**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 t**his chapter a discu**ssion is presented** of analysis options for longitudinal data with time-depen**dent covariates. Relevant** definitions are presented and explained in the context of practical applications, such as different types of time-dependent c**ovariates. 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 mod els 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 popul**ation-averaged models** include the independent GEE model and various forms of the GMM (generalized **method of moment**s) approach, including **researcher-determined types of time dependent covariates alo**ng with data-driven selection of moment conditio**ns using the Extended Classification. A practical data example is presented along with example programs for** both SAS and R.

**1 Introduction and Motivating Examples**

**Th**e term "longitudinal data” refers to data that involve the collection of the same **variables repeatedly over tim**e. 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 d**ata 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 co**mparisons 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, *C*O, USA** e-mail: trent.lalonde@unco.edu

© Springer International Publishing Switzerland 2015 **D.-G. Chen, J. Wilson (eds.)*, Innovative* S*tatistical Methods for Public Health Data,* ICSA Book Series in Stati**stics, DOI 10.1007*/9*78-3-319-18536-1\_4

**T.L. Lalonde**

**variables on responses, accounting f**or the effects of time. Lo**ngitudinal models** allow these qu**estions to be answered, while a**ccounting for the depend**ence 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 interes**ts, 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 no**t allow for a dynamic relationship between** the TIC and response.

TDC are independent variables that include both wi**thin-subject variation and** between-subject **variation, meaning** that the value of a TDC changes for a given **individual across time and can also change a**mong different subjects. A TDC can be used to m**ake comparisons across populations, to describe time trends, and also to describe dynamic relationshi**ps between the TDC and response. The focus of this chapter will be on approp**riate analysis techniques for TDC.**

**Examples of longitudinal data with TDC arise often and in many disciplines. For exam**ple, 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, dat**a were collected three time**s per day for 14 c**onsecutive days. To model the mean number of times marijuana was used, a longitudinal count model was** applied usi**ng 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 rehospit**alization within 30 days of a previous visit. R**ehospitalization within this ti**me frame is an important consid** eration for Medicare funding. Subjects for **the study were selected such that each** subject in the d**atabase had exa**ctly three hospi**tal follow-ups. Using a longitudinal logistic model, predictor**s of probabilit**y 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 1**995). It is also clear that certain TDC convey differen**t 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 t**ime but cannot be predicted as age can. In such cases **it is important to consider relationships between the TDC and the response across time.**

**Modeling Time-Dependent Covariates in Longitudinal Data Analyses**

**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 margin**al methods. The chapter

concludes with a data example exemplifying all relevant mode**ling techniques.**

**2 Classifying Time-Dependent Covariates**

Within a longi**tudinal study, a TDC can be defined as a variable that involves variatio**n both within subjects and between subjects. **The additional variation within** subjects is a source of dispersion that must be ac**counted for in longitudinal models, and can provide insi**ght into dynamic relationships b**etween 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 variabl**es and the respo**nse at a given time, the associations should not carry over** such that the treatment” at one time affect**s the response at a different t**ime. 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 “feed** back” relationships with the response. T**he 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 "tim**e-dependent confounders” (Diggle et al. 2002). **Random variables suc**h as "atmospheric pressure" or "unsolicited donations" **can change over time, but vary rando**mly with respect to a system. These types of **variables can have accumulated effects on the respon**se, 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 differen**t types of covariates can be associated with different conclusions or different appropriate estimation methods within the same models.**

***2.1 Exogeneit*y**

One of the most common distinctions made of TDC is that of exogeneity (Cham berlain 1982; Amemiya 1985; Diggle et al. 2002). An e**xogenous 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 fluctu**ating covariate that cannot be explained using other varia**bles in the study. It is most important to determ**ine exogeneity with respect to the response. A TDC is said to be exogeno**us with respect to **the response process if that t**ime-dependent variable at one time is conditionally independent of all pre**vious responses.**

**T.L. Lalonde**

Formally, let the response for subject i at time t be denoted by Yit, and let Kit denote a TDC for subject i at t**ime *t*. Then x is exogenous with respect to the response pr**ocess Y if

fx (Xit|Yil, ... , Yit; Xil, ... , Xi(*t*-1)) = fx (Xit|xil, . . . , X*i(t*-1)),

**(1)**

where fx denotes the density of x. Under the definition of Eq.(1), while Xit may be **associated with previous covariate value**s Xil, ..., X*i*(*t*-1), it will not be associated with previous or current responses Yil, ..., Yit*.* A **consequenc**e of this definition is **that the current r**esponse Yit will be independent of futu**re covariate values, even if there is an association with prior covariate values,**

E [Yit|xil, ... , Xit] = E [Yir|Xil, ... , Xi(-1)].

