 807-815 (1993).

19. Zeger, S.L., Liang, K.-Y., Albert, P.S. ' Models for longitudinal data: A generalized estimating equation approach', *Biometrics,* **44**, 1049-1060 (1988).

20. Neuhaus, J.M., Kalbfleisch, J.D., Hauck, W.W. ' A comparison of cluster-specific and population-averaged approaches for analyzing correlated binary data', *International Statistical Review,* **59**, 25-35 (1991).

21. Wei, L.J. & Stram, D.O. 'Analysing repeated measurements with possibly missing observations by modelling marginal distributions', *Statistics in Medicine,* **7**, 139-148 (1988).

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

23. Prentice, R.L. 'Correlated binary regression with covariates specific to each binary observation', *Biometrics,* **44**, 1033-1048 (1988).

24. Glynn, R.J., Laird, N.M. (1983) *Regression estimates and missing data: complete case analysis.* Unpublished manuscript, Department of Biostatistics, Harvard University

25. Vach, W. and Blettner, M. 'Biased estimation of the odds ratio in case-control studies due to the use of ad-hoc methods of correcting for missing values for confounding variables', *American Journal of Epidemiology,* **134**, 895-907 (1991).

26. Robins, J.M., Rotnitzky, A., and Zhao, L.P. ' Analysis of semiparametric regression models for repeated outcomes in the presence of missing data', *Journal of the American Statistical Association,* **90**, 106-121 (1995).

27. Bonney, G.E. 'Logistic regression for dependent binary observations', *Biometrics,* **43**, 951-973 (1987).

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

29. Williams, D.A. 'Extra-binomial variation in logistic linear models', *Applied Statistics,* **31**, 144-148 (1982).

30. Golaszewski, T., Kolbl, H., Lahousen, M., Staab, H.J. 'Granisetron, ein Antiemetikum fur die Behandlung der Zytostatika-induzierten Emesis: Ergebnisse einer praxisorientierten Studie (in German)', *Wiener Klinische Wochenschrift,* **107**, 613-621 (1995).

31. Mallet, A. 'A maximum likelihood estimation method for random coefficient regression models', *Biometrika,* **73**, 645-656 (1986).

32. Davidian, M., Gallant, A.R. 'The nonlinear mixed effects model with a smooth random effects density', *Biometrika,* **80**, 475-488 (1993).

33. Pierce, D.A. 'The asymptotic effect of substituting estimators for parameters in certain types of statistics', *Annals of Statistics,* **10**, 475-478 (1982).

34. Parke, W.R. 'Pseudo maximum likelihood estimation: The asymptotic distribution', *Annals of Statistics,* **14**, 355-357 (1986).

35. Benichou, J., Gail, M.H. 'A delta method for implicitly defined random variables', *The American Statistician,* **43**, 41-44 (1989).

