University of Pennsylvania
UPenn Biostatistics Working Papers

Year 2008 *Paper* 29

Analysis of Adverse Events in Drug Safety: A Multivariate Approach Using Stratified Quasi-least Squares

Hanjoo Kim* Justine Shults†
Scott Patterson‡ Robert Goldberg-Alberts**

*University of Pennsylvania School of Medicine, hanjoo@mail.med.upenn.edu
†Univeristy of Pennsylvania Department of Biostatistics, jshults@mail.med.upenn.edu
‡Wyeth, patters4@wyeth.com
**Wyeth, GoldbeRW@wyeth.com

Analysis of Adverse Events in Drug Safety: A Multivariate Approach Using Stratified Quasi-least Squares

Hanjoo Kim, Justine Shults, Scott Patterson, and Robert Goldberg-Alberts

Abstract

Safety assessment in drug development involves numerous statistical challenges, and yet statistical methodologies and their applications to safety data have not been fully developed, despite a recent increase of interest in this area. In practice, a conventional univariate approach for analysis of safety data involves application of the Fisher's exact test to compare the proportion of subjects who experience adverse events (AEs) between treatment groups; This approach ignores several common features of safety data, including the presence of multiple endpoints, longitudinal follow-up, and a possible relationship between the AEs within body systems. In this article, we propose various regression modeling strategies to model multiple longitudinal AEs that are biologically classified into different body systems via the stratified quasi-least squares (SQLS) method. We then analyze safety data from a clinical drug development program at Wyeth Research that compared an experimental drug with a standard treatment using SQLS, which could be a superior alternative to application of the Fisher's exact test.

Analysis of Adverse Events in Drug Safety: A Multivariate Approach Using Stratified Quasi-least Squares

Hanjoo Kim[1], Justine Shults[1,*], Scott Patterson[2], and Robert Goldberg-Alberts[3]

[1]Center for Clinical Epidemiology and Biostatistics, Department of Biostatistics, University of Pennsylvania School of Medicine, Philadelphia, Pennsylvania, 19104, U.S.A.

[2]Vaccines and Infectious Disease, Wyeth Research, Collegeville, Pennsylvania, 19426, U.S.A.

[3]Clinical Biostatistics, Wyeth Research, Collegeville, Pennsylvania, 19426, U.S.A.

**email:* jshults@mail.med.upenn.edu*

SUMMARY: Safety assessment in drug development involves numerous statistical challenges, and yet statistical methodologies and their applications to safety data have not been fully developed, despite a recent increase of interest in this area. In practice, a conventional univariate approach for analysis of safety data involves application of the Fisher's exact test to compare the proportion of subjects who experience adverse events (AEs) between treatment groups; This approach ignores several common features of safety data, including the presence of multiple endpoints, longitudinal follow-up, and a possible relationship between the AEs within body systems. In this article, we propose various regression modeling strategies to model multiple longitudinal AEs that are biologically classified into different body systems via the *stratified quasi-least squares* (SQLS) method. We then analyze safety data from a clinical drug development program at Wyeth Research that compared an experimental drug with a standard treatment using SQLS, which could be a superior alternative to application of the Fisher's exact test.

KEY WORDS: Longitudinal Data; Adverse Events; Drug Safety; Generalized Estimating Equations; Stratified Quasi-least Squares; Multiple Sources of Correlation.

1. Introduction

In the area of pharmaceutical drug safety, one of the primary goals in the analysis of adverse events (AEs) is to detect any signal for a difference between the treatment versus control arms with respect to the development of AEs. A simple and perhaps one of the most popular approaches for analysis of AEs is to compare the proportion of subjects who experience AEs between the treatment and control arms using the Fisher exact test; This involves constructing a dichotomous variable for each AE that takes value 1 for subjects who experienced that AE, and that takes value 0 otherwise. Next, at a nominal 5% level of significance, if the computed Fisher's exact p-value for each AE is greater than 0.05, the typical conclusion is that the Fisher's exact tests have provided no strong evidence, or *signal*, that the rate of AEs differs between treatment arms. (See Table 1 of Mehrota and Heyse (2004) for an example.)

Safety data, however, are often multi-dimensional, due to the fact that multiple AEs are typically assessed on subjects over time. As a result, a multivariate approach could be employed to detect signals in the differences between the treatment and control arms with respect to the development of AEs. Furthermore, popular approaches that reduce information on each AE to a dichotomous variable prior to the implementation of a univariate method such as application of the Fisher's exact test, may result in a substantial loss of information (by overly simplifying the original multi-dimensional structure of the data) which may jeopardize the validity of the statistical conclusions.