*(*2)

**Exogeneity wi**th respect to the **response has important mod**eling implications. Specifically, the def**inition implies that the response at any time may depend on prior** responses and prior values of the TDC, but will be independent of all o**ther covariate values. Th**ere is no feedback cycle of effects between responses **and exogenous**

**TDC.**

TDC that are no**t exogenous are referred to as endogenous TDC. An endogenous variable, sometimes called an internal variable, is a vari**able that is stochastically **related to other measured factors in th**e 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 sai**d to be feedback between the respo**nse and endogenous TDC**. This type of **relation**ship should be accounted for in any l**ongitudinal mod**el with TDC.

**As discussed b**y Diggle et al. (2002), e**xogeneity can be assessed by considering a regression of covariate valu**es X*i*t on both prior covariate values Xil, ..., X*i*(t-1) and also prior response values Yi1,..., *Yi*(t-1). If, after controlling for prior c**ovariate values, the current covariate value Xit shows an association with past response valu**es, the covariate shows evidence of endogeneity.

**2*.2***

***Types of Time-Dependent Covariates***

**Recent work has focused on further categorizati**on of types of TDC to facilitate **interpretations and proper estimatio**n 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 me**an of the response for subject i at time t is denoted by *Mir(B*), where *ß* is a vector of **mean parameters. This definition may** be induced by **an appropriately defined generalized linear m**odel, 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.**

**Modeling Time-Dependent Covariates in Longitudinal Data Analyses**

Lai and Small (2007) defined three types of TDC, and a fourth type was defined by Lalonde et al. (2014). Each type of T**DC 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 b**e a Type I TDC if

Talis (*Y:*

*- Mit)*

*Vs, t,*

**(3)**

L*OB;*

**where *Mis* and *Mit* are evaluated at the true parameter values *B*, 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 betwee**n the TDC and the response at differe**nt times**. A sufficient condition for a TDC to be Type I is

E [Yit|Xil, . . . , XT] = E [Yit|Xit].

**(4)**

Thus the response is independent of all TDC values at different times. The sufficient **requireme**nt of Eq. (4) would seem to be a strong**er condition than the exogeneity present**ed by Eq.(2), in that Eq. (4) requires the response at time t to be independent of all other TDC values, even those prior t*o 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 t**o be a Type II TDC if

E 2 (Y*in – Hit*) = 0 **0**

**Vs > *t.***

***(*5)**

**The expectati**on must be satisfied when s > t, but not neces**sarily when** s < *t,* **suggesting dependence between the response and covariate. In this** case the TDC **process is not associated with prior responses**, but the re**sponse process can be associated with prior TDC values. A sufficient condition for a covariate to be**

Type II is

E [Yit|Xi1, ..., XT] =E [Y:\Xil,..., Xi].

**As dis**cussed 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 suffi**cient for a TDC to be of Type II (Chamberlain 1982; Lai and Small 2007). Examples of Type **II TDCs include covariates that may h**ave a "lagged” association with the response in th**at previous TDC values can affect the response, but covariate values will not be affected by previous response values. One exam**ple is the covariate "blood **pressure m**edication" as a Type II covariate with the response "blood **pressure," as** the accumulated effects of med**ication over time can be expected to have an impact** on blood pressure at any time. A covariate is said to be a Type II TDC if

Talis (Yit

*(*Yit - *Ui*t)

= 0

Vs = t.

әрі

**62**

**T.L. Lalonde**

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 cycl**e 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 t**he probability of MI, an MI eve**nt may** elicit a change in blood **pressure medication. A covariate is said to be a Type IV TDC if**

t

.