					$\hat{\alpha}$			$\hat{\beta}$			$\hat{\gamma}$				$\hat{\sigma}$				$\hat{\Delta}$			$\hat{c}$		
α	β	σ		n	bias	sdv	cvg	bias	sdv	cvg	bias	sdv	cvg	f(0)	bias	sdv	cvg	f(0)	bias	sdv	cvg	bias	sdv	cvg
0.0	1.5	1.0	A	100	.002	.333	.962	.020	.441	.865	-.053	.248	.889	.121	-.118	.695	.927	.252	.002	.219	.680	-.012	.315	.931
0.5	1.5	1.0	A	100	.018	.364	.939	.013	.461	.849	-.045	.236	.892	.107	-.117	.683	.922	.245	-.000	.229	.694	-.014	.328	.935
1.0	1.5	1.0	A	100	.042	.431	.917	.005	.460	.890	-.042	.250	.901	.105	-.081	.663	.934	.212	.008	.235	.736	-.022	.334	.938
0.0	1.5	1.0	B	100	.011	.413	.961	-.000	.661	.937	-.021	.338	.922	.184	-.126	.880	.983	.334	.013	.323	.792	-.043	.416	.946
0.5	1.5	1.0	B	100	.052	.453	.963	-.030	.719	.928	-.001	.370	.901	.192	-.064	.951	.978	.319	.035	.377	.807	-.059	.433	.953
1.0	1.5	1.0	B	100	.082	.537	.946	-.035	.694	.929	.005	.381	.897	.185	-.083	.910	.984	.315	.029	.369	.831	-.057	.416	.941
1.0	1.5	0.0	A	100	.161	.322	.952	-.167	.366	.934	-.097	.233	.962	.128	.362	.465	.830	.533	.124	.186	.842	-.116	.274	.934
1.0	1.5	0.5	A	100	.104	.371	.946	-.075	.403	.933	-.069	.234	.926	.095	-.016	.555	.915	.438	.062	.213	.825	-.070	.306	.952
1.0	1.5	1.0	A	100	.042	.431	.917	.005	.460	.890	-.042	.250	.901	.105	-.081	.663	.934	.212	.008	.235	.736	-.022	.334	.938
1.0	1.5	1.5	A	100	-.001	.513	.899	.061	.517	.868	-.045	.254	.867	.118	-.125	.810	.915	.125	-.030	.270	.784	.003	.386	.910
1.0	1.5	2.0	A	100	.002	.615	.903	.082	.548	.874	-.031	.261	.871	.105	-.113	.925	.928	.081	-.035	.274	.834	.006	.465	.916
1.0	1.5	0.0	B	100	.217	.467	.964	-.256	.640	.941	-.062	.372	.928	.222	.427	.684	.892	.548	.181	.319	.938	-.169	.385	.971
1.0	1.5	0.5	B	100	.145	.473	.974	-.158	.629	.954	-.018	.371	.929	.184	.015	.725	.970	.483	.110	.318	.920	-.107	.367	.969
1.0	1.5	1.0	B	100	.082	.537	.946	-.035	.694	.929	.005	.381	.897	.185	-.083	.910	.984	.315	.029	.369	.831	-.057	.416	.941
1.0	1.5	1.5	B	100	.050	.641	.919	.052	.775	.872	-.014	.400	.898	.231	-.136	1.08	.940	.221	-.025	.397	.757	-.026	.498	.919
1.0	1.5	2.0	B	100	.006	.701	.893	.167	.833	.841	-.046	.589	.883	.177	-.195	1.27	.937	.162	-.084	.408	.805	.027	.516	.903
1.0	0.0	1.0	A	100	.029	.372	.911	-.033	.462	.824	-	.373	-	.257	-.027	.512	.942	.094	.014	.179	.872	-.023	.296	.941
1.0	0.8	1.0	A	100	-.012	.371	.902	.057	.380	.853	-.059	.305	.855	.187	-.144	.550	.946	.185	-.024	.186	.757	.001	.299	.915
1.0	1.5	1.0	A	100	.042	.431	.917	.005	.460	.890	-.042	.250	.901	.105	-.081	.663	.934	.212	.008	.235	.736	-.022	.334	.938
