Discussion of "Generalized Estimating Equations: Notes on the Choice of the Working Correlation Matrix"

J. Breitung[1]; N. R. Chaganty[2]; R. M. Daniel[3]; M. G. Kenward[3]; M. Lechner[4]; P. Martus[5]; R. T. Sabo[6]; Y.-G. Wang[7]; C. Zorn[8]

[1]University of Bonn, Bonn, Germany;
[2]Old Dominion University, Norfolk, VA, USA;
[3]London School of Hygiene and Tropical Medicine, London, UK;
[4]University of St. Gallen, St. Gallen, Switzerland;
[5]Charité – Universitätsmedizin Berlin, Berlin, Germany;
[6]Virginia Commonwealth University, Richmond, VA, USA;
[7]The University of Queensland, St. Lucia, Queensland, Australia;
[8]Pennsylvania State University, University Park, PA, USA

Keywords

Correlation matrix, generalized estimating equations, independence estimating equations, restriction of parameter space

Summary

Objective: To discuss generalized estimating equations as an extension of generalized linear models by commenting on the paper of Ziegler and Vens "Generalized Estimating Equations: Notes on the Choice of the Working Correlation Matrix".

Methods: Inviting an international group of experts to comment on this paper.

Results: Several perspectives have been taken by the discussants. Econometricians have established parallels to the generalized method of moments (GMM). Statisticians discussed model assumptions and the aspect of missing data. Applied statisticians commented on practical aspects in data analysis.

Conclusions: In general, careful modeling correlation is encouraged when considering estimation efficiency and other implications, and a comparison of choosing instruments in GMM and generalized estimating equations (GEE) would be worthwhile. Some theoretical drawbacks of GEE need to be further addressed and require careful analysis of data. This particularly applies to the situation when data are missing at random.

Correspondence to:
See list of authors´ adresses at the end of the article

Methods Inf Med 2010; 49: 426–432

With these comments on the paper "Generalized Estimating Equations: Notes on the Choice of the Working Correlation Matrix", written by Andreas Ziegler and Maren Vens [1], *Methods of Information in Medicine* wants to stimulate a discussion on generalized estimating equations as an extension of generalized linear models. An international group of experts have been invited by the editor of *Methods* to comment on this paper. Each of the invited commentaries forms one section of this paper.

1. Comment by J. Breitung

The paper by Ziegler and Vens [1] provides a comprehensive and up-to-date review of the problem of selecting the working correlation matrix in a GEE framework. In my comments, I will look at the problem from a different angle, as the GEE estimator is related to the Generalized Method of Moments (GMM) estimator, which is very popular in econometrics.

To be specific, consider the (grouped) Probit model that can be written as a nonlinear regression of the form

$$y_{it} = \Phi(\beta' \mathbf{x}_{it}) + u_{it},$$

where $y_{it} \in \{0, 1\}$, $\Phi(\cdot)$ denotes the normal cdf, and u_{it} is a dichotomous random variable with

$$u_{it} \mid x_{it} = \begin{cases} -\Phi(\beta' \mathbf{x}_{it}) \\ \text{with probability } 1 - \Phi(\beta' \mathbf{x}_{it}) \\ 1 - \Phi(\beta' \mathbf{x}_{it}) \\ \text{with probability } \Phi(\beta' \mathbf{x}_{it}) \end{cases}$$

In practical applications we usually have no information on the covariances $\omega_{i,ts} = E(u_{i\cdot t} u_{is})$. Therefore, a specification of the matrix $\Omega i = (\omega_{i,ts})$ has to be selected that depends on a limited number of parameters. In econometric applications, for example, it is quite common to assume a random effects model for the errors in the latent model $y_{it}^* = \beta' \mathbf{x}_{it} + \varepsilon_{it}$, where $\varepsilon_{it} = \alpha_i + v_{it}$, $\alpha_i \sim N(0, \sigma_\alpha^2)$ and $v_{it} \sim N(0, 1 - \sigma_\alpha^2)$. The observed variable y_{it} is obtained from the latent variable y^*_{it} by setting $y_{it} = 1$ if $y_{it}^* > 0$ and zero otherwise. While this random effects specification gives rise to a simple exchangeable correlation structure of the vector of latent errors $(\varepsilon_{i1}, \ldots, \varepsilon_{iT})$, the corresponding covariance matrix Ω_i of the observational equation is difficult to evaluate. The popular econometric software STATA, for example, uses numerical integration techniques to obtain the ML estimator of β and σ_α^2.

