



## RESEARCH ARTICLE

# Generalized estimating equations to estimate the ordered stereotype logit model for panel data

Martin Spiess<sup>1</sup> | Daniel Fernández<sup>2,3</sup> | Thuong Nguyen<sup>4</sup> | Ivy Liu<sup>4</sup>

<sup>1</sup>Psychological Methods and Statistics, Hamburg University, Hamburg, Germany

<sup>2</sup>Serra Hünter fellow. Department of Statistics and Operations Research, Polytechnic University of Catalonia-BarcelonaTech, 08028 Barcelona, Spain

<sup>3</sup>Centro de Investigación Biomédica en Red de Salud Mental (CIBERSAM), Instituto de Salud, Carlos III, 28029 Madrid, Spain

<sup>4</sup>School of Mathematics and Statistics, Victoria University of Wellington, Wellington, New Zealand

**Correspondence**

Martin Spiess, Psychological Methods and Statistics, Hamburg University, Von-Melle-Park 5, 20146 Hamburg, Germany.  
Email: martin.spiess@uni-hamburg.de

**Funding information**

Ministerio de Ciencia e Innovación (MCI, Spain), Grant/Award Number: RTI2018-100927-J-I00; Agencia Estatal de Investigación (AEI, Spain); European Regional Development Fund FEDER (FEDER, UE), Grant/Award Number: E2987-3648; Royal Society of New Zealand, Grant/Award Number: 2017 SGR 622; Departament d'Economia i Coneixement de la Generalitat de Catalunya (Spain)

By modeling the effects of predictor variables as a multiplicative function of regression parameters being invariant over categories, and category-specific scalar effects, the ordered stereotype logit model is a flexible regression model for ordinal response variables. In this article, we propose a generalized estimating equations (GEE) approach to estimate the ordered stereotype logit model for panel data based on working covariance matrices, which are not required to be correctly specified. A simulation study compares the performance of GEE estimators based on various working correlation matrices and working covariance matrices using local odds ratios. Estimation of the model is illustrated using a real-world dataset. The results from the simulation study suggest that GEE estimation of this model is feasible in medium-sized and large samples and that estimators based on local odds ratios as realized in this study tend to be less efficient compared with estimators based on a working correlation matrix. For low true correlations, the efficiency gains seem to be rather small and if the working covariance structure is too flexible, the corresponding estimator may even be less efficient compared with the GEE estimator assuming independence. Like for GEE estimators more generally, if the true correlations over time are high, then a working covariance structure which is close to the true structure can lead to considerable efficiency gains compared with assuming independence.

**KEYWORDS**

generalized estimating equations, Likert scale, ordered categorical variables, panel data, simulation study

## 1 | INTRODUCTION

In applications, it is often of interest to estimate the effects of predictor variables on categorical response variables. One such example could be the effect of the predictors “consulted medical specialist”, “age”, and other variables on a response observed on an ordinal scale where patients can tick one of, say four possible answers from “not at all” to “completely”, when asked how deeply they have been involved in the therapeutic decision process.

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2020 The Authors. *Statistics in Medicine* published by John Wiley & Sons, Ltd.

In the cross-sectional context, several models are available to estimate the corresponding parameters of scientific interest. Examples are cumulative, sequential, adjacent-categories, or baseline category models.<sup>1-3</sup> These models account for the ordinal or nominal character of the response by either restricting the effects of the predictor variables to be identical over the categories (i.e., the proportional odds property) or allow all regression parameters to vary freely. Stereotype models differ from the above models by allowing the effects of predictor variables to be specific for each possible category of the response variable but in a parsimonious way.<sup>4</sup>

The ordered stereotype model assumes that the effects of the predictor variables are proportionally changing depending on the response category. This is modeled by introducing a scalar parameter that is specific to each category and is multiplied with the vector of regression parameters. Hence, this model is more flexible than models that restrict the effects of the predictors to be the same for the different categories but more parsimonious than the unrestricted models with freely varying regression parameters. The ordinal nature of the response in the ordered stereotype model is guaranteed by constraints on the order of these scalar parameters.<sup>1,4-7</sup> Thus, by introducing the interaction of regression parameters and category-specific effects underlying the ordering constraints, the ordered stereotype model allows a flexible interpretation of the effects of predictors but still respects the ordinal nature of the response. In contrast to other models for ordinal responses which can be derived from latent one-dimensional processes, the ordered stereotype model can be motivated by a latent but inherently multidimensional nature of the response variable.<sup>4</sup> The price for the increased model flexibility is that estimation of the model is not straightforward due to the multiplicative structure of the parameters.

To estimate cross-sectional-ordered stereotype models, maximum likelihood<sup>4</sup> (ML) or generalized least squares<sup>8</sup> (GLS) methods have been proposed. This can be realized via general software to maximize a likelihood function,<sup>8</sup> usually using numerical derivatives, or by embedding the ordered stereotype models into a more general class of models. For example, stereotype models can be interpreted as a subclass of generalized or multivariate nonlinear models<sup>8,9</sup> and estimated, for example, using the `gnm` package<sup>9</sup> as part of the R system,<sup>10</sup> as restricted versions of multinomial models<sup>11</sup> or as special cases of the class of reduced-rank vector generalized models,<sup>12</sup> which simplifies to a reduced-rank multinomial logit model. Finally, the package `ordinalgmifs`<sup>13</sup> as part of the R system<sup>10</sup> allows fitting ordered stereotype models when the number of parameters exceeds the sample size, using the generalized monotone incremental forward stagewise method and imposing penalties to a set of specified predictors. This package can also be used in the case of non-high-dimensional data without specifying any penalty on the predictors in the model fitting process.

Estimation of ordered stereotype panel models or models for clustered data is, however, scarce. Notable exceptions propose to estimate the ordered stereotype panel model either via GLS<sup>14</sup> or ML<sup>14,15</sup> using existing software. For consistent estimation of the parameters of interest and their (co)variances, the GLS approaches require that the mean and association structure of dummy variables representing the possible values of the ordered response variable are correctly specified. For ML estimation, correct specification of the high-dimensional distribution of the response variables over time and numerical calculation of high-dimensional integrals is required. Although the latter is no obstacle anymore with today's computing capacity, the former is a strong requirement.

The generalized estimating equations (GEE) approach was proposed as an extension of generalized linear models to panel data.<sup>16</sup> Assuming a marginal generalized linear model for the one-dimensional response variable at each time point, the parameter estimators of the mean structure are consistent and asymptotically normally distributed under weak regularity conditions. A crucial condition is that the mean structure is correctly specified. Dependencies over time are modeled via a "working" correlation matrix avoiding computer-intensive estimation procedures for nonlinear models. For the estimators of the mean structure to be consistent and asymptotically normally distributed, however, correct specification of the working correlation matrix is not required.

GEE estimation of models for one-dimensional response variables at each time point has soon been generalized to vector-valued response variables. Hence, GEE estimation of models for clustered nominal response variables has been proposed<sup>17,18</sup> just as estimation of models that include clustered ordinal response variables<sup>17,19-22</sup>—usually assuming a cumulative ordinal model.<sup>2</sup> However, to the best of our knowledge, estimation of the ordered stereotype logit model for panel data that is robust against a misspecified association structure over time has not yet been considered.

Therefore, in this article, we propose estimation of the ordered stereotype model for panel data via GEE. We choose the GEE approach because it avoids the calculation of high-dimensional integrals and is thus computationally efficient. In addition, the GEE approach turned out to be an attractive approach as illustrated in several simulation studies:<sup>16,22</sup> If the true associations are weak to moderate, then GEE estimators assuming independence are not only consistent but are usually efficient relative to ML or GEE estimators that take associations into account. If the true associations are strong, however, then explicitly modeling any association structure usually leads to more efficient GEE estimators compared with assuming independence. The observed effects on the relative efficiency may be weak if specific versions of invariant

covariate designs are considered or may be amplified if the number of observations within each cluster varies.<sup>23-26</sup> In many practical cases, the efficiency loss relative to ML estimators will often be small.

In Section 2, we discuss the ordered stereotype logit model for panel data and in Section 3 its estimation. Sections 3.1 and 3.2 describe estimation of the stereotype model with the help of GEE based on a working covariance matrix. In Section 3.2, we introduce two ways of modeling the association structure: First, via working correlation matrices and second adopting a local odds ratios approach. Section 4 describes a simulation study to illustrate the finite sample properties of the GEE estimator of the stereotype model under different structures of the working covariance matrix. Application of the GEE estimator is illustrated using a dataset from a randomized clinical design to evaluate a treatment of rheumatoid arthritis<sup>27</sup> in Section 5. Finally, in Section 6, we discuss the results from the simulation study and the application and address limitations of the approach.

## 2 | THE ORDERED STEREOTYPE LOGIT MODEL FOR PANEL DATA

Let  $\tilde{Y}_{it}$  be a  $q$ -level ordered categorical (ordinal) response with possible realizations  $\tilde{y}_{it} \in \{1, \dots, q\}$  (eg, a 5-level Likert scale defined as “strongly disagree”, “disagree”, “neutral”, “agree”, and “strongly agree”) for unit  $i$ ,  $i = 1, \dots, n$ , at time  $t = 1, \dots, T$ . The ordered stereotype panel logit model for the probability that  $\tilde{Y}_{it} = k$  can be characterized by the log odds

$$\log \left[ \frac{\Pr(\tilde{Y}_{it} = k | \mathbf{x}_{it})}{\Pr(\tilde{Y}_{it} = 1 | \mathbf{x}_{it})} \right] = \alpha_k + \mathbf{x}'_{it} \phi_k \boldsymbol{\beta}, \quad k = 2, \dots, q, \quad (1)$$

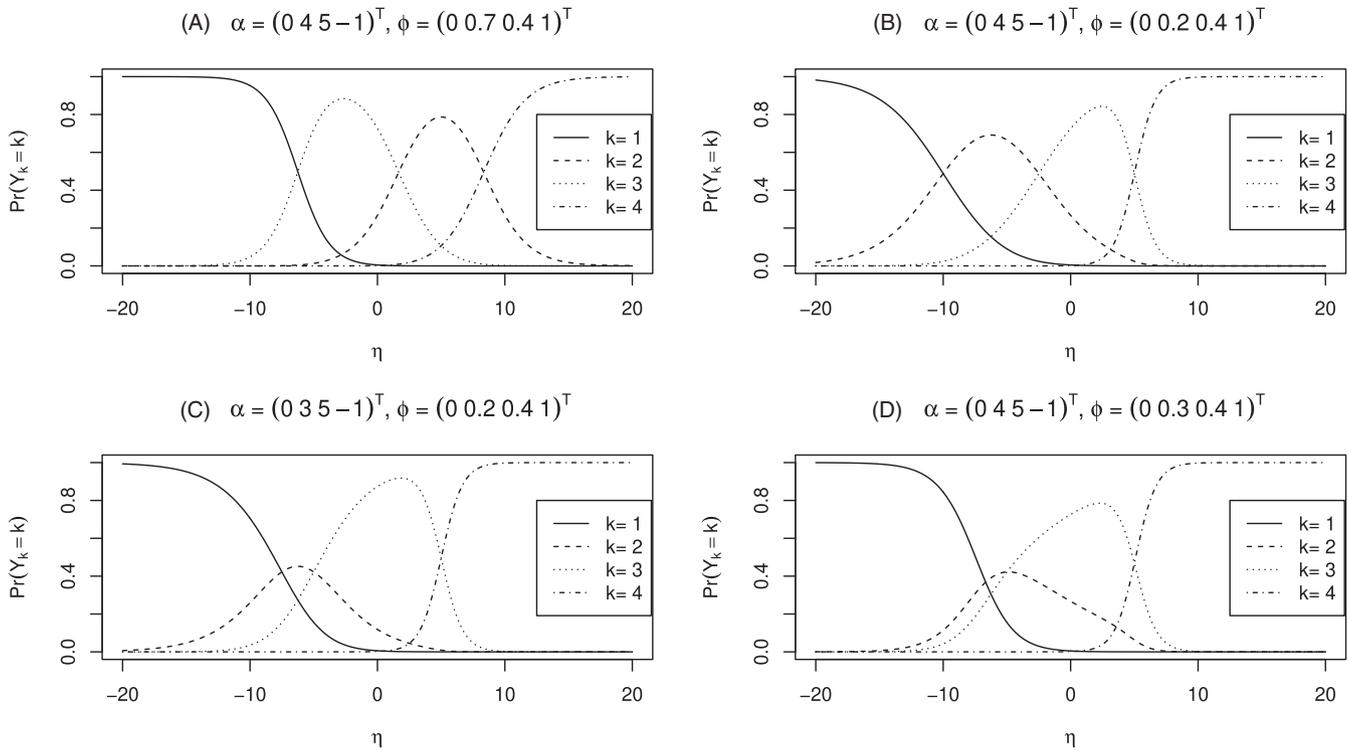
at each time point  $t$ , where  $\alpha_k$ ,  $\phi_k$ , and  $(p \times 1)$ -dimensional  $\boldsymbol{\beta}$  are usually unknown parameters, the vector  $\mathbf{x}_{it}$  is a set of  $l = 1, \dots, p$  predictor variables for unit  $i$  at time point  $t$  that may include categorical or continuous variables, is invariant over the categories but may vary over time, and  $\phi_k$  can be interpreted<sup>6</sup> as the effect of  $\mathbf{x}'_{it} \boldsymbol{\beta}$  on the probability of observing  $\tilde{Y}_{it} = k$  rather than  $\tilde{Y}_{it} = 1$  for  $k = 2, \dots, q$ . Model (1) implies

$$\begin{aligned} \Pr(\tilde{Y}_{it} = 1 | \mathbf{x}_{it}) &= 1 - \sum_{l=2}^q \Pr(\tilde{Y}_{it} = l | \mathbf{x}_{it}) = \frac{1}{1 + \sum_{l=2}^q \exp(\alpha_l + \mathbf{x}'_{it} \phi_l \boldsymbol{\beta})}, \\ \Pr(\tilde{Y}_{it} = k | \mathbf{x}_{it}) &= \frac{\exp(\alpha_k + \mathbf{x}'_{it} \phi_k \boldsymbol{\beta})}{1 + \sum_{l=2}^q \exp(\alpha_l + \mathbf{x}'_{it} \phi_l \boldsymbol{\beta})}, \quad k = 2, \dots, q, \end{aligned} \quad (2)$$

