# Modeling Time-Dependent Covariates in Longitudinal Data Analyses

**Trent L. Lalonde**

**Abstract** Often public health data contain variables of interest that change over the course of longitudinal data collection. In this chapter a discussion is presented of analysis options for longitudinal data with time-dependent covariates. Relevant definitions are presented and explained in the context of practical applications, such as different types of time-dependent covariates. The consequences of ignoring the time-dependent nature of variables in models is discussed. Modeling options for time-dependent covariate data are presented in two general classes: subject-specific models and population-averaged models. Specific subject-specific models include random-intercept models and random-slopes models. Decomposition of time-dependent covariates into "within" and "between" components within each subject-specific model are discussed. Specific population-averaged models include the independent GEE model and various forms of the GMM (generalized method of moments) approach, including researcher-determined types of time-dependent covariates along with data-driven selection of moment conditions using the Extended Classification. A practical data example is presented along with example programs for both SAS and R.

## 1 Introduction and Motivating Examples

The term "longitudinal data" refers to data that involve the collection of the same variables repeatedly over time. Typically the term is used to refer to longitudinal *panel* data, which denotes the case of collecting data repeatedly from the same subjects. This type of data is very common in practice, and allows for researchers to assess trends over time and gives power to typical population comparisons (Zeger and Liang 1992; Diggle et al. 2002; Hedeker and Gibbons 2006; Fitzmaurice et al. 2012). Specific research interests can include comparisons of mean responses at different times; comparisons of mean responses across different populations, accounting for the effects of time; and assessing the impacts of independent

T.L. Lalonde (✉)
Department of Applied Statistics and Research Methods, University of Northern Colorado, Greeley, CO, USA
e-mail: trent.lalonde@unco.edu



D.-G. Chen, J. Wilson (eds.), *Innovative Statistical Methods for Public Health Data*, ICSA Book Series in Statistics, DOI 10.1007/978-3-319-18536-1_4

variables on responses, accounting for the effects of time. Longitudinal models allow these questions to be answered, while accounting for the dependence inherent in repeated observation of the same individuals.

Independent variables in longitudinal studies can be broadly classified into one of two categories: time-independent covariates (TIC), or time-dependent covariates (TDC). The differences between these types of covariates can lead to different research interests, different analysis approaches, and different conclusions.

TIC are independent variables with no within-subject variation, meaning that the value of a TIC does not change for a given individual in a longitudinal study. This type of covariate can be used to make comparisons across populations and to describe different time trends, but does not allow for a dynamic relationship between the TIC and response.

TDC are independent variables that include both within-subject variation and between-subject variation, meaning that the value of a TDC changes for a given individual across time and can also change among different subjects. A TDC can be used to make comparisons across populations, to describe time trends, and also to describe dynamic relationships between the TDC and response. The focus of this chapter will be on appropriate analysis techniques for TDC.

Examples of longitudinal data with TDC arise often and in many disciplines. For example, Phillips et al. (2014) were interested in the associations among marijuana usage, drug craving, and motivation. For "heavy users" of marijuana between the ages of 18 and 30, data were collected three times per day for 14 consecutive days. To model the mean number of times marijuana was used, a longitudinal count model was applied using drug craving and motivation as predictors. Over the course of 14 days, both drug craving and motivation vary both within and between subjects, and therefore should be treated as TDC.

Using the Arizona state inpatient database (SID) for records of hospital visits, Lalonde et al. (2014) modeled the probability of rehospitalization within 30 days of a previous visit. Rehospitalization within this time frame is an important consideration for Medicare funding. Subjects for the study were selected such that each subject in the database had exactly three hospital follow-ups. Using a longitudinal logistic model, predictors of probability of rehospitalization included the number of diagnoses during a hospital visit, the number of procedures performed, and the length of hospital stay. Each of these predictors can vary over the three hospital follow-ups, and therefore should be treated as TDC.

It can be seen that TDC allow for different types of conclusions and relationships than TIC. For example, TDC can be involved in accumulated effects from differing values over time (Fitzmaurice and Laird 1995). It is also clear that certain TDC convey different information than others. For example, variables such as age may change over time, but change predictably. On the other hand, variables such as daily precipitation may change over time but cannot be predicted as age can. In such cases it is important to consider relationships between the TDC and the response across time.

In the following sections, the distinctions among TDC will be explored, including methods of identifying types of TDC. Modeling TDC data using conditional methods is discussed, followed by modeling using marginal methods. The chapter concludes with a data example exemplifying all relevant modeling techniques.

## 2 Classifying Time-Dependent Covariates

Within a longitudinal study, a TDC can be defined as a variable that involves variation both within subjects and between subjects. The additional variation within subjects is a source of dispersion that must be accounted for in longitudinal models, and can provide insight into dynamic relationships between a TDC and the response.

Various types of TDC can behave differently from each other. Variables such as "time of observation" or "treatment" can change through a study, but these changes are inherently deterministic. While there may be an association between such variables and the response at a given time, the associations should not carry over such that the "treatment" at one time affects the response at a different time. Subject-specific variables such as "systolic blood pressure" or "drug craving" can change over time, although not deterministically. These types of variables are associated with subject characteristics, and as such can often be involved in dynamic "feedback" relationships with the response. The response at a given time can be affected by the accumulated prior values of such a variable, and correspondingly the value of the response can affect these variables in future observations. Covariates involved in feedback have also been referred to as "time-dependent confounders" (Diggle et al. 2002). Random variables such as "atmospheric pressure" or "unsolicited donations" can change over time, but vary randomly with respect to a system. These types of variables can have accumulated effects on the response, but feedback is unlikely.

It is evident that TDC can be classified according to the nature of the relationship between the TDC and the response. It is important to identify the different types of TDC, as different types of covariates can be associated with different conclusions or different appropriate estimation methods within the same models.

### *2.1 Exogeneity*

One of the most common distinctions made of TDC is that of exogeneity (Chamberlain 1982; Amemiya 1985; Diggle et al. 2002). An exogenous variable has a stochastic process that can be determined by factors outside the system under study, and is not influenced by the individual under study. An exogenous TDC can be thought of as a randomly fluctuating covariate that cannot be explained using other variables in the study. It is most important to determine exogeneity with respect to the response. A TDC is said to be exogenous with respect to the response process if that time-dependent variable at one time is conditionally independent of all previous responses.

Formally, let the response for subject $i$ at time $t$ be denoted by $Y_{it}$, and let $x_{it}$ denote a TDC for subject $i$ at time $t$. Then $\mathbf{x}$ is exogenous with respect to the response process $\mathbf{Y}$ if

$$f_X\left(x_{it}|Y_{i1},\ldots,Y_{it};x_{i1},\ldots,x_{i(t-1)}\right)=f_X\left(x_{it}|x_{i1},\ldots,x_{i(t-1)}\right), \tag{1}$$

where $f_X$ denotes the density of $\mathbf{x}$. Under the definition of Eq. (1), while $x_{it}$ may be associated with previous covariate values $x_{i1},\ldots,x_{i(t-1)}$, it will not be associated with previous or current responses $Y_{i1},\ldots,Y_{it}$. A consequence of this definition is that the current response $Y_{it}$ will be independent of future covariate values, even if there is an association with prior covariate values,

$$\mathrm{E}\left[Y_{it}|x_{i1},\ldots,x_{iT}\right]=\mathrm{E}\left[Y_{it}|x_{i1},\ldots,x_{i(t-1)}\right]. \tag{2}$$

Exogeneity with respect to the response has important modeling implications. Specifically, the definition implies that the response at any time may depend on prior responses and prior values of the TDC, but will be independent of all other covariate values. There is no feedback cycle of effects between responses and exogenous TDC.