Following the influential work of Liang and Zeger, biometricians prefer to impose a simple (but generally misspecified) covariance structure, which is represented by the "working correlation matrix". The idea is, that if Σ_i is a suitable approximation of the unknown covariance matrix Ω_i, we can hope that the efficiency loss due to using a misspecified covariance matrix is small. Ziegler and Vens [1] discuss various special situations in which the IEE estimator with independent estimation equations is as efficient as any other GEE estimator. In other cases, the working correlation matrix has to be carefully specified in order to avoid a substantial loss of efficiency. Different statistical procedures for selecting a suitable working correlation matrix are considered by Ziegler and Vens [1].

An alternative approach to allow for an unrestricted correlation matrix[a] $\mathbf{R}_i = \mathbf{R}$ for all i is the GMM (see [2] for an introduction to GMM estimators). Let $\mathbf{Z}_i = diag(v_{it}^{1/2})\mathbf{D}_i$ and $\mathbf{u}_i = diag(v_{it}^{1/2})(\mathbf{y}_i - \mu_i)$. The GEE estimator is a (just identified) GMM estimator based on the moment conditions $E(\mathbf{Z}_i'\mathbf{R}\mathbf{u}_i) = 0$. In other words, the instruments of the GEE estimator are particular linear combinations of the instruments collected in the matrix $\mathbf{Z}_i$, where the elements of the matrix $\mathbf{R}$ provide the weights. Therefore, these moment conditions define a subset of the following comprehensive set of pT^2 moment conditions:

$$E(\mathbf{Z}_{it}u_{is}) = 0 \quad \text{for } t \in \{1, ..., T\} \text{ and } s \in \{1, ..., T\}, \qquad (1)$$

where $\mathbf{Z}_{it}$ denotes the t-th $(p \times 1)$ column of the matrix $\mathbf{Z}_i'$. The resulting set of moment conditions can be written compactly in matrix form as

$$E(\mathbf{Z}_i^{\circ}\mathbf{u}_i) = 0 \qquad (2)$$

where $\mathbf{Z}_i^{\circ} = I_T \otimes \mathbf{z}_i^{\circ}$ and $\mathbf{x}_i^{\circ} = vec(\mathbf{Z}_i)$. Since a GMM estimator based on a larger set of moments is asymptotically at least as efficient as an estimator based on a subset of moment conditions, the GMM estimator resulting from the pT^2 moment conditions yields the optimal estimator in the class of GEE estimators defined by all possible correlation matrices $\mathbf{R}$. For the panel probit model, a similar GMM estimator was suggested by Breitung and Lechner [3].

At a first glance, this approach looks very attractive as we do not need to estimate the correlation matrix $\mathbf{R}$. It is well known however that the GMM estimator has poor small sample properties if the number of moment conditions is large relative to the sample size n. Therefore, the GMM estimator should only be applied if the group size T is small relative to n.

It is interesting to note that the restricted correlation matrices considered by Ziegler and Vens [1] can also be represented by appropriate GMM estimators. For example, the exchangeable working correlation structure gives rise to a GMM estimator based on the following set of $2p$ moment conditions:

$$E\left(\sum_{t=1}^{T} \mathbf{Z}_{it}u_{it}\right) = 0 \qquad (3)$$

$$E\left(\overline{\mathbf{Z}}_i \sum_{t=1}^{T} u_{it}\right) = 0 \qquad (4)$$

where $\overline{\mathbf{Z}}_i = T^{-1}\sum_{t=1}^{T} \mathbf{Z}_{it}$, since the GEE estimator employs a linear combination of these two moment conditions. Note that these $2p$ instruments define a subset of the full set of pT^2 moment conditions considered above.

These findings suggest that the choice of the working correlation matrix is equivalent to selecting the instruments of a GMM estimator. Strategies for selecting instruments from a large set of moment conditions are suggested, inter alia, by Andrews [4], Donald and Newey [5], Doran and Schmidt [6], Hall and Peixe [7], and Hall et al. [8]. It would be interesting to compare the GMM approach to the "biometric approach" based on GEE estimators.