$\alpha_1 = 0$  and  $\phi_1 = 0$ . However, to identify the model, one additional restriction is required and we set  $\phi_q = 1$ . Note that  $\boldsymbol{\beta}$  is invariant over the categories and thus a monotone ordering constraint on  $\phi_k$ ,  $k = 1, \dots, q$ , is required for  $\tilde{Y}_{it}$  to be ordinal<sup>4</sup>—without this constraint, the model would be a model for nominal response variables. We choose<sup>28</sup>  $0 = \phi_1 \leq \phi_2 \leq \dots \leq \phi_{q-1} \leq \phi_q = 1$ . Note that model (1) could be formulated based on other restrictions as well. For example, instead of a model implying that  $\alpha_1 = 0$ , one could choose a version based on  $\sum_{k=1}^q \alpha_k = 0$ . For the version adopted here, the parameters to be estimated are  $\alpha_2, \dots, \alpha_q, \phi_2, \dots, \phi_{q-1}$ , and  $\boldsymbol{\beta}$ . It should be noted, however, that if  $\boldsymbol{\beta} = \mathbf{0}$ , then the model is not identified.<sup>29</sup>

A comparison of graph (A) of the first panel in Figure 1 with the other three graphs illustrates the effect of the non-decreasing restriction with respect to the parameters  $\phi_k$ ,  $k = 1, \dots, q$ , on  $\tilde{Y}_{it}$ : If the ordering constraints are met, then the assumed order in the values of  $\tilde{Y}_{it}$  is justified. Note that when two neighboring  $\phi$ -parameters have equal values, we could collapse the corresponding response categories into one single response category as the model still holds with the corresponding subset of  $\phi$ -parameters.<sup>1</sup>

Model (1) defines the first category at each time point as a reference category. The vector of parameters  $\boldsymbol{\beta}$  represents the category-invariant effects of  $\mathbf{x}_{it}$  on the log odds for the categories  $k = 2, \dots, q$  relative to the reference category  $\tilde{Y}_{it} = 1$ . The parameters  $\{\alpha_2, \dots, \alpha_q\}$  can be interpreted as weights assigned to the categories: The larger  $\alpha_k$ , the higher the probability of observing category  $k$  relative to the reference category. The parameter  $\alpha_k$  affects aspects like skewness and kurtosis of the probability functions  $\Pr(\tilde{Y}_{it} = k' | \mathbf{x}_{it})$ ,  $k \neq k'$  for a given  $i, t$ , and  $k > 1$  only via the restriction  $1 = \sum_{k=1}^q \Pr(\tilde{Y}_{it} = k | \mathbf{x}_{it})$ . For an illustration, compare graph (B) in the first panel of Figure 1 with graph (C) in the second panel. Finally,  $\{\phi_1, \phi_2, \dots, \phi_q\}$  are category-specific parameters that affect the probabilities of observing  $\tilde{Y}_{it} = k$  depending on  $\mathbf{x}'_{it} \boldsymbol{\beta}$ . Thus, they have a stronger effect on the position and shape of the probability functions  $\Pr(\tilde{Y}_{it} = k | \mathbf{x}_{it})$  for  $k = 2, \dots, q - 1$ . A comparison of



**FIGURE 1** Illustration of effects of  $\alpha_k$  and  $\phi_k$  as a function of  $\eta = \beta'x$

graphs (A) and (B) in the first panel and graph (D) in the second panel of Figure 1 illustrates this effect. Note that the assumptions of invariant parameters  $\alpha_k$ ,  $\phi_k$ , and  $\beta$  over time can easily be relaxed.

### 3 | GEE ESTIMATION

#### 3.1 | The estimating equations

Treating the ordered stereotype panel model as a multinomial model with restrictions on the parameters, we adopt the GEE approach<sup>16</sup> to estimate the parameters in a parsimonious way, avoiding numerical integration and correct specification of high-dimensional distributions. This approach requires only the correct specification of the mean structure of the dummy variables representing  $\tilde{Y}_{it}$ . However, since the variance estimator<sup>16</sup> for  $n \rightarrow \infty$  turned out to underestimate the true variances in small to medium-sized samples, we propose an analytical correction for finite samples.

Define

$$Y_{itk} = \begin{cases} 1 & \text{if } \tilde{Y}_{it} = k, \\ 0 & \text{else,} \end{cases}$$

for  $k = 2, \dots, q$  and let  $\mathbf{Y}_{it} = (Y_{it2}, \dots, Y_{itq})'$  and  $\mathbf{Y}_i = (\mathbf{Y}'_{i1}, \dots, \mathbf{Y}'_{iT})'$ . Define

$$\mu_{itk} = \frac{\exp(\alpha_k + \mathbf{x}'_{it}\phi_k\beta)}{1 + \sum_{l=2}^q \exp(\alpha_l + \mathbf{x}'_{it}\phi_l\beta)}, \quad k = 1, \dots, q,$$

and let  $\boldsymbol{\mu}_{it} = (\mu_{it2}, \dots, \mu_{itq})'$  and  $\boldsymbol{\mu}_i = (\boldsymbol{\mu}'_{i1}, \dots, \boldsymbol{\mu}'_{iT})'$ . The marginal distribution of  $\mathbf{Y}_{it}$  is multinomial and we assume that variables  $\mathbf{Y}_i$  are independent conditional on the predictor variables  $\mathbf{x}_{i'1}, \dots, \mathbf{x}_{i'T}$ , for all  $i, i' = 1, \dots, n$  and  $t = 1, \dots, T$ .

Let  $\boldsymbol{\theta} = (\boldsymbol{\alpha}', \boldsymbol{\phi}', \boldsymbol{\beta}')$ , where  $\boldsymbol{\alpha} = (\alpha_2, \dots, \alpha_q)'$  and  $\boldsymbol{\phi} = (\phi_2, \dots, \phi_{q-1})'$ , be the parameters of scientific interest. The GEE estimator  $\hat{\boldsymbol{\theta}}$  solves the estimating equations

$$\mathbf{0} = \sum_i \mathbf{D}_i \mathbf{W}_i^{-1} (\mathbf{Y}_i - \boldsymbol{\mu}_i), \tag{3}$$

where  $\mathbf{D}_i = \partial \boldsymbol{\mu}_i / \partial \boldsymbol{\theta}$  and  $\mathbf{W}_i = \mathbf{V}_i^{1/2} \mathbf{R} \mathbf{V}_i^{1/2}$  is the working covariance matrix. The positive definite matrix  $\mathbf{V}_i$  is the covariance matrix of variables  $\mathbf{Y}_{it}$  treated as independent over time, that is,  $\mathbf{V}_i = \text{bdiag}(\text{diag}(\boldsymbol{\mu}_{i1}) - \boldsymbol{\mu}_{i1}\boldsymbol{\mu}'_{i1}, \dots, \text{diag}(\boldsymbol{\mu}_{iT}) -$

$\mu_{iT} \boldsymbol{\mu}'_{iT}$ ), where  $\text{bdiag}()$  denotes a block-diagonal and  $\text{diag}()$  denotes a diagonal matrix.  $\mathbf{V}_i^{1/2}$  denotes the symmetric square root of  $\mathbf{V}_i$  and  $\mathbf{R}$  is a “working” correlation matrix, which depends on a “nuisance” parameter  $\xi$  that is not of scientific interest<sup>16</sup> but accounts for (linear) dependencies in  $\mathbf{Y}_{it} | \mathbf{x}_{i1}, \dots, \mathbf{x}_{iT}$  over time,  $t = 1, \dots, T$ .

Generally, estimating equations (3) need to be solved iteratively. Given estimate  $\hat{\xi}_j$  at iteration  $j$ , an estimate for  $\theta$ ,  $\hat{\theta}_j$ , is computed as<sup>16</sup>

$$\hat{\theta}_j = \hat{\theta}_{j-1} - \left[ \left( \sum_i \mathbf{D}_i \mathbf{W}_i^{-1} \mathbf{D}'_i \right)^{-1} \sum_i \mathbf{D}_i \mathbf{W}_i^{-1} (\mathbf{Y}_i - \boldsymbol{\mu}_i) \right]_{\theta = \hat{\theta}_{j-1}, \xi = \hat{\xi}_{j-1}}, \tag{4}$$

where  $-\sum_i \mathbf{D}_i \mathbf{W}_i^{-1} \mathbf{D}'_i$  approximates the derivative of  $\sum_i \mathbf{D}_i \mathbf{W}_i^{-1} (\mathbf{Y}_i - \boldsymbol{\mu}_i)$  with respect to  $\theta$ . To calculate an estimate for  $\hat{\theta}$ , we adopt the strategy of alternating between a modified Fisher scoring to estimate  $\theta$  and estimation of  $\xi$ .

If the model of the mean  $\boldsymbol{\mu}_i$  is correctly specified for all  $i = 1, \dots, n$ , that is, if  $E(Y_{itk} | \mathbf{x}_{it}) = \mu_{itk}$  at the true value  $\theta_0$ , then under some additional regularity conditions,  $\hat{\theta}$  is consistent and asymptotically normally distributed,<sup>16,30,31</sup> and its covariance matrix can be estimated by

$$\widehat{\text{Var}}(\hat{\theta} | \mathbf{X}) = \left( \sum_i \hat{\mathbf{D}}_i \hat{\mathbf{W}}_i^{-1} \hat{\mathbf{D}}'_i \right)^{-1} \sum_i \hat{\mathbf{D}}_i \hat{\mathbf{W}}_i^{-1} \hat{\mathbf{S}}_i \hat{\mathbf{W}}_i^{-1} \hat{\mathbf{D}}'_i \left( \sum_i \hat{\mathbf{D}}_i \hat{\mathbf{W}}_i^{-1} \hat{\mathbf{D}}'_i \right)^{-1}, \tag{5}$$

where

$$\hat{\mathbf{S}}_i = (\mathbf{Y}_i - \hat{\boldsymbol{\mu}}_i)(\mathbf{Y}_i - \hat{\boldsymbol{\mu}}_i)' |_{\theta = \hat{\theta}},$$

$\hat{\mathbf{D}}_i$  is  $\mathbf{D}_i$  evaluated at  $\theta = \hat{\theta}$  and  $\hat{\mathbf{W}}_i$  is the working covariance matrix evaluated at  $\theta = \hat{\theta}$  and the estimated association parameters  $\hat{\xi}$ , the latter being equal to zero if independence is assumed.

However, an outstanding characteristic of the stereotype model is that the means  $\mu_{itk}$  are functions of the product of parameters  $\phi_k$  and  $\beta$ . As a consequence,  $\mathbf{D}_i = \partial \boldsymbol{\mu}_i / \partial \theta$  is actually  $\mathbf{D}_i = \mathbf{D}_{(1)} \mathbf{D}_{(2),i}$ ,

$$\mathbf{D}_{(1)} = \partial \boldsymbol{\zeta} / \partial \theta = \begin{pmatrix} \mathbf{I}_{q-1} & \mathbf{0}_{(q-1) \times p(q-1)} \\ \mathbf{0}_{(q-2) \times (q-1)} & (\mathbf{I}_{q-2} \mathbf{0}_{(q-2) \times 1}) \otimes \boldsymbol{\beta}' \\ \mathbf{0}_{p \times (q-1)} & (\boldsymbol{\phi}' \mathbf{1})' \otimes \mathbf{I}_p \end{pmatrix},$$

$$\mathbf{D}_{(2),i} = \partial \boldsymbol{\mu}_i / \partial \boldsymbol{\zeta} = \tilde{\mathbf{X}}'_i \mathbf{V}_i, \quad \tilde{\mathbf{X}}_i = (\mathbf{1}_T \otimes \mathbf{I}_{q-1} \quad \mathbf{X}_i), \quad \mathbf{X}_i = \begin{pmatrix} \mathbf{I}_{q-1} \otimes \mathbf{x}'_{i1} \\ \vdots \\ \mathbf{I}_{q-1} \otimes \mathbf{x}'_{iT} \end{pmatrix},$$

where  $\boldsymbol{\zeta} = [\boldsymbol{\alpha}' \quad (\boldsymbol{\phi}' \mathbf{1}) \otimes \boldsymbol{\beta}']'$ ,  $\otimes$  is the Kronecker product,  $\mathbf{1}_J$  is a  $(J \times 1)$  vector of ones,  $\mathbf{I}_J$  is the  $(J \times J)$  identity matrix, and  $\mathbf{0}_{J \times K}$  is a  $(J \times K)$  matrix of zeros.

Treating the covariance matrix  $\mathbf{W}_i$  as fixed at the observed values, the derivative of estimating equations (3) with respect to  $\theta$  is

$$\begin{aligned} \frac{\partial}{\partial \theta} \sum_i \mathbf{D}_i \mathbf{W}_i^{-1} (\mathbf{Y}_i - \boldsymbol{\mu}_i) &= \frac{\partial \mathbf{D}_{(1)}}{\partial \theta} \sum_i [\mathbf{I}_{(2q-3+p)} \otimes \tilde{\mathbf{X}}'_i \mathbf{V}_i \mathbf{W}_i^{-1} (\mathbf{Y}_i - \boldsymbol{\mu}_i)] \\ &+ \sum_i \frac{\partial \mathbf{V}_i}{\partial \theta} [\tilde{\mathbf{X}}_i \mathbf{D}'_{(1)} \otimes \mathbf{V}_i \mathbf{W}_i^{-1} (\mathbf{Y}_i - \boldsymbol{\mu}_i)] - \sum_i \mathbf{D}_i \mathbf{W}_i^{-1} \mathbf{D}'_i. \end{aligned} \tag{6}$$

A major difference between the models for which GEE was proposed<sup>16</sup> and the ordered stereotype model is  $\mathbf{D}_{(1)}$ , a matrix which would be equal to the identity matrix in most cases, but is a function of parameters  $\phi$  and  $\beta$  in case of the stereotype model. Consequently, the first term on the right-hand side of (6) would be zero in standard GEE situations. Because the estimating equations (3) are solved with  $\mathbf{D}_i = \mathbf{D}_{(1)} \tilde{\mathbf{X}}'_i \mathbf{V}_i$  in case of the stereotype model, which does not imply that  $\sum_i \tilde{\mathbf{X}}'_i \mathbf{V}_i \mathbf{W}_i^{-1} (\mathbf{Y}_i - \boldsymbol{\mu}_i)$  is zero at the estimates  $\hat{\theta}$  and  $\hat{\xi}$  in finite samples, the first term on the right-hand side of (6) may not vanish. The second term on the right-hand side, which is ignored in (4) and (5), may also not vanish in small samples. Hence,  $-\sum_i \mathbf{D}_i \mathbf{W}_i^{-1} \mathbf{D}'_i$  may not be a close approximation to  $\partial [\sum_i \mathbf{D}_i \mathbf{W}_i^{-1} (\mathbf{Y}_i - \boldsymbol{\mu}_i)] / \partial \theta$  in small to medium-sized samples. Therefore, to calculate  $\hat{\theta}$  and to estimate its variance, we use the first and the second term in addition to the last term on the right-hand side of (6). Both terms are given in the Appendix A1. It should be noted, however, that although the

approximation  $-\sum_i \mathbf{D}_i \mathbf{W}_i^{-1} \mathbf{D}_i'$  is negative definite as long as  $\theta$  does not take extreme values and the working correlation matrix is positive definite, (6) may not be negative definite for values of  $\theta$  outside a close neighborhood of the solution  $\hat{\theta}$ .