*(7)*

**The exp**ectation must be satisfied for s < *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

E [Yit|Xil, ..., XIT] = E [Yit|xit, ... , X17]

Type IV TDC are associated with prior response values, but the **response at time t is only associate**d with the TDC at time t. **One exam**ple 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 associat**ed with different relationships with the **response. It is im**portant to be able to identify different types of TDCs to guide **model selection and to provide appropriate interpretations. Lai an**d Small (2007) **proposed selecting t**he type of TDC by choice of the researcher, b**ut also presented a** x2 test to compare two different selections of types for TDC. The idea is to construct **competing quadratic forms using the expressions fr**om Eqs. (3), (5), (6), and *(7*) with **zero expectation, so that additional expressions from a differ**ent selection of a type of TDC can inflate the quadratic form if those additional expressions do not, in **fact, have zero ex**pectation. However, this method will o**nly allow for comparisons** between possible selections of types of TDC, but will not make the selection for **the researcher. The Extended Classific**ation 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 m**odels, the focus of this section, are **often referred to as mixed models, random effect models,**

**Modeling Time-Dependent Covariates in Longitudinal Data Analyses**

**hierarchical models, or mixture models. Conditional models involve specification of a response model, conditional on a rand**om subject effect. This random effect is **intended to a**ccount for the clus**tering of responses** by subject, and induces “subject specific” or “cluster-specific” concl**usions 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 discussi**on 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* c**onditional correlated generalized linear model with random intercept can be written,**

**Random Component:**

Yi*|ui~ D(u*(X*it,* Zit)),

***Ui D*ula),**

Systematic and Link Components:

g*(u*(X*it, Z*it)) = x

+zv(u).

**In the expression** of the random component, represents a specific conditional **response** distribution from the exponential family, and *u*i indicates the random subject effect distributed according to *D*u wi**th parameters a. In defining conditional** models these two distributions are typically completely specified. In the **expression** of the systematic component, Zit represents a component of the random effects **design matrix, and v is a function transforming the random eff**ect 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 a**nd can be thought of as "errors" associated with the intercept *B*o. In a "random-slopes” model, pr**oducts of random effe**cts with the fixed **effects design matrix components Xit can be viewed a**s “errors” for the fixed-effects **par**ameters *Bk*, and thus allow the slopes to **vary randomly (Lalonde** et al. 2013). **Random-sl**opes models are often prese**nted as hierarchical models in which each parameter *Bk* has an associated linear model with an individual error term and can** include predictors.

***6*4**

**T.L. Lalonde**

**Here the interpretation of conditional model fixed effects can be made clear. The parameter *Bk* represents the expected change in the (transformed) mean response for a unit increase in** Xk, it for an individual subject, holding all other p**redictors fixed.** In other words, if predictor xk changes for an individual subject, *Bk* 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" co**mponents of effects associated with TDC (Neuhaus and Kal**bfleisch 1998). If a covariate includes both va**riation 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 dec**omposed into two terms: one associated with variation within subjects and the other associated with variation between subjects,**

*Bxit*

→

*Bw*(Xit - Ti) + *Bo*ž*i.*

In this expression the parameter *B*w represents the expected change in **the mean response associated with changes** of the TDC within subjects, while the pa**rameter *BB* is more of a population-averaged parameter that represents the expected change in the mean response associated with change**s of the TDC across subjects.

**3.2**

***An Issue with Estimation***

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

**N**

S(B) = Ž (OMC; )w.(Y, – (0;x) = 0,

***i=* 1**

***0P***

where the weight **matrix W; is often taken t**o be the inverse of the **variance covariance of the marginal** response (Diggle et al. 2002). Pepe and Anderson (1994) **show**ed that these estimating equations have zero expectation only if the data meet **the assumption,**

E[Y|X] = E[Y|*X),j =* 1,...,*T*],

**(8)**

for each TDC. The assumption of Eq. (8) is met trivially for TIC. Notice that **exogeneity is not a sufficient conditi**on, 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 sh**ould be **independent of covariate** values at all other times. When this assumption is satisfied,

**Modeling Time-Dependent Covariates in Longitudinal Data Analyses**

E [S(*R*)] = EE [S()|Xit, t = 1, ...,7]]

== e Ž (Oreo: ) cov.acer)1\*®–16:x)ku, 1 = 1,...,:

**Li=1**

**=**

**E**

*au(B;*x;)? 1*2B*

*[ØVic*a

*8*; x;))]^'E[(Y*; – u(B;*x;))|Xi,*t* = 1,..