2. Comment by N. R. Chaganty and R. T. Sabo

We thank the editor for inviting us to comment on the topics presented in this paper [1], and as such we are appreciative to the authors for the chance to discuss and elaborate on their ideas. We applaud the authors in their recognition that there are caveats that need to be addressed when GEEs are applied to dependent dichotomous data. These important considerations have largely been ignored by the statistical community with *very few* exceptions. Unlike the case for normally distributed random variables, for most discrete random variables the variance is a function of the mean: e.g. for dichotomous distributions the variance $p(1 - p)$ is a function of the mean p; for Poisson distributions the variance λ is identical to the mean. In the GLM framework it is the expected values (means) that are modeled against the covariates, and as such the variances are functions of the covariates through their relationships with the mean. When we generalize to the repeated-measures case, both the variance *and* correlation are functions of the marginal means. These conditions can add further restrictions to the correlation matrix *beyond* the positive definite range. See Chaganty and Joe [9] for a detailed description of the bounds and examples of positive definite correlation matrices that are unattainable for dichotomous variables. Further, since the marginal means are estimated with covariates, and since the correlation is a function of the marginal means, the correlation and/or the range of the correlation are also functions of the data. The authors of the current paper provide in Figure 1 graphs of the restricted bounds for correlated binary variables as a function of the means. See Chaganty and Mav [10] for similar graphs of the restricted bounds for correlated Poisson variables.

The correlation bounds for dichotomous variables are important in the sense that values in violation of those boundaries are not feasible, much like a negative variance or a probability that is greater than 1 or negative. Using correlation estimates that violate the restricted range thus leads to regression parameter estimates that are invalid in the sense that they are based on other parameter estimates that are not theoretically allowed. For example, one would not trust their results if they obtained a negative variance estimate; likewise, one should not trust their results if a correlation estimate is not theoretically possible. GEE as a procedure makes no attempt at ensuring that the correlation estimates for discrete random variables are within the restricted ranges. Sabo and Chaganty [11] provide examples of the potential consequences when the GEE method produces correlation estimates outside the feasible range. They also provide alternatives, advocating the use of likelihood-based procedures, such as the multivariate probit regression [12], that adhere to the added restrictions on the correlation due to the discrete nature of the random variables. In the current paper, the authors also provide three techniques for selecting covariance/

[a] Although the working correlation matrix is indexed by i it is usually assumed to be identical for all i.

correlation structures that do not violate the data-dependent bounds on the correlation. We feel that this is a major step in the right direction for the GEE literature.

However, there is a larger issue to consider in that the GEE methodology when applied to discrete data is fundamentally flawed. The base GEE procedure (i.e. what one would use in SAS Proc GENMOD or S-PLUS gee.fit) models the covariance matrix as

$$\Sigma = (A(\beta))^{1/2}\, R(\alpha)(A(\beta))^{1/2} \qquad (1)$$

where $A(\beta)$ is a diagonal matrix of variance functions (resulting from the selected link function) dependent on the regression parameter β, and $R(\alpha)$ is a symmetric working correlation matrix parameterized by α. Implicit in Equation 1 is the assumption that the variances of the response variables are functions of the mean (via matrix $A(\beta)$) but not the correlation. This is clearly an unreasonable assumption, as for correlated discrete random variables either the correlation itself or its feasible range depends upon the marginal mean and the covariates. Further, as Crowder [13] points out, the working correlation matrix used in GEE lacks a proper mathematical definition. This fact has been mentioned elsewhere [14] and elaborated upon, but bears repeating: GEE uses a working correlation structure for estimation, and is not defined as the correlation between the response variables. If the working correlation is not the true correlation, then what is it? What is the mathematical relation of the working correlation to the distribution of the repeated responses?

To expand upon this point, since there is no specified correlation between the dependent outcomes there is also no specified probability distribution. The proofs of consistency and asymptotic normality of the GEE regression parameter estimate rely on the law of large numbers and the central limit theorem. However, these theorems are not applicable, as they both require the existence of an underlying probability distribution. Some researchers have argued that quasi-likelihood theory can form the foundation for the GEE method. However, Lee and Nelder [15] point out that for correlated responses no quasi-likelihood function exists. In either case, the claims of consistency and asymptotic normality for the GEE regression parameter estimate are not on a solid foundation. Unless the GEE method can be modified or such theory can be provided to avoid or correct for these issues, much caution should be used in applying the methodology, or alternatives should be employed instead.

3. Comment by R. M. Daniel and M. G. Kenward

We would like to commend the authors of this paper [1] on their thorough treatment of the issue of selecting the working correlation matrix in the use of Generalized Estimating Equations (GEEs). In our comments we would like to extend the discussion to the problem of missing data, in particular to dropout (or attrition) in longitudinal data, a setting for which GEEs are widely used. The relevance of the missing value problem in this setting was recognized early. Liang and Zeger [17] wrote in their original paper:

For [GEE estimators] to be consistent even when R is misspecified, we require that data be missing completely at random . . . When R is the true correlation, the missing completely at random assumption can be unnecessary . . . For binary outcomes, the pattern can depend on any single previous outcome [and consistency is retained].