TDC that are not exogenous are referred to as endogenous TDC. An endogenous variable, sometimes called an internal variable, is a variable that is stochastically related to other measured factors in the study. This can also be defined as a variable generated by a process related to the individual under study. In other words, endogenous TDC are associated with an individual effect, and can often be explained by other variables in the study. When the stochastic process of an endogenous TDC can be (at least partially) explained by the response variable, there is said to be feedback between the response and endogenous TDC. This type of relationship should be accounted for in any longitudinal model with TDC.

As discussed by Diggle et al. (2002), exogeneity can be assessed by considering a regression of covariate values $x_{it}$ on both prior covariate values $x_{i1},\ldots,x_{i(t-1)}$ and also prior response values $Y_{i1},\ldots,Y_{i(t-1)}$. If, after controlling for prior covariate values, the current covariate value $x_{it}$ shows an association with past response values, the covariate shows evidence of endogeneity.

## 2.2 Types of Time-Dependent Covariates

Recent work has focused on further categorization of types of TDC to facilitate interpretations and proper estimation methods for a model. While these additional types can be interpreted generally with respect to the covariate and response, they are defined with respect to an appropriately defined marginal response distribution. Suppose the marginal mean of the response for subject $i$ at time $t$ is denoted by $\mu_{it}(\boldsymbol{\beta})$, where $\boldsymbol{\beta}$ is a vector of mean parameters. This definition may be induced by an appropriately defined generalized linear model. Four types of TDC can be defined using distinctions in the relationships between the rate of change of the mean and raw errors between the response and mean.

Lai and Small (2007) defined three types of TDC, and a fourth type was defined by Lalonde et al. (2014). Each type of TDC is related to the extent of non-exogeneity with respect to the response and can help determine appropriate analysis techniques. A covariate is said to be a Type I TDC if

$$\mathrm{E}\left[\frac{\partial \mu_{is}}{\partial \beta_j}(Y_{it} - \mu_{it})\right] = 0 \qquad \forall s, t, \tag{3}$$

where $\mu_{is}$ and $\mu_{it}$ are evaluated at the true parameter values $\boldsymbol{\beta}$, and $j$ is the index of the TDC in question. The expectation must be satisfied for all combinations of times $s$ and $t$, suggesting there is no relationship between the TDC and the response at different times. A sufficient condition for a TDC to be Type I is

$$\mathrm{E}\left[Y_{it}|x_{i1}, \ldots, x_{iT}\right] = \mathrm{E}\left[Y_{it}|x_{it}\right]. \tag{4}$$

Thus the response is independent of all TDC values at different times. The sufficient requirement of Eq. (4) would seem to be a stronger condition than the exogeneity presented by Eq. (2), in that Eq. (4) requires the response at time $t$ to be independent of all other TDC values, even those prior to $t$. Variables that involve predictable changes over time, such as age or time of observation, are typically treated as Type I TDCs. A covariate is said to be a Type II TDC if

$$\mathrm{E}\left[\frac{\partial \mu_{is}}{\partial \beta_j}(Y_{it} - \mu_{it})\right] = 0 \qquad \forall s \geq t. \tag{5}$$

The expectation must be satisfied when $s \geq t$, but not necessarily when $s < t$, suggesting dependence between the response and covariate. In this case the TDC process is not associated with prior responses, but the response process can be associated with prior TDC values. A sufficient condition for a covariate to be Type II is

$$\mathrm{E}\left[Y_{it}|x_{i1}, \ldots, x_{iT}\right] = \mathrm{E}\left[Y_{it}|x_{i1}, \ldots, x_{it}\right].$$

As discussed in Lai and Small (2007), this is similar but not equivalent to exogeneity with respect to the response process. It can be shown that exogeneity is sufficient for a TDC to be of Type II (Chamberlain 1982; Lai and Small 2007). Examples of Type II TDCs include covariates that may have a "lagged" association with the response in that previous TDC values can affect the response, but covariate values will not be affected by previous response values. One example is the covariate "blood pressure medication" as a Type II covariate with the response "blood pressure," as the accumulated effects of medication over time can be expected to have an impact on blood pressure at any time. A covariate is said to be a Type III TDC if

$$\mathrm{E}\left[\frac{\partial \mu_{is}}{\partial \beta_j}(Y_{it} - \mu_{it})\right] = 0 \qquad \forall s = t. \tag{6}$$

For Type III TDC, there is no assumption of independence between responses and covariate values at different times. Thus a Type III TDC may involve a feedback cycle between the covariate and response, in which covariate values can be affected by previous response values. One example is the covariate "blood pressure medication" as a Type III covariate with the response "myocardial infarction." While it is expected that medication can impact the probability of MI, an MI event may elicit a change in blood pressure medication. A covariate is said to be a Type IV TDC if

$$\mathrm{E}\left[\frac{\partial \mu_{is}}{\partial \beta_j}(Y_{it} - \mu_{it})\right] = 0 \qquad \forall s \leq t. \tag{7}$$

The expectation must be satisfied for $s \leq t$, but not necessarily when $s > t$, suggesting dependence between the response and covariate. For a Type IV TDC, a covariate can be associated with previous response values, but the response is not associated with previous covariate values. A sufficient condition for a covariate to be Type IV is

$$\mathrm{E}\left[Y_{it}|x_{i1}, \ldots, x_{iT}\right] = \mathrm{E}\left[Y_{it}|x_{it}, \ldots, x_{iT}\right].$$

Type IV TDC are associated with prior response values, but the response at time $t$ is only associated with the TDC at time $t$. One example is the covariate "blood pressure" as a Type IV covariate with the response "weight." While there is an association between weight and blood pressure, the direction of the effect seems to be that weight impacts blood pressure, but the reverse is unlikely.

Different types of TDCs are associated with different relationships with the response. It is important to be able to identify different types of TDCs to guide model selection and to provide appropriate interpretations. Lai and Small (2007) proposed selecting the type of TDC by choice of the researcher, but also presented a $\chi^2$ test to compare two different selections of types for TDC. The idea is to construct competing quadratic forms using the expressions from Eqs. (3), (5), (6), and (7) with zero expectation, so that additional expressions from a different selection of a type of TDC can inflate the quadratic form if those additional expressions do not, in fact, have zero expectation. However, this method will only allow for comparisons between possible selections of types of TDC, but will not make the selection for the researcher. The Extended Classification method, described in Sect. 4.3, presents such a process.