**L*i*=1**

== [ (249:-))"Lov\_(QCP:x)+x0]

**L*i*=1**

**=**

**0.**

In this derivation, the step o**f removing the derivative term from the inner expectation** depends on the assumption of Eq.(8). Depending on the choice of the weight matrix Wi, the **estimating equations may require combinations** of the first te**rm (the derivative of the systematic component) with the second term (the raw error term)** across different times. Specifically, this will be the c**ase 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 w**eight matrix is applied. Notice Eq.(8) is a sufficient con**dition 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 *z*ero expectation, leading to bias and loss of efficiency in parameter estimates (Pepe and Ander**son 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 u**sing random** effects. Instead a marginal model involves specification of the marginal response **distribution, or at least moments of the response dis**tribution (McCullagh and Nelder 1989; Hardin and Hilbe 2003; Diggle et al. 2002). This type of model is associated with “popul**ation-averaged” interpretations, or standard regression interpretations. Parameters in margin**al longitudinal models p**rovide a comparison of the mean response between two populations with different average v**alues of the predictor of **interest. While marginal** conclusions can be obtained thro**ugh conditional models, the term “margin**al model” will be used to refer to a model specifically **intended for marginal expression and interpretatio**ns (Lee and Nelder 2004). A marginal correlated generalized linear model c**an be written,**

***66***

**T.L. Lalonde**

**Random Component:**

Yit ~

*D (u(*X*it), V(u*(x*i*))),

Marginal Mean:

In(*u* (Xi)) = x *B.*

For this type of model D is a**ssumed to be a distribution from the exponential family** of distributions, but may not be fully specified w**ithin a marginal model. Instead, the mean *u(Xit*) and varian**ce V(*u(*X*i*t)) (with poss**ible over dispersion parameter ) are supplied by the researcher. While there are many marginal methods for estimating parameter**s 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, p**arameter estimates are** obtained by solving the equations,

S(B) = ¿ (04E01 x)) [V:69(8;x))+(Y– u(B;x.) = 0,

**i=**

**1**

**where the variance-covariance structure is specified through a working correlation structure R*i*(Q),**

Vi(2(Xi)) = A[/R;(Q)A?*/*?.

**Pepe and Anderso**n (1994) argued that the **structure of the GEE requires satisfaction of the assum**ption,

E[Y:\Xi] = E[YiX*j, j* = 1, ..., *7*],

so that t**he GEE will have zero expectation. As with conditional model estimation, this as**sumption is met trivially for TIC. When the assumption is met, the first term of the GEE can be factored out of the expectation of S(*B*), producing unbiased **estimating equations. When the assum**ption is not met, the GEE will not have **zero exp**ectation unless the wor**king correlation str**ucture is selected so that all **components of the GEE involve only a single observation time. This is achieved by a**

**Modeling Time-Dependent Covariates in Longitudinal Data Analyses**

**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 whe**n applying GEE to longitudinal data with TDC. **However, the result**s of Fitzmaurice (1995) sugg**est there may be meaningful losses** in efficiency depending on the strength of the **auto-correlation. Additionally, the** approach of applying independe**nt 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 *B* can be estimated by **minimizing a qu**adratic form Q*(B*) with appropriately chosen components,

*Q(B*) *= GT (*B;Y,X)W-'G*(B*;Y,X),

**where G(*B*;Y, X) is a vect**or of “moment conditions" with zero expectation and W is a correspondingly chosen weight matrix. For longitudinal data situations, G **is ty**pically **constructed as an average of vectors of “valid moment conditions” for** each subject,

G*C*B; Y,X) = 18:(8;Y,X).

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

W = Cov *(gi(****B;*Y, X)).**

The challenge in applying the GMM is to determine appropriate components of the subject-leve**l vectors of valid moments** conditions gi. 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 elem**ent of g*i(B*;Y, X) according to the expected nature of each TDC. Specifically, **the expectations associate**d with Type I, Type I, Type III, and Type IV TDC, Eqs. (3), (5), (6), and *(7*), respectively, define combinatio**ns of times at which**

**68**

**T.L. Lalonde**

**components of potential moment conditions will be independent. When components are i**ndependent 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,**

*8it(B*;Y,X) = (

) Vis – *Hli*t).