Denote the derivative (6) as  $\mathbf{A}_n$ . Then the GEE estimator of the ordered stereotype panel model can be calculated by replacing  $-\sum_i \mathbf{D}_i \mathbf{W}_i^{-1} \mathbf{D}_i'$  in (4) by  $\mathbf{A}_n$ . In case of convergence problems, approximation  $-\sum_i \mathbf{D}_i \mathbf{W}_i^{-1} \mathbf{D}_i'$  may be used instead of  $\mathbf{A}_n$  until convergence and then be replaced by  $\mathbf{A}_n$  for a few additional iteration steps until the convergence criterion is again met. This strategy requires usually more iterations but will more likely converge to a solution of (4). The variance of  $\hat{\theta}$  is estimated by replacing  $-\sum_i \mathbf{D}_i \mathbf{W}_i^{-1} \mathbf{D}_i'$  in (5) by  $\mathbf{A}_n$ .

The ordinal nature of the stereotype model is guaranteed by the restriction  $0 \leq \phi_2 \leq \dots \leq \phi_{q-1} \leq 1$ , where at least two inequalities must be strict inequalities. These restrictions are neither necessary for the model to be a valid model nor is it required for the algorithm to solve the estimating equations. The ordering only ensures that the categorical response variable is ordered for the dataset at hand. The computed estimates  $\hat{\phi}_k$  are realizations of estimators for which only asymptotic properties are available. In the GEE case, the asymptotic distribution of the estimators is the normal distribution. Thus, in finite samples confidence intervals for  $\phi_k$  may overlap,<sup>1,4</sup> may cover negative values, or values larger than one even if the above restrictions are met for the underlying true data generating process. Consequently, it can also not be ruled out that the above ordering is violated during some iterations in the estimation process. Thus, in applications it might be of interest to test whether the response is in fact ordinal.<sup>4,8</sup> Therefore, we did not enforce the restrictions on the parameter estimates  $\hat{\phi}_k, k = 2, \dots, q - 1$ , by the estimation algorithm.<sup>32</sup>

### 3.2 | The working covariance matrix

For a consistent estimation of the mean structure parameter  $\theta$ , a correct specification of the underlying association structure of  $\mathbf{Y}_i | \mathbf{X}_i$  is not necessary and, if the observations are assumed to be correlated,  $\hat{\xi}$  needs only be consistent<sup>16</sup> for some  $\xi$ . However, the choice of the working association structure may have an effect on the relative efficiency of  $\hat{\theta}$ . To the best of our knowledge, there do not appear to exist simulation studies that consider GEE estimation of stereotype panel models, but simulation studies based on various other models suggest that if the true associations are weak or moderate, which is the case in many applications, then the choice of the working association structure seems to play only a minor role especially in sufficiently large samples.<sup>18,33</sup> If associations are strong, then efficiency improvements can be substantial.<sup>16,18,25,26,34</sup> On the other hand, it has been shown, that these effects depend on the amount of variation in cluster sizes and the covariate design, where for specific designs of invariant predictors, the choice of the working association structure may play only a minor or no role at all.<sup>18,23-26</sup>

In the context of models with binary response variables and working correlation matrices, it has been pointed out that the range of possible correlations is restricted<sup>35</sup> and alternatives to a working correlation matrix have been proposed based on global or local odds ratios to model associations.<sup>18,36</sup> It is, however, not quite clear what the practical consequences of violations of the corresponding restrictions are. Adopting an approach to estimate the correlation structure parameters  $\xi$  in a binary model in a simulation study, it turned out that violations of the restrictions were associated with convergence problems, like multiple solutions of the estimating equations or negative estimated variances for single parameters of scientific interest.<sup>37</sup> In this study, extreme situations with sample sizes as small as  $n = 20$  and  $T = 4$  were simulated. Another simulation study with  $n$  ranging from 50 to 1000 and  $T = 5$  reported rather high percentages of violations of up to 47% in a binary model but convergence problems only in one out thousands of simulated datasets and no negative variance estimates.<sup>38</sup> Therefore and because of its simplicity and computational efficiency, we adopt the working correlation matrix approach,<sup>16</sup> but compare it with an approach using local odds ratios.<sup>18,39</sup>

#### 3.2.1 | Working correlation matrix

In the cross-sectional case with ordered or unordered categorical variables, the covariance matrix of the binary indicators  $(Y_{ik}, Y_{ik'})$  is given by

$$\hat{\mathbf{V}}_i = \text{bdiag}[\text{diag}(\hat{\boldsymbol{\mu}}_{i1}) - \hat{\boldsymbol{\mu}}_{i1} \hat{\boldsymbol{\mu}}_{i1}', \dots, \text{diag}(\hat{\boldsymbol{\mu}}_{iT}) - \hat{\boldsymbol{\mu}}_{iT} \hat{\boldsymbol{\mu}}_{iT}'],$$

where  $\hat{\mathbf{V}}_i = \hat{\mathbf{V}}_i(\hat{\theta})$  and  $\hat{\boldsymbol{\mu}}_i = \hat{\boldsymbol{\mu}}_i(\hat{\theta})$ . In the case of panel data, however, variables  $Y_{itk}$  and  $Y_{it'k'}$ ,  $t, t' = 1, \dots, T$  and  $k, k' = 2, \dots, q$ , are not independent for different  $t, t'$ .

Assuming independence in the ordinal model for panel data amounts to adopting the identity matrix as the working correlation and  $\hat{\mathbf{V}}_i$  as the working covariance matrix. The GEE estimator calculated under the independence assumption will be denoted as  $GEE_I$ .

In the panel data context, this assumption is not plausible and, if wrong, may lead to inefficient estimators. There are several possible ways to proceed, depending on the permitted flexibility of the working correlation matrix to be estimated. Note that the  $(q - 1)T \times (q - 1)T$  working correlation matrix consists of  $T(T - 1)/2$  off-diagonal  $(q - 1) \times (q - 1)$  matrices, whose structures need to be modeled. Thus, the number of parameters to be estimated may vary between one and  $(q - 1)^2 T(T - 1)/2$ . We consider two basic strategies to generate a working correlation matrix that differ with respect to the flexibility of the working correlation structure and the number of parameters to be estimated. Both simplify to the above block-diagonal matrix  $\hat{\mathbf{V}}_i$  under the independence assumption. Under both strategies, four different association structures are modeled, that is, an equicorrelation structure, an autoregressive structure of order one (AR(1)), a Toeplitz structure and an unstructured correlation matrix.

In all cases, we first calculate an unstructured covariance matrix of the residuals over all units,

$$\hat{\mathbf{U}} = \begin{pmatrix} \hat{\mathbf{U}}_{11} & \dots & \hat{\mathbf{U}}_{1T} \\ \vdots & \ddots & \vdots \\ \hat{\mathbf{U}}_{T1} & \dots & \hat{\mathbf{U}}_{TT} \end{pmatrix} = \frac{1}{n} \sum_{i=1}^n \hat{\mathbf{S}}_i,$$

with  $[(q - 1) \times (q - 1)]$ -blocks  $\hat{\mathbf{U}}_{tt'}$  representing moment estimates of covariances  $\text{cov}(Y_{itk}, Y_{it'k'})$  if  $t \neq t'$  and variances and covariances if  $t = t'$ . A correlation matrix  $\hat{\mathbf{R}}$  common to all units that captures associations over time is then calculated as

$$\hat{\mathbf{R}} = \text{bdiag} \left( \hat{\mathbf{U}}_{11}^{-1/2}, \dots, \hat{\mathbf{U}}_{TT}^{-1/2} \right) \hat{\mathbf{U}} \text{bdiag} \left( \hat{\mathbf{U}}_{11}^{-1/2}, \dots, \hat{\mathbf{U}}_{TT}^{-1/2} \right),$$

where  $\hat{\mathbf{U}}_{tt}^{-1/2}$  denotes the inverse of the symmetric square root of  $\hat{\mathbf{U}}_{tt}$ , the  $t$ th  $(T \times T)$  matrix on the diagonal of  $\hat{\mathbf{U}}$ . Thus,

$$\hat{\mathbf{R}} = \begin{pmatrix} \hat{\mathbf{R}}_{11} & \dots & \hat{\mathbf{R}}_{1T} \\ \vdots & \ddots & \vdots \\ \hat{\mathbf{R}}_{T1} & \dots & \hat{\mathbf{R}}_{TT} \end{pmatrix},$$

with elements  $\hat{\mathbf{R}}_{tt'} = \hat{\mathbf{U}}_{tt}^{-1/2} \hat{\mathbf{U}}_{tt'} \hat{\mathbf{U}}_{t't}^{-1/2}$ . The diagonal block matrices reduce to  $[(q - 1) \times (q - 1)]$ -identity matrices.

The two basic strategies of generating working correlation matrices differ systematically with respect to the treatment of the off-diagonal elements of the  $[(q - 1) \times (q - 1)]$  matrices  $\hat{\mathbf{R}}_{tt'}$ . Note that due to the symmetry of correlation matrices we only need to consider either the block matrices  $\hat{\mathbf{R}}_{tt'}$  in the upper ( $t < t'$ ) or  $\hat{\mathbf{R}}_{t't}$  the lower ( $t > t'$ ) triangular part of  $\hat{\mathbf{R}}$ .

Under the restrictive within-blocks strategy, we calculate the mean of all off-diagonal elements of those block matrices  $\hat{\mathbf{R}}_{tt'}$ ,  $t < t'$ , which are, given the adopted working correlation structure, assumed to be equal using the Fisher (and its inverse) transformation. If an unstructured correlation matrix is assumed, then all matrices  $\hat{\mathbf{R}}_{tt'}$  are considered independently from each other. In this case, for each block matrix, an individual mean of the off-diagonal elements is calculated and used as an estimate of the off-diagonal correlations in this block. For example, a  $(3 \times 3)$  block  $\hat{\mathbf{R}}_{tt'}$  with correlations  $r_{tk't'k'}$ , where  $t < t'$ ,

$$\hat{\mathbf{R}}_{tt'} = \begin{pmatrix} - & r_{t2t'3} & r_{t2t'4} \\ r_{t3t'2} & - & r_{t3t'4} \\ r_{t4t'2} & r_{t4t'3} & - \end{pmatrix}, \quad \text{becomes} \quad \bar{\mathbf{R}}_{tt'} = \begin{pmatrix} - & \bar{r}_{tt'} & \bar{r}_{tt'} \\ \bar{r}_{tt'} & - & \bar{r}_{tt'} \\ \bar{r}_{tt'} & \bar{r}_{tt'} & - \end{pmatrix}$$

under the restrictive strategy, where  $\bar{r}_{tt'}$  is the mean of the correlations  $r_{tk't'k}$ ,  $k \neq k'$ . If an AR(1)- or a Toeplitz structure is assumed, then the mean over all off-diagonal elements of all block matrices  $\hat{\mathbf{R}}_{tt'}$ ,  $t < t'$ , for fixed  $\Delta = |t - t'|$  is calculated, where  $\Delta = 1, 2, \dots, T - 1$ . Assuming an equicorrelation structure amounts to calculating the mean of all off-diagonal elements of all block matrices  $\hat{\mathbf{R}}_{tt'}$ ,  $t < t'$ .

Calculation of the diagonal elements of each  $\hat{\mathbf{R}}_{tt'}$  under the restrictive within-blocks strategy depends on the adopted working correlation structure. If an unstructured, a Toeplitz or an equicorrelation structure is assumed, then the diagonal elements of the blocks  $\bar{\mathbf{R}}_{tt'}$  of the working correlation matrix are the means of the diagonal elements of all blocks  $\hat{\mathbf{R}}_{tt'}$ ,  $t < t'$ , assumed to be identical. If an AR(1)-structure is assumed then the diagonal elements are given by  $\hat{\xi}^\Delta$ , where  $\hat{\xi}$  is the solution to  $\min_{\xi} \sum_{t=1}^{T-1} \sum_{t'=t+1}^T \sum_{k=2}^q (r_{tk'tk} - \xi^\Delta)^2$ .

Under the flexible within-blocks strategy, the step of generating identical off-diagonal elements within the blocks  $\bar{\mathbf{R}}_{it'}$  is avoided. The off-diagonal elements within blocks  $\bar{\mathbf{R}}_{it'}$  are given by  $r_{tk't'}$ . Similarly, the diagonal elements within blocks  $\bar{\mathbf{R}}_{it'}$  are not restricted to be equal if the working correlation matrix that is assumed to be unstructured has a Toeplitz or an equicorrelation structure. In case of an AR(1)-structure, the diagonal elements are calculated as under the restrictive within-blocks strategy and thus are all equal. The number of correlation structure parameters varies considerably, depending on the within-blocks strategy and the working association structure. For example, the assumption of a flexible within-blocks unrestricted structure implies that all elements in all blocks  $\bar{\mathbf{R}}_{it'}$  may differ. Hence, the number of correlation parameters to be estimated is  $(q-1)^2 T(T-1)/2$ . On the other hand, if an equicorrelation structure under the flexible within-blocks strategy is assumed, then all blocks  $\bar{\mathbf{R}}_{it'}$  are equal. In that case, only  $(q-1)^2$  parameters are required. Under the restrictive within-blocks strategy, assuming an equicorrelation matrix requires estimation of only two parameters.

Working correlation matrices  $\hat{\mathbf{R}}$  calculated assuming independence, an unstructured, Toeplitz, AR(1) or equicorrelation matrix will be indicated by subscripts  $I, U, T, A$ , and  $E$ , respectively. Superscripts  $R$  and  $F$  will denote the restrictive and the flexible within-blocks version. Finally, individual working covariance matrices are generated by

$$\widehat{\mathbf{W}}_{l,i}^m = \widehat{\mathbf{V}}_i^{1/2} \widehat{\mathbf{R}}_l^m \widehat{\mathbf{V}}_i^{1/2},$$

where  $l \in \{I, U, T, A, E\}$ ,  $m \in \{R, F\}$ , and  $\widehat{\mathbf{V}}_i^{1/2} = \text{bdiag}(\widehat{\mathbf{V}}_{i1}^{1/2}, \dots, \widehat{\mathbf{V}}_{iT}^{1/2})$ . GEE estimators will be indicated correspondingly as  $\text{GEE}_l^m$ , where the superscript is ignored in case of the  $\text{GEE}_l$  estimator.

### 3.2.2 | Association structure based on local odds ratios

To avoid possible problems of inconsistencies of the marginal distributions  $\boldsymbol{\mu}_i$  and the distribution implied by the working covariance matrix if a working correlation matrix is adopted, we in addition adopt a local odds ratios approach to model the association structure.<sup>18</sup>

In a first step,  $(q-1)^2 T(T-1)/2$  local odds ratios  $\vartheta_{tk't'}$  are estimated, where  $t < t'$ ,  $t = 1, \dots, T-1$ ,  $t' = 2, \dots, T$ , and  $k, k' = 2, \dots, q$ , based on the corresponding  $(q-1) \times (q-1)$  blocks of a  $[(q-1)T \times (q-1)T]$  contingency table of variables  $\mathbf{Y}_i$ ,  $i = 1, \dots, n$ , ignoring any predictor variables and assuming independence over the  $T(T-1)/2$  blocks.