1.0	2.3	1.0	A	100	.103	.541	.919	-.038	.578	.930	-.046	.232	.936	.079	-.010	.854	.940	.264	.048	.316	.725	-.057	.398	.943
1.0	0.0	1.0	B	100	.131	.537	.938	-.179	.776	.811	-	.378	-	.276	.055	.770	.968	.173	.071	.311	.869	-.096	.412	.948
1.0	0.8	1.0	B	100	.033	.486	.942	.002	.686	.848	.029	.416	.860	.203	-.155	.814	.978	.283	-.010	.325	.827	-.024	.391	.942
1.0	1.5	1.0	B	100	.082	.537	.946	-.035	.694	.929	.005	.381	.897	.185	-.083	.910	.984	.315	.029	.369	.831	-.057	.416	.941
1.0	2.3	1.0	B	100	.140	.617	.946	-.077	.768	.971	.011	.392	.968	.175	.026	1.06	.983	.349	.088	.437	.823	-.106	.480	.946
0.0	1.5	1.0	B	100	.011	.413	.961	-.000	.661	.937	-.021	.338	.922	.184	-.126	.880	.983	.334	.013	.323	.792	-.043	.416	.946
0.0	1.5	1.0	C	100	.001	.409	.957	.018	.635	.939	-.010	.398	.934	.188	-.104	.902	.978	.302	.017	.308	.772	-.033	.407	.948
0.0	1.5	1.0	D	100	-.005	.406	.967	-.039	.656	.909	-.023	.348	.926	.176	-.074	.884	.977	.305	.032	.282	.765	-.045	.401	.940
0.0	1.5	1.0	E	100	-.011	.406	.976	-.041	.642	.904	-.028	.316	.906	.169	-.058	.909	.964	.290	.034	.313	.771	-.041	.400	.962
1.0	1.5	1.0	B	100	.082	.537	.946	-.035	.694	.929	.005	.381	.897	.185	-.083	.910	.984	.315	.029	.369	.831	-.057	.416	.941
1.0	1.5	1.0	C	100	.067	.506	.950	-.000	.687	.943	.088	.424	.927	.181	-.062	.863	.990	.293	.031	.333	.790	-.038	.400	.956
1.0	1.5	1.0	D	100	.102	.573	.957	-.069	.720	.939	-.008	.404	.932	.216	.001	.972	.983	.287	.051	.380	.816	-.057	.410	.957
1.0	1.5	1.0	E	100	.072	.558	.953	-.066	.761	.917	-.019	.381	.898	.197	-.021	.928	.974	.268	.055	.411	.806	-.061	.449	.944
1.0	1.5	1.0	A	100	.042	.431	.917	.005	.460	.890	-.042	.250	.901	.105	-.081	.663	.934	.212	.008	.235	.736	-.022	.334	.938
1.0	1.5	1.0	A	400	-.006	.244	.876	.033	.280	.852	-.021	.133	.881	.008	-.090	.477	.936	.112	-.013	.165	.824	.004	.185	.904
1.0	1.5	1.0	A	1600	-.008	.125	.940	.019	.145	.952	-.006	.063	.932	.000	-.028	.222	.971	.008	-.007	.086	.941	.004	.096	.941
1.0	1.5	1.0	A	6400	-.002	.062	.954	.004	.072	.953	-.002	.032	.939	.000	-.006	.102	.948	.000	-.002	.042	.956	.003	.047	.954
1.0	1.5	1.0	B	100	.082	.537	.946	-.035	.694	.929	.005	.381	.897	.185	-.083	.910	.984	.315	.029	.369	.831	-.057	.416	.941
1.0	1.5	1.0	B	400	.032	.294	.905	-.021	.419	.836	-.040	.223	.900	.081	-.069	.630	.894	.196	.017	.242	.724	-.027	.238	.908
1.0	1.5	1.0	B	1600	.012	.164	.914	-.004	.243	.905	-.018	.112	.906	.003	-.030	.380	.945	.050	.005	.147	.889	-.011	.132	.916
1.0	1.5	1.0	B	6400	.003	.083	.945	-.001	.143	.948	-.003	.053	.933	.000	-.004	.173	.956	.001	.000	.074	.939	-.003	.066	.942