Before pursuing this issue we need to make clear the meaning of 'Missing Completely at Random' (MCAR) and contrast it with 'Missing at Random' (MAR), terms now very familiar in the missing value literature. Put simply, the missing data mechanism is MCAR if the probability of an observation being missing does not depend on observed or unobserved measurements, and MAR if, conditional on the observed data, the missing value mechanism does not depend on the unobserved data (Rubin [19]). There are further subtleties to these definitions which we do not need to explore here. There are three key points to make about the MCAR/MAR distinction. First, MCAR is highly implausible in longitudinal studies with dropout and, while still unlikely to fully hold in most settings, MAR is less implausible, and there is a sense in which an analysis that holds under MAR is likely to be less biased, even when this assumption does not hold exactly, than one that requires MCAR. There are of course important exceptions to this. Second, as made clear in Liang and Zeger's [17] quote above, MCAR will be required in general for simple GEE's to produce consistent estimators. Under MAR there are special conditions that lead to consistent estimation, and we look more closely at these below. Third, under a correctly specified model, likelihood based inferences will be valid under MAR. In the light of these points a range of methods have been proposed for making corrections to simple GEE estimators to provide consistency under MAR, principally using either appropriate weighting of the equations or some form of multiple imputation. These methods are reviewed in Molenberghs and Kenward ([18], chapters 10 and 11).

We are concerned here with the role of the working correlation matrix R in this issue. From an intuitive point of view, the validity of the likelihood analyses under MAR in the dropout setting, comes from (among other things) the role of the likelihood in correctly specifying the future statistical behavior of those who drop out. Implicitly the correct conditional distribution of future observations given past ones is used, and under MAR, this distribution is consistently estimated from those who remain. More generally the validity of analyses under MAR derives from properly estimating the moments of these conditional distributions. With likelihood all moments are (implicitly) estimated, while valid semiparametric methods, including those based on GEE, require a restricted set of conditional moments only to be estimated [21]. Only in special circumstances will simple GEE reduce (implicitly) to the appropriate conditional moments.

We now restrict ourselves to the binary setting, as in the paper, and assume that the means model is saturated, and that the working correlation matrix is the correct one. Then it can be shown that, if there are only two repeated measurements, simple GEE will lead to consistent estimators under MAR dropout [16]. This result can be exploited in two different ways when there are more than two times, in each case

using a local dependence property that requires consideration only of neighboring pairs of measurements. 1) Seaman and Copas [20] demonstrate validity of simple GEE in these circumstances when the MAR mechanism is restricted to dependence on a single preceding measurement only; realistically this would be the previous one. That is, the local dependence applies to the missing value mechanism. 2) Daniel [16] shows that validity holds when the data themselves follow a Markov structure, that is, the local dependence applies to the outcomes. Simulations indicate that estimates can still be quite robust to departures from these two assumptions, but an appropriate R is still required. With GEE the implied regression of future measurements on the past is linear, and presumably in those settings where bias is negligible this regression is providing an acceptable approximation to the actual regression relationship.

In summary, when there are missing data, the role of the working correlation matrix becomes potentially critical to ensure validity under MAR. Dependence among the outcomes, in terms of appropriate conditional moments, needs to be represented at least approximately correctly in the estimating process. Even with access to the true correlation matrix, consistency is only guaranteed in very limited settings.

4. Comment by M. Lechner

It is a pleasure to have the opportunity to comment on the very thoughtful paper by Andreas Ziegler and Maren Vens [1]. Being an econometrician and not an epidemiologist, the first observation when reading this and many related papers is about the lack of interaction between the literature written in those fields.

Indeed, the literature on Generalized Estimation Equations (GEE) is closely related to the literature on Generalized Methods of Moments estimation (Hansen [22]). Of particular relevance here is a special case of GMM called Conditional Methods of Moments estimation (Chamberlain [23], Newey [24], and the survey by Newey [25]). CME is based on almost the same general structure as GEE although different names are used for the various objects in GEE and CME. In CME we have to (correctly) specify the conditional (on exogenous covariates, X) expectation of the dependent variable and then find 'instruments', i.e. particular linear or nonlinear functions of the covariates uncorrelated with those conditional moments. If we do so in an asymptotically optimal way, we obtain an estimator exploiting the information in those moments efficiently in large samples. If we use suboptimal instruments instead, then under weak conditions CME remains consistent, but is less precise[b]. As shown for example by Bertschek and Lechner [26], for the panel probit model the optimal instruments depend on the with-in cluster correlation structure. This result is of course also true for other nonlinear models.