In a second step, one out of four versions of Goodman's<sup>40</sup> row and column (RC) effects model is estimated, where the log of the local odds ratios is modeled via

$$\log \vartheta_{tk't'} = \tilde{\xi}_{tt'} (v_{t(k-1)}^{tt'} - v_{tk}^{tt'}) (v_{t'(k'-1)}^{tt'} - v_{t'k'}^{tt'}),$$

$\tilde{\xi}_{tt'}$  is a parameter representing the association strength between  $\mathbf{Y}_{it}$  and  $\mathbf{Y}_{it'}$ , and  $v_{tk}^{tt'}$  are scores assigned to the  $k$ th response category at time point  $t$  if the time pair  $t, t'$  is considered.<sup>18</sup> The four different models imply four different association structures.

A parsimonious uniform structure is realized by assigning fixed unit-spaced values to the score parameters  $v_{tk}^{tt'}$  and assuming  $\tilde{\xi} = \tilde{\xi}_{tt'}$  for all  $t, t'$ .<sup>18</sup> This implies equal association strength over categories and time,  $\log \vartheta_{tk't'} = \tilde{\xi}$ . Allowing  $\log \vartheta_{tk't'} = \tilde{\xi}_{tt'}$ , that is, the association strength to vary over the  $T(T-1)/2$  blocks, but be invariant over categories within blocks, models an association structure denoted as category exchangeability.<sup>18</sup>

Restricting  $\tilde{\xi} = \tilde{\xi}_{tt'}$ ,  $v_k = v_{tk}^{tt'}$  for all  $t, t', k$  and assuming that  $v_k$  are also unknown parameters to be estimated, implies an association structure denoted as time exchangeability.<sup>18</sup> This model allows the association strength to vary with the involved categories but assumes symmetry in the sense that  $\vartheta_{tk't'} = \vartheta_{t'k't}$ , and allows no time dependence in the variables  $\mathbf{Y}_{it}$ . A fourth model denoted as RC model<sup>18</sup> allows in addition the association strength to vary over time. This model is given by  $\log \vartheta_{tk't'} = \tilde{\xi}_{tt'} (v_{k-1}^{tt'} - v_k^{tt'}) (v_{k'-1}^{tt'} - v_{k'}^{tt'})$ , where the parameters  $v_k^{tt'}$  are unknown and need to be estimated. Thus, this structure is very flexible but requires estimation of many parameters.

In a third step, based on the local odds ratios, the individual bivariate probabilities are estimated<sup>18</sup> and then adapted to be consistent with the estimated marginal distributions of  $\mathbf{Y}_i$ ,  $i = 1, \dots, n$ , using the iterative proportional fitting (IPF) algorithm.<sup>41</sup> Note that the association structure is estimated ignoring the predictors and hence this step can be performed before the iterations to calculate the GEE estimator start. The IPF step, however, is applied in each iteration. If a model for ordinal responses is estimated and it can be assumed that the true association structure is close to being time and category exchangeable, then the uniform association structure may be



appropriate.<sup>18</sup> Furthermore, results from a simulation study imply that convergence problems can be expected to increase with increasing flexibility of the association structure. The RC structure is only recommended for large sample sizes,<sup>18</sup> like  $n = 500$ . The results of this three-step approach are individual estimated working covariance matrices  $\widehat{\mathbf{W}}_i$ .

In the simulation study to be described in the next section, we compare GEE estimators based on a working correlation matrix with GEE estimators using the local odds ratios approach. The two approaches do not generate working covariance matrices that are easily comparable. However, with respect to their flexibility, the GEE estimator based on the uniform structure can be compared with the GEE estimator assuming the equicorrelation structure under the restrictive within-blocks strategy and the GEE estimator assuming time exchangeability with the GEE estimator based on the equicorrelation structure under the flexible within-blocks strategy. The GEE estimator assuming category exchangeability can be compared with the GEE estimator with an unrestricted correlation structure under a restrictive within-blocks strategy and the GEE estimator calculated under an RC strategy can be compared with a GEE estimator with an unrestricted correlation structure under a flexible within-blocks strategy. The GEE estimators using the uniform, the category exchangeable, the time exchangeable, and the RC structure will be denoted by  $GEE_{UN}$ ,  $GEE_{CE}$ ,  $GEE_{TE}$ , and  $GEE_{RC}$ , respectively.

## 4 | SIMULATION STUDY

### 4.1 | Design of the study

In this section we present the results of a simulation study to evaluate the finite sample properties of the GEE estimator adopting different working covariance structures. All programs were written and run under R versions<sup>10</sup> 3.5.3 to 3.6.2 and are available from the corresponding author upon request.

For the main part of the study, we generated data for  $n = 50, 100, 200, 300, 500$  and  $n = 1000$  units, observed at  $T = 5$  time points. For every unit we simulated a response variable with  $q = 4$  possible values at each time point in each of 1000 simulations. To approximate an applied situation we generated three predictor variables from different distributions with low to medium associations between them and over time. The predictor values were held fixed over the simulations. The first predictor was generated from a normal distribution and was correlated over time within each unit with correlation 0.3. The second predictor was a binary variable which was correlated with the first predictor with correlation of approximately 0.35 and correlations close to zero over time. The third predictor was generated from a  $\chi^2$ -distribution with three degrees of freedom, again correlated over time with correlation 0.4, but independent from the first two predictor variables. The true values  $\alpha = (0.3 \ 0 \ 0.4)'$ ,  $\phi = (0.3 \ 0.7)'$ , and  $\beta = (-.5 \ 1 \ -.1)'$  were chosen such that the fraction of observations in each category of the response variable at all time points was well above zero.

To study the effect of time-invariant and unit-invariant predictor variables on the relative efficiency of GEE estimators, we generated datasets as described above but with three predictors generated independently from each other from a standard normal distribution. The first variable was generated independently over time and units, the second was invariant over time but varying over units, and the third predictor was invariant over units but varied over time. As true values for  $\beta$  we chose  $\beta = (1 \ 1 \ 1)'$ .

To simulate the response variables from the required distributions, variable values were generated in every simulation in two steps. First,  $nT$  (pseudo) random numbers were generated from a multivariate normal distribution with mean zero and one of two different  $(T \times T)$ -dimensional covariance matrices independently for each category  $k = 1, \dots, q$ . The covariance matrices were both positive definite and equal to Toeplitz correlation matrices with elements  $(1.0, 0.4, 0.3, 0.2, 0.1)$  under the low-correlation condition and  $(1.0, 0.9, 0.8, 0.7, 0.6)$  under the high-correlation condition. In the next step, the generated values were transformed into quantiles of a type I extreme value (or Gumbel) distribution with location parameter equal to zero and scale parameter equal to one by a probability integral transformation. Given predictor variables and true values for  $\alpha$ ,  $\phi$ , and  $\beta$ , the values of the linear predictor as given by the right-hand side of (1) was calculated. Then, for each  $i$  and  $t$ , the category for which this value was maximum was selected to be the observed value of  $\widetilde{Y}_{it}$ . The probability of observing the maximum given the predictor variables can be shown to be equal to (2) if the error variables are independent and follow a type I extreme value distribution.<sup>42</sup>

It should be noted that the correlations under the low- and the high-correlation conditions are not equal to the correlations of the observed  $Y_{itk}, Y_{it'k'} | \mathbf{X}_i$ . Instead they are the correlations of corresponding response variables given predictor

variables in a latent model. The correlations in the observed variables  $Y_{itk}, Y_{it'k'} | \mathbf{X}_i$  will be somewhat lower and are not decreasing according to the same relationships as in the latent model.

Both the low- and the high-correlation conditions were realized for each sample size condition. Estimators adopted in the main study are  $GEE_I, GEE_E^F, GEE_E^R, GEE_A^F, GEE_A^R, GEE_T^F, GEE_T^R, GEE_U^F, GEE_U^R, GEE_{UN}, GEE_{CE}, GEE_{TE}$  and  $GEE_{RC}$  under each condition. The effects of time- and unit-invariant predictors are studied based on  $GEE_I, GEE_T^R$  and  $GEE_U^R$  estimators.

## 4.2 | Notes on the estimation strategy

As noted in the literature,<sup>6,8</sup> the starting values should be chosen carefully to prevent the estimation algorithm from nonconvergence. Thus, we adopted a three-step strategy to generate starting values. In a first step, starting values for a multinomial panel model are generated. As suggested by a reviewer, we used  $\tilde{\tau}_k = \log(\sum_t \sum_i Y_{itk}) / (\sum_t \sum_i Y_{it1})$ ,  $k = 2, \dots, q$ , as starting values for the constants. Starting values for the  $p$  regression coefficients restricted to be identical for all categories and time points,  $\tilde{\gamma}$ , were generated randomly from the uniform(-1,1) distribution and then transformed such that  $|\mathbf{x}'_{it} \tilde{\gamma}| < 3$  for all  $i, t$ . In the second step, a GEE estimator for a multinomial panel model is estimated.<sup>8</sup> If the working correlation matrix approach was chosen to estimate the stereotype model, then the corresponding working correlation matrix is adopted to estimate the multinomial model. In case of the local odds ratios approach, we adopt the time-exchangeable association structure in this second step. Let the estimates from this step be denoted as  $\hat{\kappa} = (\hat{\tau}' \hat{\gamma}')'$  and its estimated covariance matrix be  $\widehat{\text{Cov}}(\hat{\kappa})$ . Starting values for  $\theta$  are then generated in a third step as the solution to  $\min_{\theta} (\hat{\kappa} - \boldsymbol{\nu})' \widehat{\text{Cov}}(\hat{\kappa})^{-1} (\hat{\kappa} - \boldsymbol{\nu})$ , where  $\boldsymbol{\nu} = \boldsymbol{\nu}(\theta) = (\boldsymbol{\alpha}' (\boldsymbol{\phi}' \mathbf{1}) \otimes \boldsymbol{\beta}')'$ . The maximum number of iterations in the second and third step and the estimation procedure for the stereotype model was restricted to 200. Iterations stop if the maximum absolute difference of estimates in consecutive iterations is below  $10^{-6}$ .

The working covariance matrices based on the approach using local odds ratios are generated using adapted functions from the R package<sup>10</sup> `multgee`<sup>39</sup> called with the default values. A strong constraint of this approach is that individual covariance matrices need to be inverted. To prevent the algorithm from too many nonconverging runs, we implemented a two-step approach. If an individual estimated covariance matrix is singular then, in a first step, diagonal values close to zero are forced toward larger values. If an inverse of the resulting matrix does still not exist, a generalized inverse of the original matrix is calculated. Both events are indicated by a warning message.

To fit the stereotype model to the arthritis dataset, we generated the starting values as described above for all but one GEE estimator: for the  $GEE_{CE}$  estimator we used the  $GEE_{UN}$  estimates as starting values. Because the starting values for the  $\boldsymbol{\beta}$  parameters are randomly selected, we needed several draws for some of the GEE estimators to converge. Furthermore, the strategy of first using the approximation to the first derivative and then, after convergence, to adopt the derivative derived in Section 3.1 and iterate again until convergence, turned out to be tolerant with respect to the starting values. Hence this strategy is utilized for the real data example.

## 4.3 | Simulation results

In this section, we present detailed results only for  $n = 200$  and  $n = 500$ . Results for  $n = 50, 100, 300$  and  $n = 1000$  can be found in the supplementary material.

The estimation results are collected over 1000 simulations under each of the conditions defined by the samples sizes and the association strength (low- and high-correlation conditions). Pseudo random numbers were generated using the same seed under all realized conditions. To compare the results, we calculated the estimated bias as the difference between the true values and the means of the estimates over the simulations, the SDs of the estimates relative to the SD of the  $GEE_I$  estimates over the simulations, and the actual coverage rates of the true values for symmetric 0.95-level confidence intervals assuming normality. Estimated biases for  $n = 200$  and  $n = 500$  under the low- and high-correlation conditions can be found in Tables 1 and 2, relative SDs in Tables 3 and 4, respectively. The actual coverage rates are presented in Figures 2 and 3. The results in these tables and figures are based on 1000 simulations for most estimators. Exceptions are the  $GEE_I$ , the  $GEE_{UN}$ , the  $GEE_{CE}$ , and the  $GEE_{TE}$  estimators with 997 to 999 successful simulations if  $n = 200$ . Results for the  $GEE_{RC}$  estimator under the high-correlation condition are omitted because the estimation algorithm converged only in 48 and 55 out of 1000 simulations.

**TABLE 1** Estimated bias over 1000 simulations for  $n = 200$  and  $n = 500$  under the low-correlation condition, rounded to three decimal places

	GEE <sub>I</sub>	GEE <sub>E</sub> <sup>F</sup>	GEE <sub>E</sub> <sup>R</sup>	GEE <sub>A</sub> <sup>F</sup>	GEE <sub>A</sub> <sup>R</sup>	GEE <sub>T</sub> <sup>F</sup>	GEE <sub>T</sub> <sup>R</sup>	GEE <sub>U</sub> <sup>F</sup>	GEE <sub>U</sub> <sup>R</sup>	GEE <sub>UN</sub>	GEE <sub>CE</sub>	GEE <sub>TE</sub>	GEE <sub>RC</sub>
<i>n = 200</i>													
$\alpha_2$	0.010	0.008	0.009	0.008	0.009	0.008	0.008	0.007	0.011	0.010	0.011	0.009	0.011
$\alpha_3$	0.013	0.011	0.011	0.011	0.011	0.010	0.011	0.009	0.014	0.013	0.016	0.012	0.016
$\alpha_4$	0.008	0.004	0.005	0.004	0.005	0.004	0.005	0.003	0.008	0.008	0.012	0.006	0.013
$\phi_2$	-0.006	-0.005	-0.006	-0.006	-0.005	-0.005	-0.005	-0.008	-0.006	-0.006	-0.007	-0.004	-0.006
$\phi_3$	-0.001	-0.002	-0.001	-0.003	-0.002	-0.001	-0.001	-0.001	-0.001	-0.001	-0.001	-0.001	0.001
$\beta_1$	-0.007	-0.005	-0.006	-0.006	-0.006	-0.005	-0.006	-0.004	-0.004	-0.005	0.004	-0.002	0.013
$\beta_2$	0.012	0.012	0.012	0.012	0.012	0.012	0.012	0.011	0.010	0.009	-0.011	0.009	-0.025
$\beta_3$	-0.002	-0.001	-0.001	-0.001	-0.001	-0.001	-0.001	0.000	0.000	-0.001	0.001	-0.001	0.003
<i>n = 500</i>													
$\alpha_2$	0.003	0.002	0.002	0.002	0.002	0.002	0.002	0.002	0.003	0.002	0.003	0.002	0.003
$\alpha_3$	0.005	0.003	0.003	0.004	0.004	0.004	0.004	0.003	0.005	0.005	0.006	0.004	0.007
$\alpha_4$	0.008	0.005	0.006	0.006	0.006	0.006	0.006	0.006	0.007	0.007	0.009	0.006	0.010
$\phi_2$	0.001	0.001	0.001	0.002	0.002	0.001	0.001	0.001	0.001	0.001	0.001	0.002	0.001
$\phi_3$	0.002	0.002	0.002	0.002	0.002	0.002	0.002	0.001	0.002	0.002	0.002	0.002	0.002
$\beta_1$	-0.003	-0.002	-0.002	-0.003	-0.003	-0.002	-0.002	-0.002	-0.002	-0.002	0.001	0.000	0.007
$\beta_2$	0.003	0.004	0.004	0.004	0.004	0.004	0.004	0.003	0.003	0.002	-0.005	0.001	-0.016
$\beta_3$	-0.002	-0.001	-0.001	-0.001	-0.001	-0.001	-0.001	-0.001	-0.001	-0.001	-0.001	-0.001	0.001

Note: True values are  $\alpha = (0.3 \ 0 \ 0.4)'$ ,  $\phi = (0.3 \ 0.7)'$ ,  $\beta = (-.5 \ 1 \ -.1)'$ .  $n = 200$ : Results for GEE<sub>I</sub> and GEE<sub>UN</sub> are based on 999, those for GEE<sub>CE</sub> on 998 simulations.

