

MARGINAL ANALYSIS OF LONGITUDINAL COUNT DATA IN LONG SEQUENCES: METHODS AND APPLICATIONS TO A DRIVING STUDY¹

BY ZHIWEI ZHANG, PAUL S. ALBERT AND BRUCE SIMONS-MORTON

*Eunice Kennedy Shriver National Institute of Child Health
and Human Development*

Most of the available methods for longitudinal data analysis are designed and validated for the situation where the number of subjects is large and the number of observations per subject is relatively small. Motivated by the Naturalistic Teenage Driving Study (NTDS), which represents the exact opposite situation, we examine standard and propose new methodology for marginal analysis of longitudinal count data in a small number of very long sequences. We consider standard methods based on generalized estimating equations, under working independence or an appropriate correlation structure, and find them unsatisfactory for dealing with time-dependent covariates when the counts are low. For this situation, we explore a within-cluster resampling (WCR) approach that involves repeated analyses of random subsamples with a final analysis that synthesizes results across subsamples. This leads to a novel WCR method which operates on separated blocks within subjects and which performs better than all of the previously considered methods. The methods are applied to the NTDS data and evaluated in simulation experiments mimicking the NTDS.

1. Introduction. In this paper we consider the analysis of longitudinal data that arise in the form of a small number of long sequences. Our interest in this problem is motivated by the Naturalistic Teenage Driving Study (NTDS), an observational study of teenage driving performance and characteristics [Simons-Morton et al. (2011a, 2011b)]. In the study, 42 newly licensed teenage drivers in Virginia were monitored continuously during their first 18 months of independent driving using in-vehicle data recording systems. The instrumentation included accelerometers, video cameras, a global positioning system, a front radar and a lane tracker. The NTDS is a first of its kind, at least for teenage drivers in the United States. The study provides valuable information on risky driving behavior, which can be assessed in terms of elevated gravitational force (g -force) events (rapid start, hard

Received February 2011; revised September 2011.

¹Supported by the Intramural Research Program of the National Institutes of Health (NIH), Eunice Kennedy Shriver National Institute of Child Health and Human Development. The computation was facilitated by the Biowulf cluster computer system made available by the Center for Information-Technology at the NIH.

Key words and phrases. Correlation, generalized estimating equation, multiple outputation, overdispersion, random effect, separated blocks, within-cluster resampling.

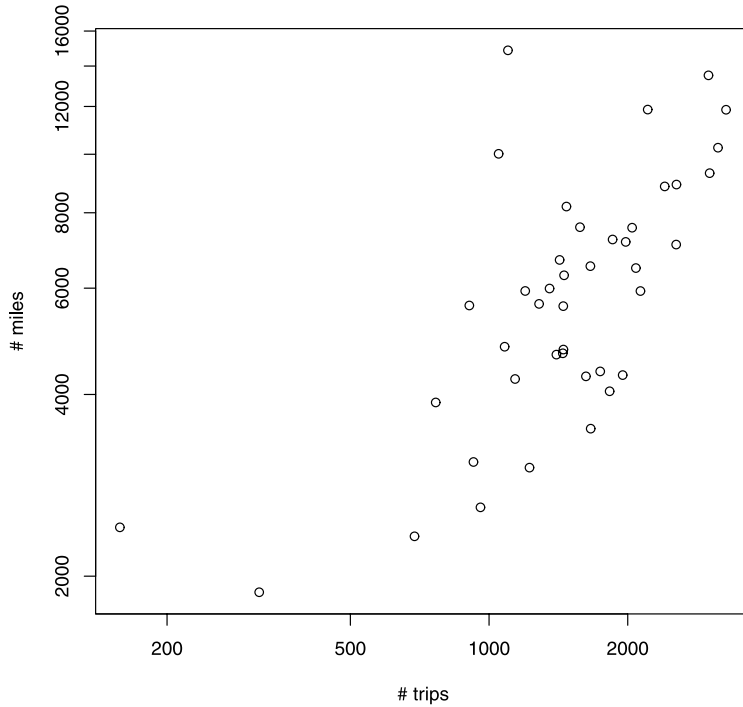


FIG. 1. Summary of NTDS data by subject: the number of trips made and the total number of miles driven, both plotted on the logarithmic scale.

stop, hard turn and yaw). Counts of g -force events are available for each trip (defined as ignition on to ignition off), and their incidence rates represent different aspects of risky driving behavior. The NTDS data set comprises more than 68,000 trips by the 42 teen subjects, with an average of 1,626 trips per subject. Figure 1 provides a summary of the NTDS data by subject in terms of the number of trips made and the total number of miles driven. An important goal in our analysis of the NTDS data is to understand how risky driving is associated with subject-level covariates (i.e., individual characteristics such as gender) as well as trip-level or time-dependent covariates (e.g., time since licensure, presence of passengers).

There is an extensive literature on longitudinal data analysis; see, for example, Diggle et al. (2002), Fitzmaurice et al. (2008) and McCulloch, Searle and Neuhaus (2008). Common approaches to longitudinal data analysis include random effect models [Laird and Ware (1982), McCulloch, Searle and Neuhaus (2008)] and generalized estimating equations (GEE) [Liang and Zeger (1986), Zeger and Liang (1986), Zeger, Liang and Albert (1988)]. In general, the two approaches have different interpretations (subject-specific versus marginal), although that distinction is not important when the log link is used for count data. For analyzing the NTDS data, it is natural to consider generalized linear mixed models (GLMM)

with appropriate random effects to account for population heterogeneity, serial correlation and/or possible overdispersion [Chan and Ledolter (1995), McCulloch (1997), McCulloch, Searle and Neuhaus (2008)]. However, a realistic GLMM for the NTDS data would involve several random components, including a latent process that induces serial correlation (see Section 5), and the computational demand of such a GLMM analysis can be prohibitive with thousands of trips per subject. Even if the computation is feasible, the resulting inference may be sensitive to modeling assumptions that are difficult to verify. The computational burden can be reduced by resorting to a marginal analysis using the GEE approach, especially under a working independence assumption. The GEE approach only requires specification of the first two moments and not the entire distribution. The robust variance estimate can be used to make asymptotically valid inference that only requires correct specification of the mean structure, assuming that the number of subjects is large relative to the number of observations per subject. The exact opposite situation occurs in the NTDS, and Albert and McShane (1995) have shown that the robust variance estimate can perform poorly in such a situation and that the model-based variance estimate may be preferable. Numerous methods have been proposed to improve upon the robust variance estimate, including jackknife [Paik (1988), Lipsitz, Laird and Harrington (1990)], bias correction [Mancl and DeRouen (2001)] and window subsampling techniques [Sherman (1996), Heagerty and Lumley (2000)].

Given the well-known advantages and potential issues of the GEE approach, it is natural to ask how to perform a simple and valid marginal analysis of the NTDS data with reasonable efficiency and robustness. We attempt to address this question in the present paper by examining some existing methods and developing new ones. We find that it is generally helpful to include a fixed effect for each subject in the GEE model. Once the subject effects are estimated, they can be treated as the response variable in a subsequent linear regression analysis for estimating the effects of subject-level covariates. These fixed effects also help with trip-level covariates by removing the correlation due to population heterogeneity. If the counts are large (with a marginal mean of 1, say), the effects of trip-level covariates can be estimated from a conventional GEE analysis using an estimated covariance matrix together with the robust variance estimate. For small counts (with a marginal mean of 0.1), the conventional approach is not satisfactory, and we explore a within-cluster resampling (WCR) approach [Hoffman, Sen and Weinberg (2001), Follmann, Proschan and Leifer (2003)]. The WCR approach was originally proposed to deal with informative cluster sizes, which is not of concern in the NTDS. Our motivation for considering WCR is to improve the performance of GEE methods by altering the data structure. To this end, we consider extensions of WCR that reduce serial correlation within clusters or increase the number of (approximately) independent clusters. This leads to a WCR method involving separated blocks which performs better than the conventional methods.

The rest of the paper is organized as follows. In the next section we set up the notation, formulate the problem, and discuss potential issues. Then we examine the standard GEE methods in Section 3 and explore some WCR methods in Section 4. The NTDS data are analyzed in Section 5 using the appropriate methods. The paper concludes with a discussion in Section 6.

2. Formulation. Let Y_{ij} ($i = 1, \dots, n$; $j = 1, \dots, k_i$) denote the number of events that occur during the j th trip by the i th subject. For reasons that will become clear later, we distinguish subject-level covariates such as gender from trip-level covariates such as time since licensure, writing Z_i for the former group of covariates and X_{ij} for the latter. Note that X_{ij} may include interactions between subject-level and trip-level covariates. With m_{ij} denoting the mileage of the j th trip by the i th subject, the marginal model of interest to us may be written as

$$(1) \quad E(Y_{ij}) = m_{ij} \exp(v + \alpha' Z_i + \beta' X_{ij}),$$

where v , α and β are unknown parameters, the latter two being of primary importance. Without loss of generality, we treat the covariates as fixed.

Estimation of the parameters in model (1) is challenged by the fact that the Y_{ij} from the same subject tend to be correlated due to considerable variability between drivers as well as serial correlation (over time) within drivers. These sources of correlation are often accounted for using random effects [e.g., Zeger (1988), Davis, Dunsmuir and Wang (2000)]. For example, one might postulate that, conditional on the random effects b_i , c_{ij} and e_{ij} , the Y_{ij} are independent and each follows a Poisson distribution with conditional mean

$$(2) \quad E(Y_{ij} | b_i, c_{ij}, e_{ij}) = m_{ij} \exp(v^* + \alpha' Z_i + \beta' X_{ij} + b_i + c_{ij} + e_{ij}),$$

where b_i induces some heterogeneity between drivers (beyond that explained by Z_i), c_{ij} generates serial correlation among trips close in time, and e_{ij} accounts for any additional overdispersion relative to the Poisson model. It is often assumed that the c_{ij} for a given subject arise from a subject-specific stochastic process c_i through the relationship $c_{ij} = c_i(t_{ij})$, where t_{ij} denotes the time of the j th trip. Note that model (2) is consistent with model (1) because it implies the same mean structure after integrating out the random effects, provided the distributions of the random effects do not depend on the covariates.