Table 1: Finite sample properties of maximum likelihood estimates. f(0)= relative frequency of estimates equal to 0. Rows in italics are repeated several times to facilitate comparisons. Results are based on 1000 repetitions of a Monte Carlo study.

α	β	σ		n	$\tilde{\alpha}$	$\hat{\tilde{\alpha}}$ bias	sdv	cov
0.0	1.5	1.0	A	100	0.000	.000	.261	.945
0.5	1.5	1.0	A	100	0.419	.001	.267	.938
1.0	*1.5*	*1.0*	*A*	*100*	*0.833*	*.006*	*.279*	*.943*
0.0	1.5	1.0	B	100	0.000	.009	.321	.950
0.5	1.5	1.0	B	100	0.419	.016	.310	.957
1.0	*1.5*	*1.0*	*B*	*100*	*0.833*	*.024*	*.324*	*.948*
1.0	1.5	0.0	A	100	1.000	.084	.256	.948
1.0	1.5	0.5	A	100	0.948	.046	.259	.951
1.0	*1.5*	*1.0*	*A*	*100*	*0.833*	*.006*	*.279*	*.943*
1.0	1.5	1.5	A	100	0.713	-.012	.305	.945
1.0	1.5	2.0	A	100	0.610	-.005	.314	.934
1.0	1.5	0.0	B	100	1.000	.094	.314	.948
1.0	1.5	0.5	B	100	0.948	.056	.311	.958
1.0	*1.5*	*1.0*	*B*	*100*	*0.833*	*.024*	*.324*	*.948*
1.0	1.5	1.5	B	100	0.713	.010	.340	.944
1.0	1.5	2.0	B	100	0.610	-.005	.350	.954
1.0	0.0	1.0	A	100	0.833	-.003	.257	.942
1.0	0.8	1.0	A	100	0.833	-.013	.257	.940
1.0	*1.5*	*1.0*	*A*	*100*	*0.833*	*.006*	*.279*	*.943*
1.0	2.3	1.0	A	100	0.833	.018	.304	.948
1.0	0.0	1.0	B	100	0.833	.038	.310	.949
1.0	0.8	1.0	B	100	0.833	.010	.306	.956
1.0	*1.5*	*1.0*	*B*	*100*	*0.833*	*.024*	*.324*	*.948*
1.0	2.3	1.0	B	100	0.833	.027	.347	.954
0.0	1.5	1.0	B	100	0.000	.009	.321	.950
0.0	1.5	1.0	C	100	0.000	.001	.306	.939
0.0	1.5	1.0	D	100	0.000	-.010	.320	.966
0.0	1.5	1.0	E	100	0.000	-.004	.314	.950
1.0	*1.5*	*1.0*	*B*	*100*	*0.833*	*.024*	*.324*	*.948*
1.0	1.5	1.0	C	100	0.833	.012	.318	.954
1.0	1.5	1.0	D	100	0.833	.005	.322	.962
1.0	1.5	1.0	E	100	0.833	.025	.313	.951
1.0	*1.5*	*1.0*	*A*	*100*	*0.833*	*.006*	*.279*	*.943*
1.0	1.5	1.0	A	400	0.833	-.011	.141	.947
1.0	1.5	1.0	A	1600	0.833	-.005	.072	.947
1.0	1.5	1.0	A	6400	0.833	-.002	.035	.957
1.0	*1.5*	*1.0*	*B*	*100*	*0.833*	*.024*	*.324*	*.948*
1.0	1.5	1.0	B	400	0.833	.003	.170	.954
1.0	1.5	1.0	B	1600	0.833	.001	.085	.949
1.0	1.5	1.0	B	6400	0.833	.001	.041	.954

Table 2: Finite sample properties of estimates of $\tilde{\alpha}$. Rows in italics are repeated several times to facilitate comparisons. Results are based on 1000 repetitions of a Monte Carlo study.