Abbreviation: GEE, generalized estimating equations.

The results in Table 1 do not imply systematic biases in the different GEE estimators under the low-correlation condition. The estimated biases are almost of the same order for all GEE estimators with a tendency of larger estimated biases for single elements of the GEE<sub>RC</sub> estimator. As expected, the estimated biases tend to be smaller for the larger sample size. The pattern of results for the high-correlation condition presented in Table 2 is very similar to the pattern in Table 1 for all but the GEE<sub>RC</sub> estimator. Again, the estimated biases are not large in general and are of the same order for the different GEE estimators. Under the larger sample size  $n = 500$ , they tend to be smaller compared with  $n = 200$ .

Table 3 presents SDs of the various GEE estimators relative to the GEE<sub>I</sub> estimator under the low-correlation condition. The results suggest that all GEE estimators assuming some association tend to be more efficient than the GEE estimator assuming independence. These efficiency gains are not large and the differences between the various GEE estimators are small. Comparing the results of GEE estimators based on a working correlation matrix implies that the variances of those using a flexible correlation structure tend to be larger compared with those using a restrictive structure. The most efficient estimator, at least with respect to the  $\beta$  parameters, of those that are based on local odds ratios tends to be the GEE<sub>CE</sub> estimator which allows time varying association parameters. The most efficient estimator tends to be the GEE<sub>T</sub><sup>R</sup> estimator which is able to reproduce a correlation structure of decreasing correlations over increasing time distances, allows different correlations  $r_{tk't'k}$  and  $r_{tk't'k'}$  but at the same time requires estimation of less parameters compared with GEE<sub>T</sub><sup>F</sup>, GEE<sub>U</sub><sup>R</sup>, or GEE<sub>U</sub><sup>F</sup> estimators. The AR(1) structure is less flexible than a Toeplitz structure, which explains why the GEE<sub>A</sub><sup>R</sup> estimator is less efficient than the GEE<sub>T</sub><sup>R</sup> estimator given the simulated true correlation structure.

Table 4 implies that the efficiency gains can be considerably larger if the true correlations are higher at least for sample sizes  $n = 200$  and  $n = 500$ . Again, however, the GEE estimators based on a flexible correlation structure tend to have larger variances compared with those calculated under a restrictive strategy. These differences tend to increase

**TABLE 2** Estimated bias over 1000 simulations for  $n = 200$  and  $n = 500$  under the high-correlation condition, rounded to three decimal places

	$GEE_I$	$GEE_E^F$	$GEE_E^R$	$GEE_A^F$	$GEE_A^R$	$GEE_T^F$	$GEE_T^R$	$GEE_U^F$	$GEE_U^R$	$GEE_{UN}$	$GEE_{CE}$	$GEE_{TE}$
$n = 200$												
$\alpha_2$	0.015	0.011	0.012	0.009	0.010	0.009	0.010	0.009	0.014	0.012	0.015	0.014
$\alpha_3$	0.022	0.013	0.014	0.013	0.012	0.012	0.013	0.015	0.018	0.018	0.022	0.017
$\alpha_4$	0.015	0.003	0.005	0.004	0.003	0.003	0.004	0.005	0.010	0.011	0.015	0.008
$\phi_2$	-0.008	-0.008	-0.007	-0.010	-0.011	-0.008	-0.009	-0.010	-0.013	-0.006	-0.006	-0.009
$\phi_3$	0.003	0.000	0.001	-0.001	-0.002	0.000	-0.001	-0.003	-0.004	0.003	0.002	0.000
$\beta_1$	-0.011	-0.007	-0.007	-0.005	-0.004	-0.005	-0.004	0.006	0.005	-0.010	0.005	-0.006
$\beta_2$	0.016	0.017	0.017	0.013	0.012	0.012	0.012	-0.011	-0.006	0.015	-0.015	0.014
$\beta_3$	-0.004	-0.001	-0.001	-0.001	0.000	0.000	0.000	0.002	0.002	-0.003	0.000	-0.002
$n = 500$												
$\alpha_2$	0.001	-0.002	-0.002	-0.002	-0.002	-0.002	-0.002	-0.003	0.000	-0.001	0.000	0.000
$\alpha_3$	0.005	0.000	0.000	-0.001	-0.001	-0.001	0.000	-0.001	0.002	0.002	0.004	0.003
$\alpha_4$	0.009	0.003	0.003	0.002	0.002	0.003	0.003	0.004	0.005	0.006	0.008	0.006
$\phi_2$	-0.001	0.002	0.002	0.000	-0.001	0.001	0.000	0.000	-0.001	0.001	0.000	0.000
$\phi_3$	0.000	0.002	0.002	0.001	0.001	0.001	0.001	0.000	-0.001	0.002	0.001	0.001
$\beta_1$	-0.006	-0.003	-0.003	-0.003	-0.003	-0.002	-0.002	0.002	0.001	-0.005	0.001	-0.003
$\beta_2$	0.003	0.007	0.006	0.006	0.006	0.005	0.005	-0.004	-0.002	0.006	-0.006	0.002
$\beta_3$	-0.002	-0.001	-0.001	0.000	0.000	0.000	0.000	0.000	0.000	-0.001	0.000	-0.001

Note: True values are  $\alpha = (0.3 \ 0 \ 0.4)'$ ,  $\phi = (0.3 \ 0.7)'$ ,  $\beta = (-.5 \ 1 \ -.1)'$ .  $n = 200$ : Results for  $GEE_I$  are based on 999, those for  $GEE_{UN}$  and  $GEE_{TE}$  on 998 and for  $GEE_{CE}$  on 997 simulations. Results for  $GEE_{RC}$  are omitted due to massive convergence problems.

Abbreviation: GEE, generalized estimating equations.

from the GEE estimators assuming an equi-, AR(1), Toeplitz to the GEE estimators assuming an unrestricted working correlation matrix. Like under the low-correlation condition, GEE estimators using local odds ratios seem to be less efficient compared with those based on a working correlation matrix. The most efficient estimator is again the GEE estimator assuming a restrictive Toeplitz structure. The closer the structure of the adopted working correlation matrix to the true underlying correlation structure, the more efficient the GEE estimator, although the extra flexibility of the unrestricted version of the corresponding estimator does not pay, very likely due to the additional number of parameters to be estimated. In summary, the pattern of results is similar, but more pronounced than under the low-correlation condition if  $n = 200$ .

The coverage rates presented in the left column of Figure 2 imply that under the low-correlation condition most of the coverage rates are within a symmetric 0.99-probability interval (dashed lines) around the value 0.95 based on 1000 independent Bernoulli trials. Hence, with but one exception, there is no evidence of systematic over- or undercoverage. The exception is  $GEE_U^F$  which involves the estimation of many correlations and is associated with systematically lower—and for four parameters too low—coverage rates compared with the other GEE estimators. Note that the results for  $GEE_{RC}$  are only shown for the low-correlation condition with a slightly too low coverage rate for one  $\beta$  parameter. Under the high-correlation condition (right column) again  $GEE_U^F$  shows undercoverage but so do the estimators  $GEE_{UN}$  and  $GEE_{CE}$  for one  $\beta$  parameter each, although less pronounced. The coverage rates for all other estimators are in an acceptable range.

Figure 3 presents the results for sample size  $n = 500$ . In this case, the coverage rates of all GEE estimators under the low- and high-correlation conditions are in an acceptable range. Under the high-correlation condition, no results are shown for  $GEE_{RC}$  due to massive convergence problems.

The results based on  $n = 50, 100, 300$  and  $n = 1000$  can be found in the supplementary material. Results for  $GEE_{RC}$  are only available for sample sizes of  $n > 100$  under the low-correlation condition. In samples of sizes  $n = 50$  and

**TABLE 3** SD of estimates relative to assuming independence over 1000 simulations for  $n = 200$  and  $n = 500$  under the low-correlation condition, rounded to three decimal places

	$GEE_E^F$	$GEE_E^R$	$GEE_A^F$	$GEE_A^R$	$GEE_T^F$	$GEE_T^R$	$GEE_U^F$	$GEE_U^R$	$GEE_{UN}$	$GEE_{CE}$	$GEE_{TE}$	$GEE_{RC}$
$n = 200$												
$\alpha_2$	0.984	0.981	0.976	0.977	0.975	0.975	0.987	0.980	0.989	0.983	0.994	0.985
$\alpha_3$	0.985	0.978	0.971	0.966	0.970	0.966	0.978	0.969	0.982	0.969	0.993	0.990
$\alpha_4$	0.985	0.976	0.980	0.972	0.979	0.967	0.985	0.965	0.980	0.962	0.992	0.971
$\phi_2$	0.957	0.957	0.961	0.960	0.958	0.951	0.998	0.963	0.996	1.012	0.983	1.002
$\phi_3$	0.948	0.945	0.964	0.955	0.954	0.941	0.979	0.949	1.004	1.020	0.974	0.971
$\beta_1$	0.966	0.966	0.975	0.964	0.970	0.955	0.999	0.961	0.968	0.962	0.982	0.981
$\beta_2$	0.961	0.963	0.983	0.975	0.972	0.963	0.990	0.967	0.978	0.967	0.979	0.972
$\beta_3$	0.953	0.947	0.959	0.950	0.954	0.942	0.973	0.944	0.952	0.934	0.980	0.954
$n = 500$												
$\alpha_2$	0.997	0.994	0.993	0.991	0.991	0.989	1.001	0.995	0.993	0.991	1.001	1.009
$\alpha_3$	0.984	0.982	0.983	0.982	0.981	0.978	0.988	0.979	0.983	0.978	0.997	0.993
$\alpha_4$	0.978	0.976	0.980	0.977	0.973	0.971	0.985	0.973	0.976	0.970	0.993	0.989
$\phi_2$	0.963	0.959	0.958	0.955	0.953	0.948	0.961	0.950	0.998	1.004	0.988	0.979
$\phi_3$	0.961	0.958	0.971	0.966	0.962	0.955	0.965	0.955	0.996	1.002	0.989	0.993
$\beta_1$	0.974	0.973	0.980	0.976	0.971	0.968	0.974	0.970	0.971	0.967	0.978	0.972
$\beta_2$	0.961	0.960	0.970	0.966	0.955	0.952	0.963	0.952	0.966	0.960	0.980	0.971
$\beta_3$	0.965	0.962	0.964	0.961	0.958	0.955	0.969	0.958	0.966	0.958	0.988	0.969

Note: True values are  $\alpha = (0.3 \ 0 \ 0.4)'$ ,  $\phi = (0.3 \ 0.7)'$ ,  $\beta = (-.5 \ 1 \ -.1)'$ .  $n = 200$ : Results for  $GEE_U$  and  $GEE_{UN}$  are based on 999, those for  $GEE_{CE}$  on 998 simulations.

Abbreviation: GEE, generalized estimating equations.

$n = 100$ , the model for the association structure was only weakly or not identified. In larger samples, the number of successful simulations varied between 52 and 55 under the high-correlation condition. Hence, these results will not be considered.

In general, under the low- and high-correlation condition, the estimated biases tend to decrease as the sample size increases. It should be noted, however, that the results for  $n = 50$  are more volatile than for the other sample sizes and are based on only 867 to 925 successful simulations under the low-correlation condition and 601 to 875 successful simulations under the high-correlation condition for the GEE estimators based on a working correlation matrix. The number of successful simulations for estimators based on local odds ratios varied between 809 and 866 under the low and 568 and 747 under the high-correlation condition. Hence, for  $n = 50$ , conclusions about finite sample properties can only be drawn with great caution. A general statement that seems to be warranted even in this case is that under the high-correlation condition considerable efficiency gains are possible if an association structure is modeled, but coverage rates may only be acceptable for very restrictive association structures, like those adopted for the  $GEE_U^R$  or  $GEE_{UN}$  estimators.

The number of successful simulations for  $n = 100$  under the low (high)-correlation condition varied between 989 and 998 (973 and 997) if a working correlation structure is adopted and between 982 and 992 (914 and 956) if local odds ratios are modeled. Therefore, these results may be interpreted but with caution.

Under both the low- and high-correlation conditions, the order of the estimated biases are roughly equal for the different GEE estimators. There are only single slightly larger estimated biases for  $n = 100$  and the  $GEE_T$ ,  $GEE_U^F$  and  $GEE_U^R$ ,  $GEE_{CE}$  and  $GEE_{RC}$  estimators.