## 3 Subject-Specific Modeling

Longitudinal data models can be thought of as belonging to two classes of estimation: conditional models and marginal models. Conditional models, the focus of this section, are often referred to as mixed models, random effect models,

hierarchical models, or mixture models. Conditional models involve specification of a response model, conditional on a random subject effect. This random effect is intended to account for the clustering of responses by subject, and induces "subject-specific" or "cluster-specific" conclusions from the model. Because parameters must condition on the random effect, parameters are interpreted as expected changes for a specific (average) subject and not a comparison between populations. For a discussion of subject-specific and population-averaged models, see Zeger et al. (1988), Neuhaus et al. (1991), and Zeger and Liang (1992).

## 3.1 Conditional Model Decomposition

Conditional correlated generalized linear models have been covered extensively in the literature (Lee and Nelder 1996; Diggle et al. 2002; Hedeker and Gibbons 2006; Lee et al. 2006; McCulloch et al. 2008; Fitzmaurice et al. 2012). A conditional correlated generalized linear model with random intercept can be written,

Random Component:

$$Y_{it}|u_i \sim \mathscr{D}(\mu(\mathbf{x}_{it}, \mathbf{z}_{it})),$$
$$u_i \sim \mathscr{D}_u(\boldsymbol{\alpha}),$$

Systematic and Link Components:

$$g(\mu(\mathbf{x}_{it}, \mathbf{z}_{it})) = \mathbf{x}_{it}^T\boldsymbol{\beta} + \mathbf{z}_{it}^T\mathbf{v}(\mathbf{u}).$$

In the expression of the random component, $\mathscr{D}$ represents a specific conditional response distribution from the exponential family, and $u_i$ indicates the random subject effect distributed according to $\mathscr{D}_u$ with parameters $\boldsymbol{\alpha}$. In defining conditional models these two distributions are typically completely specified. In the expression of the systematic component, $\mathbf{z}_{it}$ represents a component of the random effects design matrix, and $v$ is a function transforming the random effect to a range on the continuum (Lee and Nelder 1996, 2001). This model is referred to as a "random intercept" model because the random effects are additively included in the systematic component and can be thought of as "errors" associated with the intercept $\beta_0$. In a "random-slopes" model, products of random effects with the fixed-effects design matrix components $x_{it}$ can be viewed as "errors" for the fixed-effects parameters $\beta_k$, and thus allow the slopes to vary randomly (Lalonde et al. 2013). Random-slopes models are often presented as hierarchical models in which each parameter $\beta_k$ has an associated linear model with an individual error term and can include predictors.

Here the interpretation of conditional model fixed effects can be made clear. The parameter $\beta_k$ represents the expected change in the (transformed) mean response for a unit increase in $x_{k,it}$ for an individual subject, holding all other predictors fixed. In other words, if predictor $\mathbf{x}_k$ changes for an individual subject, $\beta_k$ represents the expected impact on the mean response.

In the presence of TDC, the standard conditional models are often adjusted to allow for both "within" and "between" components of effects associated with TDC (Neuhaus and Kalbfleisch 1998). If a covariate includes both variation within subjects and variation between subjects, it is believed these two distinct sources of variation can be associated with different effects. The term in the model representing each TDC can be decomposed into two terms: one associated with variation within subjects and the other associated with variation between subjects,

$$\beta x_{it} \quad \rightarrow \quad \beta_W(x_{it} - \bar{x}_{i.}) + \beta_B \bar{x}_{i.}.$$

In this expression the parameter $\beta_W$ represents the expected change in the mean response associated with changes of the TDC within subjects, while the parameter $\beta_B$ is more of a population-averaged parameter that represents the expected change in the mean response associated with changes of the TDC across subjects.

## 3.2 An Issue with Estimation

Estimation of parameters in conditional models typically proceeds by using likelihood-based methods (McCullagh and Nelder 1989; Lee et al. 2006). Standard maximum likelihood estimating equations are of the form,

$$\mathbf{S}(\boldsymbol{\beta}) = \sum_{i=1}^{N} \left( \frac{\partial \mu(\boldsymbol{\beta}; \mathbf{x}_i)}{\partial \boldsymbol{\beta}} \right)^T \mathbf{W}_i(\mathbf{Y}_i - \mu(\boldsymbol{\beta}; \mathbf{x}_i)) = \mathbf{0},$$

where the weight matrix $\mathbf{W}_i$ is often taken to be the inverse of the variance–covariance of the marginal response (Diggle et al. 2002). Pepe and Anderson (1994) showed that these estimating equations have zero expectation only if the data meet the assumption,

$$\mathrm{E}[Y_{it}|X_{it}] = \mathrm{E}[Y_{it}|X_{ij}, j = 1, \ldots, T], \tag{8}$$

for each TDC. The assumption of Eq. (8) is met trivially for TIC. Notice that exogeneity is not a sufficient condition, as Eq. (2) implies that the response at one time will be independent of an exogenous covariate's values at future times. Equation (8), on the other hand, suggests the response at one time should be independent of covariate values at all other times. When this assumption is satisfied,

$$
\begin{aligned}
&\mathrm{E}\left[\mathbf{S}(\boldsymbol{\beta})\right] = \mathrm{E}\left[\mathrm{E}\left[\mathbf{S}(\boldsymbol{\beta})|x_{it}, t = 1, \ldots, T\right]\right] \\
&= \mathrm{E}\left[\mathrm{E}\left[\sum_{i=1}^{N}\left(\frac{\partial \mu(\boldsymbol{\beta}; \mathbf{x}_i)}{\partial \boldsymbol{\beta}}\right)^T \left[\phi \mathbf{V}_i(\lambda(\boldsymbol{\beta}; \mathbf{x}_i))\right]^{-1} \left(\mathbf{Y}_i - \mu(\boldsymbol{\beta}; \mathbf{x}_i)\right)|x_{it}, t = 1, \ldots, T\right]\right] \\
&= \mathrm{E}\left[\sum_{i=1}^{N}\left(\frac{\partial \mu(\boldsymbol{\beta}; \mathbf{x}_i)}{\partial \boldsymbol{\beta}}\right)^T \left[\phi \mathbf{V}_i(\lambda(\boldsymbol{\beta}; \mathbf{x}_i))\right]^{-1} \mathrm{E}\left[\left(\mathbf{Y}_i - \mu(\boldsymbol{\beta}; \mathbf{x}_i)\right)|x_{it}, t = 1, \ldots, T\right]\right] \\
&= \mathrm{E}\left[\sum_{i=1}^{N}\left(\frac{\partial \mu(\boldsymbol{\beta}; \mathbf{x}_i)}{\partial \boldsymbol{\beta}}\right)^T \left[\phi \mathbf{V}_i(\lambda(\boldsymbol{\beta}; \mathbf{x}_i))\right]^{-1} \times \mathbf{0}\right] \\
&= \mathbf{0}.
\end{aligned}
$$

In this derivation, the step of removing the derivative term from the inner expectation depends on the assumption of Eq. (8). Depending on the choice of the weight matrix $\mathbf{W}_i$, the estimating equations may require combinations of the first term (the derivative of the systematic component) with the second term (the raw error term) across different times. Specifically, this will be the case for any non-diagonal weight matrix. The assumption presented by Pepe and Anderson (1994) requires that the derivative and raw residual terms are independent at any two time points combined by the weight matrix.