							$\hat{\tilde{\alpha}}$				
α	β	σ		n	α	$\tilde{\alpha}$	$\tilde{\alpha}_1$	α^B	α_I^M	α_E^M	α_A^M
0.0	1.5	1.0	A	100	.038	.055	.045	.061	.067	.064	.056
0.5	1.5	1.0	A	100	.254	.375	.164	.384	.400	.366	.390
1.0	*1.5*	*1.0*	*A*	*100*	*.758*	*.878*	*.511*	*.899*	*.900*	*.867*	*.889*
0.0	1.5	1.0	B	100	.039	.050	.045	.051	.059	.056	.055
0.5	1.5	1.0	B	100	.189	.272	.164	.269	.267	.269	.281
1.0	*1.5*	*1.0*	*B*	*100*	*.589*	*.763*	*.511*	*.754*	*.729*	*.758*	*.757*
1.0	1.5	0.0	A	100	.964	.992	.650	.990	.990	.986	.987
1.0	1.5	0.5	A	100	.994	.979	.601	.980	.981	.964	.976
1.0	*1.5*	*1.0*	*A*	*100*	*.758*	*.878*	*.511*	*.899*	*.900*	*.867*	*.889*
1.0	1.5	1.5	A	100	.518	.698	.404	.697	.708	.672	.699
1.0	1.5	2.0	A	100	.348	.510	.323	.511	.521	.512	.513
1.0	1.5	0.0	B	100	.887	.955	.650	.948	.958	.946	.954
1.0	1.5	0.5	B	100	.814	.916	.601	.889	.902	.896	.911
1.0	*1.5*	*1.0*	*B*	*100*	*.589*	*.763*	*.511*	*.754*	*.729*	*.758*	*.757*
1.0	1.5	1.5	B	100	.397	.602	.404	.570	.554	.571	.589
1.0	1.5	2.0	B	100	.207	.439	.323	.421	.387	.430	.440
1.0	0.0	1.0	A	100	.925	.956	.511	.959	.958	.960	.956
1.0	0.8	1.0	A	100	.839	.907	.511	.925	.931	.928	.925
1.0	*1.5*	*1.0*	*A*	*100*	*.758*	*.878*	*.511*	*.899*	*.900*	*.867*	*.889*
1.0	2.3	1.0	A	100	.652	.829	.511	.835	.844	.737	.832
1.0	0.0	1.0	B	100	.759	.831	.511	.827	.824	.842	.832
1.0	0.8	1.0	B	100	.661	.802	.511	.794	.786	.814	.806
1.0	*1.5*	*1.0*	*B*	*100*	*.589*	*.763*	*.511*	*.754*	*.729*	*.758*	*.757*
1.0	2.3	1.0	B	100	.493	.743	.511	.721	.708	.686	.722
0.0	1.5	1.0	B	100	.039	.050	.045	.051	.059	.056	.055
0.0	1.5	1.0	C	100	.043	.061	.045	.057	.062	.048	.046
0.0	1.5	1.0	D	100	.024	.034	.045	.030	.043	.043	.039
0.0	1.5	1.0	E	100	.033	.050	.045	.209	.062	.041	.048
1.0	*1.5*	*1.0*	*B*	*100*	*.589*	*.763*	*.511*	*.754*	*.729*	*.758*	*.757*
1.0	1.5	1.0	C	100	.572	.776	.511	.639	.752	.707	.766
1.0	1.5	1.0	D	100	.534	.735	.511	.493	.703	.754	.725
1.0	1.5	1.0	E	100	.603	.782	.511	.949	.791	.764	.791

Table 3: Comparison of finite sample power. α_I^M: working correlation matrix according to independence; α_E^M: exchangeable working correlation matrix; α_A^M: working correlation matrix according to an AR(1) process. For $\tilde{\alpha}_1$ averages are reported summarizing all simulations with equal power. Rows in italics are repeated several times to facilitate comparisons. Results are based on 1000 repetitions of a Monte Carlo study.

marginal model								
α	β	σ	α_1^M	α_2^M	α_3^M	α_4^M	α_5^M	α_6^M
0.0	1.0	1.0	0.000	0.000	0.000	0.000	0.000	0.000
0.5	1.0	1.0	0.414	0.467	0.502	0.525	0.539	0.548
1.0	1.0	1.0	0.833	0.946	1.021	1.068	1.096	1.112
1.0	0.0	1.0	0.833	0.833	0.833	0.833	0.833	0.833
1.0	2.0	1.0	0.833	1.015	1.163	1.279	1.366	1.429
1.0	1.0	0.5	0.948	1.114	1.224	1.288	1.318	1.327
1.0	1.0	2.0	0.610	0.653	0.679	0.694	0.704	0.710
transition model								
α	β	σ	α_1^T	α_2^T	α_3^T	α_4^T	α_5^T	α_6^T
0.0	1.0	1.0	0.000	0.000	0.000	0.000	0.000	0.000
0.5	1.0	1.0	0.414	0.363	0.322	0.288	0.263	0.246
1.0	1.0	1.0	0.833	0.734	0.653	0.589	0.543	0.512
1.0	0.0	1.0	0.833	0.734	0.645	0.564	0.490	0.423
1.0	2.0	1.0	0.833	0.731	0.658	0.609	0.582	0.576
1.0	1.0	0.5	0.948	0.910	0.875	0.848	0.831	0.824
1.0	1.0	2.0	0.610	0.463	0.369	0.303	0.256	0.222
random intercept model								
α	β	σ	α_1^R	α_2^R	α_3^R	α_4^R	α_5^R	α_6^R
0.0	1.0	1.0	0.000	0.000	0.000	0.000	0.000	0.000
0.5	1.0	1.0	0.640	0.709	0.750	0.774	0.787	0.795
1.0	1.0	1.0	1.268	1.403	1.481	1.525	1.548	1.558
1.0	0.0	1.0	1.000	1.000	1.000	1.000	1.000	1.000
1.0	2.0	1.0	1.781	2.030	2.179	2.274	2.332	2.361
1.0	1.0	0.5	1.222	1.398	1.500	1.550	1.566	1.564
1.0	1.0	2.0	1.324	1.396	1.435	1.457	1.470	1.477

Table 4: Stochastic limits of time-varying parameter estimates in alternative models in dependence of parameters of the true model. The values of γ, Δ and c are 0.5, 0.0 and 0.0, respectively.