Under the low-correlation condition, systematic, if only small to medium, efficiency gains can be observed from about  $n = 100$  on if an association structure is explicitly modeled. This is different under the high-correlation condition, where the efficiency gains of GEE estimators based on working association structures closer to the true structure are more pronounced and tend to be much larger, with an advantage of the GEE estimators based on a working

**TABLE 4** SD of estimates relative to assuming independence over 1000 simulations for  $n = 200$  and  $n = 500$  under the high-correlation condition, rounded to three decimal places

	$GEE_E^F$	$GEE_E^R$	$GEE_A^F$	$GEE_A^R$	$GEE_T^F$	$GEE_T^R$	$GEE_U^F$	$GEE_U^R$	$GEE_{UN}$	$GEE_{CE}$	$GEE_{TE}$
$n = 200$											
$\alpha_2$	0.953	0.950	0.936	0.934	0.942	0.935	0.959	0.944	0.966	0.962	0.973
$\alpha_3$	0.916	0.910	0.899	0.896	0.899	0.892	0.908	0.895	0.925	0.911	0.955
$\alpha_4$	0.865	0.858	0.853	0.851	0.848	0.839	0.860	0.844	0.870	0.846	0.927
$\phi_2$	0.718	0.713	0.733	0.724	0.701	0.693	0.730	0.706	0.910	0.918	0.810
$\phi_3$	0.699	0.695	0.688	0.680	0.661	0.659	0.681	0.665	0.921	0.925	0.822
$\beta_1$	0.739	0.735	0.737	0.731	0.711	0.703	0.714	0.698	0.779	0.737	0.876
$\beta_2$	0.794	0.794	0.776	0.774	0.759	0.757	0.762	0.748	0.832	0.796	0.894
$\beta_3$	0.698	0.693	0.699	0.692	0.674	0.668	0.683	0.669	0.727	0.694	0.871
$n = 500$											
$\alpha_2$	0.970	0.967	0.953	0.951	0.951	0.949	0.958	0.950	0.971	0.964	0.986
$\alpha_3$	0.912	0.911	0.918	0.918	0.907	0.905	0.915	0.906	0.926	0.919	0.955
$\alpha_4$	0.861	0.859	0.858	0.857	0.844	0.841	0.853	0.843	0.865	0.847	0.931
$\phi_2$	0.715	0.713	0.685	0.685	0.675	0.673	0.691	0.683	0.919	0.921	0.811
$\phi_3$	0.740	0.738	0.711	0.703	0.691	0.685	0.706	0.695	0.926	0.926	0.883
$\beta_1$	0.707	0.708	0.716	0.716	0.686	0.687	0.694	0.688	0.739	0.724	0.867
$\beta_2$	0.766	0.767	0.779	0.778	0.743	0.743	0.751	0.747	0.800	0.782	0.878
$\beta_3$	0.706	0.704	0.679	0.676	0.660	0.657	0.667	0.658	0.734	0.691	0.878

Note: True values are  $\alpha = (0.3 \ 0 \ 0.4)'$ ,  $\phi = (0.3 \ 0.7)'$ ,  $\beta = (-.5 \ 1 \ -.1)'$ .  $n = 200$ : Results for  $GEE_T$  are based on 999, those for  $GEE_{UN}$  and  $GEE_{TE}$  on 998 and for  $GEE_{CE}$  on 997 simulations. Results for  $GEE_{RC}$  are omitted due to massive convergence problems.

Abbreviation: GEE, generalized estimating equations.

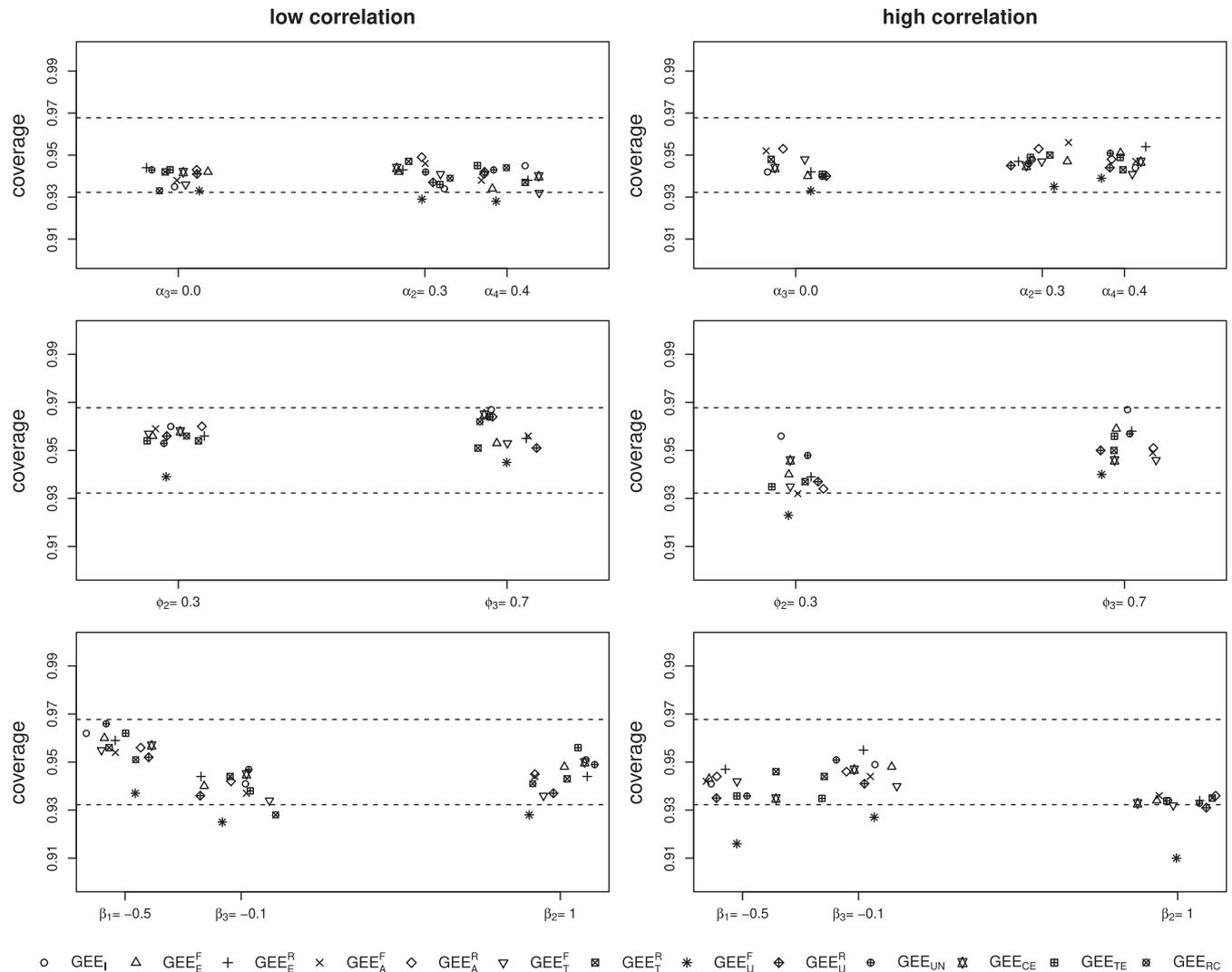
correlation structure compared with those using local odds ratios. One explanation for these latter findings for the local odds ratios versions may be the fact that the association parameters are estimated before any GEE estimates are calculated ignoring any predictor variables. The corresponding association structure may thus deviate markedly from the true structure. A general tendency under the high-correlation condition is that the flexible versions of the working correlation structures lead to less efficient estimators as the restrictive versions but these differences seem to vanish with increasing sample size.

For most GEE estimators, the coverage rates are acceptable or close to acceptable from  $n = 100$  on under both correlation conditions. This tends to be true for all estimators from  $n = 300$  on under the low but also under the high-correlation condition. Below this sample size, the very flexible  $GEE_U^F$  estimator is associated with coverage rates that are systematically too low.

Taken together, the results imply that at least for the situations simulated, the proposed GEE estimators allow valid inferences from approximately  $n = 100$  on if the working association structure does not require the estimation of too many parameters.

From the above findings, one may expect that for situations with invariant predictors, the general pattern of results with respect to efficiency of the estimators can be inferred from comparisons of results based on sample sizes  $n = 200$  and  $n = 500$  and estimators  $GEE_I$ ,  $GEE_T^R$ , and  $GEE_U^R$ . Thus, Table 5 presents biases and relative SDs of the estimates for the low-correlation condition, Table 6 those for the high-correlation condition from the simulations based on invariant predictors.

The estimated biases differ depending on the sample size but also on the predictor variable. As before, estimated biases are smaller for larger sample sizes. In addition, if  $n = 200$ , the estimated bias for  $\beta_2$  weighting the time-invariant predictor tends to be larger than for the other two  $\beta$  parameters. With respect to the relative SDs, the results in Tables 5 and 6 suggest that the efficiency gain of  $GEE_T^R$  or  $GEE_U^R$  compared with  $GEE_I$  is largest for  $\beta_1$  weighting the freely varying predictor



**FIGURE 2** Actual coverage of true  $\alpha_k$ ,  $\phi_k$  and  $\beta_l$ , 0.95-level confidence intervals,  $n = 200$ ; dashed lines represent symmetric 0.99-probability intervals around 0.95. High-correlation condition: Results for  $GEE_{RC}$  are omitted due to massive convergence problems

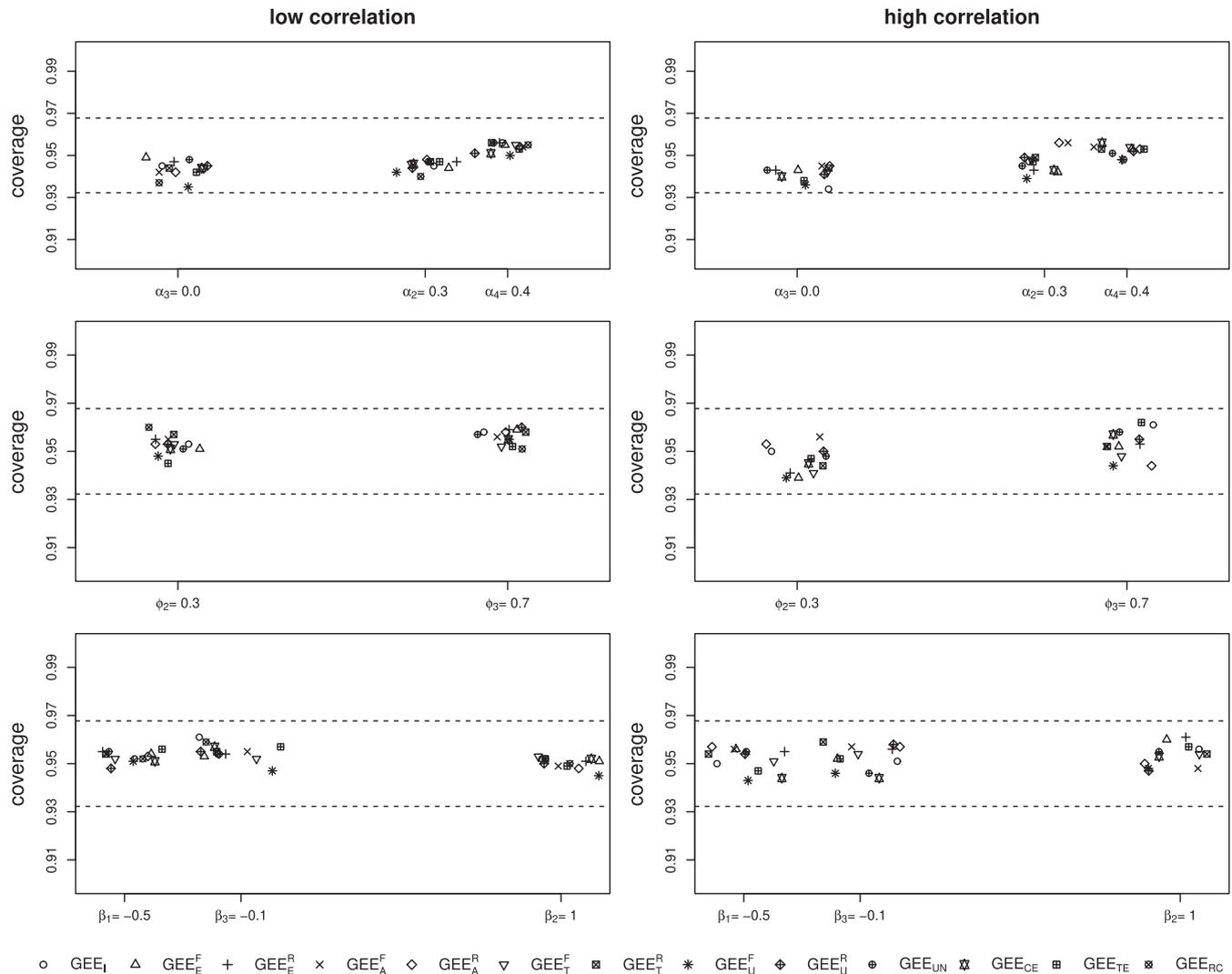
and smallest for  $\beta_2$  weighting the time-invariant variable. This effect seems to be strongest under the high-correlation condition. Put differently, for time-invariant predictors and lesser so for unit-invariant predictors, adopting a working covariance structure which is closer to the true structure may not lead to more efficient or only slightly more efficient estimators.

The coverage rates over at least 998 simulations are almost all in an acceptable range. Only the coverage rates for  $\alpha_2$  of  $GEE_T$  and  $GEE_U^R$  under the low-correlation condition, 0.930 and 0.929, are slightly too low (see Table 7).

## 5 | THE ARTHRITIS DATASET

The data to illustrate the application of the GEE estimator is from a randomized clinical trial designed to evaluate the effectiveness of the drug Auranofin relative to a placebo therapy for the treatment of rheumatoid arthritis<sup>17,27</sup> and was taken from R package<sup>10</sup> `multgee`.<sup>39</sup> The dataset can be retrieved by calling `data(arthritis)` after `multgee` is loaded\*. The dataset includes 302 units out of which  $n = 289$  were completely observed at  $T = 3$  time points, that is, 1, 3, and 5 months after treatment.

\*Data availability statement: Data sharing is not applicable to this article as no new data were created or analyzed in this study.



**FIGURE 3** Actual coverage of true  $\alpha_k$ ,  $\phi_k$  and  $\beta_l$ , 0.95-level confidence intervals,  $n = 500$ ; dashed lines represent symmetric 0.99-probability intervals around 0.95. High-correlation condition: Results for  $GEE_{RC}$  are omitted due to massive convergence problems

The ordinal response variable is self-assessment of rheumatoid arthritis surveyed with a 5-point scale from 1 = “very poor” to 5 = “very good” at a baseline before the trial and at the three follow-up observation points. Self-assessment with  $q = 5$  possible values was converted into four 0-1 variables with “very poor” as the reference category. It turned out, however, that the fraction of observed “very good” answers at the baseline was very small leading to convergence problems. Thus, we combined the fourth and the fifth category for the baseline predictors and the response variable.

Baseline assessment variables were treated as predictors ( $b_2, \dots, b_4$ ) and the follow-up assessments as the response variable.<sup>17,18</sup> Additional predictors that entered the model are two time-dummies,  $t_2$  and  $t_3$  (reference category is the first month follow-up) and the binary treatment indicator  $tr$  with the placebo group being the reference group.

For 13 out of the 302 units in the dataset, the variable self-assessment over the follow-up period was not completely observed. Baseline self-assessment variables were completely observed though. Assuming<sup>17,18</sup> that the missing values are missing completely at random,<sup>43</sup> we ignored the incompletely observed cases in our analysis.