The standard conditional generalized linear models induce a block-diagonal variance–covariance structure for the marginal response, and thus the condition of Eq. (8) must be satisfied if the standard weight matrix is applied. Notice Eq. (8) is a sufficient condition for a covariate to be a Type I TDC. For other types of TDC, the condition is likely not satisfied. If the condition is not satisfied, the likelihood-based estimating equations will not have zero expectation, leading to bias and loss of efficiency in parameter estimates (Pepe and Anderson 1994; Diggle et al. 2002).

## 4 Population-Averaged Modeling

Unlike the conditional models of Sect. 3, marginal models for longitudinal data do not involve specification of a conditional response distribution using random effects. Instead a marginal model involves specification of the marginal response distribution, or at least moments of the response distribution (McCullagh and Nelder 1989; Hardin and Hilbe 2003; Diggle et al. 2002). This type of model is associated with "population-averaged" interpretations, or standard regression interpretations. Parameters in marginal longitudinal models provide a comparison of the mean response between two populations with different average values of the predictor of interest. While marginal conclusions can be obtained through conditional models, the term "marginal model" will be used to refer to a model specifically intended for marginal expression and interpretations (Lee and Nelder 2004). A marginal correlated generalized linear model can be written,

Random Component:

$$Y_{it} \sim \mathscr{D}\left(\mu(\mathbf{x}_{it}), \phi V(\mu(\mathbf{x}_{it}))\right),$$

Marginal Mean:

$$\ln(\mu(\mathbf{x}_{it})) = \mathbf{x}_{it}^T \boldsymbol{\beta}.$$

For this type of model $\mathscr{D}$ is assumed to be a distribution from the exponential family of distributions, but may not be fully specified within a marginal model. Instead, the mean $\mu(\mathbf{x}_{it})$ and variance $V(\mu(\mathbf{x}_{it}))$ (with possible over dispersion parameter $\phi$) are supplied by the researcher. While there are many marginal methods for estimating parameters in a longitudinal generalized linear model, this chapter will focus on two methods: the generalized estimating equations (GEE) and the generalized method of moments (GMM).

## 4.1 Generalized Estimating Equations

The GEE approach to model fitting has been covered extensively in the literature (Liang and Zeger 1986; Zeger and Liang 1986; Liang et al. 1992; Ziegler 1995; Hardin and Hilbe 2003; Diggle et al. 2002). Briefly, parameter estimates are obtained by solving the equations,

$$\mathbf{S}(\boldsymbol{\beta}) = \sum_{i=1}^{N} \left( \frac{\partial \mu(\boldsymbol{\beta}; \mathbf{x}_i)}{\partial \boldsymbol{\beta}} \right)^T [\phi \mathbf{V}_i(\lambda(\boldsymbol{\beta}; \mathbf{x}_i))]^{-1} (\mathbf{Y}_i - \mu(\boldsymbol{\beta}; \mathbf{x}_i)) = \mathbf{0},$$

where the variance–covariance structure is specified through a working correlation structure $\mathbf{R}_i(\boldsymbol{\alpha})$,

$$\mathbf{V}_i(\lambda(\mathbf{x}_{it})) = \mathbf{A}_i^{1/2} \mathbf{R}_i(\boldsymbol{\alpha}) \mathbf{A}_i^{1/2}.$$

Pepe and Anderson (1994) argued that the structure of the GEE requires satisfaction of the assumption,

$$\mathrm{E}[Y_{it}|X_{it}] = \mathrm{E}[Y_{it}|X_{ij}, j = 1, \ldots, T],$$

so that the GEE will have zero expectation. As with conditional model estimation, this assumption is met trivially for TIC. When the assumption is met, the first term of the GEE can be factored out of the expectation of $\mathbf{S}(\boldsymbol{\beta})$, producing unbiased estimating equations. When the assumption is not met, the GEE will not have zero expectation unless the working correlation structure is selected so that all components of the GEE involve only a single observation time. This is achieved by a

diagonal working correlation structure, so Pepe and Anderson (1994) recommended use of the independent working correlation structure in the presence of TDC. However, Fitzmaurice (1995) noted that using the independent working correlation structure when it is not appropriate can lead to substantial losses of efficiency.

Together these results have been taken as instructions to use the independent working correlation structure when applying GEE to longitudinal data with TDC. However, the results of Fitzmaurice (1995) suggest there may be meaningful losses in efficiency depending on the strength of the auto-correlation. Additionally, the approach of applying independent GEE makes no distinction among different types of TDC, or even between exogenous and endogenous covariates. An approach using the GMM addresses these issues.

## *4.2 Generalized Method of Moments*

The GMM is a minimum-quadratic method of estimating parameters (Hansen 1982; Hansen et al. 1996; Hansen 2007). Model parameters $\boldsymbol{\beta}$ can be estimated by minimizing a quadratic form $Q(\boldsymbol{\beta})$ with appropriately chosen components,

$$Q(\boldsymbol{\beta}) = \mathbf{G}^T(\boldsymbol{\beta}; \mathbf{Y}, \mathbf{X})\mathbf{W}^{-1}\mathbf{G}(\boldsymbol{\beta}; \mathbf{Y}, \mathbf{X}),$$

where $\mathbf{G}(\boldsymbol{\beta}; \mathbf{Y}, \mathbf{X})$ is a vector of "moment conditions" with zero expectation and $\mathbf{W}$ is a correspondingly chosen weight matrix. For longitudinal data situations, $\mathbf{G}$ is typically constructed as an average of vectors of "valid moment conditions" for each subject,

$$\mathbf{G}(\boldsymbol{\beta}; \mathbf{Y}, \mathbf{X}) = \frac{1}{N}\sum_{i-1}^{N} \mathbf{g}_i(\boldsymbol{\beta}; \mathbf{Y}, \mathbf{X}).$$

When presenting the GMM, Hansen (1982) argued that the optimal choice for the weight matrix is the inverse of the variance–covariance structure of the subject-level vector of valid moment conditions,

$$\mathbf{W} = \mathrm{Cov}\left(\mathbf{g}_i(\boldsymbol{\beta}; \mathbf{Y}, \mathbf{X})\right).$$

The challenge in applying the GMM is to determine appropriate components of the subject-level vectors of valid moments conditions $\mathbf{g}_i$. In some data applications, the valid moment conditions can involve transformations of the raw residuals using appropriately chosen instrumental variables (Wooldridge 2008). In the situation of longitudinal data with TDC, Lai and Small (2007) proposed defining each element of $\mathbf{g}_i(\boldsymbol{\beta}; \mathbf{Y}, \mathbf{X})$ according to the expected nature of each TDC. Specifically, the expectations associated with Type I, Type II, Type III, and Type IV TDC, Eqs. (3), (5), (6), and (7), respectively, define combinations of times at which

components of potential moment conditions will be independent. When components are independent and the expectation of Eqs. (3), (5), (6), and (7) is zero, the argument of the expectation can be treated as one component of the vector of valid conditions,

$$g_{ik}(\boldsymbol{\beta}; \mathbf{Y}, \mathbf{X}) = \left(\frac{\partial \mu_{is}}{\partial \beta_j}\right)(y_{it} - \mu_{it}).$$

The type of TDC will determine which combinations of times form valid moment conditions. For all predictors in the model, the concatenation of all valid moment conditions will form the vector $\mathbf{g}_i$ for each subject. Notice that this method avoids choosing a general weight matrix to apply across all covariates, as with likelihood-based estimation or with the GEE. Instead, the GMM allows expressions to be constructed separately for each TDC, which provides the ability to treat each covariate according to its type. This eliminates a major restriction from both likelihood-based methods and the GEE.