For the NTDS data, it seems reasonable to equip model (2) with the following distributional assumptions. One might assume that $b_i \sim N(0, \sigma_b^2)$, $e_{ij} \sim N(0, \sigma_e^2)$, and c_i is a zero-mean Gaussian process with a covariance structure given by

$$(3) \quad \text{cov}\{c_i(t_1), c_i(t_2)\} = \sigma_c^2 \exp(-\gamma |t_1 - t_2|).$$

The random effects b_i and e_{ij} and the process c_i are assumed independent of each other, although the $c_{ij} = c_i(t_{ij})$ are necessarily correlated within each subject. The parameter $\gamma > 0$ determines how rapidly the serial correlation decreases with the

gap time. The process c_i is known as the Ornstein–Uhlenbeck process [Uhlenbeck and Ornstein (1930)], and the above distributional assumptions will be collectively referred to as the Gauss–Ornstein–Uhlenbeck–Poisson (GOUP) model. It is possible to perform a maximum likelihood analysis under the GOUP model [e.g., Chan and Ledolter (1995)]. However, such an analysis can be computationally demanding because of the serial correlation, and the resulting inference may be sensitive to the distributional assumptions involved. The methods to be considered for our marginal analysis will not require the full strength of the GOUP model, and they may or may not involve the distributional assumptions in the GOUP model. Some of the methods we consider do incorporate such distributional assumptions into estimating equations through a working covariance matrix, in which case we also assess the robustness of the resulting inference against misspecification of the correlation structure. Our goal is to make valid and reasonably efficient inference on α and β in model (1) with minimal dependence on the distributional assumptions in the GOUP model.

Without the distributional assumptions, model (2) does play a crucial role in this paper, as a way to conceptualize the different sources of variability. We assume that the different subjects are independent of each other and that the random effects b_i and $(e_{ij})_{j=1}^{k_i}$ and the random process c_i are independent of each other within each subject. It follows that the Y_{ij} ($j = 1, \dots, k_i$) are conditionally independent given b_i and c_i . Another important implication is that

$$(4) \quad E(Y_{ij}|b_i) = m_{ij} \exp(v_i + \beta' X_{ij}),$$

where

$$(5) \quad v_i = v^* + \alpha' Z_i + b_i + \log[E\{\exp(c_{ij} + e_{ij})\}] = \text{const.} + \alpha' Z_i + b_i.$$

This shows that the correlation due to b_i can be removed by treating subject as a fixed effect. Obviously, this approach would not work in the usual asymptotic theory assuming a large number of subjects and a limited number of observations per subject. However, it might be appropriate when the number of subjects is small and the number of observations per subject is large as in the NTDS. Note that the subject-specific model (4) does not directly involve α , which has been absorbed into the subject-specific intercept v_i ; thus, α cannot be estimated directly by fitting model (4). Nonetheless, equation (5) suggests that once v_i has been estimated, say, by \hat{v}_i , it should be possible to estimate α from a subject-level linear model regressing \hat{v}_i on Z_i .

3. Standard GEE methods.

3.1. *GEE with working independence.* Assuming working independence, a standard GEE analysis can be performed with or without fixed subject effects (FSE), using either the robust variance estimate or the model-based variance estimate. This gives rise to four possible methods for estimating β . As discussed

earlier, α is not directly estimable from a GEE analysis with FSE but can be recovered from a subsequent linear regression analysis based on (5). Let \hat{v}_i denote the GEE estimate of v_i , and assume that \hat{v}_i is approximately unbiased, that is, $E(\hat{v}_i|v_i) \approx v_i$. This, together with (5), implies that

$$E(\hat{v}_i) = E\{E(\hat{v}_i|v_i)\} \approx E(v_i) = \text{const.} + \alpha' Z_i.$$

An approximately unbiased estimate of α can then be obtained from a subject-level linear model regressing \hat{v}_i on Z_i . A standard least squares (LS) algorithm can be used to fit this model if the \hat{v}_i can be treated as independent. In general, the \hat{v}_i are correlated due to correlated errors in GEE estimation, even though the v_i are indeed independent of each other. This correlation can be taken into account using an iteratively reweighted least squares (IRLS) algorithm described in Appendix A. Thus, under working independence, α can also be estimated using four different methods (LS and IRLS with FSE, robust and model-based without FSE).

These methods are evaluated and compared in a simulation study mimicking the NTDS. Specifically, each simulated data set consists of 40 subjects and 60,000 trips (1500 per subject). The study duration is rescaled to the unit interval, over which the trips are uniformly distributed. The offset $\log(m_{ij})$ follows a normal distribution with mean 1 and variance 1, Z_i is generated as a Bernoulli variable with success probability 0.5, and X_{ij} is taken to be the trip time (i.e., $X_{ij} = t_{ij}$). Given the covariates, the outcome Y_{ij} is generated according to the Goup model described in Section 2, with $\sigma_b^2 = \sigma_c^2 = \sigma_e^2 = 1$ and $\gamma = 50$ or 300, corresponding to longer- and shorter-lived serial correlation, respectively. These values are based on the NTDS data (see Section 5) and the wide range for γ reflects a large amount of variability in its estimation (see Section 3.2). We set $\alpha = \beta = 0$ for simplicity because the results change little over a range of realistic values for these parameters. We choose v^* in (2) such that the marginal mean of Y_{ij} equals a specified value (0.1 or 1). This range for $E(Y_{ij})$ covers most kinematic measures in the NTDS with varying thresholds. Some measures are associated with larger counts, which are generally easier to deal with. In each scenario (combination of parameter values), 1,000 replicate samples are generated and analyzed using the methods described in the preceding paragraph.

Table 1 compares the four methods for estimating α in terms of empirical bias, standard deviation, median standard error and coverage probability of intended 95% confidence intervals. All methods in Table 1 are nearly unbiased, suggesting that bias is not of concern here. In terms of efficiency, the most important factor is clearly the use of FSE, which consistently results in smaller standard deviations. With FSE in the model, one might expect the IRLS estimate of α , which accounts for the correlation among the \hat{v}_i , to be more efficient than the LS estimate. However, their difference seems rather small in Table 1, for two reasons. First, the variability due to estimating the v_i is dominated by the variability in the v_i , resulting in weak correlation among the \hat{v}_i in this particular example. Second, a referee

TABLE 1

Estimation of α , the effect of a subject-level covariate, using GEE methods with working independence. The methods are described in Section 3.1 and Appendix A, and compared in terms of empirical bias, standard deviation (SD), median standard error (SE) and coverage probability (CP) of intended 95% confidence intervals. Each entry is based on 1,000 replicates

Mean count	Serial correlation	Include FSE?	Further option	Emp. bias	Emp. SD	Median SE	Emp. CP
1	Short	No	Robust	0.00	0.39	0.33	0.90
		No	Model-based	0.00	0.40	0.03	0.14
		Yes	LS	0.00	0.33	0.32	0.95
		Yes	IRLS	0.01	0.32	0.31	0.95
	Long	No	Robust	0.00	0.42	0.34	0.90
		No	Model-based	-0.01	0.42	0.03	0.11
		Yes	LS	-0.01	0.33	0.33	0.94
		Yes	IRLS	0.00	0.32	0.32	0.95
0.1	Short	No	Robust	0.02	0.41	0.33	0.90
		No	Model-based	0.01	0.40	0.04	0.17
		Yes	LS	-0.01	0.35	0.33	0.94
		Yes	IRLS	-0.01	0.36	0.32	0.93
	Long	No	Robust	-0.01	0.42	0.34	0.89
		No	Model-based	0.00	0.40	0.04	0.15
		Yes	LS	0.01	0.35	0.33	0.94
		Yes	IRLS	0.00	0.33	0.32	0.95

pointed out that any correlation that does exist among the \hat{v}_i should be approximately exchangeable, in which case the LS estimate is still efficient. The precision in estimating α appears insensitive to the length of the serial correlation (specified through γ), although the FSE estimates do seem to become more variable for low counts [$E(Y_{ij}) = 0.1$]. The use of FSE also helps with variance estimation. Without FSE, the model-based variance estimate is clearly disastrous, as expected, and even the robust variance estimate is unsatisfactory, with sub-nominal coverage ($\approx 90\%$). Including FSE in the GEE model leads to reasonable variance estimates and nearly correct coverage, under both (LS and IRLS) approaches. Thus, it seems that the key to valid and efficient inference about α in similar situations is to include FSE in a GEE analysis followed by a subject-level linear regression analysis for the estimated FSE.

Estimation of β is a different story, as shown in the top section of Table 2. As in the case of estimating α , bias is not a major issue in estimating β . The precision in estimating β depends mostly on the length of the serial correlation, with better precision for shorter-lived serial correlation, and seems insensitive to other factors (e.g., FSE, mean count). This is clearly different from the situation in Table 1, and an intuitive explanation is the following. In general, estimating the effect of a subject-level covariate is essentially comparing one group of subjects with another,

TABLE 2

Estimation of β , the effect of a trip-level covariate, using standard GEE methods. The methods are described in Sections 3.1–3.3 and compared in terms of empirical bias, standard deviation (SD), median standard error (SE), coverage probability (CP) of intended 95% confidence intervals and the percentage of estimates that are not available (NA) due to numerical problems such as noninvertible matrices. In each scenario, 1,000 replicates are generated, and the NA estimates are counted and then excluded in calculating the other summary statistics