l*obi*

The type of TDC will determine whic**h combinations of times form valid moment con**ditions. For all predictors in the model, the co**ncatenation of all valid** moment conditions will form the vector gi for each subject. Notice that this method **avoids choosing a general weight matrix to apply across all covariates, as with** likelihood-base**d estimation or with the G**EE. Instead, the G**MM allows expressions to be constructed separately for each TDC, which provide**s the ability to treat **each covariate according to i**ts type. This el**iminates 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-deter**mined types, the Extended Classification p**rocess 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 *B,*** are obtained using GEE with the independent **working correlation structure. Value**s of both the **derivative component and raw** residual component of pote**ntial moment conditions can** be calculated for TDC Xj,

â. \_ ðîs

*ab;'* f*t*=*yt - hea,*

where *û,* represents a vector of predicted mean respo**nses at time *t* across all** subjects, evaluated using Bo. After standardizing both vectors to obtain *dsji* and iti, the association between these components is **then evaluated using Pearson** correlation,

*@sjt =*

Elā*sji – āsj) (Tai* – 7.) *vēlāsji – ās*)? (12 – 7,72

**Modeling Time-Dependent Covariates in Longitudinal Data Analyses**

**69**

and standardized for comparison. Assuming all fourth moments of *Ô*s*it* **exist and** are finite,

*posit* =

*=*

-

*Pajt VÛ22/N*

~ N(0,1),

where *Û22 = (1/N*) *sii)?* (ti)?, and *N* is the total number of subjects. Sig nificantly correla**ted components show evidence of non**-independence between the **derivative and raw residual terms, and therefore the associat**ed product o**f derivative and raw err**or should no**t reasonably have zero expectation and can be omitted as a potential valid moment conditi**on. To account for the large number of hypothesis **tests involved in the Extended Classification process, p-values f**or all correlation **tests can** be collectively e**valuated (Conn**eely and Boehnke 2007).

**The method of Extended Classification removes the potentia**lly subjective deci **si**on of the type of each TDC by the researcher and allows the data to **determine** appropriate valid moment conditions. Extended Classification also al**lows for more** than four discrete types, admitting all po**ssible combinations of times instead of the four cases corresponding t**o the four types. The Extended Classif**ication 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*(B)*. 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 sepa**rate steps to address the weight matrix and moment conditions**. Using initial values *B*o), **an estimate of the weight matrix Wo is** obtained and substituted into the quadratic form,

*Ots(B)* = G*T(B*;Y, X)WO G(*B*;Y,X).

**Th**e quadratic form is then **minimized to obtain final parameter estimates . The TSGMM process appears to be the most commonly applied method in the literature. The IGMM process involves an iterative re**peat of the steps in **TSGMM. After the** quadratic form Cts has been minimized to obtain updated parameter estimates *(*1), **the estimat**e of the weight matrix is updated, providing W(1). The process then **iterates between u**pdating using the quadratic form and updating **W) using the resulting estimates,**

ce+1) = ar*g*m*i*n [GT(B;Y, X)W, GCB;Y,X)].

***70***

**T.L. Lalonde**

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

**Time**

Re-admit No

**Total** *75*6

**1** 231 46.48 % **266 53.52%**

**2** *272* 54.73 % ***2*2*5* 45*.*27 %**

**3** 253 50.91 % **244 49.09%**

**Yes**

**73*5***

The process completes on sufficient convergence of *Bl*o. **The IGMM process** appears to be the least commonly used method in the lite**rature, 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,**

*Qcu(B*) = G*+(B*;Y,X) (W*(*))-G*(B*;Y,X).

**Estimates are obtained by a single minimization of** O*cu.*

**5 Data Example**

In order to exemplify the impl**ementation and interpretation associated with the** models discussed in Sects. 3 and 4, an analysis is presented using the A**rizona 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 i**ncludes 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 hospi**talization (“LOS”), the existence of coronary atheroscleros**is (“C.A."), and indicators for time 2 and time 3. Table 1 pro**vides the percentage of the patients who were readmitted t**o the hospital within 30 days of **discharge against** the percentages of the patien**ts 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-int**ercept logistic regression with deco**mposition of TDC (RS); random-slope logistic regression with decomposition of TDC (RS); GEE logistic regression with independent working correlation structure (IGE**E); TSGMM **logistic regression** with the type of each TDC selected by the researcher (GMM Types); and TSGMM logist**ic regression using extended classif**ication *(*GMM-EC). **The GMM models w**ill be fit using the TSGMM.