### 4.3 GMM with Extended Classification

As an alternative to constructing subject vectors of valid moment conditions using researcher-determined types, the Extended Classification process can be used (Lalonde et al. 2014). Through this process, for each TDC the data will be used to determine the specific combinations of times that will construct valid moment conditions for all subjects.

First initial parameter estimates $\hat{\boldsymbol{\beta}}_0$ are obtained using GEE with the independent working correlation structure. Values of both the derivative component and raw residual component of potential moment conditions can be calculated for TDC $\mathbf{x}_j$,

$$\hat{d}_{sj} = \frac{\partial \hat{\boldsymbol{\mu}}_s}{\partial \beta_j},$$

$$\hat{r}_t = \mathbf{y}_t - \hat{\boldsymbol{\mu}}_t,$$

where $\hat{\boldsymbol{\mu}}_t$ represents a vector of predicted mean responses at time $t$ across all subjects, evaluated using $\hat{\boldsymbol{\beta}}_0$. After standardizing both vectors to obtain $\tilde{d}_{sji}$ and $\tilde{r}_{ti}$, the association between these components is then evaluated using Pearson correlation,

$$\hat{\rho}_{sjt} = \frac{\sum(\tilde{d}_{sji} - \bar{\tilde{d}}_{sj})(\tilde{r}_{ti} - \bar{\tilde{r}}_t)}{\sqrt{\sum(\tilde{d}_{sji} - \bar{\tilde{d}}_{sj})^2 \sum(\tilde{r}_{ti} - \bar{\tilde{r}}_t)^2}},$$

and standardized for comparison. Assuming all fourth moments of $\hat{\rho}_{sjt}$ exist and are finite,

$$\rho^*_{sjt} = \frac{\hat{\rho}_{sjt}}{\sqrt{\hat{\mu}_{22}/N}} \sim \mathcal{N}(0,1),$$

where $\hat{\mu}_{22} = (1/N)\sum_i (\tilde{d}_{sji})^2(\tilde{r}_{ti})^2$, and $N$ is the total number of subjects. Significantly correlated components show evidence of non-independence between the derivative and raw residual terms, and therefore the associated product of derivative and raw error should not reasonably have zero expectation and can be omitted as a potential valid moment condition. To account for the large number of hypothesis tests involved in the Extended Classification process, p-values for all correlation tests can be collectively evaluated (Conneely and Boehnke 2007).

The method of Extended Classification removes the potentially subjective decision of the type of each TDC by the researcher and allows the data to determine appropriate valid moment conditions. Extended Classification also allows for more than four discrete types, admitting all possible combinations of times instead of the four cases corresponding to the four types. The Extended Classification process has shown to be effective in determining appropriate types of TDC, with results similar or superior to those of subjectively chosen types (Lalonde et al. 2014).

## 4.4 Minimization For GMM

To complete GMM estimation it is necessary to minimize the constructed quadratic form $Q(\boldsymbol{\beta})$. Minimization of the quadratic form has been described using three methods: Two-step GMM (TSGMM), iterated GMM (IGMM), and continuously updating GMM (CUGMM) (Hansen et al. 1996).

TSGMM includes separate steps to address the weight matrix and moment conditions. Using initial values $\hat{\boldsymbol{\beta}}_{(0)}$, an estimate of the weight matrix $\hat{\mathbf{W}}_{(0)}$ is obtained and substituted into the quadratic form,

$$Q_{TS}(\boldsymbol{\beta}) = \mathbf{G}^T(\boldsymbol{\beta};\mathbf{Y},\mathbf{X})\hat{\mathbf{W}}^{-1}_{(0)}\mathbf{G}(\boldsymbol{\beta};\mathbf{Y},\mathbf{X}).$$

The quadratic form is then minimized to obtain final parameter estimates $\hat{\boldsymbol{\beta}}$. The TSGMM process appears to be the most commonly applied method in the literature. The IGMM process involves an iterative repeat of the steps in TSGMM. After the quadratic form $Q_{TS}$ has been minimized to obtain updated parameter estimates $\hat{\boldsymbol{\beta}}_{(1)}$, the estimate of the weight matrix is updated, providing $\hat{\mathbf{W}}_{(1)}$. The process then iterates between updating $\hat{\boldsymbol{\beta}}_{(i)}$ using the quadratic form and updating $\hat{\mathbf{W}}_{(i)}$ using the resulting estimates,

$$\hat{\boldsymbol{\beta}}_{(i+1)} = argmin\left[\mathbf{G}^T(\boldsymbol{\beta};\mathbf{Y},\mathbf{X})\hat{\mathbf{W}}^{-1}_{(i)}\mathbf{G}(\boldsymbol{\beta};\mathbf{Y},\mathbf{X})\right].$$

**Table 1** Cross-classification of rehospitalization by time

| | | Time | | | |
|---|---|---|---|---|---|
| | | 1 | 2 | 3 | Total |
| Re-admit | No | 231 | 272 | 253 | 756 |
| | | 46.48 % | 54.73 % | 50.91 % | |
| | Yes | 266 | 225 | 244 | 735 |
| | | 53.52 % | 45.27 % | 49.09 % | |

The process completes on sufficient convergence of $\hat{\boldsymbol{\beta}}_{(i)}$. The IGMM process appears to be the least commonly used method in the literature, and is associated with convergence problems (Hansen et al. 1996; Hansen 2007). CUGMM proceeds by treating the weight matrix as a function of the model parameters,

$$Q_{CU}(\boldsymbol{\beta}) = \mathbf{G}^T(\boldsymbol{\beta}; \mathbf{Y}, \mathbf{X}) \left(\mathbf{W}(\boldsymbol{\beta})\right)^{-1} \mathbf{G}(\boldsymbol{\beta}; \mathbf{Y}, \mathbf{X}).$$

Estimates are obtained by a single minimization of $Q_{CU}$.