Table 8 presents the results of estimating the ordered stereotype logit model assuming independence, adopting a restrictive equi- and a restrictive unstructured working correlation matrix. Estimation results for all other GEE estimators can be found in the supplementary material. Table 9 shows the working correlation matrices over time and categories for the  $GEE_E^R$  (lower triangular matrix) and the  $GEE_U^R$  (upper triangular matrix) estimators.

A comparison of the estimation results in Table 8 and the tables in the supplementary material suggests that, given all necessary assumptions are at least approximately met, the general conclusions with respect to  $\alpha_3$ ,  $\alpha_4$ ,  $\phi_3$ ,  $\beta_{t2}$ ,  $\beta_{trt}$ ,  $\beta_{b3}$ ,



**TABLE 5** Estimated biases and SDs of estimates relative to assuming independence (relative SD) over 1000 simulations for  $n = 200$  and  $n = 500$  under the low-correlation condition, rounded to three decimal places

	Estimated bias						Relative SD			
	$n = 200$			$n = 500$			$n = 200$		$n = 500$	
	$GEE_I$	$GEE_T^R$	$GEE_U^R$	$GEE_I$	$GEE_T^R$	$GEE_U^R$	$GEE_T^R$	$GEE_U^R$	$GEE_T^R$	$GEE_U^R$
$\alpha_2$	0.009	0.010	0.011	0.000	0.000	0.001	0.994	0.999	0.997	0.999
$\alpha_3$	0.010	0.011	0.012	0.001	0.001	0.002	1.004	1.006	1.000	1.002
$\alpha_4$	0.003	0.004	0.006	0.003	0.003	0.004	0.994	0.994	1.001	1.004
$\phi_2$	-0.003	-0.002	-0.002	-0.003	-0.002	-0.002	0.941	0.948	0.967	0.967
$\phi_3$	-0.001	-0.002	-0.002	-0.001	-0.001	-0.002	0.949	0.947	0.966	0.967
$\beta_1$	0.008	0.008	0.006	0.003	0.003	0.002	0.986	0.992	0.978	0.982
$\beta_2$	0.016	0.013	0.014	0.004	0.003	0.003	1.006	1.010	0.992	0.994
$\beta_3$	0.013	0.014	0.013	0.000	0.000	0.000	1.000	0.998	0.989	0.991

Note: True values are  $\alpha = (0.3 \ 0 \ 0.4)'$ ,  $\phi = (0.3 \ 0.7)'$ ,  $\beta = (1 \ 1 \ 1)'$ . Predictor variables are independent normally distributed, freely varying ( $x_1$ ), invariant over time ( $x_2$ ) and invariant over units ( $x_3$ ).  $n = 200$ : Results for  $GEE_I$  and  $GEE_U^R$  based on 998, those for  $GEE_T^R$  on 999 simulations.

Abbreviation: GEE, generalized estimating equations.

**TABLE 6** Estimated biases and SDs of estimates relative to assuming independence (relative SD) over 1000 simulations for  $n = 200$  and  $n = 500$  under the high-correlation condition, rounded to three decimal places

	Estimated bias						Relative SD			
	$n = 200$			$n = 500$			$n = 200$		$n = 500$	
	$GEE_I$	$GEE_T^R$	$GEE_U^R$	$GEE_I$	$GEE_T^R$	$GEE_U^R$	$GEE_T^R$	$GEE_U^R$	$GEE_T^R$	$GEE_U^R$
$\alpha_2$	0.016	0.016	0.019	0.002	0.003	0.005	0.988	0.994	0.996	0.994
$\alpha_3$	0.018	0.021	0.024	0.003	0.004	0.007	1.000	1.002	0.994	0.993
$\alpha_4$	0.010	0.011	0.017	0.005	0.006	0.008	1.030	1.033	1.014	1.012
$\phi_2$	-0.001	0.000	0.000	-0.003	-0.002	-0.002	0.765	0.774	0.802	0.805
$\phi_3$	0.000	-0.001	-0.001	0.000	-0.001	-0.002	0.751	0.759	0.748	0.752
$\beta_1$	0.012	0.013	0.006	0.006	0.007	0.003	0.861	0.870	0.846	0.846
$\beta_2$	0.029	0.017	0.019	0.010	0.006	0.007	1.018	1.007	0.965	0.965
$\beta_3$	0.010	0.013	0.005	0.003	0.004	0.002	0.960	0.972	0.946	0.946

Note: True values are  $\alpha = (0.3 \ 0 \ 0.4)'$ ,  $\phi = (0.3 \ 0.7)'$ ,  $\beta = (1 \ 1 \ 1)'$ . Predictor variables are independent normally distributed, freely varying ( $x_1$ ), invariant over time ( $x_2$ ) and invariant over units ( $x_3$ ).  $n = 200$ : Number of successful simulations for  $GEE_I$ -998 and  $GEE_T^R$ -999.

Abbreviation: GEE, generalized estimating equations.

and  $\beta_{b4}$  do not depend on the working correlation matrix. For all working correlation matrices, the 0.95-level confidence intervals for  $\alpha_3$ ,  $\alpha_4$ , and  $\beta_{t2}$  cover the value zero. Thus, there is no evidence that these parameters have an effect on the response variable. On the other hand, the 0.95-confidence intervals for  $\phi_3$ ,  $\beta_{tt}$ ,  $\beta_{b3}$ , and  $\beta_{b4}$  do not cover zero. With respect to  $\alpha_2$ ,  $\phi_2$ ,  $\beta_{t3}$ , and  $\beta_{b2}$  the results are inconsistent. However, given that the predictors are all either unit- or time invariant and that the estimated elements of the working correlation matrices are rather low, the choice of the working correlation matrix cannot be expected to have a major effect on the efficiency of the estimators. Since the sample is of medium size, choosing a GEE estimator based on a very flexible correlation matrix or on local odds ratios may, however, not allow substantially more precise inferences compared with the  $GEE_E^R$  estimator. Hence, we will interpret the estimation results based on the  $GEE_E^R$  estimator.

All the 0.95-confidence intervals for the  $\alpha$  parameters cover zero, which implies that there is no evidence that they differ from zero. The results for  $\phi_2$  and  $\phi_3$  imply that they presumably differ from  $\phi_1 = 0$  and  $\phi_4 = 1$ . Due to the overlapping confidence intervals, there seems to be no strong evidence for  $\phi_2$  and  $\phi_3$  to be different. Hence, the second and third category could probably be combined into one. The estimates for  $\phi_2$  and  $\phi_3$ , however, do conform with the presupposed ordinal nature of the response variable, that is,  $0 < \hat{\phi}_2 < \hat{\phi}_3 < 1$ .

**TABLE 7** Actual coverage rates over  $n_{\text{sim}}$  successful simulations for  $n = 200$  and  $n = 500$  rounded to three decimal places

	Low-correlation condition						High-correlation condition					
	$n = 200$			$n = 500$			$n = 200$			$n = 500$		
	GEE <sub>I</sub>	GEE <sub>T</sub> <sup>R</sup>	GEE <sub>U</sub> <sup>R</sup>	GEE <sub>I</sub>	GEE <sub>T</sub> <sup>R</sup>	GEE <sub>U</sub> <sup>R</sup>	GEE <sub>I</sub>	GEE <sub>T</sub> <sup>R</sup>	GEE <sub>U</sub> <sup>R</sup>	GEE <sub>I</sub>	GEE <sub>T</sub> <sup>R</sup>	GEE <sub>U</sub> <sup>R</sup>
$\alpha_2$	0.930	0.935	0.929	0.944	0.943	0.942	0.944	0.941	0.942	0.945	0.949	0.950
$\alpha_3$	0.948	0.945	0.946	0.948	0.947	0.945	0.944	0.946	0.946	0.951	0.935	0.934
$\alpha_4$	0.942	0.943	0.944	0.955	0.950	0.953	0.933	0.942	0.940	0.950	0.955	0.958
$\phi_2$	0.948	0.957	0.954	0.955	0.954	0.957	0.956	0.945	0.943	0.950	0.942	0.944
$\phi_3$	0.949	0.947	0.945	0.955	0.949	0.947	0.951	0.949	0.943	0.959	0.955	0.960
$\beta_1$	0.943	0.939	0.939	0.934	0.935	0.934	0.945	0.944	0.937	0.943	0.955	0.954
$\beta_2$	0.952	0.953	0.950	0.955	0.954	0.955	0.952	0.956	0.960	0.949	0.955	0.958
$\beta_3$	0.959	0.959	0.962	0.947	0.952	0.952	0.956	0.959	0.952	0.949	0.947	0.944
$n_{\text{sim}}$	998	999	998	1000	1000	1000	998	999	1000	1000	1000	1000

Note: True values are  $\alpha = (0.3 \ 0 \ 0.4)'$ ,  $\phi = (0.3 \ 0.7)'$ ,  $\beta = (1 \ 1 \ 1)'$ . Predictor variables are independent normally distributed, freely varying ( $x_1$ ), invariant over time ( $x_2$ ), and invariant over units ( $x_3$ ).

Abbreviation: GEE, generalized estimating equations.

**TABLE 8** Arthritis dataset: Parameter estimates (est), SEs, lower (low), and upper (up) 0.95-level confidence intervals assuming normality, working correlation matrix independence, equicorrelation and unstructured,  $n = 289$ 

	GEE <sub>I</sub>				GEE <sub>E</sub> <sup>R</sup>				GEE <sub>U</sub> <sup>R</sup>			
	Est	SE	Low	Up	Est	SE	Low	Up	Est	SE	Low	Up
$\alpha_2$	0.800	0.552	-0.283	1.883	0.746	0.453	-0.141	1.633	0.788	0.451	-0.095	1.672
$\alpha_3$	0.824	0.674	-0.496	2.145	0.755	0.543	-0.310	1.820	0.804	0.542	-0.258	1.865
$\alpha_4$	-0.485	0.852	-2.156	1.185	-0.654	0.723	-2.070	0.762	0.615	0.730	-2.045	0.815
$\phi_2$	0.349	0.167	0.021	0.677	0.349	0.169	0.018	0.680	0.339	0.170	0.005	0.672
$\phi_3$	0.623	0.102	0.422	0.823	0.612	0.122	0.373	0.851	0.605	0.123	0.365	0.846
$\beta_{12}$	-0.130	0.269	-0.656	0.397	-0.114	0.260	-0.624	0.397	0.113	0.258	-0.619	0.393
$\beta_{13}$	0.505	0.254	0.007	1.004	0.538	0.266	0.017	1.059	0.526	0.263	0.012	1.041
$\beta_{1rt}$	1.191	0.471	0.267	2.115	1.240	0.410	0.437	2.043	1.216	0.409	0.415	2.017
$\beta_{b2}$	1.271	0.900	-0.494	3.036	1.458	0.765	-0.042	2.957	1.467	0.768	-0.037	2.972
$\beta_{b3}$	2.449	0.937	0.613	4.285	2.602	0.751	1.130	4.075	2.552	0.755	1.072	4.033
$\beta_{b4}$	5.331	1.637	2.123	8.538	5.356	1.482	2.451	8.262	5.312	1.458	2.454	8.171

Finally, the results for the regression parameters suggest that there seems to be no effect of the first relative to the second follow-up time point, but a positive effect of the third follow-up time point. This result implies that a difference in self-assessment may be observed at the earliest after just over 3 months.

The treatment with Auranofin seems to have a positive effect on self-assessment of rheumatoid arthritis, the estimated increase in the log odds is 1.240 for the fourth category relative to the placebo group. The effect of the baseline self-assessment 2 = "poor" seems to have no different effect on self-assessment at the follow-up time points than the effect of the self-assessment "very poor." However, baseline self-assessments of 3 = "fair" or 4 = "good or very good" before the trial, seem to have a positive effect on later self-assessments. The estimated increase in the log odds for the fourth category is 5.356 relative to the reference category.

**TABLE 9** Arthritis dataset: Working correlation matrix of  $GEE_E^R$  (lower triangular) and  $GEE_U^R$  (upper triangular)

1-month follow-up			3-month follow-up			5-month follow-up		
1	0	0	0.116	-0.107	-0.107	0.221	-0.047	-0.047
0	1	0	-0.107	0.116	-0.107	-0.047	0.221	-0.047
0	0	1	-0.107	-0.107	0.116	-0.047	-0.047	0.221
0.191	-0.078	-0.078	1	0	0	0.236	-0.078	-0.078
-0.078	0.191	-0.078	0	1	0	-0.078	0.236	-0.078
-0.078	-0.078	0.191	0	0	1	-0.078	-0.078	0.236
0.191	-0.078	-0.078	0.191	-0.078	-0.078	1	0	0
-0.078	0.191	-0.078	-0.078	0.191	-0.078	0	1	0
-0.078	-0.078	0.191	-0.078	-0.078	0.191	0	1	0

## 6 | DISCUSSION

Estimation of the ordered stereotype logit model in the cross-sectional context is not as straightforward as estimation of sequential or cumulative models. This is due to the intrinsic nonlinearity in the parameters reflecting the increased flexibility of the stereotype model compared with the sequential or cumulative model and justifies simulations as a valuable tool to study the finite sample properties. However, this nonlinearity of the model parameters makes the interpretation of estimation results more difficult, which may also explain why the stereotype model is not very common in applications.

Ordered stereotype panel models have been estimated using the ML or GLS method, but to the best of our knowledge, simulation results studying the finite sample properties seem not to have been published. ML methods require correct specification of the multivariate distribution of the vectors of binary responses indicating the observed category over all observed time points. If the distributional assumptions are correct, then ML estimators are known to be consistent, asymptotically normally distributed and allow asymptotically precise inferences. In this case, ML estimation would be the preferred approach. GLS estimation requires correct specification of the mean, variances, and covariances only, but the estimator tends to have larger variances.

The GEE approach requires only correct specification of the means of the vectors of binary variables, the association structure or higher moments need not be correctly specified. This robustness property comes at a price, which is a loss of efficiency compared with ML estimators. GLS estimation can be made robust by allowing a misspecified association structure as well, but would then also lead to a loss of efficiency and would require a robust estimator of the covariance matrix of the GLS estimator.

In this article we propose a GEE approach with a finite sample correction to estimate ordered stereotype panel logit models with a possibly misspecified association structure. The simulation study and the application to a panel dataset from a randomized clinical trial design to assess the effect of a drug on self-assessment of rheumatoid arthritis at three follow-up observation points after treatment show that GEE estimation of this model is feasible. In fact, the results from the simulation study imply that the finite sample properties of GEE estimators as reported in the literature for other models apply to these models in samples of medium to large size as well and inferences tend to be statistically valid.