	$\beta = 1.0$				$\beta = 2.0$			
	A_3	A_6	B_3	B_6	A_3	A_6	B_3	B_6
marginal model (α^M)	0.925	0.995	0.904	0.941	0.971	1.084	0.937	0.962
transition model (α^T)	0.751	0.679	0.771	0.736	0.767	0.729	0.786	0.752
random intercept model (α^R)	1.407	1.443	1.376	1.386	1.957	2.030	1.894	1.924
binomial count model (α^B)	0.909	0.968	0.889	0.915	0.927	0.999	0.896	0.914

Table 5: Stochastic limits of time-constant parameter estimates in alternative models. The true values of the parameters are $\alpha = 1.0$, $\sigma = 1.0$, $\gamma = 0.5$, $c = 0.0$, and $\Delta = 0.0$. A_3: T=3 cycles and no drop outs; A_6: T=6 cycles and no drop outs; B_3: T=3 cycles and 29% drop outs after each cycle; B_6: T=6 cycles and 29% drop outs after each cycle.

The-rapy	Sex	# patients	vomiting 1st cycle	# cycles 1	2	3	4	5	6	age <45	45-60	>60	diagnoses
CMF	f	167	17%	36	17	23	25	17	49	58	74	35	mamma(167)
CC	f	40	20%	15	4	6	3	1	11	2	16	22	ovar(40)
EC	f	40	35%	12	7	8	6	2	5	10	21	9	mamma(40)
FEC	f	20	20%	2	4	2	1	3	8	9	6	5	mamma(20)
Cis	f	61	61%	21	17	11	2	5	5	18	23	20	ovar(34), uterus(6), others(21)
Cis	m	61	38%	23	16	9	12	0	1	13	31	17	head/neck(24), lung(18), others(19)

Table 6: Patient characteristics for patient groups defined by regimen and sex and investigated in the example. CMF = Cyclophosphamid + Methotrexat + 5-FU, CC = Carboplatin + Cyclophosphamid, EC = Epirubicin + Cyclophosphamid, FEC = 5-FU + Epirubicin + Cyclophosphamid, Cis = Cisplatin

Example: Part 1

	estimate	sdv	p-value
α(f)	0.89	0.48	0.064
α(45-60)	0.26	0.39	0.326
α(>60)	0.72	0.48	
β	2.05	0.67	
γ	0.38	0.26	
σ	0.66	1.02	
Δ	0.47	0.31	
c	0.83	0.43	
$\tilde{\alpha}$(f)	0.81	0.31	0.010
$\tilde{\alpha}$(45-60)	0.24	0.35	0.275
$\tilde{\alpha}$(>60)	0.65	0.41	

Example: Part 2

	estimate	sdv	p-value
α(CC)	0.11	0.46	< 0.0001
α(EC)	1.06	0.45	
α(FEC)	1.21	0.54	
α(Cis)	2.70	0.52	
α(45-60)	0.34	0.32	0.545
α(>60)	0.15	0.37	
β	0.97	0.34	
γ	0.27	0.29	
σ	1.54	0.37	
Δ	0.07	0.10	
c	2.43	0.39	
$\tilde{\alpha}$(CC)	0.08	0.35	< 0.0001
$\tilde{\alpha}$(EC)	0.79	0.30	
$\tilde{\alpha}$(FEC)	0.89	0.38	
$\tilde{\alpha}$(Cis)	1.94	0.27	
$\tilde{\alpha}$(45-60)	0.26	0.24	0.546
$\tilde{\alpha}$(>60)	0.11	0.28	

Table 7: Application of estimation procedures described in Sections 3 and 4. p-values are only given for risk estimates.