## 5 Data Example

In order to exemplify the implementation and interpretation associated with the models discussed in Sects. 3 and 4, an analysis is presented using the Arizona SID (Lalonde et al. 2014).The dataset contains patient information from Arizona hospital discharges for the 3-year period from 2003 through 2005, for individuals admitted to a hospital exactly four times. The dataset includes 1,625 patients with three observations; each observation corresponds to a rehospitalization. It is of interest to model the probability of returning to a hospital within 30 days using the predictors: total number of diagnoses ("Diagnoses"), total number of procedures performed ("Procedures"), length of patient hospitalization ("LOS"), the existence of coronary atherosclerosis ("C.A."), and indicators for time 2 and time 3. Table 1 provides the percentage of the patients who were readmitted to the hospital within 30 days of discharge against the percentages of the patients who were not readmitted for each of their first three hospitalizations.

All four predictors as well as the two time indicators will be TDC. Results of modeling the probability of rehospitalization within 30 days will be presented using the five models: random-intercept logistic regression with decomposition of TDC (RS); random-slope logistic regression with decomposition of TDC (RS); GEE logistic regression with independent working correlation structure (IGEE); TSGMM logistic regression with the type of each TDC selected by the researcher (GMM-Types); and TSGMM logistic regression using extended classification (GMM-EC). The GMM models will be fit using the TSGMM.

The RI logistic regression model can be written with a decomposition of all TDC, except for the time indicators,

$$\text{logit}(\pi_{it}) = \beta_0 + \sum_{k=1}^{4} (\beta_{kW}(x_{k,it} - \bar{x}_{k,i.}) + \beta_{kB}\bar{x}_{k,i.}) + \beta_{t2}\text{Time2} + \beta_{t3}\text{Time3} + \gamma_{0i},$$

where $\pi_{it}$ indicates the probability of rehospitalization within 30 days for subject $i$ at time $t$, and $\gamma_{0i}$ indicates the random subject effect. The model can be fit using SAS or R with the commands provided in Sect. 7.1.

The RS logistic regression model can be written similarly, including a random slope for the length of stay predictor. This will allow the effect of length of stay on probability of rehospitalization to vary randomly among subjects,

$$\begin{aligned}\text{logit}(\pi_{it}) = \beta_0 + \sum_{k=1}^{4} (\beta_{kW}(x_{k,it} - \bar{x}_{k,i.}) + \beta_{kB}\bar{x}_{k,i.}) \\ + \beta_{t2}\text{Time2} + \beta_{t3}\text{Time3} + \gamma_{0i} + \gamma_{1i}\text{LOS}_{it},\end{aligned}$$

where $\gamma_{1i}$ represents the random variation in the slope for length of stay. The model can be fit using SAS or R with the commands provided in Sect. 7.2.

The IGEE logistic model will be written without the decomposition of TDC, and without random subject effects,

$$\text{logit}(\pi_{it}) = \beta_0 + \sum_{k=1}^{4} \beta_{kW}x_{k,it} + \beta_{t2}\text{Time2} + \beta_{t3}\text{Time3}.$$

This GEE model can be fit with the independent working correlation structure using SAS or R with the commands provided in Sect. 7.3.

The systematic and link components for the GMM-Types model will look identical to that of the IGEE model. For the GMM-Types model, specific types will be assumed for each TDC. Both time indicators will be treated as Type I TDC, as is common for such deterministic variables. Both "length of stay" and "existence of coronary atherosclerosis" will be treated as Type II TDC, as it is reasonable to assume an accumulated effect on the response from these two variables, but it is unlikely that the response at one time will affect future values of these covariates. Both "number of diagnoses" and "number of procedures" will be treated as Type III TDC, as it is reasonable to assume feedback between the probability of rehospitalization within 30 days and these two counts.

For the GMM-EC model there will be no assumptions of specific types of TDC. Instead the extended classification process will be used to select appropriate valid moment conditions to be used in the GMM quadratic form. These GMM methods are not yet available in SAS; R functions written by the author can be requested.

Results of fitting all five models are presented in Table 2. First consider the results of the conditional models. For the model including a random intercept, the variation associated with that intercept (0.1472) is significant, suggesting there is significant individual variation in the baseline probability of rehospitalization within 30 days. For all models the time indicators have significant negative coefficients, which implies the chance of rehospitalization within 30 days is significantly lower for later follow-up visits. This is suggestive of either a patient fatigue effect in which an individual tires of visiting the hospital, or the positive impact of multiple visits on curing an illness.

The decomposed TDC in this model provide interesting interpretations. The "between" components of the TDC provide population-averaged types of conclusions. For example, there is evidence that subjects with higher average length of stay tend to have a higher probability of rehospitalization (0.0736), perhaps an indication of more serious illnesses. The "within" components provide interpretations of individual effects over time. For example, there is evidence that an increase in the number of diagnoses for an individual is associated with a higher probability of rehospitalization (0.0780), perhaps an indication of identifying additional illnesses.

Results for the model including a random-slope for length of stay are similar. Within the RS model, the variation in the length of stay slope (0.0025) is significant, indicating meaningful individual variation in the effect of length of stay on the probability of rehospitalization. The variation in the intercept (0.1512) remains significant. Two changes are evident when compared to the random-intercept model. First, the random-slope model shows a significant positive association with length of stay *within* subjects, suggesting an increase in length of stay over time is associated with a higher probability of rehospitalization within 30 days. Second, the RS model shows a significant positive association with existence of coronary atherosclerosis *between* subjects, suggesting an increase in the probability of rehospitalization within 30 days for subjects who eventually develop coronary atherosclerosis.

Next consider the results of the marginal models. For all three of the models IGEE, GMM-Types, and GMM-EC, the parameter associated with length of stay is positive and significant. This indicates that, when comparing two populations with different average lengths of stay, the population with the higher length of stay has a higher probability of rehospitalization within 30 days. Notice that while all three marginal models show a negative effect for the number of procedures, significance is identified with GMM but not with GEE. This is to be expected, as GMM is intended to improve the efficiency over the conservative IGEE process. Also notice that the signs of significant "between" effects for the conditional models are similar to those of the corresponding effects in the marginal models. This is also to be expected, as "between" effects produce conclusions similar to the population-averaged marginal model conclusions.

Overall fit statistics are provided but may not provide meaningful information for selection between conditional and marginal models. Selecting the most appropriate model is often based on researcher intentions. The IGEE model is a safe choice,

**Table 2** Conditional and marginal logistic regression models

| Parameter estimates and significance | | | | | | | |
|---|---|---|---|---|---|---|---|
| | RI—within | RI—between | RS—within | RS—between | IGEE | GMM-types | GMM-EC |
| Diagnoses | 0.0780*** | 0.0444 | 0.0686** | 0.0362 | 0.0648*** | 0.0613*** | 0.0573*** |
| Procedures | 0.0188 | −0.0824** | 0.0092 | −0.0915** | −0.0306 | −0.0458* | −0.0366• |
| LOS | 0.0008 | 0.0736*** | 0.0200* | 0.0952*** | 0.0344*** | 0.0530*** | 0.0497*** |
| C.A. | −0.2607* | 0.2223 | −0.2646* | 0.3050* | −0.1143 | −0.0536 | −0.0802 |
| Time 2 | −0.3730*** | | −0.4061*** | | −0.3876*** | −0.4004*** | −0.3933*** |
| Time 3 | −0.2130** | | −0.2357** | | −0.2412*** | −0.2417*** | −0.2633*** |
| Intercept | 0.1472** | | 0.1512• | | | | |
| Slope | – | | 0.0025* | | | | |
| Gen $\chi^2/DF$ | 0.98 | | 0.96 | | | | |
| QIC | | | | | 6648.52 | | |
| QICu | | | | | 6646.56 | | |

0 ***0.001 **0.01 *0.05 •0.10

but generally lacks the power of the GMM models. The conditional models are an appropriate choice when subject-specific designs and conclusions are of interest, but also impose the assumption of a block-diagonal marginal variance–covariance structure.