Mean count	Serial correlation	Include FSE?	Variance estimate	Emp. bias	Emp. SD	Median SE	Emp. CP	%NA
<i>Working independence</i>								
1	Short	No	Robust	0.00	0.12	0.09	0.91	0.0
		No	Model-based	0.00	0.12	0.05	0.67	0.0
		Yes	Robust	-0.01	0.12	0.09	0.90	0.0
		Yes	Model-based	0.00	0.12	0.04	0.47	0.0
	Long	No	Robust	0.01	0.21	0.16	0.92	0.0
		No	Model-based	-0.01	0.21	0.05	0.41	0.0
		Yes	Robust	-0.01	0.21	0.16	0.91	0.0
		Yes	Model-based	-0.01	0.21	0.04	0.28	0.0
0.1	Short	No	Robust	0.00	0.13	0.10	0.90	0.0
		No	Model-based	-0.01	0.13	0.07	0.74	0.0
		Yes	Robust	0.00	0.13	0.10	0.90	0.0
		Yes	Model-based	0.00	0.13	0.06	0.65	0.0
	Long	No	Robust	0.02	0.21	0.17	0.90	0.0
		No	Model-based	0.00	0.21	0.07	0.49	0.0
		Yes	Robust	-0.01	0.21	0.17	0.91	0.0
		Yes	Model-based	-0.01	0.21	0.06	0.42	0.0
<i>Estimated covariance matrix</i>								
1	Short	Yes	Robust	0.00	0.07	0.07	0.95	0.0
		Yes	Model-based	-0.01	0.07	0.07	0.95	0.0
	Long	Yes	Robust	-0.01	0.12	0.12	0.94	0.0
		Yes	Model-based	0.00	0.12	0.11	0.90	0.1
0.1	Short	Yes	Robust	0.00	0.09	0.10	0.92	4.6
		Yes	Model-based	0.00	0.10	0.10	0.95	4.9
	Long	Yes	Robust	0.00	0.15	0.16	0.90	3.5
		Yes	Model-based	-0.01	0.15	0.12	0.87	3.9
<i>Misspecified covariance matrix</i>								
1	Varying	Yes	Robust	-0.02	0.09	0.09	0.93	0.0
		Yes	Model-based	-0.03	0.09	0.08	0.92	0.0
0.1	Varying	Yes	Robust	-0.02	0.11	0.12	0.90	3.7
		Yes	Model-based	-0.02	0.11	0.11	0.93	2.9

while estimating the effect of a trip-level covariate is essentially comparing one group of trips with another group of trips by the same subjects. With $X_{ij} = t_{ij}$, estimation of β is basically a comparison of later trips with earlier trips, and it

seems natural that serial correlation has a larger impact on this comparison than on a comparison of boys with girls, say. The latter comparison, on the other hand, is more likely to benefit from the use of FSE to separate the relevant information (i.e., overall incidence rates of individual drivers) from the noise (i.e., within-subject variation). In Table 2, none of the four methods is satisfactory in terms of coverage, though the robust variance estimate performs better than the model-based one. Several alternatives to the robust variance estimate have also been explored with little success (see Section 6).

3.2. *GEE based on an estimated covariance matrix.* We now consider GEE methods incorporating the covariance matrix for the Y_{ij} . Under the GOUP model described in Section 2, it is straightforward to show, by appealing to well-known properties of normal and Poisson distributions, that

$$(6) \quad \begin{aligned} E(Y_{ij}) &= m_{ij} \exp\{v^* + \alpha' Z_i + \beta' X_{ij} + (\sigma_b^2 + \sigma_c^2 + \sigma_e^2)/2\} =: \mu_{ij}, \\ \text{var}(Y_{ij}) &= \mu_{ij} + \mu_{ij}^2 \{\exp(\sigma_b^2 + \sigma_c^2 + \sigma_e^2) - 1\}, \end{aligned}$$

$$(7) \quad \text{cov}(Y_{ij}, Y_{ij'}) = \mu_{ij} \mu_{ij'} [\exp\{\sigma_b^2 + \sigma_c^2 \exp(-\gamma |t_{ij} - t_{ij'}|)\} - 1],$$

($j \neq j'$).

These expressions provide the marginal covariance matrix relevant in a GEE analysis without FSE. With FSE in the model, we need to condition on b_i and the relevant formulas become

$$(8) \quad E(Y_{ij}|b_i) = m_{ij} \exp(v_i + \beta' X_{ij}) =: \mu_{ij|b_i},$$

$$\text{var}(Y_{ij}|b_i) = \mu_{ij|b_i} + \mu_{ij|b_i}^2 \{\exp(\sigma_c^2 + \sigma_e^2) - 1\},$$

$$(9) \quad \text{cov}(Y_{ij}, Y_{ij'}|b_i) = \mu_{ij|b_i} \mu_{ij'|b_i} [\exp\{\sigma_c^2 \exp(-\gamma |t_{ij} - t_{ij'}|)\} - 1],$$

($j \neq j'$),

with v_i defined in Section 2. Note that σ_b^2 is not involved in the covariance matrix in a GEE analysis with FSE. Unknown parameters in the covariance matrix (σ_c^2 , σ_e^2 , γ and possibly σ_b^2) can be estimated by applying moment methods and nonlinear regression techniques to residuals from a preliminary GEE analysis with working independence (see Appendix B for details). This preliminary GEE analysis can be performed with or without FSE, regardless of the primary GEE analysis for estimating β . With FSE in the preliminary GEE analysis, the aforementioned techniques do not provide an estimate of σ_b^2 . If desired, an estimate of σ_b^2 can be obtained from the IRLS estimation of α or simply as the error variance in the LS analysis, ignoring the fact that \hat{v}_i is an error-prone estimate of v_i (see Section 3.1 and Appendix A for details). Thus, the covariance matrix can be estimated using three methods: FSE-LS, FSE-IRLS and no FSE, with the first two differing only in the estimate of σ_b^2 .

TABLE 3

Estimation of parameters in the covariance matrix $(\sigma_b^2, \sigma_c^2, \sigma_e^2, \gamma)$ using moment and nonlinear regression methods. The methods are described in Appendix B and compared in terms of empirical bias, standard deviation (SD) and the percentage of estimates that are not available (NA) due to convergence failure in nonlinear regression. In each scenario, 1,000 replicates are generated, and the NA estimates are counted and then excluded in calculating the other summary statistics

Serial correlation	Method	Emp. bias				Emp. SD				% NA
		σ_b^2	σ_c^2	σ_e^2	γ	σ_b^2	σ_c^2	σ_e^2	γ	
<i>Mean count = 10</i>										
Short ($\gamma = 300$)	FSE-LS	-0.01	-0.03	-0.02	1.1E+01	0.23	0.10	0.11	1.4E+02	0.1
	FSE-IRLS	-0.02	-0.03	-0.02	1.1E+01	0.23	0.10	0.11	1.4E+02	0.1
	No FSE	-0.26	0.01	-0.06	4.1E+02	0.25	0.29	0.22	5.9E+03	6.8
Long ($\gamma = 50$)	FSE-LS	0.03	-0.10	-0.02	2.0E+01	0.24	0.07	0.08	4.9E+01	0.2
	FSE-IRLS	0.02	-0.10	-0.02	2.0E+01	0.24	0.07	0.08	4.9E+01	0.2
	No FSE	-0.30	0.01	-0.06	1.8E+02	0.26	0.40	0.17	2.1E+03	5.8
<i>Mean count = 1</i>										
Short ($\gamma = 300$)	FSE-LS	-0.01	-0.03	0.00	1.2E+02	0.23	0.21	0.23	2.9E+03	0.3
	FSE-IRLS	-0.02	-0.03	0.00	1.2E+02	0.23	0.21	0.23	2.9E+03	0.3
	No FSE	-0.29	0.01	-0.06	2.6E+02	0.23	0.24	0.21	4.5E+03	8.7
Long ($\gamma = 50$)	FSE-LS	0.03	-0.10	-0.02	3.4E+01	0.25	0.10	0.21	8.5E+01	0.4
	FSE-IRLS	0.01	-0.10	-0.02	3.4E+01	0.24	0.10	0.21	8.5E+01	0.4
	No FSE	-0.30	0.05	-0.08	2.1E+02	0.28	0.97	0.20	1.5E+03	17.5
<i>Mean count = 0.1</i>										
Short ($\gamma = 300$)	FSE-LS	0.28	-0.03	-0.13	6.3E+02	2.97	0.51	1.18	4.1E+03	7.9
	FSE-IRLS	0.25	-0.03	-0.13	6.3E+02	2.97	0.51	1.18	4.1E+03	7.9
	No FSE	-0.28	0.08	-0.12	6.5E+02	0.35	0.39	0.39	3.6E+03	30.4
Long ($\gamma = 50$)	FSE-LS	0.14	-0.04	-0.14	1.0E+03	1.65	0.51	1.17	1.4E+04	12.9
	FSE-IRLS	0.08	-0.04	-0.14	1.0E+03	1.24	0.51	1.17	1.4E+04	12.9
	No FSE	-0.34	0.14	-0.14	2.2E+03	0.33	0.58	0.35	3.3E+04	37.3

Table 3 shows the performance of the above three methods for estimating parameters in the covariance structure, in terms of bias, standard deviation and convergence. The methods are evaluated in the same scenarios as the previous simulation experiments, with the addition of a higher mean count (10) to show a more complete picture. For a mean count of 10, the FSE methods are nearly unbiased for the variance components but biased for γ in a way that results in underestimation of the length of the serial correlation. The bias for γ is larger for longer-lived serial correlation. The no FSE method is more biased for γ (in the same direction) and even visibly biased for σ_b^2 . With decreasing counts, the bias problem becomes more severe, especially for γ , to the extent that estimation of the serial correlation is meaningless for a mean count of 0.1. The problem is compounded by a large amount of variability in estimating γ . Another issue with low counts is conver-

gence in the nonlinear regression. For a mean count of 0.1, the convergence failure rates are approximately 8% and 13% (with short- and long-lived serial correlation, resp.) for the FSE methods and even worse for the no FSE method. The FSE methods appear more reasonable than the no FSE method, and we will use FSE-LS, which is simpler than FSE-IRLS, in the subsequent simulations.

GEE methods based on covariance matrices estimated through FSE-LS are evaluated in the same setting described in Section 3.1. When the nonlinear regression in FSE-LS fails to converge, the data-driven initial values will be taken as the final estimates [see equation (15) in Appendix B]. With thousands of trips per subject, inverting a covariance matrix is time-consuming, and iterating until convergence in the usual Fisher scoring algorithm [Fitzmaurice et al. (2008), Section 3.2.4] would be prohibitive. We therefore settle for a one-step estimate of the regression coefficients based on just one iteration in the Fisher scoring algorithm. Specifically, we start with a GEE analysis with working independence, use the residuals to estimate parameters in the covariance matrix, and update the regression coefficients just once using the estimated covariance matrix. Under this approach, it is necessary to include FSE in each GEE analysis; otherwise meaningful simulation results cannot be produced. This is not surprising because inversion of large covariance matrices can be difficult when the correlation is strong (without FSE) and poorly estimated. This is not a serious limitation because the use of FSE leads to better efficiency and coverage anyway, which we have observed in simulations where the true covariance matrix is used in GEE (results not shown). Once FSE are included, numerical problems become much less frequent (3–5% for a mean count of 0.1, $\leq 1\%$ for higher counts).

We have found that incorporating the covariance matrix into the estimating equation does little to improve the estimation of α in terms of efficiency and coverage. We therefore omit the results for estimating α and henceforth focus on estimating β . The results for the latter, reported in the middle section of Table 2, show that better efficiency is attained using an estimated covariance matrix than under working independence. In terms of coverage, the robust variance estimate performs well for higher counts but not for lower ones, especially when the serial correlation is long-lived. The model-based variance estimate performs well for short-lived serial correlation and poorly for long-lived serial correlation. The former case shows that underestimating the length of the serial correlation has little effect when the serial correlation is already negligibly short. Although not shown in Table 2, the model-based variance estimate has been found to work well in all cases if the true covariance matrix is used. Thus, the root of the problem appears to be the difficulty in estimating γ .