My methodological background in econometrics and lack of practical experience in actually estimating GEE models (GEEs are rarely used in empirical studies in economics), leads me to the following two considerations on how to obtain more precise GEE estimators in practice that go somewhat beyond the paper by Ziegler and Vens [1]. In their paper they give hints on how to choose the optimal correlation matrix given i) that the conditional expectation of y_{it} is the key input in GEE, and ii) that the approximate working covariance matrix Σ_i is used. Under those conditions and the given parameterization of the working covariance matrix the remaining issue is how to choose the working correlation structure R_i that appears in Σ_i.

Considering (ii), we observe that as R_i usually does not depend on covariates (at least in the examples given by Ziegler and Vens), it will not always be possible to exploit the information that is contained in the model about the conditional expectations optimally. To see this, consider the dichotomous model as a convenient example. In addition to the notation already introduced, define $g^{(2)}(x_{it}\beta, x_{is}\beta, \rho_{ts}) \equiv E(y_{it}\, y_{is} | \mathrm{X}_i) = P(y_{it} = 1, y_{is} = 1 | \mathrm{X}_i)$, where $g^{(2)}(\cdot)$ denotes another link function and ρ_{ts} an unknown coefficient. It also implicitly assumes that $E(y_{it}\, y_{is} | \mathrm{X}_i)$ depends on X only via the linear indices of the relevant periods and the pairwise correlation coefficient, as it would, for example, be true in the probit model. Of course, this assumption can easily be generalized. For the probit model, $g_{ts}^{(2)}(\cdot)$ denotes the cumulative distribution function of the bivariate standard normal distribution evaluated at $x_{it}\beta$ and $x_{is}\beta$, and ρ_{ts} denotes the corresponding correlation coefficient of the error terms in the latent model (that are assumed to be jointly normally distributed in the probit model). Hence, we get following expression for the variance-covariance matrix of the conditional moments:

$$
\begin{aligned}
&\mathrm{Covar}(y_{it}, y_{is} | \mathrm{X}_i) \\
&= E\{[y_{it} - g(x_{it}\beta)][y_{is} - g(x_{is}\beta)] | X_i\} \\
&= \begin{cases} g(x_{it}\beta)[1 - g(x_{it}\beta)] & \text{if } t = s \\ g^{(2)}(x_{it}\beta, x_{is}\beta, \rho_{ts}) - g(x_{it}\beta)g(x_{is}\beta) & \text{if } t \neq s. \end{cases}
\end{aligned}
$$

For the case of the panel probit model, Bertschek and Lechner [26] show that this expression for the conditional covariance matrix can easily be exploited to obtain an estimator that is asymptotically efficient based on the information that $E(y_{it}|X_i) = g(x_{it}\beta)$. The required estimation to obtain a weighting matrix that achieves efficiency either involves bivariate probit estimation or simple nearest neighbor (or any other nonparametric) smoothing, both of which are rather simple, fast (in particular given modern day computing power) and well established estimation methods.

The other issue the authors take for granted is that we are exploiting the first conditional moments, $E(y_{it}|X_i) = g(x_{it}\beta)$, only. However, in particular in the case of a dichotomous model, an assumption on the link function for $E(y_{it}y_{is}|X_i)$ may come very naturally and may not be considered restrictive at all. In terms of the probit models, $E(y_{it}|X_i) = g(x_{it}\beta)$ would involve the assumption that the marginal distributions of the error terms in the latent model are normally distributed, whereas $E(y_{it}y_{is}|X_i) = g^{(2)}(x_{it}\beta, x_{is}\beta, \rho_{ts})$ would follow from the additional assumption that all pairs of error terms are jointly bivariate normally distributed. There are not many applications in which the first assumption is deemed to be credible but not the second. The additional information provided by these moment conditions is obtained

[b] All these results are valid only asymptotically, at least for non-linear conditional expectations.

exactly from those intra-cluster dependencies, which are also the motivation for choosing the initial working correlation matrix. I do not know of any paper that has investigated the issues related to such an approach, for example like i) how important is the gain in precision by adding these extra moments? ii) do the small sample properties of the estimators deteriorate due to the increase in the dimension of the estimation problem?, or iii) how important is the correlation structure of the expanded weighting matrix necessary for the extra conditions? Nevertheless, if an increase in precision is the name of the game, this strategy seems to be worthwhile, in addition to the sensible proposals made by Ziegler and Vens [1].