Example: Part 1

	α(f)		α(40-60)		α(>60)		β	γ	σ	Δ	further					
	est	sdv	est	sdv	est	sdv	est	est	est	est	estimates					
	0.89	0.48	0.26	0.39	0.72	0.48	2.05	0.38	0.66	0.47						
a)	0.83	0.53	0.27	0.36	0.69	0.48	2.22	0.39	0.46	-	$\hat{c}_1$=-0.78	$\hat{c}_2$=-1.51	$\hat{c}_3$=-1.54	$\hat{c}_4$=-2.46	$\hat{c}_5$=-3.47	$\hat{c}_6$=-2.82
b)	0.81	0.46	0.23	0.35	0.66	0.45	-	0.27	1.05	0.30	$\hat{\beta}_2$=2.51	$\hat{\beta}_3$=1.53	$\hat{\beta}_4$=2.49	$\hat{\beta}_5$=2.14	$\hat{\beta}_6=\infty$	
c1)	1.41	0.68	0.51	0.54	1.17	0.72	1.60	0.24	-	0.25	$\hat{p}_1$=0.64	$\hat{p}_2$=0.36	$\hat{\sigma}_1$=0	$\hat{\sigma}_2$=0	$\hat{\mu}_2$ =2.69	
c2)	1.41	0.68	0.51	0.54	1.17	0.72	1.60	0.24	-	0.25	$\hat{p}_{-1}$=0.64	$\hat{p}_0$=0.00	$\hat{p}_1 = 0.36$	$\hat{\sigma} = 1.35$		
d)	0.94	0.41	0.30	0.38	0.72	0.47	4.28	0.42	0.00	1.23	$\hat{\sigma}_2$=1.47	$\hat{\sigma}_3$=3.45	$\hat{\sigma}_4$=1.35	$\hat{\sigma}_5$=1.38	$\hat{\sigma}_6$=0.02	

Example: Part 2

	α(CC)		α(EC)		α(FEC)		α(Cis)		α(40-60)		α(>60)		β	γ	σ	Δ	further					
	est	sdv	est	sdv	est	sdv	est	sdv	est	sdv	est	sdv	est	est	est	est	estimates					
	0.11	0.46	1.06	0.45	1.21	0.54	2.70	0.52	0.34	0.32	0.15	0.37	0.97	0.27	1.54	0.07						
a)	0.07	0.43	0.98	0.43	1.14	0.52	2.50	0.54	0.34	0.29	0.15	0.35	1.17	0.29	1.38	-	$\hat{c}_1$=2.22	$\hat{c}_2$=2.47	$\hat{c}_3$=2.15	$\hat{c}_4$=2.49	$\hat{c}_5$=2.23	$\hat{c}_6$=1.84
b)	0.09	0.44	1.01	0.44	1.18	0.52	2.61	0.52	0.34	0.31	0.15	0.36	-	0.35	1.45	0.04	$\hat{\beta}_2$=1.02	$\hat{\beta}_3$=1.27	$\hat{\beta}_4$=0.44	$\hat{\beta}_5$=1.18	$\hat{\beta}_6$=1.60	
c1)	0.32	0.43	1.03	0.40	1.23	0.51	2.81	0.50	0.24	0.35	0.04	0.41	0.87	0.19	-	0.12	$\hat{p}_1$=0.43	$\hat{p}_2$=0.57	$\hat{\sigma}_1$=0	$\hat{\sigma}_2$=2.70	$\hat{\mu}_2 = -0.48$	
c2)	0.35	0.39	0.98	0.38	1.18	0.48	2.91	0.46	0.28	0.34	0.08	0.38	0.84	0.17	-	0.13	$\hat{p}_{-1}$=0.25	$\hat{p}_0$=0.63	$\hat{p}_1 = 0.13$	$\hat{\sigma} = 2.96$		
d)	0.06	0.39	0.85	0.42	1.13	0.47	2.29	0.57	0.36	0.27	0.15	0.32	0.94	0.35	1.25	-0.03	$\hat{\sigma}_2$=0.69	$\hat{\sigma}_3$=0.94	$\hat{\sigma}_4$=0.90	$\hat{\sigma}_5$=1.05	$\hat{\sigma}_6$=0.93	

Table 8: Results of sensitivity analyses. est=estimate; sdv=estimated standard deviation. The first row repeats the results of Table 7.