3.3. *GEE based on a misspecified covariance matrix.* It is natural to ask how the GEE methods based on the Goup model would perform if the model is misspecified. For the NTDS, the Goup model seems plausible for the most part. It is

possible, however, that the length of the serial correlation, which is assumed constant, might change over time. To formalize this notion of changing length, note that the covariance structure given by (3) can be generalized into

$$(10) \quad \text{cov}\{c_i(t_1), c_i(t_2)\} = \sigma_c^2 \exp\left(-\int_{t_1}^{t_2} \gamma(t) dt\right), \quad t_1 < t_2,$$

for a function $\gamma : [0, 1] \rightarrow (0, \infty)$. For a constant γ , this reduces to the original GOUP model. In the next set of simulation experiments, we will generate data using the generalized GOUP model with γ taken to be a straight line from $\gamma(0) = 300$ to $\gamma(1) = 50$. Thus, instead of working exclusively with short- or long-lived serial correlation, we allow it to be short-lived at the beginning and to gradually become long-lived at the end. This is motivated by the conjecture that teenage driving will tend to be haphazard at the beginning and will become more consistent over time. Unfortunately, it appears difficult to verify this conjecture directly using the data. Indeed, even when γ is a constant, we have seen in Table 3 that it can be really difficult to estimate, especially when the counts are low. It seems reasonable to expect that a time-varying γ will be even more difficult to estimate. We therefore take a sensitivity analysis approach to this issue. The bottom section of Table 2 shows the performance in the present setting of GEE methods based on the original GOUP model. The point estimates are more (or less) efficient in this setting than if the serial correlation is consistently long-lived (or short-lived). The coverage property of the model-based variance estimate is also intermediate between the two extremes. As before, the robust variance estimate works well for large counts but not for small ones.

3.4. *Summary.* To summarize, valid inference on α can be made through a GEE analysis with FSE (regardless of the covariance matrix) followed by a linear regression analysis, while inference on β requires more work. For large counts, valid inference on β can be made by estimating the covariance matrix and using the robust variance estimate; in this case, the inference seems robust with respect to the serial correlation structure. For small counts, the robust variance estimate is unsatisfactory and the model-based variance estimate performs well only for short-lived serial correlation. Because the serial correlation is difficult to estimate with small counts, one would not know in practice whether the model-based variance estimate is appropriate for a particular application. Therefore, it is necessary to develop methods that perform well (or better) for small counts with short- and long-lived serial correlation. This is the focus of the next section, where we explore WCR-type methods.

4. WCR and extensions.

4.1. *The standard WCR approach.* WCR is a resampling approach for marginal analysis of clustered data, originally designed to deal with informative

cluster sizes [Hoffman, Sen and Weinberg (2001), Follmann, Proschan and Leifer (2003)]. In its original form, the approach can be applied to the present problem as follows. Suppose we sample a trip from each subject randomly and form a subset of nonclustered data

$$(11) \quad \{(Z_{iJ_i}, X_{iJ_i}, Y_{iJ_i}), i = 1, \dots, n\},$$

where J_i is uniformly distributed on $\{1, \dots, k_i\}$, independently across i . There is no correlation in the subsample (11), which can therefore be analyzed using a quasi-Poisson model with mean structure given by (1). This can be repeated a large number (say, L) of times to yield $\{(\hat{\theta}_l, \hat{\Sigma}_l), l = 1, \dots, L\}$, where $\hat{\theta}_l$ denotes the point estimate of $\theta = (\alpha, \beta)$ based on the l th subsample, and $\hat{\Sigma}_l$ is the associated variance estimate. Then the WCR estimate of θ is given by

$$(12) \quad \hat{\theta}_{\text{WCR}} = \frac{1}{L} \sum_{l=1}^L \hat{\theta}_l,$$

and the associated variance estimate is

$$(13) \quad \hat{\Sigma}_{\text{WCR}} = \frac{1}{L} \sum_{l=1}^L \hat{\Sigma}_l - \frac{1}{L-1} \sum_{l=1}^L (\hat{\theta}_l - \hat{\theta}_{\text{WCR}})(\hat{\theta}_l - \hat{\theta}_{\text{WCR}})'$$

The first term on the right-hand side of (13) measures the variability within subsamples, while the second term measures the variability between subsamples.

In general, for any WCR method to work, it is essential that valid estimates $(\hat{\theta}_l, \hat{\Sigma}_l)$ be obtained from each subsample. Note that the estimates $(\hat{\theta}_l, \hat{\Sigma}_l)$ are ideally distributed. If $\hat{\theta}_l$ is biased for θ , then $\hat{\theta}_{\text{WCR}}$ given by (12) is also biased by the same amount. Similarly, if $\hat{\Sigma}_l$ underestimates the sampling variance of $\hat{\theta}_l$, then $\hat{\Sigma}_{\text{WCR}}$ given by (13) will underestimate the sampling variance of $\hat{\theta}_{\text{WCR}}$ by a similar amount, because the second term on the right-hand side of (13) will be approximately unbiased for large L . Furthermore, in terms of relative bias, the underestimation problem of $\hat{\Sigma}_{\text{WCR}}$ is generally worse than that of $\hat{\Sigma}_l$, simply because the true variance of $\hat{\theta}_{\text{WCR}}$ is smaller than that of $\hat{\theta}_l$. In light of these observations, we will study WCR methods in two steps, always considering a single subsample before moving to repeated sampling.

4.2. *WCR with simple random sampling (WCR-SRS).* There is no reason why a WCR subsample must consist of exactly one observation from each subject. In fact, to fit a model with FSE, there has to be more than one observation from each subject. Without FSE, the standard WCR approach is feasible but, as we shall see, does not work well. We therefore consider an extended WCR approach where a simple random sample (SRS) is taken from each subject to form an WCR subsample. Thus, instead of a single number J_i , we take an SRS $\{J_i^{(1)}, \dots, J_i^{(R)}\}$ from the index set $\{1, \dots, k_i\}$ for each i . The resulting subsample is

$$\{(Z_{iJ_i^{(r)}}, X_{iJ_i^{(r)}}, Y_{iJ_i^{(r)}}), i = 1, \dots, n, r = 1, \dots, R\},$$

for which a standard GEE analysis can be performed, under working independence or using an estimated covariance matrix, with or without FSE, using the robust or model-based variance estimate. This can be repeated a large number of times, and the resulting estimates can be combined using (12) and (13). If the covariance matrix for the Y_{ij} is used, it will be estimated once from the full sample and the estimate will be applied to all subsamples.

This WCR-SRS approach is evaluated in Table 4. Given the findings in the last section, this evaluation is focused on small counts. Under working independence, only the robust variance estimate is considered because the model-based variance estimate is clearly invalid. As before, no meaningful results are available when an

TABLE 4

Estimation of β , the effect of a trip-level covariate, using the WCR-SRS approach with different values of R (number of trips to sample from each subject) and L (number of repetitions in WCR).

The methods are described in Sections 4.1 and 4.2 and compared in terms of empirical bias, standard deviation (SD), median standard error (SE), coverage probability (CP) of intended 95% confidence intervals and the percentage of estimates that are not available (NA). The mean count is fixed at 0.1. In each scenario, 1,000 replicates are generated, and the NA estimates are counted and then excluded in calculating the other summary statistics

Serial correlation	Include FSE?	Variance estimate	R	L	Emp. bias	Emp. SD	Median SE	Emp. CP	%NA
<i>Working independence</i>									
Short	No	Robust	1	1	-15.40	357.22	1.59	0.64	0.4
			5	1	-0.02	1.44	0.99	0.86	0.0
			25	1	0.03	0.64	0.52	0.91	0.0
			100	1	-0.02	0.37	0.28	0.91	0.0
			500	1	-0.01	0.20	0.15	0.92	0.0
	Yes	Robust	100	500	0.00	0.13	0.00	0.16	0.0
			5	1	0.09	1.59	1.16	0.87	0.2
			25	1	0.01	0.68	0.50	0.87	0.1
			100	1	-0.02	0.37	0.28	0.90	0.1
			500	1	0.00	0.18	0.14	0.91	0.0
100	500	0.00	0.12	0.00	0.18	0.0			
Long	No	Robust	1	1	32.12	2011.24	1.64	0.64	0.8
			5	1	-0.08	1.43	1.00	0.89	0.0
			25	1	-0.01	0.72	0.53	0.89	0.0
			100	1	0.01	0.42	0.31	0.92	0.0
			500	1	0.00	0.24	0.20	0.91	0.0
	Yes	Robust	100	500	-0.01	0.21	0.10	0.52	0.0
			5	1	-0.62	19.13	1.13	0.85	0.1
			25	1	-0.02	0.71	0.52	0.89	0.1
			100	1	0.00	0.39	0.30	0.90	0.0
			500	1	-0.01	0.24	0.20	0.90	0.0
100	500	-0.01	0.21	0.10	0.56	0.0			

TABLE 4
(Continued)