The most powerful and appropriate choice appears to be the GMM method that avoids the necessary condition of Eq. (8) presented by Pepe and Anderson (1994), and allows for TDC to be treated differently from each other. In this sense the Extended Classification method provides the most flexibility, as moment conditions are selected individually based on empirical evidence from the dataset. In this data example the results of both the GMM-Types and GMM-EC models are quite similar, yielding the same signs of parameter estimates and similar significance levels, which suggests the researcher-selected types of covariates are probably appropriate according to the dataset.

## 6 Discussion

TDC occur commonly in practice, as data collected for longitudinal studies often change over time. There are numerous ways to classify TDC. The most common type of classification is as exogenous versus endogenous covariates. Exogenous covariates vary according to factors external to the system under consideration, while endogenous covariates show association with other recorded variables. It is most important to identify exogeneity with respect to the response variable.

TDC more recently have been classified according to four "types" that reflect the nature of the association between the TDC and the response. While these definitions are related to exogeneity, they do not represent the same characteristics. Instead, the different types of TDC reflect different levels of association between covariates and responses at different times, with the most substantial relationship a "feedback" loop between covariates and response at different times.

Existing methods for modeling longitudinal data with TDC can be split into two classes: conditional models and marginal models. Conditional models incorporate random effects into the systematic component of the model to account for the autocorrelation in responses. To accommodate TDC, individual regression terms can be decomposed into contributions from variation "within" subjects and variation "between" subjects. When maximum-likelihood-type methods are applied to estimate parameters in conditional models, there is an implicit assumption of independence between the response at one time and covariate values at other times. If this assumption is not met, the likelihood estimating equations will not have zero expectation because of off-diagonal components of the response variance–covariance structure, which can bias parameter estimates.

Marginal models, on the other hand, define a marginal response (quasi-) distribution through specification of a marginal mean and a marginal variance–covariance structure. The most commonly used such method is the GEE. To accommodate TDC, it has been recommended that the independent working correlation structure

is applied when using GEE. This recommendation is made to avoid satisfying a necessary condition for the GEE to have zero expected value, as individual estimating equations that combine components at different times may not have zero expectation due to dependence between responses and covariates at different times. However, the use of independent GEE can lead to meaningful losses in efficiency if the autocorrelation is substantial.

An alternative to both conditional models and GEE estimation is the use of the GMM. The GMM can be used to treat each TDC differently, depending on the type of covariate, and to avoid issues with estimating equations constructed from non-independent components. The GMM can be applied by allowing the researcher to identify the type of each TDC, or the Extended Classification can be used to allow the data to determine the nature of the relationship between each TDC and the response. In the future, the GMM with Extended Classification should be improved and utilized as a standard method for analysis of longitudinal data with TDC.

## 7 Example SAS and R Commands

### *7.1 Random-Intercept Models*

The random-intercept (RI) model discussed in Section LABEL can be fit using the following SAS commands.

```
/* PROC GLIMMIX DOES NOT REQUIRE INITIAL VALUES */
PROC GLIMMIX DATA=ASID_DATA;
    CLASS subject_id;
    MODEL readmission(event = '1') = diagnoses_w diagnoses_b
                                procedures_w procedures_b
                                LOS_w LOS_b
                                CA_w CA_b
                                time2 time3
                                / DIST=BINARY LINK=LOGIT
                                     DDFM=BW SOLUTION;
    RANDOM INTERCEPT/ subject=subject_id;
RUN;

/* PROC NLMIXED REQUIRES INITIAL VALUES:USE INDEPENDENT GEE */
PROC NLMIXED DATA=ASID_DATA QPOINTS=30;
    PARMS beta0= beta1= beta2= beta3= beta4= beta5=
                  beta6= beta7= beta8= beta9= beta10=;
    eta = u + beta0 + beta1*diagnoses_w
                    + beta2*diagnoses_b
                    + beta3*procedures_w
                    + beta4*procedures_b
                    + beta5*LOS_w + beta6*LOS_b
                    + beta7*CA_w + beta8*CA_b
```

```
                    + beta9*time2 + beta10*time3;
    exp_eta = exp(eta);
    pi = ((exp_eta)/(1+exp_eta));
    MODEL readmission ~ BINARY(pi);
    RANDOM u ~ NORMAL(0, sigmau*sigmau)SUBJECT=subject_id;
RUN;
```

Alternatively, the model can be fit using R with the following commands.

```
install.packages("lme4")
library(lme4)
# USE start=c(diagnoses_w=, ... ) OPTION TO SPECIFY
    INITIAL VALUES #
# USE INDEPENDENT GEE FOR INITIAL VALUES #
R_Int = glmer(readmission ~ diagnoses_w+diagnoses_b
             +procedures_w+procedures_b+LOS_w+LOS_b
                  +CA_w+CA_b
             +time2+time3 + (1|subject_id),family=binomial,
             REML=FALSE,data=ASID_DATA)
summary(R_Int)
```

## *7.2 Random-Slope Models*

The random-slope (RS) model discussed in Section LABEL can be fit using the following SAS commands.

```
/* PROC GLIMMIX DOES NOT REQUIRE INITIAL VALUES */
PROC GLIMMIX DATA=ASID_DATA;
    CLASS subject_id;
    MODEL readmission(event = '1') = diagnoses_w diagnoses_b
                                     procedures_w procedures_b
                                     LOS_w LOS_b
                                     CA_w CA_b
                                     time2 time3
                                     / DIST=BINARY LINK=LOGIT
                                          DDFM=BW SOLUTION;
    RANDOM INTERCEPT LOS / subject=subject_id;
run;

/* PROC NLMIXED REQUIRES INITIAL VALUES:
    USE INDEPENDENT GEE */
PROC NLMIXED DATA=ASID_DATA QPOINTS=30;
    PARMS beta0= beta1= beta2= beta3= beta4= beta5=
                  beta6= beta7= beta8= beta9= beta10=;
    eta = u + beta0 + beta1*diagnoses_w + beta2*diagnoses_b
                    + beta3*procedures_w + beta4*procedures_b
                    + beta5*LOS_w + beta6*LOS_b
```

```
                    + beta7*CA_w + beta8*CA_b
                    + beta9*time2 + beta10*time3
                    + rb1*LOS;
    exp_eta = exp(eta);
    pi = ((exp_eta)/(1+exp_eta));
    MODEL readmission ~ BINARY(pi);
    RANDOM u rb1 ~ NORMAL([0, 0], [s2u, 0, s2f])
    SUBJECT=subject_id;
RUN;
```