5. Comment by P. Martus

Let me first thank the editor for inviting me to comment on the paper of Ziegler and Vens [1]. The problem of choosing the "best" working correlation matrix in a GEE analysis is of high importance and the paper is providing very useful criteria for this choice. Most important, it becomes clear that in general these matrices should not be chosen too complicated. I want to comment on several aspects:

There is a very basic example which may illustrate for readers not so familiar with GEE the consequences of IEE analyses: Just consider the problem of estimating the expectation of clustered normally distributed data without any further covariates. Basically we can choose some type of weighted mean or just the unweighted mean. If we choose to estimate the weights from the data, we are in exactly the same situation as in GEEs with the nondiagonal working correlation matrix to be estimated. If we choose not to use any weights we are applying IEE. Obviously, in case of small cluster effects or approximately equal cluster sizes (one of both criteria is sufficient) the weighted and the non-weighted mean will not differ considerably as can be shown with elementary calculations. However, very different sized clusters in combination with large cluster effects will clearly lead to a considerable loss of efficiency. If we use the unweighted mean, large clusters will dominate the estimate even so they do not provide substantially more information than smaller ones.

Note that in both cases GEEs/IEEs would provide consistent estimates of standard errors as the observed cluster effect is taken into account in the midth term of the sandwich estimator (3), namely $\mathbf{\Omega}_i$.

A second remark is what to do if results vary considerably depending on the chosen working correlation matrix. If the cluster size and structure is similar for each cluster and the number of observations is not too large, one could argue that it is useful to inspect the fully structured correlation matrix and then decide on using a simpler one which is mostly similar to the estimate of the unstructured one. However, I was confronted with a real data example were IEE; exchangeable and AR(1) working correlation matrices led to very similar results whereas the fully structured approach revealed completely different parameter estimates. This was probably due to some outliers, and I would not recommend this approach. (Data are available from the author by request.)

Nevertheless, the choice of the working correlation matrix has some analogies to the problem of variable selection in a regression model or to the choice of the correct model structure in confirmatory factor analysis. We can start with the identity matrix ("zero structure" IEE) and proceed to more complicated working correlation structures (i.e. exchangeable, blockwise exchangeable, fully structured) analogously to forward variable selection and of course it is principally possible to do it backwards. However, the example described in the preceding paragraph demonstrates that "backward selection" which could start with the fully structured working correlation matrix might not be the best way due to instabilities of the estimated correlations.

Finally, it is clear from the idea of GEEs, especially IEEs, that the correlations within clusters are nuisance parameters, the aim of the analysis is not to provide interpretations of these parameters. From my point of view the logical consequence of this fact should be that the criterion for the choice of the working correlation matrix (or more precisely its structure) should be that it provides stable estimates of the mean structure and the standard errors. Thus, the comparison of "observed" vs. "predicted" correlations should not be the primary criterion. I prefer approaches which investigate the stability of the mean structure using cross validation or bootstrap and based on these results decide to use the working correlation matrix.

6. Comment by Y.-G. Wang

This paper [1] provides a review on selection of a working correlation structure when applying the well known approach of GEE, and, in particular, identifies situations when the independence model is appropriate based on both statistical and biological arguments. Overall, I find their conclusions/recommendations reasonable.

The merits of the independence model are also described by a few other authors (e.g., Fitzmaurice [27], and the references there). The most appealing advantage is that we do not need to model the correlation structure or estimate any correlation parameters. It is therefore very useful to obtain the estimates using an independence model. Estimates from other working models should be compared and reasons for large differences should be explored as any other improved estimates should only differ by $O_p(1/n^{1/2})$, where n is the total sample size.

Model selection criteria are used to assist and guide our modeling process and making final recommendations. In the case of multiple covariates consisting of both cluster-level and within-cluster covariates, efficiency loss from the independence model is often substantial unless the within cluster correlations are small. In general, even for balanced designs, modeling correlation structure should be encouraged, and in many cases, some simple correlation structure such as AR(1) or exchangeable may be adequate in capturing most efficiency over the independence model. This will also give us some rough estimates of the correlations. Although the correlations may be regarded as nuisance parameters, they often have implications in interpreting statistical results and designing similar studies.

The other important issue is that correlation cannot be ignored when calculating the standard errors. The performance of the robust estimator relies on the sample size being moderate. Careful modeling of correlation structure will result in more reliable estimates of parameters and possibly their standard errors. Improving the robust variance estimator is also important which perhaps we should pay more attention to.