In practice, AEs are often recorded by self-report from study participants and/or investigators during scheduled visits, in which participants may report that they have experienced multiple AEs of different types since their prior visit. Reported AEs will then typically be classified in terms of affected body systems in accordance with accepted dictionaries of preferred AE vocabulary such as COSTART (Chow and Liu, 2003, p563) and Med-

DRA (MedDRA, 2004). For example, some of the preferred terms for body systems in the COSTART dictionary include the digestive, nervous, metabolic and nutritional, and respiratory systems. Each body system in COSTART has its own distinct set of AEs that are described using the COSTART preferred terminology, e.g. the AEs associated with the digestive system include vomiting, nausea, diarrhea, and constipation.

Despite a recent increase in interest and attention paid to the issue of pharmaceutical drug safety, few methods are available that propose either the novel implementation of existing methods, or new statistical methodologies, for analysis of safety data. One recent approach addresses the issue of multiplicity using a three-level Bayesian hierarchical mixture model to directly calculate the posterior probability that the (random) log odds ratio for development of AEs between the treatment and control groups is greater than one (Berry and Berry, 2004); These authors considered three types of AEs that had been identified by Mehrota and Heyse (2004). Goldberg-Alberts and Page (2006) fit a log-linear model to the total number of AEs in order to explore the strength of association between multiple AEs. In addition, Schildcrout et. al. (2008) focused on modeling longitudinal clinical laboratory data in the presence of dropout, follow-up frequency, and treatment discontinuation, based on both the generalized estimating equation (Liang and Zeger, 1986) and maximum likelihood approaches. Each of the methods just cited addresses different issues in the analysis of drug safety data that represent important and meaningful first steps toward a clearer understanding of the issues involved in and optimum approaches for analysis of safety data.

In this paper, we make an additional contribution to the relatively open field of analysis of safety data. We are mainly concerned with a modeling strategy for the analysis of AEs by incorporating multiple sources of correlation (that result from the fact that multiple AEs are measured on subjects over time) using our proposed method, *stratified quasi-least squares* (SQLS). SQLS generalizes the method of quasi-least squares (QLS) (Chaganty and

Shults, 1999), a two-stage approach for analysis of longitudinal data with an assumed working correlation structure in the framework of generalized estimating equations (GEE) (Liang and Zeger, 1986).

Prior to the development of SQLS, QLS had been extended for analysis of longitudinal data with multiple outcomes, but this prior extension was limited to either totally balanced data for two sources of correlation (Chaganty and Naik, 2002), or data with two or more sources of correlation that were balanced within subjects (Shults and Morrow, 2002; Shults, Whitt and Kumanyika, 2004; Shults and Ratcliffe, 2007). For example, Table 1 displays a data structure for multi-outcome longitudinal data, for which up to three outcomes were measured on each cluster i at a total of 5 measurement occasions. If we consider only measurements from cluster 1, then the data are totally balanced because 5 measurements were collected on each of 3 outcomes. Next, if we consider measurements from clusters 3 and n only, then the data are balanced within subjects because the same number of measurements was collected on all measured outcomes *within each cluster*. Finally, if we consider only measurements from cluster 2, then the data are unbalanced because the number of measurements varied between outcomes for this cluster.

Prior approaches that extended QLS for analysis of data with multiple sources of correlation (Shults and Morrow, 2002; Chaganty and Naik, 2002; Shults, Whitt and Kumanyika, 2004; Shults and Ratcliffe, 2007) described the pattern of association amongst multiple longitudinal outcomes by specifying a working correlation structure that was constructed as the Kronecker product of correlation structures that would be appropriate for each source of correlation, if that were the only source of correlation in the data. For example, for analysis of multiple outcomes that are measured over time, a plausible working correlation structure would be the Kronecker product of an exchangeable correlation structure (for the multiple

Table 1
Example of data format for longitudinal data with 3 outcomes at 5 time points