Serial correlation	Include FSE?	Variance estimate	R	L	Emp. bias	Emp. SD	Median SE	Emp. CP	%NA	
<i>Estimated covariance matrix</i>										
Short	Yes	Robust	5	1	0.03	1.44	0.89	0.75	0.4	
			25	1	0.02	0.59	0.48	0.85	1.4	
			100	1	0.01	0.30	0.29	0.91	3.3	
		500	1	-0.01	0.15	0.15	0.92	4.6		
		100	500	0.00	0.09	0.21	0.70	2.6		
		Model-based	5	1	-0.09	1.49	1.19	0.91	0.2	
	25	1	0.02	0.58	0.55	0.94	1.1			
	100	1	0.00	0.29	0.30	0.94	2.2			
	500	1	0.00	0.15	0.15	0.97	4.1			
	100	500	0.01	0.10	0.12	0.68	1.4			
	Long	Yes	Robust	5	1	-0.04	1.48	0.93	0.76	0.4
				25	1	0.00	0.61	0.51	0.87	1.2
100				1	0.00	0.31	0.30	0.90	2.3	
500			1	-0.01	0.18	0.19	0.90	3.1		
100			500	0.00	0.15	0.24	0.69	2.8		
Model-based			5	1	-0.02	1.67	1.22	0.92	0.4	
25		1	0.01	0.59	0.54	0.93	1.9			
100		1	-0.01	0.32	0.31	0.94	2.4			
500		1	0.00	0.18	0.17	0.92	3.2			
100		500	0.00	0.16	0.16	0.72	2.6			

estimated covariance matrix is used without FSE. For each method, we begin by analyzing a single subsample, with $L = 1$ and different values of R (size of the SRS from each subject). Table 4 shows poor performance for $R = 1$ (i.e., standard WCR), apparently due to erratic estimates, which are likely to arise when the subsample is very small. Even for $R = 5$, the point estimate is visibly biased in some cases, probably because the subsample is still small. The coverage probability generally increases with R , at least for $R \leq 100$. The best coverage is seen when an estimated covariance matrix is used together with FSE and the model-based variance estimate. As explained in Section 3.2, this method tends to work well when the serial correlation is weak, which is precisely what happens to a randomly selected subset of trips, which tend to be farther apart from each other than in the full sample. This explains why better coverage is seen here than in the corresponding portion of Table 2. Of course, this trend will be reversed when the number of selected trips gets too large, which explains the slight drop in coverage probability when R increases further to 500. Overall, it appears that $R = 100$ would be a reasonable choice in terms of bias and coverage. Efficiency is not a major consideration here because the efficiency loss due to $R \leq k_i$ can be recovered through

repeated sampling. Table 4 also shows the performance of the WCR-SRS method with $L = 500$ and $R = 100$, whose efficiency levels appear similar to the corresponding results in Table 2, as expected. For the working independence methods, the WCR variance estimate is seriously biased downward and the coverage probability is far below the nominal level, confirming the conjecture in Section 4.1 that the variance underestimation problem of $\widehat{\Sigma}_{\text{WCR}}$ is generally worse than that of $\widehat{\Sigma}_l$ in terms of relative bias. For methods based on an estimated covariance matrix, the WCR variance estimate is not necessarily biased downward but the coverage properties are not good. This behavior is probably due to erratic estimates which arise frequently in repeated analyses of subsamples when a poorly estimated covariance matrix is used.

Table 4 does suggest a potential benefit of WCR, that is, the ability to produce a sparse subset of trips with weaker serial correlation. This advantage can be exploited further by selecting trips that are separated from each other by a specified amount. Specifically, for a given positive integer S , we can choose a trip randomly among the first $S + 1$ trips and every $(S + 1)$ st trip from there on, until the end of the sequence. This can be done for each subject and the process can be repeated many times as before. This approach does lead to good coverage in single outputation when an estimated covariance matrix is used together with FSE and the model-based variance estimate (results not shown). However, upon moving to WCR, this approach exhibits the same strange behavior as seen in Table 4 (results not shown), probably for the same reason.

4.3. *WCR with separated blocks (WCR-SB)*. The previous experiments with WCR show that variance underestimation can be a serious problem, especially when the variance of $\widehat{\theta}_{\text{WCR}}$ is much smaller than that of $\widehat{\theta}_l$. The latter difference can be reduced by including more observations in each subsample, and the question then becomes how to ensure reasonable performance of the variance estimate based on a subsample. To this end, we propose the following WCR approach involving separated blocks. With FSE in the model, the only source of correlation is serial correlation, which is weaker for trips farther apart. Thus, with an appropriate amount of separation, trips from the same subject can be treated as approximately independent. Let B denote the block size (i.e., the number of consecutive trips to be analyzed together as a block) and S the separation (i.e., the number of trips used to separate the blocks). For a given subject, one possible set of separated blocks can be formed by taking the first B trips, skipping the next S trips, taking the next B , skipping the next S , and so on. Each group of B trips sampled together will be treated as a block and the different blocks will be treated as independent in the subsequent GEE analysis. A random shift can be added to this sampling process, and the random sampling process can be repeated many times as before.

Preliminary simulation results, as well as those in Table 4, suggest that a reasonable block size would be $B = 100$. Simple calculations based on (8) and (9) show that, with $S = 50$, the correlation between trips in different blocks is typically

TABLE 5

Estimation of β , the effect of a trip-level covariate, using the WCR-SB approach with working independence and the robust variance estimate, implemented with $B = 100$ (block size), $S = 50$ (separation) and $L = 1$ (single subsample) or 50. The method is evaluated in terms of empirical bias, standard deviation (SD), median standard error (SE), coverage probability (CP) of intended 95% confidence intervals and the percentage of estimates that are not available (NA). The mean count is fixed at 0.1. In each scenario, 1,000 replicates are generated, and the NA estimates are counted and then excluded in calculating the other summary statistics

Serial correlation	L	Emp. bias	Emp. SD	Median SE	Emp. CP	%NA
Short	1	0.00	0.15	0.13	0.94	0.0
	50	0.00	0.12	0.10	0.92	0.0
Long	1	0.00	0.22	0.19	0.92	0.0
	50	0.01	0.20	0.17	0.92	0.0
Varying	1	0.00	0.18	0.15	0.92	0.1
	50	0.00	0.15	0.13	0.92	0.0

well below 0.05 in all scenarios considered in this paper. With these choices for B and S , the WCR-SB approach is evaluated through simulations and the results are presented in Table 5. We focus on working independence to avoid numerical stability issues, and only consider the robust variance estimate. The method appears to have better coverage when applied to a single set of separated blocks than to the full sample (Table 2), apparently owing to a larger number of approximately independent clusters (400 vs. 40). Note that the efficiency based on a single outputation is quite close to that in a full sample analysis. Considering this and the computational demand, the WCR version is implemented with $L = 50$ repetitions, and the resulting WCR-SB estimates are virtually as efficient as the full sample estimates. In the case of long-lived serial correlation, better coverage (92%) is achieved here using the working independence method than in any of the previous simulations. The improvement (2% over Table 2) is not dramatic but still substantial, and similar phenomena have been observed consistently in other simulation experiments under similar scenarios. Table 5 also shows the performance of the WCR-SB method when the serial correlation is generated according to (10) rather than (3). The only noticeable effect of this change is an intermediate level of efficiency (between the extreme cases of short- and long-lived serial correlation).

5. Analysis of NTDS data. Our analysis of the NTDS data concerns the following elevated g -force events: rapid start (longitudinal acceleration $\geq 0.35g$), hard stop (longitudinal deceleration $\geq 0.40g$), hard left/right turn (lateral deceleration/acceleration $\geq 0.45g$), yaw (≥ 5 degrees within 3 seconds), as well as a composite measure defined as the totality of these 5 types of events. Yaw is a measure of correction after a turn and is calculated as the absolute change in angle be-

tween an initial turn and the correction. For each specific g -force event, the threshold (in parenthesis) is chosen to allow meaningful differences between drivers and different driving conditions to be revealed and estimated with a sufficient number of events. The mean counts based on the chosen thresholds are 0.13 (rapid start), 0.18 (hard stop), 0.20 (hard left turn), 0.15 (hard right turn), 0.04 (yaw) and 0.70 (composite measure). The mean count for yaw is relatively low, perhaps too low for the methods discussed in this paper, but the corresponding threshold is already the lowest for which we have data available. This limitation should be kept in mind when interpreting the results.

From the viewpoint of behavior science, it is reasonable to expect some serial correlation because risky driving may be related to the weather and the driver's mental and physical conditions (e.g., mood and fatigue). All of these factors could result in serial correlation unless they are all included in the model, which is impossible in the NTDS due to the lack of this information. To explore the nature of the serial correlation in the NTDS data, a marginal model with FSE and all relevant covariates (to be described later) is fit for the composite measure under working independence. Pairwise products of standardized residuals are calculated for consecutive trips (because the number of all possible pairs is too large). After sorting by gap time, these pairs are grouped into 100 bins for further reduction. Within each bin, we calculate the median gap time and the mean product of standardized residuals, and the results are plotted in Figure 2 together with a lowess smooth. Figure 2 suggests that some serial correlation is present in the data, although a precise characterization of the serial correlation seems difficult. Under the GOUP model and using the techniques described in Appendix B, the estimated variance components are generally close to 1 (for the composite measure and other g -force events), while the estimate of γ varies wildly and depends heavily on the initial value.

Figure 3 presents the estimated incidence rate (IR) per 100 miles of each g -force event as a function of time since licensure. The IR is defined by (1) with m_{ij} replaced by 100, Z_i empty, and X_{ij} consisting of indicators of calendar month and a set of regression splines for t_{ij} (one knot for every 3 months). Calendar month is included in the model to adjust for a possible seasonal effect, which may confound with the effect of time since licensure because the enrollment of subjects was not uniform with respect to calendar month, with more subjects enrolled in the fall and fewer in the spring. The estimates shown in Figure 3 represent geometric means of the monthly estimates. Note that the IR considered here is a marginal quantity which involves the marginal intercept ν and which should therefore be estimated in a model without FSE. The estimates in Figure 3 are obtained under working independence, because the GOUP model without FSE is problematic to fit and its performance is not well understood (see Section 3.2). Under the regression spline model, there is some ambiguity as to whether the IR for the composite measure changes over time. The p -value for the regression splines is 0.49 based on the robust variance estimate under working independence, although the validity of this

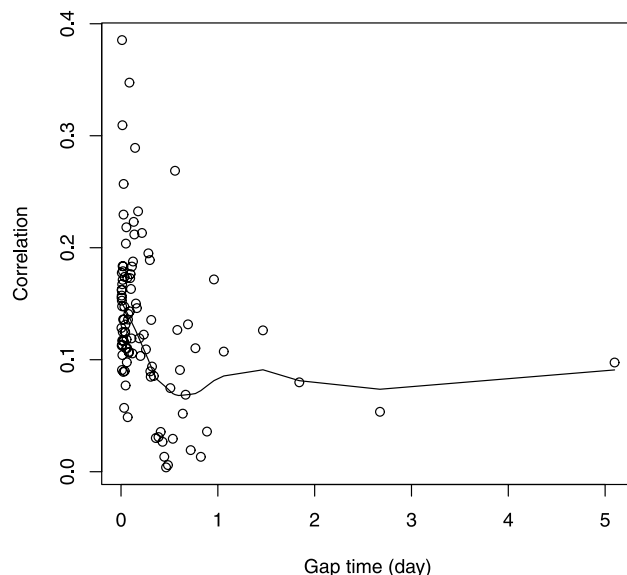


FIG. 2. *Serial correlation for the composite measure in the NTDS. Pairwise products of standardized residuals are calculated for consecutive trips and, after sorting by gap time, grouped into 100 bins. Each circle represents the median gap time and the mean product of standardized residuals in a bin, and the solid line is a lowess smooth.*

test may be questionable (see Section 3). Under the GOUP model with FSE, the IR for the composite measure does appear to change over time ($p = 0.0014$ for the robust variance estimate, $p < 0.0001$ for the model-based variance estimate). The latter observation is in contrast to a previous report that risky driving largely remains constant [Simons-Morton et al. (2011a)], and the difference is probably due to the lower thresholds used here, which lead to more events and perhaps more relevant information. Despite the ambiguity about statistical significance, the main conclusion from our marginal analysis is fairly consistent with the findings of Simons-Morton et al. (2011a). The latter reference reports IRs for adults/parents that are consistently and substantially lower than those for novice teenagers driving the same vehicles. In Figure 3, the estimated IR of the composite measure for teenage drivers increases over the first 6 months, declines over the next few months, and then increases again, suggesting the establishment of a risky driving style. This general impression is reinforced by the plots for the specific g -force events, which show some temporal changes but no clear trends.