**Modeling Time-Dependent Covariates in Longitudinal Data Analyses**

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

logit(Tli*t) = Bo* + )

(*Bew*(Xk,it – Ik*, i.*) + *Bkbłki.)* + *B*rzTime2+ *B*13 Time3 + Yoi,

***k=***

**1**

**wher**e *Ti*t indicates the probability of rehospitalization within 30 days for subject ***i* at time t, an**d yoi indicates the random subject effect. The model can be fit using **SAS or R w**ith 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 re**hospitalization to vary r**andomly among subjects,

**4**

logit(Tä) = Bo +

*(B*xw (Xk, is –āk.i) + B*xb*ā*k.i.)*

***k*=1**

+*B1*2 Time2 + *B*13 Time3 + Voi + 71LOSit,

where yli 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 rand**om subject effects,

logit(*fit) = Bo* +

*Bx*wX*k,it + B1*2 Time2 + *B*:3 Time3.

***k=***

**1**

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

**The system**atic 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 Ty**pe I TDC, as **is common for such deterministic v**ariables. Both “length o**f stay” and “existence of coronary atherosclerosis" w**ill be treated as Type II TDC, as it i**s reasonable to assume an a**ccumulated effect on th**e response from these two var**iables, but **it is unlikely that the response at one time will affect future val**ues of these **covariates. B**oth “number of diagnoses” and “number of procedures" will be **treated** as Type III TDC, as it is **reasonable to assume feedback between the prob**ability 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 classificati**on process will be used to select app**ropriate valid moment conditions to be used in the GMM quadratic form. These G**MM methods

are not yet available in SAS; R funct**ions written by the author can be requested.**

**T.L. Lalonde**

**Results of fitting all five models are presented in T**able 2. Firs**t consider the resul**ts of the conditional models. For the model includ**ing a random intercept, the variation associat**ed with that intercept (0.14*7*2) is significant, suggesting there is **significant individual variation in the baseline probabi**lity of rehospita**lization within** 30 days. For all models the t**ime 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 prov**ide interesting interpretations. The** "between” components of the TDC provide popu**lation-averaged types of conclu sions. For example**, there is evidence that subject**s with higher average length of stay** tend to have a higher probability of rehospitalization (0.0*7*36), perhaps an indication **of more serious illnesses. The "within” components provide interpretations of individual effects over time. For exam**ple, there i**s evidence that an increase in the number of diagnoses for an individual is associated with a higher probability of** rehospitalization (0.0780), per**haps an indicati**on of identifying additional illnesses.

**Result**s for the model including a random-slope for length o**f stay are similar.** Within the RS model, the variation in the length of stay slope (0.0025) is significant, **indicating meaningful individual variation i**n the effect of leng**th of stay on the probability of rehospitalization. The variation in the int**ercept (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 i**n the probability o**f rehospitalization within 30 days for subjects who eventually develop coronary atherosclerosis.**

**Next consider the results of the margina**l models. For all three of the **models IGEE, GMM-T**ypes, and GMM-EC, the p**arameter associated with length of stay is positive and significant. This indicates that, when comparing two populations with different avera**ge lengths of stay, the population with the higher length o**f stay has** a higher probability o**f rehospitalizatio**n 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 IGE**E process. Also notice tha**t 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 popu**lation-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 app**ropriate model is often based on researcher intentions. The IGEE m**odel is a safe choice,

**Table 2 Conditional and marginal logistic regression models Parameter estimates and significance**

**RI—within RI—between RS—within RS—between Diagnoses 0*.*0780\*\*\* | 0.0444** | 0.0686\*\* 0.0362 U **Procedures**

**-0.0824\*\* 0.0092 -0.0915\*\***

**LOS**

0.0008 0*.0*736\*\*\* | 0.0200\* 0.0952\*\*\* C.A. -0.2607\* 0.2223 -0.2646\* 0.3050\* **Time 2 -0.3730\*\*\***