Cluster i	Outcome j	Time Points 1	2	3	4	5
	1	y_{111}	y_{112}	y_{113}	y_{114}	y_{115}
1	2	y_{121}	y_{122}	y_{123}	y_{124}	y_{125}
	3	y_{131}	y_{132}	y_{133}	y_{134}	y_{135}
	1	y_{211}	·	y_{213}	y_{214}	y_{215}
2	2	y_{221}	y_{222}	y_{223}	y_{224}	y_{225}
	3	y_{231}	·	y_{233}	·	y_{235}
	1	y_{311}	y_{312}	y_{313}	y_{314}	·
3	2	y_{321}	y_{322}	y_{323}	y_{324}	·
	3	y_{331}	y_{332}	y_{333}	y_{334}	·
⋮	⋮			⋮		
	1	y_{n11}	y_{n12}	y_{n13}	y_{n14}	y_{n15}
n	2	·	·	·	·	·
	3	y_{n31}	y_{n32}	y_{n33}	y_{n34}	y_{n35}

· represents missing observation.

outcomes) and a first-order autoregressive (AR1) structure (for the outcomes over time). SQLS extends these prior approaches for analysis of unbalanced data by assuming that the working correlation structure for each subject is a sub-matrix of a larger Kronecker product structure; This assumption is appropriate for studies that planned for balanced data, but for which some measurements were missing.

Unbalanced data are common in safety studies. For example, the following situations led to unbalanced safety data in the Depression Research Unit (DRU) at the University of Pennsylvania (Dr. Jay Amsterdam, Head of DRU, personal correspondence): Investigators who were embarrassed to ask about sexual side effects of depression had missing data for this type of AE. In a study of SSRI anti-depressants, information was collected on gastrointestinal AEs, but no information was obtained regarding gynecomastia (breast enlargement) due to sensitivity on the part of investigators to ask about this condition; Furthermore, no patients volunteered information on this condition. In some studies, information was missed due

to confusion regarding the difference between illness symptoms (e.g. decreased libido as a symptom of depression) versus AE (e.g. erectile dsyfunction as a result of medication).

In addition to extending QLS for analysis of unbalanced data with multiple sources of correlation, SQLS also allows for construction of working correlation structures that are stratified according to a third variable of interest. For example, for analysis of multiple AEs that are measured over time, as mentioned above, a plausible working correlation structure is Exchangeable(ρ)$\otimes$AR1(α), where $\otimes$ represents the Kronecker product, and ρ and α represent the correlations parameters for the exchangeable and AR1 structures, respectively. However, each body system has a distinct set of AEs, so that fitting a common exchangeable structure for all AEs may not be a reasonable choice. For example, although it might be reasonable to assume that the pairwise correlations between AEs are the same within each body system, we might anticipate a greater degree of similarity of occurrence of AEs within some body systems in comparison with others. SQLS therefore allows the parameters for the exchangeable structure to vary according to body system, so that the exchangeable structure for multiple AEs is the exchangeable(ρ_j) for the jth body system.

Our outline for the manuscript is as follows. In §2, we describe our motivating example from a study of safety data from a clinical drug development program at Wyeth Research. In §3 we then describe our development of SQLS for the analysis of the AE data introduced in §2. In §4 we next analyze the AEs data via SQLS. Finally, we provide a brief discussion and some concluding remarks in §5.

2. Wyeth Safety Data

Our goal in this report is to describe our approach to the analysis of safety data and to demonstrate the potential approach. Safety data from positive controls were obtained from a clinical drug development program at Wyeth. Data were blinded to the drug, adverse event, and body system. Simulation was used to produce an adverse event profile for an hypothetical comparator drug in order to evaluate the potential approach to analysis.

Table 2

Example of 15 selected adverse events from the Wyeth control data set. AE_{ij} represents the j^{th} adverse event in the ith body system. Rates were computed by dividing the total number of subjects who experienced AE_{ij} by 412, the number of subjects at the start of the trial.

Body System	AE_{ij}	Y_{ij}^*	Rates
1	AE_{11}	81	0.197
1	AE_{12}	112	0.272
1	AE_{13}	84	0.204
1	AE_{14}	11	0.027
1	AE_{15}	3	0.022
1	AE_{16}	9	0.002
1	AE_{17}	1	0.007
1	AE_{18}	3	0.009
2	AE_{21}	49	0.119
2	AE_{22}	80	0.194
2	AE_{23}	21	0.051
2	AE_{24}	7	0.017
2	AE_{25}	8	0.019
2	AE_{26}	18	0.044
2	AE_{27}	1	0.002

* Total number of subjects who experienced AE_{ij}.

The Wyeth safety study involved 5 weekly scheduled visits for each participant. The first two weeks corresponded to the on-therapy period and the remaining weeks corresponded to the follow-up period. The durations of AEs were recorded either during the physical examination and clinical evaluation of a subject during his/her scheduled visit, or through a self-report from the subject. Table 2 displays the total number of subjects (out of 412 subjects in the control arm at the start of the trial) who had experienced an adverse event among 15 selected AEs (8 AEs for body system 1 and 7 AEs for body system 2).