As noted by a referee, some instances of risky driving (e.g., mistakes due to the lack of skills) may depend more on the amount of practice, which can be quantified by accumulated mileage, than on time since licensure. This raises the question of which measure is more appropriate to adjust for. Accumulated mileage is a measure of driving experience, while time since licensure reflects both experience and maturity. From a public health perspective, we believe that time since licensure

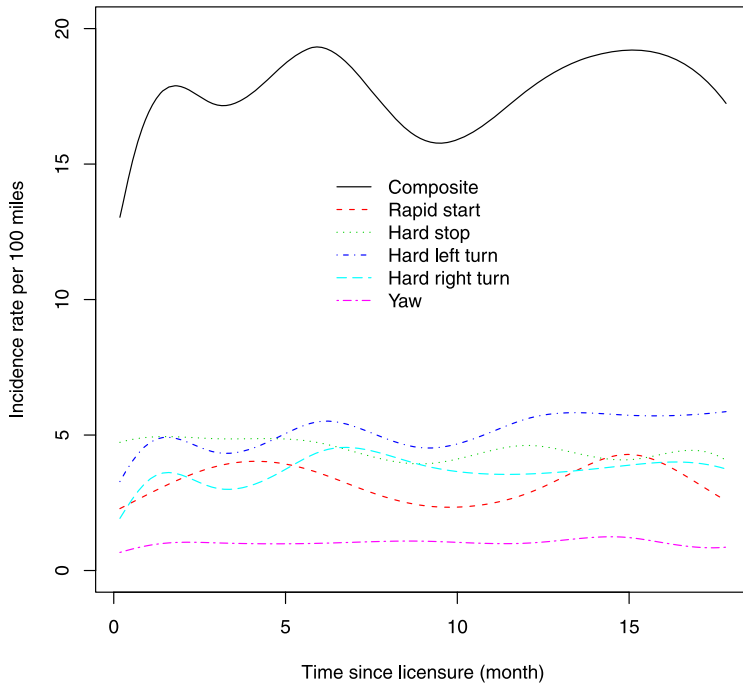


FIG. 3. Incidence rates of various g -force events (rapid start, hard stop, hard left/right turn, yaw, and the composite measure) as functions of time since licensure, estimated using regression splines (one node for every 3 months).

is more relevant to policy makers. Nonetheless, all analyses in this section have been repeated with time since licensure replaced by accumulated mileage, and the results are very similar and therefore omitted. With regard to Figure 3, this similarity suggests that persistent risky driving among teenage drivers is not only due to inexperience but also to immaturity (i.e., aspects of adolescent development that motivate novice young drivers to seek excitement and under-recognize risks).

Each g -force event is further analyzed in a larger model that simultaneously adjusts for the following covariates: driver's gender, risky friends, time of day, passenger condition, calendar month, and time since licensure. The risky behavior of a teenage driver's friends is assessed using a 7-item index asking "How many of your friends would you estimate... smoke cigarettes, drink alcohol, get drunk at least once a week, use marijuana, drive after having two or more drinks in the previous hour, exceed speed limits, and do not use safety belts (none, a few, some, most, all)." The assessment was made at 4 time points (baseline, 6, 12 and 18 months), the 4 scores were averaged for each driver, and the average score was then dichotomized according to the median split among all drivers in the study. Night driving was determined from video data by visual observation of the ambient natural lighting at the start of the trip. Late night was defined as 10 pm to 6 am,

and night trips that did not start at late night were considered early night trips. The presence and relative age (adult or teen) of each passenger was determined by the coders examining the video data. In summary, gender and risky friends enter the marginal model as subject-level binary covariates, time of day is a trip-level covariate with 3 levels (day, early night and late night), passenger condition is also a trip-level covariate with 3 levels (none, teens, at least one adult), calendar month is a trip-level covariate with 12 levels, and time since licensure is represented by the same set of regression splines as before. No signs of model misspecification are observed in residual plots (not shown).

Table 6 presents estimates of incidence rate ratios (IRR) that quantify the association of each g -force event with the aforementioned covariates (except calendar month and time since licensure). An IRR is just the result of exponentiating the corresponding regression coefficient in the marginal model. The estimates are obtained using standard GEE methods (working independence or the GOUP model, robust or model-based variance estimate) as well as the proposed WCR-SB method (working independence, $B = 100$, $S = 50$, $L = 50$). All of these methods include FSE. In a sensitivity analysis, a different separation size ($S = 100$) for the WCR-SB method is used to produce similar results, which are therefore omitted. Since the point estimates depend primarily on the working correlation structure, only 2 sets of point estimates are shown in Table 6. For the subject-level covariates, the LS method is used to obtain 2 sets of 95% confidence intervals (for working independence and the GOUP model). For the trip-level covariates, Table 6 presents 4 sets of 95% confidence intervals corresponding to the robust variance estimate and the WCR-SB method under working independence as well as the robust and model-based variance estimates under the GOUP model. Table 6 contains some unrealistically large values (>100 , labeled as ∞) for the IRRs associating risky friends and gender with yaw and rapid start, probably due to very few events in the subgroup of teenage drivers acting as the denominator.

Despite some numerical differences between the different methods, the results in Table 6 indicate clearly that teenage risky driving is not associated with gender, positively associated with risky friends, negatively associated with early night, and negatively associated with the presence of passengers (more strongly for adults than for teens). The evidence is not conclusive with regard to late night, whose association with risky driving is significant in some cases but not in others. The association of risky friends with risky driving points to potentially destructive effects of risk-accepting social norms and social identity on risky behavior. Both the perception and the actuality of having risk-taking friends could contribute to the perceived acceptability of risky behavior. The association of early night with risky driving suggests that teens do recognize the danger of night driving (at least partially) and drive more carefully (or rather, less carelessly) at night. There is no contradiction between this finding and the apparent ambiguity about late night because late night trips often take place under unusual circumstances. Finally, it is worth noting that passengers, especially adult passengers, appear protective with

TABLE 6

Analysis of NTDS data using standard GEE methods (working independence or the GOUN model, robust or model-based variance estimate) and the proposed WCR-SB method (working independence, $B = 100, S = 50, L = 50$). Each outcome variable (rapid start, hard stop, hard left/right turn, and a composite measure) is analyzed separately in a marginal model that simultaneously adjusts for gender, risky friends, time of day, passenger condition, calendar month and time since licensure (the last two as confounders). The results are summarized in terms of incidence rate ratios (IRRs) and 95% confidence intervals. The symbol ∞ denotes a value that is unrealistically large (> 100)

Covariate	Comparison	GOUN model			Working independence		
		IRR	Model-based	Robust	IRR	Robust	WCR-SB
<i>Composite measure (mean count = 0.70)</i>							
Gender	Male vs. female	1.16	(0.67, 1.99)		1.07	(0.64, 1.79)	
Risky friends	More vs. fewer	1.91	(1.11, 3.29)		1.82	(1.08, 3.05)	
Time of day	Early night vs. day	0.64	(0.61, 0.68)	(0.59, 0.71)	0.73	(0.64, 0.83)	(0.69, 0.78)
	Late night vs. day	0.76	(0.69, 0.84)	(0.66, 0.88)	0.89	(0.68, 1.16)	(0.80, 1.00)
Passengers	Teen vs. none	0.81	(0.78, 0.85)	(0.74, 0.90)	0.78	(0.72, 0.86)	(0.75, 0.83)
	Adult vs. none	0.39	(0.34, 0.45)	(0.34, 0.46)	0.32	(0.26, 0.40)	(0.28, 0.40)
<i>Rapid start (mean count = 0.13)</i>							
Gender	Male vs. female	0.12	(0.00, 5.69)		0.12	(0.00, 5.86)	
Risky friends	More vs. fewer	12.33	(0.26, ∞)		11.50	(0.26, ∞)	
Time of day	Early night vs. day	0.89	(0.83, 0.95)	(0.70, 1.14)	0.89	(0.74, 1.07)	(0.83, 0.95)
	Late night vs. day	0.89	(0.79, 1.00)	(0.68, 1.15)	0.93	(0.77, 1.12)	(0.78, 1.03)
Passengers	Teen vs. none	0.83	(0.78, 0.88)	(0.68, 1.01)	0.81	(0.70, 0.92)	(0.75, 0.87)
	Adult vs. none	0.32	(0.25, 0.41)	(0.21, 0.50)	0.34	(0.26, 0.44)	(0.26, 0.44)
<i>Hard stop (mean count = 0.18)</i>							
Gender	Male vs. female	0.78	(0.44, 1.40)		0.74	(0.42, 1.30)	
Risky friends	More vs. fewer	1.43	(0.80, 2.56)		1.35	(0.77, 2.38)	
Time of day	Early night vs. day	0.68	(0.63, 0.73)	(0.60, 0.77)	0.70	(0.62, 0.80)	(0.65, 0.75)
	Late night vs. day	0.71	(0.61, 0.82)	(0.55, 0.91)	0.70	(0.56, 0.87)	(0.61, 0.80)
Passengers	Teen vs. none	0.90	(0.84, 0.96)	(0.82, 1.00)	0.88	(0.79, 0.97)	(0.81, 0.95)
	Adult vs. none	0.51	(0.42, 0.63)	(0.41, 0.64)	0.38	(0.29, 0.49)	(0.32, 0.47)

respect to risky driving. Further research is warranted to confirm, better characterize and utilize such protective effects.