-0.4061\*\*\*

**Time 3** -0.2130\*\*

**-0.23*5*7\*\***

**Interce**pt 0.14*7*2\*\*

**0.1512\***

Slope

0.0025\*

Gen *xP/DF* 0.98

**0.96**

QIC

**QICU**

0 \*\*\*0.001 \*\*0.01 \*0.05 0.10

**IGEE**

0*.*0648\*\*\*

**-0.0306**

| 0.0344\*\*\*

-0.1143

-0.3876\*\*\*

**-0.2412\*\*\***

**GMM-types G*M*M-EC**

0.0613\*\***\* 0.0573\*\*\* -0.0**458\* -0.0366

0.0530\*\*\* 0.04*9*7\*\*\* -0.0536 -0.0802 -0.4004\*\*\* -**0.3933\*\*\* -0.2417\*\*\* -0.2633\*\*\***

**Modeling Time-Dependent Covariates in Longitudinal Data Analyses**

**6648.52**

**6646.56**

**T.L. Lalonde**

but generally lacks the power of the GMM models. The co**nditional models are an** approp**riate 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 condit**ion of Eq.(8) presented by Pepe and Anderson (1994), **and allows for TDC to be treated differ**ently from each other. In this sense the **Extended Classifica**tion method provides the most flexibility, as moment conditions are selected individually based on empirical evidence from the dataset. In this d**ata example the results of both the GMM-Types and GMM-EC models are quite similar,** yielding the **same signs of parameter estimat**es and similar significance levels, **whi**ch suggests the researcher-selected types o**f covariates are** probably appro**priate 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 classificat**ion is as exogenous versus endogenous covariates. Exogenous covariates vary according to factors extern**al to the system under con**sideration, while endogenous covariates show associatio**n with other recorded variables. It is **most important to identify exogeneity with respect to the response variable.**

**TDC more recently have been classified according t**o four "types" that reflect the **natur**e 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 substan**tial relationship a “feedback” loop **between covariates and response at different times.**

Existing methods for modeling longitudinal data with TDC can be spli**t into two classes: conditional models and marginal models. Conditional models incorporate random effects into the systematic component of t**he model to account for the **autocorrelation in responses. To accommodate TDC, individual regression terms can be decomposed into contributions from variati**on within” subjects and vari ation “between” subjects. When **maximum-likeli**hood-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 va**lues at other times. **If this assum**ption is not met, the likelihood es**timating equations will not have** zero expectation because of off-d**iagonal components of the response variance covariance structure, which can bias parameter estimates.**

Marginal models, on the other hand, define a marginal response (quasi-) distribu **tion 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 bee**n recommended that the independent working correlation structure**

**Modeling Time-Dependent Covariates in Longitudinal Data Analyses**

**is applied when using GEE. This recommendation is made to avoid satisfying a necessar**y condition for the GEE to **have zero expected value, as individual** estimating equations that combine components at diffe**rent 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 natur**e of the relationship between each T**DC and the response. In t**he future, the GMM with Extended Classifica**tion 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) mo**del discussed in Section LABEL can be fit us**ing the following SAS commands.** */\** PROC GLIMMIX DOES NOT REQUIRE INITIAL VALUES *\*/* **PROC GLIMMIX DATA=ASID DATA;**

CLASS subject\_id; MODEL readmission(event = *'*1') = diagn**oses\_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 *V*ALUES :USE INDEPENDENT GEE \**/* PROC NLMIXED DATA=ASID\_DATA QPOINTS=30;

**PARMS beta0= betal= beta2= beta3= beta4= beta5=**

beta6= beta7*=* beta8= beta9= beta10=; **eta = u + beta0 + beta1 diagnoses\_w**

**+ beta2\*diagnoses\_b + beta 3\* procedures\_w + beta4 procedures\_b + beta5\*LOS\_W + beta6\*LOS\_b + beta*7\**CA W + beta 8\*CA b**

**T.L. Lalonde**

**+ beta9\*time2 + beta10\*time3; exp**\_eta = exp(eta); pi = ((exp\_eta*)* / (1+exp\_eta)*);* MODEL r**eadmission « BINA**RY (pi);

RANDOM u ~ NORMAL (0, sigmau\*sigmau) SUBJECT=subject\_id; RUN;

**Alternat**ively, the model can be fit u**sing R with the following commands.**

install.packages ("lme4") library (lme4) # USE start=c(diagn**oses\_w=***,* ...) OPTION TO SPECIFY