As we know, correlation modeling is conditional on correct specification of the mean and variance function. To what extent the mean and variance functions are well modeled will have a great impact on results of correlation structure selection. In practice, these model components are often tangled together, which makes the modeling even more complicated [28].

Finally, I would like to congratulate the authors for an interesting review paper on this topic.

7. Comment by C. Zorn

Ziegler and Vens have done a substantial service for applied researchers who use generalized estimating equation (GEE) models in their work [1]. The question of the optimal specification of the working correlation matrix is one that has vexed analysts for more than two decades. By outlining the relevant theoretical findings on this question, and providing a clear set of guidelines and criteria, the authors have contributed significantly to the usability of GEE models in clinical and epidemiological (as well as other scientific) settings.

At the same time, it is important to note that the use of GEE models has also seen significant growth in the social and behavioral sciences as well, and that – while the recommendations of the authors are generally sound – analysts in those disciplines confront somewhat different challenges. In particular, it is frequently the case that social science applications of GEEs (including those in clinical and social psychology, demography, and medical sociology) face higher degrees of imbalance in cluster sizes and/or the absence of mean balance in covariates; as the authors note, such conditions are particularly likely to hold in observational studies common in those fields.

In such instances, the critical question is not the decision between IEE and GEE, but rather the criteria by which the structure of $\mathbf{R}_i(\alpha)$ is chosen. On the one hand, Ziegler and Vens' recommendations regarding sensitivity analyses and their suggestion that one combine biological/substantive and statistical motivations for that specification are undoubtedly good practice. Conversely, reliance on unconditional estimates of within-cluster dependence (original cite 11) runs the risk of substantially overstating the degree of intracluster dependence in such studies. This is because, unlike in experimental or crossover trials, studies relying on observational data often see substantial differences between intracluster correlations before and after conditioning on observed covariates. To the extent that controlling for right-hand-side variables will tend to reduce the marginal intracluster correlations in such instances, such an unconditional approach will tend to overstate the need for higher degrees of dependency in working correlations.

A recent survey [29] comparing population-averaged and conditional (mixed-effects) models notes that "if the focus of the analysis is the estimation of mean effects as well as the estimation of the inference of the coefficients in the model ... then estimating the population-average model via GEE provides a compelling alternative" (1). While the advantages of such approaches do not turn critically on the choice of the working correlation matrix, both efficiency and biological theory often support its selection on informed grounds. By providing a set of guidelines for that selection, the authors have made an important and lasting contribution to those models' application.

References

1. Ziegler A, Vens M. Generalized Estimating Equations: Notes on the Choice of the Working Correlation Matrix. Methods Inf Med 2010; 49 (5): 421–425.
2. Hayashi F. Econometrics: Princeton University Press; 2000.
3. Breitung J, Lechner M. Some GMM Estimation Methods and Specification Tests for Nonlinear Models. In: Matyas L, Sevestre P (eds). The Econometrics of Panel Data. 2nd edition. Dordrecht: Kluwer; 1996. pp 583–612.
4. Andrews DWK. Consistent Moment Selection Procedures for Generalized Methods of Moments Estimation. Econometrica 1999; 67: 543–564.
5. Donald SG, Newey WK. Choosing the number of instruments. Econometrica 2001; 69: 1161–1192.
6. Doran H, Schmidt P. GMM Estimators with Improved Finite Sample Properties Using Principal Components of the Weighting Matrix, with an Application to the Dynamic Panel Data Model. Journal of Econometrics 2006; 133: 387–409.
7. Hall AR, Peixe FPM. A consistent method for the selection of relevant instruments. Econometric Reviews 2003; 3: 269–287.
8. Hall AR, Inoue A, Jana K, Shin C. Information in generalized method of moments estimation and entropy-based moment selection. Journal of Econometrics 2007; 138: 488–512.
9. Chaganty NR, Joe H. Range of correlation matrices for dependent Bernoulli random variables. Biometrika 2006; 93 (1): 197–206.
10. Chaganty NR, Mav D. Estimation methods for analyzing longitudinal data occurring in biomedical research. In: Khattree R, Naik DN (eds). Computational Methods in Biomedical Research. London, Boca Raton, FL: Chapman and Hall, CRC; 2007; 12: 371–400.
11. Sabo RT, Chaganty NR. What can go wrong when ignoring correlation bounds in the use of generalized estimating equations. Statistics in Medicine 2010. DOI: 10.1002/sim.4013.
12. Ashford JR, Sowden RR. Multi-variate probit analysis. Biometrics 1970; 26 (3): 535–546.
13. Crowder M. On the use of working correlation matrix in using generalised linear models for repeated measures. Biometrika 1995; 82: 407–410.
14. Chaganty NR, Joe H. Efficiency of generalized estimating equations for binary responses. Journal of the Royal Statistical Society, B 2004; 66 (4): 851–860.
15. Lee Y, Nelder JA. Conditional and marginal models: Another view. Statistical Science 2004; 19 (2): 219–238.
16. Daniel RM. On Aspects of Robustness and Sensitivity in Missing Data Methods. Unpublished PhD thesis. London School of Hygiene and Tropical Medicine, UK, 2009.
17. Liang K-Y, Zeger SL. Longitudinal data analysis using generalized linear models. Biometrika 1986; 73: 13–22.
18. Molenberghs G, Kenward MG. Missing Data in Clinical Studies. Chichester: Wiley; 2007.
19. Rubin DB. Inference and missing data. Biometrika 1976; 63: 581–592.
20. Seaman S, Copas A.J. Doubly robust generalised estimating equations for longitudinal data. Statistics in Medicine 2009; 28: 937–955.
21. Tsiatis AA. Semiparametric Theory and Missing Data. New York: Springer; 2006.
22. Bertschek I, Lechner M. Convenient estimators for the panel probit model. Journal of Econometrics 1998; 87: 329–371
23. Chamberlain G. Asymptotic efficiency in estimation with conditional moment restrictions. Journal of Econometrics 1987; 34: 305–334
24. Hansen L. Large sample properties of generalized methods of moments estimators. Econometrica 1982; 50: 1029–1055.
25. Newey WK. Effcient estimation of models with conditional moment restrictions. In: Maddala G,