6. Discussion. Even with a great variety of methods available for longitudinal data analysis, it can still be difficult to analyze longitudinal data in long sequences, especially when the serial correlation is strong and long-lived. In this paper, we examine standard GEE methods and propose new ones for marginal analysis of longitudinal count data in a small number of very long sequences. The methods are evaluated and compared in simulation experiments mimicking the NTDS, and the main findings can be summarized as follows. We consider the use of FSE in

TABLE 6
(Continued)

Covariate	Comparison	GOUP model			Working independence		
		IRR	Model-based	Robust	IRR	Robust	WCR-SB
<i>Hard left turn (mean count = 0.20)</i>							
Gender	Male vs. female	1.36	(0.72, 2.55)		1.29	(0.69, 2.38)	
Risky friends	More vs. fewer	2.41	(1.28, 4.53)		2.20	(1.19, 4.07)	
Time of day	Early night vs. day	0.53	(0.49, 0.58)	(0.45, 0.63)	0.64	(0.52, 0.79)	(0.58, 0.71)
	Late night vs. day	0.78	(0.67, 0.92)	(0.63, 0.97)	1.00	(0.68, 1.46)	(0.83, 1.18)
Passengers	Teen vs. none	0.81	(0.75, 0.88)	(0.70, 0.94)	0.76	(0.68, 0.84)	(0.70, 0.82)
	Adult vs. none	0.38	(0.30, 0.49)	(0.28, 0.53)	0.32	(0.24, 0.43)	(0.25, 0.40)
<i>Hard right turn (mean count = 0.15)</i>							
Gender	Male vs. female	1.39	(0.72, 2.69)		1.31	(0.69, 2.49)	
Risky friends	More vs. fewer	2.04	(1.05, 3.95)		1.92	(1.01, 3.65)	
Time of day	Early night vs. day	0.63	(0.58, 0.67)	(0.50, 0.79)	0.69	(0.55, 0.85)	(0.62, 0.76)
	Late night vs. day	0.83	(0.74, 0.93)	(0.46, 1.48)	0.92	(0.58, 1.46)	(0.75, 1.14)
Passengers	Teen vs. none	0.64	(0.60, 0.69)	(0.54, 0.76)	0.62	(0.55, 0.70)	(0.56, 0.69)
	Adult vs. none	0.26	(0.20, 0.33)	(0.18, 0.37)	0.22	(0.16, 0.31)	(0.17, 0.30)
<i>Yaw (mean count = 0.04)</i>							
Gender	Male vs. female	0.07	(0.00, ∞)		0.07	(0.00, ∞)	
Risky friends	More vs. fewer	∞	(0.15, ∞)		∞	(0.17, ∞)	
Time of day	Early night vs. day	0.92	(0.76, 1.10)	(0.65, 1.29)	0.94	(0.77, 1.15)	(0.80, 1.11)
	Late night vs. day	0.88	(0.63, 1.23)	(0.42, 1.86)	0.92	(0.59, 1.45)	(0.71, 1.14)
Passengers	Teen vs. none	1.19	(1.02, 1.39)	(0.90, 1.57)	1.17	(1.00, 1.37)	(1.02, 1.33)
	Adult vs. none	0.54	(0.32, 0.92)	(0.21, 1.39)	0.44	(0.27, 0.73)	(0.27, 0.68)

this particular situation, a simple technique with important practical implications. It allows the effects of subject-level covariates to be estimated easily from a linear regression analysis for the estimated FSE, and it also helps with trip-level covariates by removing the correlation due to population heterogeneity. For trip-level covariates, we find that a standard GEE analysis under working independence can lead to inefficient estimates and serious undercoverage, and that both problems can be alleviated by incorporating a properly specified correlation structure. The latter approach works well for large counts but not for small counts, mainly because the serial correlation is hard to estimate with small counts. We therefore explore an alternative approach (WCR) for the case of small counts. The original version of WCR and an extension to simple random sampling seem unsatisfactory, however, because of numerical instability in repeated analyses of small samples and a bias magnification effect of WCR that results in variance underestimation. To address these issues, we propose an WCR-SB approach that involves separated blocks and that performs better than all of the previously considered methods (WCR and standard GEE).

In Table 6 the WCR-SB analyses are not dramatically different from the other analyses. This is reassuring to the scientists, but it also suggests that the NTDS data set is not ideal for demonstrating the advantage of the WCR-SB approach. To put the latter point in perspective, we note that better performance in the frequentist sense does not imply better results in every realization. The performance of the WCR-SB approach has been evaluated in simulation experiments which are, in fact, designed to mimic the NTDS. Most of the uncertainty in designing the simulation experiments is associated with the length of the serial correlation, which is very difficult to estimate with small counts, as shown in Table 3. It is certainly possible that the serial correlation in the NTDS is not sufficiently long-lived for the new method to make a material difference. In that case, a better example than the NTDS might be a proposed follow-up study that involves 100 teenage drivers to be followed for at least 3 years. The larger amount of data from the latter study will afford a better understanding of both the nature and the length of the serial correlation. Furthermore, as teenage drivers gain experience and perspective over time, their driving behavior may become less haphazard and more stable, in which case the serial correlation will gradually become longer-lived toward the end of the (longer) study duration. Thus, in terms of the length of the serial correlation, the follow-up study will provide a better opportunity than the NTDS to demonstrate the advantage of the WCR-SB approach.

The WCR-SB approach is designed for longitudinal data in a small number of long sequences. The strength of this approach relative to standard GEE methods is the ability to handle strong and long-lived serial correlation when the mean count is low. The weaknesses of the WCR-SB approach include increased computational demand (relative to standard GEE methods) and the need for information about the length of the serial correlation (in order to specify the separation size S). Because γ can be very difficult to estimate, one may need to consult the subject-matter scientist or perform a sensitivity analysis with different values of S , as we did in analyzing the NTDS data (see Section 5).

As mentioned earlier at the end of Section 3.1, we have actually explored several existing alternatives to the robust variance estimate in the present situation. No appreciable improvement has been achieved using a standard bootstrap procedure (i.e., sampling subjects with replacement) and jackknife methods [Paik (1988), Lipsitz, Laird and Harrington (1990)]. A block bootstrap procedure [Künsch (1989)] appears to work well for short-lived serial correlation but not for long-lived serial correlation. In a simple setting that permits closed-form calculations of the finite-sample target of the robust variance estimate (with sample quantities replaced by their population counterparts) under working independence, we have found that the target can be far below the sampling variance of the point estimate observed in simulations, suggesting that the asymptotic theory may not provide a reasonable approximation in this situation. Given that, it seems unlikely that a substantial improvement can be made using the available bias correction

methods [e.g., [Mancl and DeRouen \(2001\)](#)]. Another possible approach to variance estimation is window subsampling [[Sherman \(1996\)](#), [Heagerty and Lumley \(2000\)](#), [Oman et al. \(2007\)](#)]. A practical difficulty with this approach is specification of the window size, which has been found to have a large impact on the variance estimate. [Heagerty and Lumley \(2000\)](#) have suggested taking the maximum among variance estimates based on several window sizes, but it remains unclear how to choose the set of window sizes for the maximization. Maximizing over a large set of window sizes can lead to a very conservative variance estimate, as the maximum among several underestimators need not be an underestimator itself.

Under the WCR approach, we have considered various sampling schemes based on time. An interesting possibility, suggested by a referee, is to sample trips using a mechanism that involves other covariates than time and possibly the outcome. In the latter case, appropriate adjustments will be necessary to account for the outcome-dependent nature of the subsample. Further research is needed to explore the potential benefits of the more sophisticated sampling schemes.

This article has been focused on valid inference in the sense of (nearly) correct coverage. We have not discussed the issue of generalization from a sample of subjects to the target population, which is always important and especially so when the number of subjects is small. Such generalizations should be easier to justify when the within-subject variability is large relative to the between-subject variability, which appears to be the case in the NTDS.

APPENDIX A: IRLS ESTIMATION OF α FOLLOWING A GEE ANALYSIS WITH FSE

Write $\underline{\nu} = (\nu_1, \dots, \nu_n)'$ and $\hat{\underline{\nu}} = (\hat{\nu}_1, \dots, \hat{\nu}_n)'$, and let $\hat{\Sigma}_{\hat{\underline{\nu}}|\underline{\nu}}$ denote the variance estimate from the GEE analysis. As the notation indicates, $\hat{\Sigma}_{\hat{\underline{\nu}}|\underline{\nu}}$ estimates the conditional variance $\Sigma_{\hat{\underline{\nu}}|\underline{\nu}} = \text{var}(\hat{\underline{\nu}}|\underline{\nu})$ because the GEE model with FSE is conditional on $\underline{\nu}$. The marginal variance of $\hat{\underline{\nu}}$ is given by

$$\Sigma_{\hat{\underline{\nu}}} = \text{var}(\hat{\underline{\nu}}) = \text{var}\{E(\hat{\underline{\nu}}|\underline{\nu})\} + E\{\text{var}(\hat{\underline{\nu}}|\underline{\nu})\} \approx \sigma_b^2 \mathbf{I}_n + E\Sigma_{\hat{\underline{\nu}}|\underline{\nu}},$$

where \mathbf{I}_n denotes the $n \times n$ identity matrix. A natural estimate of $\Sigma_{\hat{\underline{\nu}}}$ can be obtained as

$$(14) \quad \hat{\Sigma}_{\hat{\underline{\nu}}} = \hat{\sigma}_b^2 \mathbf{I}_n + \hat{\Sigma}_{\hat{\underline{\nu}}|\underline{\nu}}$$

provided a reasonable estimate $\hat{\sigma}_b^2$ is available. This suggests the following IRLS algorithm. We start by obtaining an initial estimate of α , say, the LS estimate. Then we proceed to the following steps:

Step 1. Calculate a moment estimate of σ_b^2 , say, the mean squared residual minus the mean diagonal element of $\hat{\Sigma}_{\hat{\underline{\nu}}}$;

Step 2. Substitute the estimate of σ_b^2 into (14) and re-estimate α from a weighted least squares analysis based on the updated $\hat{\Sigma}_{\hat{\underline{\nu}}}$.

These steps can be iterated a number of times or until convergence.