Alternatively, the model can be fit using R with the following commands.

```
install.packages("lme4")
library(lme4)
# USE start=c(diagnoses_w=, ... ) OPTION TO SPECIFY
    INITIAL VALUES #
# USE INDEPENDENT GEE FOR INITIAL VALUES #
R_Slopes = glmer(readmission ~ diagnoses_w+diagnoses_b
             +procedures_w+procedures_b+LOS_w+LOS_b+CA_w+CA_b
             +time2+time3 + (1|subject_id)+(0+LOS|subject_id),
             family=binomial, REML=FALSE,
             start=c(diagnoses_w=, . . .),data=ASID_DATA)
summary(R_Int)
```

## *7.3 Independent GEE*

```
PROC GENMOD DATA=ASID_DATA;
    CLASS subject_id;
    MODEL readmission = diagnoses procedures LOS CA
                        time2 time3
                          / DIST=BINOMIAL LINK=LOGIT;
    REPEATED SUBJECT = id / TYPE=IND;
RUN;
```

Alternatively, the model can be fit using R with the following commands.

```
install.packages("geepack")
library(geepack)
Ind_GEE = geeglm(readmission ~ diagnoses+procedures+LOS+CA
                       +time2+time3, family=binomial,
                       id=subject_id,corstr="independence",
                       data=ASID_DATA)
summary(Ind_GEE)
```

## References

Amemiya, T.: Advanced Econometrics. Harvard University Press, Cambridge (1985)
Chamberlain, G.: The general equivalence of granger and sims causality. Econometrica **50**, 569–582 (1982)
Conneely, K.N., Boehnke, M.: So many correlated tests, so little time! rapid adjustment of p-values for multiple correlated tests. Am. J. Hum. Genet. **81**, 1158–1168 (2007)
Diggle, P.J., Heagerty, P., Liang, K.Y., Zeger, S.L.: Analysis of Longitudinal Data, 2nd edn. Oxford University Press, Oxford (2002)
Fitzmaurice, G.M.: A caveat concerning independence estimating equations with multiple multivariate binary data. Biometrics **51**, 309–317 (1995)
Fitzmaurice, G.M., Laird, N.M.: Regression models for a bivariate discrete and continuous outcome with clustering. J. Am. Stat. Assoc. **90**(431), 845–852 (1995)
Fitzmaurice, G.M., Laird, N.M., Ware, J.H.: Applied Longitudinal Analysis, 2nd edn. Wiley, Hoboken (2012)
Hansen, L.P.: Large sample properties of generalized method of moments estimators. Econometrica **50**(4), 1029–1054 (1982)
Hansen, L.P.: Generalized Method of Moments Estimation, pp. 1–14. Department of Economics, University of Chicago, Chicago (2007)
Hansen, L.P., Heaton, J., Yaron, A.: Finite-sample properties of some alternative gmm estimators. J. Bus. Econ. Stat. **14**(3), 262–280 (1996)
Hardin, J.W., Hilbe, J.M.: Generalized Estimating Equations. Chapman & Hall, New York (2003)
Hedeker, D., Gibbons, R.D.: Longitudinal Data Analysis. Wiley-Interscience, Hoboken (2006)
Lai, T.L., Small, D.: Marginal regression analysis of longitudinal data with time-dependent covariates: a generalized method-of-moments approach. J. Roy. Stat. Soc. Ser. B **69**(1), 79–99 (2007)
Lalonde, T.L., Nguyen, A.Q., Yin, J., Irimate, K., Wilson, J.R.: Modeling correlated binary outcomes with time-dependent covariates. J. Data Sci. **11**, 715–738 (2013)
Lalonde, T.L., Wilson, J.R., Yin, J.: Gmm logistic regression models for longitudinal data with time dependent covariates and extended classifications. Stat. Med. **33**, 4756–4769 (2014)
Lee, Y., Nelder, J.A.: Hierarchical generalized linear models. J. Roy. Stat. Soc. Ser. B (Methodological) **58**(4), 619–678 (1996)
Lee, Y., Nelder, J.A.: Hierarchical generalised linear models: a synthesis of generalised linear models, random-effect models and structured dispersions. Biometrika **88**(4), 987–1006 (2001)
Lee, Y., Nelder, J.A.: Conditional and marginal models: another view. Stat. Sci. **19**(2), 219–228 (2004)
Lee, Y., Nelder, J.A., Pawitan, Y.: Generalized Linear Models with Random Effects, 1st edn. Chapman & Hall/CRC, Boca Raton (2006)
Liang, K.Y., Zeger, S.L.: Longitudinal data analysis using generalized linear models. Biometrika **73**, 13–22 (1986)
Liang, K.Y., Zeger, S.L., Qaqish, B.: Multivariate regression analyses for categorical data. J. Roy. Stat. Soc. Ser. B **54**(1), 3–40 (1992)
McCullagh, P., Nelder, J.A.: Generalized Linear Models, 2nd edn. Chapman & Hall, London (1989)
McCulloch, C.E., Searle, S.R., Neuhaus, J.M.: Generalized, Linear, and Mixed Models, 2nd edn. Wiley Series in Probability and Statistics. Wiley, Hoboken (2008)
Neuhaus, J.M., Kalbfleisch, J.D.: Between- and within-cluster covariate effects in the analysis of clustered data. Biometrics **54**, 638–645 (1998)
Neuhaus, J.M., Kalbfleisch, J.D., Hauck, W.W.: A comparison of cluster-specific and population-averaged approaches for analyzing correlated binary data. Int. Stat. Rev. **59**(1), 25–35 (1991)
Pepe, M.S., Anderson, G.L.: A cautionary note on inference for marginal regression models with longitudinal data and general correlated response data. Commun. Stat. Simul. Comput. **23**, 939–951 (1994)

Phillips, M.M., Phillips, K.T., Lalonde, T.L., Dykema, K.R.: Feasibility of text messaging for ecologocial momentary assessment of marijuana use in college students. Psychol. Assess. **26**(3), 947–957 (2014)

Wooldridge, J.M.: Introductory Econometrics: A Modern Approach, 4th edn. Cengage Learning, South Melbourne (2008)

Zeger, S.L., Liang, K.Y.: Longitudinal data analysis for discrete and continuous outcomes. Biometrics **42**, 121–130 (1986)

Zeger, S.L., Liang, K.Y.: An overview of methods for the analysis of longitudinal data. Stat. Med. **11**(14–15), 1825–1839 (1992)

Zeger, S.L., Liang, K.Y., Albert, P.S.: Models for longitudinal data: a generalized estimating equation approach. Biometrics **44**(4), 1049–1060 (1988)

Ziegler, A.: The different parametrizations of the gee1 and gee2. In: Seeber, G.U.H., et al. (eds.) Statistical Modelling, pp. 315–324. Springer, New York (1995)