One challenging issue in the analysis of safety data is the low frequency of occurrence

of some AEs. For example, in Table 2, the rate for all AEs is below 6%, except for AE_{11}, AE_{12}, AE_{13}, AE_{21}, and AE_{22}. Rare AEs are often investigated on a case by case basis by the study clinician(s), as opposed to making a formal statistical comparison. For example, one case of headache in the treatment group versus zero headaches in the control group might be simply described in the report of safety data. In this paper, we focused on AEs that occurred frequently enough to warrant a formal statistical comparison. We considered AEs that occurred in at least 10 percent of study participants, although admittedly a different threshold could have been applied.

3. Stratified Quasi-least Squares

3.1 *General Setup and Notation*

We let y_{ijkl} represent the kth outcome (AE), in the jth stratum (body system), at the lth time, for the ith cluster (subject), where $i = 1, \ldots, n_j$, $j = 1, \ldots, g$, $k = 1, \ldots, n_{ij}$, and $l = 1, \ldots, n_{ijk}$. Let y_{ijk} be a $n_{ijk} \times 1$ vector of the kth outcome in the jth stratum such that $y_{ijk} = (y_{ijk1}, y_{ijk2}, \ldots, y_{ijkn_{ijk}})'$. Further, let $y_{ij} = (y'_{ij1}, y'_{ij2}, \ldots, y'_{ijn_{ij}})'$, i.e. y_{ij} is the vector of outcomes in the jth strata for the ith cluster that has been sorted according to indices k followed by l. For example, for $n_{ij} = 2$, $n_{ij1} = 3$, and $n_{ij2} = 4$,

$$y_{ij} = (y_{ij11}, y_{ij12}, y_{ij13}, y_{ij21}, y_{ij22}, y_{ij23}, y_{ij24})'. \tag{1}$$

We assume that the outcomes y_{ijkl} have mean and variance given by $\mathrm{E}(y_{ijkl}) = \mu_{ijkl}$ and $\mathrm{Var}(y_{ijkl}) = \phi h(\mu_{ijkl})$ respectively, where $\phi > 0$ is a known or unknown scale (dispersion) parameter and $h(\cdot)$ is the variance function. We also assume that each y_{ijkl} is associated with a vector of covariates $x_{ijkl} = (x_{ijkl1}, \ldots, x_{ijklp})'$ and unknown regression parameters $\beta = (\beta_1, \ldots, \beta_p)$ through an invertible link function $g(\cdot)$ such that $\mu_{ijkl} = g^{-1}(x'_{ijkl}\beta)$.

Next, let $z_{ijkl} = (y_{ijkl} - \mu_{ijkl}) / \sqrt{h(\mu_{ijkl})}$ represent the Pearson residual that corresponds to y_{ijkl}. Let z_{ijk} be the vector of Pearson residuals for each AE $k = 1, \ldots, n_{ij}$ for the jth stratum, and let

$$z_{ij} = (z'_{ij1}, z'_{ij2}, \ldots, z'_{ijn_{ij}})', \tag{2}$$

be the vector of Pearson residuals that has been sorted by index k followed by l for the ith cluster in the jth stratum.

We further assume that measurements from two different vectors y_{ij} (i.e. from different body systems on subjects) will be independent, but that measurements within vectors y_{ij} (i.e. within body systems) will be correlated, due to similarity among the multiple AEs at each measurement occasion, and across time points. (Note that if we do not stratify on body system, then we will not need to assume independence across body systems.) As in GEE, we decompose the covariance matrix of y_{ij} as

$$\Sigma_{ij} = \phi A_{ij}^{1/2} R_{ij}(\rho_j, \alpha) A_{ij}^{1/2} \tag{3}$$

where $A_{ij} = \text{diag}(h(\mu_{ij11}), \ldots, h(\mu_{ijn_{ij}n_{ijk}}))$, and $R_{ij}(\rho_j, \alpha)$ is known as the *working correlation matrix* that describes the pattern of association among the repeated measurements on each subject i for stratum j. We further decompose $R_{ij}(\rho_j, \alpha) = R_{ij}(\rho_j) \otimes R_i(\alpha)$, where $R_{ij}(\rho_j)$ is our reasonable guess of the true correlation structure to describe the pattern of association between the multiple AEs in stratum j at each measurement occasion, and $R_i(\alpha)$ is our reasonable guess of the common correlation structure to describe the pattern of association amongst the repeated AE measurements over time.