APPENDIX B: ESTIMATION OF PARAMETERS IN THE
COVARIANCE MATRIX

Let $\hat{\mu}_{ij}$ denote the fitted values from a preliminary GEE analysis without FSE. Then (6) shows that a moment estimate of the sum of σ_b^2 , σ_c^2 and σ_e^2 can be obtained as

$$\log \left[1 + \frac{1}{N} \sum_{i=1}^n \sum_{j=1}^{k_i} \left\{ \frac{(Y_{ij} - \hat{\mu}_{ij})^2}{\hat{\mu}_{ij}^2} + \frac{1}{\hat{\mu}_{ij}} \right\} \right],$$

where $N = \sum_{i=1}^n k_i$ is the total number of observations. Further, equation (7) suggests that σ_b^2 , σ_c^2 and γ can be estimated from a nonlinear regression analysis with $(Y_{ij} - \hat{\mu}_{ij})(Y_{ij'} - \hat{\mu}_{ij'})/(\hat{\mu}_{ij}\hat{\mu}_{ij'})$ as the response variable and with mean function

$$\exp\{\sigma_b^2 + \sigma_c^2 \exp(-\gamma|t_{ij} - t_{ij'}|)\} - 1.$$

The above quantities could be modified using bias correction adjustments [e.g., [Davis, Dunsmuir and Wang \(2000\)](#), Section 3.2] and/or smoothing techniques prior to the nonlinear regression analysis. However, such modifications have not been found helpful in our experiments. Ideally, the nonlinear regression analysis should include all pairs of trips within subjects. However, this may be impractical for the NTDS data because a subject with thousands of trips would contribute millions of pairs, resulting in too many data points. As a compromise, we propose to base the nonlinear regression analysis on two types of pairs: consecutive trips (with $j' = j + 1$) and symmetric trips (with $j + j' = k_i + 1$). The pairs of consecutive trips are informative about short-lived serial correlation, while the pairs of symmetric trips help characterize the overall correlation over the entire range of the gap time $|t_{ij} - t_{ij'}|$.

Similar techniques can be used for a preliminary GEE analysis with fixed subject effects. With $\hat{\mu}_{ij|b_i}$ denoting the fitted values with fixed subject effects, the sum of σ_c^2 and σ_e^2 can be estimated by

$$\log \left[1 + \frac{1}{N} \sum_{i=1}^n \sum_{j=1}^{k_i} \left\{ \frac{(Y_{ij} - \hat{\mu}_{ij|b_i})^2}{\hat{\mu}_{ij|b_i}^2} + \frac{1}{\hat{\mu}_{ij|b_i}} \right\} \right],$$

while the separate values of σ_c^2 and γ can be estimated from another nonlinear regression analysis with $(Y_{ij} - \hat{\mu}_{ij|b_i})(Y_{ij'} - \hat{\mu}_{ij'|b_i})/(\hat{\mu}_{ij|b_i}\hat{\mu}_{ij'|b_i})$ as the response variable and with mean function

$$\exp\{\sigma_c^2 \exp(-\gamma|t_{ij} - t_{ij'}|)\} - 1.$$

These are justified by (8) and (9), respectively.

Initial values of the unknown parameters are required for fitting the above nonlinear regression models. In our experience, a reasonable way to obtain initial values appears to be the following linear regression method. Consider the case with

FSE, and note that equation (9) can be rewritten as

$$(15) \quad \log \log E \left\{ \frac{(Y_{ij} - \mu_{ij|b_i})(Y_{ij'} - \mu_{ij'|b_i})}{\mu_{ij|b_i} \mu_{ij'|b_i}} + 1 \right\} = \log \sigma_c^2 - \gamma |t_{ij} - t_{ij'}|.$$

To take advantage of this relationship, we can allocate the pairs of trips into bins according to the value of $|t_{ij} - t_{ij'}|$ and treat each bin as approximately homogeneous with respect to the gap time. Within each bin, we can replace $\mu_{ij|b_i}$ with $\hat{\mu}_{ij|b_i}$ and expectation with sample average on the left-hand side of (15), replace $|t_{ij} - t_{ij'}|$ with a typical value (say, the median) on the right-hand side, and run a linear regression analysis with each bin as a data point. Initial estimates of σ_c^2 and γ can then be obtained by exponentiating the estimated intercept and negating the estimated slope, respectively. This linear regression approach would not work for the case without FSE, for which we could use as initial estimates the final estimates from a GEE analysis with FSE.

Acknowledgments. We thank the Editor, Associate Editor and two anonymous referees for constructive comments that have greatly improved the manuscript.

REFERENCES

- ALBERT, P. S. and MCSHANE, L. M. (1995). A generalized estimating equations approach for spatially correlated binary data: Applications to the analysis of neuroimaging data. *Biometrics* **51** 627–638.
- CHAN, K. S. and LEDOLTER, J. (1995). Monte Carlo EM estimation for time series models involving counts. *J. Amer. Statist. Assoc.* **90** 242–252. [MR1325132](#)
- DAVIS, R. A., DUNSMUIR, W. T. M. and WANG, Y. (2000). On autocorrelation in a Poisson regression model. *Biometrika* **87** 491–505. [MR1789805](#)
- DIGGLE, P. J., HEAGERTY, P. J., LIANG, K.-Y. and ZEGER, S. L. (2002). *Analysis of Longitudinal Data*, 2nd ed. *Oxford Statistical Science Series* **25**. Oxford Univ. Press, Oxford. [MR2049007](#)
- FITZMAURICE, G., DAVIDIAN, M., VERBEKE, G. and MOLENBERGHS, G. (2008). *Longitudinal Data Analysis: A Handbook of Modern Statistical Methods*. Chapman and Hall/CRC, New York.
- FOLLMANN, D., PROSCHAN, M. and LEIFER, E. (2003). Multiple outputation: Inference for complex clustered data by averaging analyses from independent data. *Biometrics* **59** 420–429. [MR1987409](#)
- HEAGERTY, P. J. and LUMLEY, T. (2000). Window subsampling of estimating functions with application to regression models. *J. Amer. Statist. Assoc.* **95** 197–211. [MR1803149](#)
- HOFFMAN, E. B., SEN, P. K. and WEINBERG, C. R. (2001). Within-cluster resampling. *Biometrika* **88** 1121–1134. [MR1872223](#)
- KÜNSCH, H. R. (1989). The jackknife and the bootstrap for general stationary observations. *Ann. Statist.* **17** 1217–1241. [MR1015147](#)
- LAIRD, N. M. and WARE, J. H. (1982). Random-effects models for longitudinal data. *Biometrics* **38** 963–974.
- LIANG, K. Y. and ZEGER, S. L. (1986). Longitudinal data analysis using generalized linear models. *Biometrika* **73** 13–22. [MR0836430](#)
- LIPSITZ, S. R., LAIRD, N. M. and HARRINGTON, D. P. (1990). Using the jackknife to estimate the variance of regression estimators from repeated measures studies. *Comm. Statist. Theory Methods* **19** 821–845.

- MANCL, L. A. and DEROUEN, T. A. (2001). A covariance estimator for GEE with improved small-sample properties. *Biometrics* **57** 126–134. [MR1833298](#)
- MCCULLOCH, C. E. (1997). Maximum likelihood algorithms for generalized linear mixed models. *J. Amer. Statist. Assoc.* **92** 162–170. [MR1436105](#)
- MCCULLOCH, C. E., SEARLE, S. R. and NEUHAUS, J. M. (2008). *Generalized, Linear, and Mixed Models*, 2nd ed. Wiley, Hoboken, NJ. [MR2431553](#)
- OMAN, S. D., LANDSMAN, V., CARMEL, Y. and KADMON, R. (2007). Analyzing spatially distributed binary data using independent-block estimating equations. *Biometrics* **63** 892–900. [MR2395808](#)
- PAIK, M. C. (1988). Repeated measurement analysis for nonnormal data in small samples. *Communications in Statistics: Simulations* **17** 1155–1171.
- SHERMAN, M. (1996). Variance estimation for statistics computed from spatial lattice data. *J. Roy. Statist. Soc. Ser. B* **58** 509–523. [MR1394363](#)
- SIMONS-MORTON, B. G., OUMET, M. C., ZHANG, Z., LEE, S. E., KLAUER, S. E., WANG, J., ALBERT, P. S. and DINGUS, T. A. (2011a). Risky driving and crash rates among novice teenagers and their parents. *American Journal of Public Health* **101** 2362–2367.
- SIMONS-MORTON, B. G., OUMET, M. C., ZHANG, Z., LEE, S. E., KLAUER, S. E., WANG, J., CHEN, R., ALBERT, P. S. and DINGUS, T. A. (2011b). Naturalistic assessment of risky driving and crash/near crashes among novice teenagers: the effect of passengers. *Journal of Adolescent Health* **49** 587–593.
- UHLENBECK, G. E. and ORNSTEIN, L. S. (1930). On the theory of Brownian Motion. *Phys. Rev.* **36** 823–841.
- ZEGER, S. L. (1988). A regression model for time series of counts. *Biometrika* **75** 621–629. [MR0995107](#)
- ZEGER, S. L. and LIANG, K. Y. (1986). Longitudinal data analysis for discrete and continuous outcomes. *Biometrics* **42** 121–130.
- ZEGER, S. L., LIANG, K.-Y. and ALBERT, P. S. (1988). Models for longitudinal data: A generalized estimating equation approach. *Biometrics* **44** 1049–1060. [MR0980999](#)

Z. ZHANG
 DIVISION OF BIOSTATISTICS
 CENTER FOR DEVICES
 AND RADIOLOGICAL HEALTH
 FOOD AND DRUG ADMINISTRATION
 10903 NEW HAMPSHIRE AVE.
 SILVER SPRING, MARYLAND 20993
 USA
 E-MAIL: zhiwei.zhang@fda.hhs.gov

P. S. ALBERT
 BIostatISTICS AND BIOINFORMATICS BRANCH
 DIVISION OF EPIDEMIOLOGY, STATISTICS
 AND PREVENTION RESEARCH
 EUNICE KENNEDY SHRIVER NATIONAL INSTITUTE
 OF CHILD HEALTH AND HUMAN DEVELOPMENT
 6100 EXECUTIVE BLVD.
 BETHESDA, MARYLAND 20892
 USA
 E-MAIL: albertp@mail.nih.gov

B. G. SIMONS-MORTON
 PREVENTION RESEARCH BRANCH
 DIVISION OF EPIDEMIOLOGY, STATISTICS AND PREVENTION RESEARCH
 EUNICE KENNEDY SHRIVER NATIONAL INSTITUTE
 OF CHILD HEALTH AND HUMAN DEVELOPMENT
 6100 EXECUTIVE BLVD.
 BETHESDA, MARYLAND 20892
 USA
 E-MAIL: mortonb@mail.nih.gov