The comparison of different GEE estimators in the simulation study implies that, similar to results known from binary models, adopting a working correlation matrix that explicitly models associations compared with the identity matrix may lead to substantial efficiency improvements for the estimators if the true correlations are high. In accordance with known results we found additional efficiency gains if a working correlation matrix is chosen whose structure is close to the true structure. On the other hand and again similar to results found for binary models, if the true correlations are low, then differences between the GEE estimators explicitly allowing for dependencies and even in comparison with the GEE estimator calculated under independence are small. However, the simulation results in this article additionally illustrate a conflict of objectives between the flexibility of the working correlation matrix and the consequences of a weakly identified estimation problem. Possible efficiency gains by choosing a flexible working correlation matrix may be lost or even reversed if the number of association parameters is large relative to the number of observations.

Comparisons of GEE estimators based on a working correlation matrix with those using local odds ratios did not reveal any basic problems of the former that could not be explained by weakly identified models in small samples. Instead, in our simulations we found that GEE estimates based on local odds ratios tend to have larger variances compared with those that use a working correlation matrix, which may be due to the fact that for the former the association structure parameters are estimated ignoring any predictor variables.

The results for the arthritis dataset illustrate once more that if the correlations are low, then the differences between various versions of the GEE estimator relative to the estimator assuming independence can be expected to be small.

As in other models for categorical responses, all possible categories of the response variable should be observed with sufficient frequencies and variation in the predictors should be sufficiently large to avoid weakly identified models and problems of nonconvergence. The greater flexibility of the stereotype model relative to other models requires larger datasets for identification. Hence, based on our simulation results, we suggest to fit an ordered stereotype panel model to datasets with sample sizes below  $n = 100$  only if there are good reasons for doing so. In that case, however, we recommend to adopt only restrictive versions of the proposed GEE estimators, like the estimator based on a restrictive working equicorrelation structure. Similarly and more general, if the true correlations are assumed to be low, then a working correlation matrix with a small number of parameters to be estimated may be chosen. In this case, the efficiency loss compared with the GEE estimator assuming independence tends to be small, but the costs of choosing a GEE estimator which is too flexible in terms of bias, variance and undercoverage may be high. In medium to large samples, say from  $n = 100$  on, with high true correlations, however, efficiency gains by adopting a toeplitz or AR(1)-structure if appropriate may be considerable. Based on the simulation results reported in this article, we do not recommend the use of the flexible versions of the working correlation matrices as long as sample sizes are below  $n = 300$ , depending on the type of the working correlation matrix. For example,  $n = 300$  could be sufficient to adopt the flexible version of a working equicorrelation matrix but not for the flexible version of the unrestricted correlation matrix.

Summing up, the GEE estimator with the correction terms proposed in this article seems to be an interesting candidate to estimate the flexible ordered stereotype panel logit model even if the association structure is misspecified. In addition to the above-mentioned limitations and with respect to the arthritis dataset, it should be noted that for GEE estimators to be consistent and asymptotically normal in general, missing values must be missing completely at random. If that is not the case, then one could resort to either a weighting or a multiple imputation strategy. If in addition to the assumption that the means are correctly specified, all the required assumptions about higher moments of the distribution of the response variables are met, particularly with regard to the association structure, then the stereotype panel model may be estimated assuming a random effects model and adopting the ML approach. This approach would allow missing values to be missing at random, although standard estimation procedures will not be able to handle missing values in predictor variables properly. In addition, if the pattern of missing values is nonmonotone then the assumption of missing values being missing at random may be questionable, and ML estimation with standard software will very likely lead to invalid inferences. Finally, note that the GEE estimator proposed in this article is restricted to designs where units are observed at the same fixed time points. The approach does, for example, not apply if in longitudinal studies, predictors and responses of units are recorded based on varying time intervals with possibly different frequencies.

## ACKNOWLEDGMENTS

This work has been supported by grant RTI2018-100927-J-I00 administrated by Ministerio de Ciencia e Innovación (MCI, Spain), by the Agencia Estatal de Investigación (AEI, Spain), and by the European Regional Development Fund FEDER (FEDER, UE), by Marsden grant E2987-3648 administrated by the Royal Society of New Zealand), and by grant 2017 SGR 622 (GRBIO) administrated by the Departament d'Economia i Coneixement de la Generalitat de Catalunya (Spain).

## ORCID

Martin Spiess  <https://orcid.org/0000-0003-1855-643X>

Daniel Fernández  <https://orcid.org/0000-0003-0012-2094>

## REFERENCES

1. Agresti A. *Analysis of Ordinal Categorical Data*. 2nd ed. New York, NY: John Wiley; 2010.
2. Liu I, Agresti A. The analysis of ordered categorical data: an overview and a survey of recent developments. *TEST*. 2005;14(1):1-73.

3. Tutz G. *Regression for Categorical Data*. Cambridge, MA: Cambridge University Press; 2011.
4. Anderson JA. Regression and ordered categorical variables. *J R Stat Soc Ser B Stat Methodol*. 1984;46(1):1-22.
5. Holtbrügge W, Schumacher M. A comparison of regression models for the analysis of ordered categorical data. *J R Stat Soc Ser C Appl Stat*. 1991;40(2):249-259.
6. Greenland S. Alternative models for ordinal logistic regression. *Stat Med*. 1994;13:1665-1677.
7. Lunt M. Prediction of ordinal outcomes when the association between predictors and outcome differs between outcome levels. *Stat Med*. 2005;24:1357-1369.
8. Kuss O. On the estimation of the stereotype regression model. *Comput Stat Data Anal*. 2006;50(8):1877-1890.
9. Turner H, Firth D. Generalized nonlinear models in R: an overview of the GNM package. R package version 1.1-0; 2018. <https://cran.r-project.org/package=gnm>. Accessed December 25, 2019.
10. R Core Team. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing. Vienna, Austria; 2019. <https://www.r-project.org/>. Accessed December 25, 2019.
11. Lunt M. Stereotype ordinal regression. *Stata Tech Bull*. 2001;10(61):12-18.
12. Yee TW, Hastie TJ. Reduced-rank vector generalized linear models. *Stat Model*. 2003;3(1):15-41.
13. Archer KJ, Hou J, Zhou Q, Ferber K, Layne JG, Gentry AE. ordinalgmifs: an R package for ordinal regression in high-dimensional data settings. *Cancer Inf*. 2014;13:187-195.
14. Kuss O. Modelling physicians' recommendations for optimal medical care by random effects stereotype regression. In: Verbeke G, Molenberghs G, Aerts M, Fieuws S, eds. *Proceedings of the 18th International Workshop on Statistical Modelling Leuven*. Leuven, Belgium: Katholieke Universiteit Leuven; 2003:245-249.
15. Johnson TR. Discrete choice models for ordinal response variables: a generalization of the stereotype model. *Psychometrika*. 2007;72(4):489-504.
16. Liang K-Y, Zeger SL. Longitudinal data analysis using generalized linear models. *Biometrika*. 1986;73(1):13-22.
17. Lipsitz S, Kim K, Zhao L. Analysis of repeated categorical data using generalized estimating equations. *Stat Med*. 1994;13:1149-1163.
18. Touloumis A, Agresti A, Kateri M. GEE for multinomial responses using a local odds ratios parameterization. *Biometrics*. 2013;69:633-640.
19. Qu Y, Piedmonte MR, Medendorp SV. Latent variable models for clustered ordinal data. *Biometrics*. 1995;51(1):268-275.
20. Heagerty P, Zeger S. Marginal regression models for clustered ordinal measurements. *J Am Stat Assoc*. 1996;91(435):1024-1036.
21. Spiess M. Estimation of a two-equation panel model with mixed continuous and ordered categorical outcomes and missing data. *J R Stat Soc Ser C Appl Stat*. 2006;55(4):525-538.
22. Nores ML, del Pilar Díaz M. Some properties of regression estimators in GEE models for clustered ordinal data. *Comput Stat Data Anal*. 2008;52:3877-3888.
23. Mancl LA, Leroux BG. Efficiency of regression estimates for clustered data. *Biometrics*. 1996;52(2):500-511.
24. Spiess M, Hamerle A. On the properties of GEE estimators in the presence of invariant covariates. *Biom J*. 1996;38:931-940.
25. Sutradhar BC, Das K. On the accuracy of efficiency of estimating equation approach. *Biometrics*. 2000;56(2):622-625.
26. Wang YG, Carey V. Working correlation structure misspecification, estimation and covariate design: implications for generalised estimating equations performance. *Biometrika*. 2003;90(1):29-41.
27. Bombardier C, Ware J, Russell IJ, Larson M, Chalmers A, Read JL. Auranofin therapy and quality of life in patients with rheumatoid arthritis. *Am J Med*. 1986;81(4):565-578.
28. Fernández D, Liu I. A goodness-of-fit test for the ordered stereotype model. *Stat Med*. 2016;35(25):4660-4696.
29. McCullagh P. Discussion of professor Anderson's paper. *J R Stat Soc Ser B Stat Methodol*. 1984;46(1):22-23.
30. Liang KY, Zeger SL. Inference based on estimating functions in the presence of nuisance parameters. *Stat Sci*. 1995;10:158-173.
31. Tsiatis AA. *Semiparametric Theory and Missing Data*. New York, NY: Springer; 2006.
32. Fahrmeir L, Tutz G. *Multivariate Statistical Modelling Based on Generalized Linear Models*. 2nd ed. New York, NY: Springer; 2001: 103-104.
33. Diggle P, Heagerty P, Liang K-Y, Zeger S. *Analysis of Longitudinal Data*. 2nd ed. New York, NY: Oxford University Press; 2002.
34. Zeger SL, Liang K-Y. Longitudinal data analysis for discrete and continuous outcomes. *Biometrics*. 1986;42(1):121-130.
35. Prentice RL. Correlated binary regression with covariates specific to each binary observation. *Biometrics*. 1988;44:1033-1048.
36. Lumley T. Generalized estimating equations for ordinal data: a note on working correlation structures. *Biometrics*. 1996;52(1):354-361.
37. Sharples K, Breslow N. Regression analysis of correlated binary data: some small sample results for the estimating equation approach. *J Stat Comput Simul*. 1992;42:1-20.
38. Spiess M, Hamerle A. Regression models with correlated binary responses: a comparison of different methods in finite samples. *Comput Stat Data Anal*. 2000;33(4):439-455.
39. Touloumis A. R package multgee: a generalized estimating equations solver for multinomial responses. *J Stat Softw*. 2015;64(8):1-14. <http://www.jstatsoft.org/v64/i08/> Accessed December 25, 2019.
40. Goodman LA. The analysis of cross-classified data having ordered and/or unordered categories: association models, correlation models, and asymmetry models for contingency tables with or without missing entries. *Ann Stat*. 1985;13(1):10-69.
41. Deming WE, Stephan FF. On a least squares adjustment of a sampled frequency table when the expected marginal totals are known. *Ann Math Stat*. 1940;11:427-444.
42. McFadden D. Conditional logit analysis of qualitative choice behavior. In: Zarembka P, ed. *Frontiers in Econometrics*. New York, NY: Academic Press; 1974:105-142.

43. Little RJA, Rubin DB. *Statistical Analysis with Missing Data*. 2nd ed. New York, NY: John Wiley; 2002.
44. Graybill FA. *Matrices with Applications in Statistics*. 2nd ed. Belmont, CA: Wadsworth; 1969.
45. McDonald RP. The McDonald-Swaminathan matrix calculus: clarifications, extensions, and illustrations. *Gen Syst*. 1976;21:87-94.
46. McDonald RP, Swaminathan H. A simple calculus with applications to multivariate analysis. *Gen Syst*. 1973;18:37-54.

### SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of this article.

**How to cite this article:** Spiess M, Fernández D, Nguyen T, Liu I. Generalized estimating equations to estimate the ordered stereotype logit model for panel data. *Statistics in Medicine*. 2020;39:1919–1940. <https://doi.org/10.1002/sim.8520>

### APPENDIX A DERIVATIVES

Let  $\mathbf{e}_i = (e_{i12}, \dots, e_{i1q}, \dots, e_{iT2}, \dots, e_{iTq})'$ ,  $\mathbf{e}_{it} = (e_{it2}, \dots, e_{itq})'$ , where  $e_{itk}$  is the  $tk$ th element of  $\mathbf{W}_i^{-1}(\mathbf{Y}_i - \boldsymbol{\mu}_i)$  and note that

$$\mathbf{V}_{it}^{-1} = \text{diag}(\boldsymbol{\mu}_{it}^{-1}) + (1 - \mathbf{1}'_{(q-1)}\boldsymbol{\mu}_{it})^{-1}\mathbf{1}_{(q-1)}\mathbf{1}'_{(q-1)}.$$

<sup>44</sup>For the second term in (6) we get, using the calculus of,<sup>45,46</sup>

$$\sum_i \frac{\partial \mathbf{V}_i}{\partial \boldsymbol{\theta}} (\tilde{\mathbf{X}}_i \mathbf{D}'_{(1)} \otimes \mathbf{V}_i \mathbf{e}_i) = \mathbf{D}_{(1)} \left( \sum_i \mathbf{D}_{(2),i} [\text{diag}(\mathbf{e}_i / \boldsymbol{\mu}_i) - \text{bdiag}(\mathbf{e}'_{i1} \boldsymbol{\mu}_{i1} \mathbf{V}_{i1}^{-1}, \dots, \mathbf{e}'_{iT} \boldsymbol{\mu}_{iT} \mathbf{V}_{iT}^{-1})] \mathbf{D}'_{(2),i} \right) \mathbf{D}'_{(1)}.$$

Let  $\mathbf{u} = (u_{21}, \dots, u_{(q-1)p}, u_{qp}, \dots, u_{21}, \dots, u_{(q-1)p}, u_{qp})' = \sum_i \mathbf{X}'_i \mathbf{V}_i \mathbf{e}_i$ , which is a  $[(q - 1)p] \times 1$  vector, and define

$$\mathbf{B} = \begin{pmatrix} u_{21} & u_{22} & \dots & u_{2p} \\ u_{31} & u_{32} & \dots & u_{3p} \\ \vdots & \vdots & \ddots & \vdots \\ u_{(q-1)1} & u_{(q-1)2} & \dots & u_{(q-1)p} \end{pmatrix}$$

then the first term in (6) reduces to

$$\frac{\partial \mathbf{D}_{(1)}}{\partial \boldsymbol{\theta}} \sum_i (\mathbf{I}_{(q-1)(p+1)} \otimes \tilde{\mathbf{X}}'_i \mathbf{V}_i \mathbf{W}_i^{-1} (\mathbf{Y}_i - \boldsymbol{\mu}_i)) = \begin{pmatrix} \mathbf{0}_{(q-1) \times (q-1)} & \mathbf{0}_{(q-1) \times (q-2)} & \mathbf{0}_{(q-1) \times p} \\ \mathbf{0}_{(q-2) \times (q-1)} & \mathbf{0}_{(q-2) \times (q-2)} & \mathbf{B}_{(q-2) \times p} \\ \mathbf{0}_{p \times (q-1)} & \mathbf{B}'_{p \times (q-2)} & \mathbf{0}_{p \times p} \end{pmatrix}.$$