**INITIAL *VA*LUES #** # USE INDEPENDENT GEE FOR INITIAL VALUES # R\_Int = glmer (readmission ~ diag**noses\_w+diagnoses\_b**

**+procedures\_w+procedures\_b**+LOS\_W+LOS\_b

**+CAW+CA D** +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 *V*ALUES *\*/* PROC GLIMMIX DATA*=*ASID\_DATA*;*

**CLASS subject\_id;** MODEL readmission(event = '1') = diag**noses\_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 *V*ALUES:

USE INDEPENDENT GEE \**/* **PROC NLMIXED DATA=ASID DATA QP**OINTS=30;

**PARMS beta0= betal= beta2= beta3= beta4= beta5=**

**beta6= beta*7=* beta8= beta9= beta10=;** eta = u + betaO + betal di**agnoses\_w + beta2+diagnoses\_b**

**+ beta3\*procedures\_w + beta4\*procedures\_b + beta5\*LOS\_W + beta6\*LOS\_b**

**Modeling Time-Dependent Covariates in Longitudinal Data Analyses**

**+**

**+**

**+**

**+ beta*7\**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); **RAN**DOM u rbl - 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 *V*ALUES # R\_Slo**pes = glmer (readmission ~ diagnoses\_w+diagnoses\_b**

**+procedures\_w+pr**ocedures\_b+LOS\_w+LOS\_b+CA\_W+CA\_b +time2+time3 + (1 subject\_id) + (O+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 followin**g commands.**

**install.packages ("geepac**k") **library (geepack)**

Ind\_GEE = geeglm(re**admission - diagnoses+procedures+LOS+CA**

+time2+time3, family=binomial, id=subject\_id, corstr="independence",

**data=ASID\_DATA)** summary (Ind\_GEE)

**78**

**T.L. Lalonde**

**References**

**Amemiya, T.: Advanced Econometrics. Harvard University Press, Cam**bridge (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 multi**

**var**iate binary da**ta. 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) Fitzmau**rice, G.M., Laird, N.M., Ware, J.H.: Applied Longitudinal Analysis, 2nd edn. Wile*y,*

Hoboken (2012) **Hansen, L.P.: Large sample properties of generalized method of moments estimators. Economet**

rica 50(4), 1029–1054 (1982) **Hansen, L.P.: Generalized Method of Moments Estimation, pp. 1–14. Department of Economics,**

**Univers**ity 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) **Lal**onde, 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–*7*38 (2013) Lalonde, T.L., Wilson, J.R., Yin, J.: Gmm logistic regression models f**or longitudinal data with

**time dependent covariates and extended classifications. Stat. Med. 33, 4756-4769 *(*2014)** Lee, Y., Nelder, J.A.: Hierarchical gen**eralized linear models. J. R**oy. Stat. Soc. Ser. B (Method

ological) 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.: Longit**udinal data analysis using generalized linear models. Biometrika**

*7*3, 13–22 (1986) Liang, K.Y., Zeger, S.L., Qaqish, B.: Mu**ltivariate regression analyses for cate**gorical data. J. Roy.

Stat. Soc. Ser. B 54(1), 3–40 (1992) McCullagh, P., Nelder, J*.*A.: **Generalized Linear Mo**dels, 2nd edn. Chapman & Hall, London

**(1989)** McCulloch, C.E., Searle, S.R., Neuhaus, J.*M*.; **Generalized, Linear, and *M*ixed Mod**els, 2nd edn.

Wiley Series in Probability and Statistics. Wiley**, Hoboken (2008) Neuhaus, J.M., Kalbfle**isch, 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. *5*9(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. Simu**l. Comput. 23, **939-951 (1994)**

**Modeling Time-Dependent Covariates in Longitudinal Data Analyses**

Phillips, M.M., Phillips, K.T.**, Lalonde, T.L., Dykema, K.R.: Feasibility of text messaging for**

**ecologocial momentary assessment of marijuana u**se in college students. Psychol. **Assess.**

**26(3), 947*–95*7 *(2*014)** Wooldridge, J.M.: Introductor**y Econometri**cs: 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 me**thods 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), 104**9–1060 (1988) **Ziegler, A.: The different parametrizatio**ns of the geel and gee2. In: Seeber, G.U.H., et al. (eds.)

**Statistical** Modelling, pp. 315–324. Springer, N**ew Yo*r*k (1995)**