Rao C, Vinod H (eds.). Handbook of Statistics, vol 11, chap 16. North-Holland, Amsterdam; 1993.
26. Newey WK. Efficient instrumental variables estimation of nonlinear models. Econometrica 1990; 59: 809–837.
27. Fitzmaurice GM. A caveat concerning independence estimating equations with multivariate binary data. Biometrics 1995; 51: 309–317.
28. Wang Y-G, Hin L-Y. Modeling strategies in longitudinal data analysis: Covariate, variance function and correlation structure selection. Computational Statistics and Data Analysis; 2009. Doi: 10.1016/j.csda.2009.11.006.
29. Hubbard AE, Ahern J, Fleischer N, Laan M, Lippman S, Jewell N, Bruckner T, Satariano W. To GEE or Not to GEE: Comparing Population Average and Mixed Models for Estimating the Associations Between Neighborhood Risk Factors and Health. Epidemiology 2010; 21: 467–474.

Addresses of the authors:

Prof. Dr. Jörg Breitung
University of Bonn
Department of Economics
Institute of Econometrics
Adenauerallee 24–42
53113 Bonn
Germany
E-mail: breitung@uni-bonn.de

N. Rao Chaganty
Professor of Statistics
Department of Mathematics and Statistics
Old Dominion University
Norfolk, VA 23529
USA
E-mail: rchagant@odu.edu

Rhian M. Daniel, MA MSc PhD
Research Fellow
London School of Hygiene and Tropical Medicine
Keppel Street
London WC1E 7HT
UK
E-mail: Rhian.Daniel@lshtm.ac.uk

Michael G. Kenward
Professor of Biostatistics
Room 129
London School of Hygiene and Tropical Medicine,
Keppel Street
London WC1E 7HT
UK
E-mail: mike.kenward@lshtm.ac.uk

Prof. Dr. Michael Lechner
Professor of Econometrics
Swiss Institute for Empirical Economic Research (SEW)
University of St. Gallen
Varnbüelstrasse 14
9000 St. Gallen
Switzerland
E-mail: Michael.Lechner@unisg.ch

Prof. Dr. Peter Martus
Charité – Universitätsmedizin Berlin
Institute for Biostatistics and Clinical Epidemiology
Hindenburgdamm 30, Haus 1
12205 Berlin
Germany
E-mail: peter.martus@charite.de

Roy T. Sabo
Assistant Professor
Department of Biostatistics
Virginia Commonwealth University
Richmond, VA 23298
USA
E-mail: rsabo@vcu.edu

You-Gan Wang
Professor of Applied Statistics
Centre for Applications in Natural Resource Mathematics (CARM)
School of Mathematics and Physics
The University of Queensland
St. Lucia, Queensland 4072
Australia
E-mail: you-gan.wang@uq.edu.au

Christopher Zorn
Liberal Arts Research Professor
Department of Political Science
Pennsylvania State University
University Park, PA 16803
USA
E-mail: zorn@psu.edu