3.2 *SQLS for Unbalanced Longitudinal Data with Multiple Outcomes*

The KP correlation structure in (3) is a popular choice for analysis of multi-outcome longitudinal data because it forces the correlation between measurements to be smaller when they have more disagreement with respect to the sources of correlation in the data, which is often biologically plausible in longitudinal studies. For example, the KP structure has been implemented for analysis of Gaussian data by Galecki (1994), Lu and Zimmerman (2005), Naik and Rao (2001), Roy and Khattree (2005), Roy (2006), and Roy and Leiva (2007). Moreover, the Kronecker product has convenient mathematical properties, in particular a simple expression for its inverse. However, for unbalanced data, the working correlation structure can no longer be expressed as a KP structure. As a result, the correlation structures for unbalanced data lose mathematical tractability, which is why previous implementations of QLS for multi-outcome longitudinal data focused on balanced data. (For example, Shults, Whitt and Kumanyika (2004) presented a straightforward algorithm for implementation of a KP structure for QLS that utilized the square root of the inverse of the KP structure.)

Here we introduce an alternative way to express the working correlation structure so that it is a function of the KP structure, even for unbalanced data. Our approach will hinge on considering a study that planned for balanced data, with a KP structure to describe the pattern of multiple AEs over time amongst subjects with complete data. The correlation structure of subjects with missing measurements will then a sub-matrix of the larger KP structure, that has been constructed by removing rows and columns from the larger KP matrix that correspond to the missing measurements.

To formalize this discussion, we first consider positive definite and symmetric matrices $R_j(\rho_j)$ and $R(\alpha)$ of dimensions $m_j \times m_j$ and $m \times m$ respectively, where m_j is the maximum

number of types of outcomes across all clusters in the jth stratum, and m is the maximum number of time points across all clusters and strata. Then it is well known that $R_j(\rho_j) \otimes R(\alpha)$ is also a positive definite and symmetric matrix of dimension $m_j m \times m_j m$. Next, let W be the matrix that is constructed by removing the jth row and the jth column of the $R_j(\rho_j) \otimes R(\alpha)$ matrix. Using the following argument, it is also easy to show that W is positive definite: For a non-zero vector x of the same dimension as W, we have that $x'Wx = y'(R_j(\rho_j) \otimes R(\alpha))y$ where y is the vector x, but with a zero replaced in the jth position in x. Since $y \neq 0$ and $R_j(\rho_j) \otimes R(\alpha)$ is positive definite, it follows that

$$x'Wx = y'(R_j(\rho_j) \otimes R(\alpha))y > 0. \tag{4}$$

Therefore, W is positive definite.

Next, let K_j be the set of all indices jkl such that

$$K_j = \{j11, \ldots, j1m, j21, \ldots, j2m, \ldots, jm_j1, \ldots, jm_jm\}.$$

Further, let $I_{ij} = \{jkl : jkl \notin K_j\}$ for each cluster i in the jth stratum, i.e. I_{ij} denotes the missing indices for the ith cluster in the jth stratum that are not contained in the overall index set K_j. Denote e_{ijkl} as the $m_j m \times 1$ elementary vector with a one in the jklth row and zeros elsewhere. Next, let E_i be an $m_j m \times (m_j m - \text{card}(I_{ij}))$ matrix defined as $E_i = (e_{ij11}, e_{ij12}, \ldots, e_{1jm_jm})$ for all $jkl \in K_j \setminus I_{ij}$, where $\text{card}(I_{ij})$ denotes the number of elements in the set I_{ij}. For example, suppose the maximum number of types of outcomes and time points across all subjects are 2 and 4 respectively, i.e. $m_j = 2$, and $m = 4$. Then $K_j = \{j11, j12, j13, j14, j21, j22, j23, j24\}$. If the jth stratum in the ith cluster has a missing observation in the 2nd outcome at the 3rd time point, i.e. y_{ij23} is not observed, then

$$E_{ij} = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}_{8\times 7}$$

where each column vector corresponds to e_{ij11}, e_{ij12}, e_{ij13}, e_{ij14}, e_{ij21}, e_{ij22}, and e_{ij24}, respectively. Note that the 7th observation y_{i123} was missing, so that E_{ij} was constructed by removing the 7th column of an 8×8 identity matrix. Finally, let W_{ij} be the matrix that is constructed by removing any missing rows and their corresponding columns of the $R_{ij}(\rho_j) \otimes R_i(\alpha)$ matrix. Then it is easy to verify that

$$W_{ij} = E'_{ij}(R_j(\rho_j) \otimes R(\alpha))E_{ij} \tag{5}$$

which is a positive definite and symmetric matrix by (4). That W_{ij} is invertible which follows from the fact that this matrix is positive definite. However, the inverse will not have a simple expression when E_{ij} is not an identity matrix, i.e. when the data are unbalanced.

3.3 *Stratified Quasi-least Squares*

GEE estimates the correlation parameters via the method of moments. SQLS (and QLS), on the other hand, is a two-stage approach that estimates correlation parameters using estimating equations that are orthogonal to the estimating equation for β. One advantage of SQLS is that it is straightforward to apply, even for a relatively complex correlation structure.

For estimating β, SQLS solves the following estimating equation at given values of α and

ρ_j for $j = 1, \ldots, g$:

$$\sum_{i,j} \left(\frac{\partial \mu_{ij}}{\partial \beta} \right)' A_{ij}^{-1/2} W_{ij}^{-1} A_i^{-1/2} (y_{ij} - \mu_{ij}) = 0 \tag{6}$$

where W_{ij} is given in (5).

Stage one estimating equations for α and ρ_j are obtained by taking the partial derivatives with respect to α and ρ_j respectively in the generalized error sum of squares: $Q = \sum_{i,j} z_{ij}' W_{ij}^{-1} z_{ij}$, where z_{ij} is the vector of Pearson residuals defined in (2). Thus, the stage 1 estimating equations for α are given by

$$\begin{aligned} \frac{\partial Q}{\partial \alpha} &= \frac{\partial}{\partial \alpha} \left\{ \sum_{i,j} z_{ij}' W_{ij}^{-1} z_{ij} \right\} \\ &= \sum_{i,j} z_{ij}' W_{ij}^{-1}(\alpha) \frac{\partial W_{ij}}{\partial \alpha} W_{ij}^{-1}(\alpha) z_{ij} \\ &= -\sum_{i,j} z_{ij}' W_{ij} \left[E_{ij}' \left(R_i(\rho_i) \otimes \frac{\partial R(\alpha)}{\partial \alpha} \right) E_{ij}' \right] W_{ij} z_{ij} \\ &= 0, \end{aligned} \tag{7}$$

where the third equality follows from that fact that

$$\frac{\partial W_{ij}^{-1}}{\partial \alpha} = -W_{ij}^{-1} \frac{\partial W_{ij}}{\partial \alpha} W_{ij}^{-1} \tag{8}$$

for any α such that W_{ij} is positive definite. Similarly, the stage 1 estimating equation for ρ_j is given by

$$\frac{\partial Q}{\partial \rho_j} = \sum_{i=1}^{n_j} z_{ij}' W_{ij} \left[E_{ij}' \left(\frac{\partial R_j(\rho_j)}{\partial \rho_j} \otimes R(\alpha) \right) E_{ij}' \right] W_{ij} z_{ij} = 0, \tag{9}$$

for $j = 1, \ldots, g$, using (8). It is important to note that the summation in the stage 1 estimating equation for α ranges from $i = 1, \ldots, n_j$ for all strata $j = 1, \ldots, g$, whereas the summation for ρ_j ranges from $i = 1, \ldots, n_j$.

SQLS alternates between solving the GEE estimating equation (6) and the stage 1 estimating equations (7, 9) until there is convergence in the estimates. However, the stage 1 estimates for α and ρ_j are asymptotically biased. Bias correction is made via the stage 2

estimating equations to obtain consistent estimates for α and ρ_j, $i = 1, \ldots, g$. (See Chaganty and Shults (1999) Theorem 3.2). The stage 2 estimating equations for this bias correction are given by

$$\sum_{i,j} \text{tr}\left\{\frac{\partial \tilde{W}_{ij}^{-1}}{\partial \tilde{\alpha}} W_{ij}\right\} = 0 \text{ and } \sum_{i=1}^{n_j} \text{tr}\left\{\frac{\partial \tilde{W}_{ij}^{-1}}{\partial \tilde{\rho}_j} W_{ij}\right\} = 0 \tag{10}$$

for α and ρ_j, respectively, where "tr" is the trace, and $\tilde{W}_{ij}$ is the W_{ij} matrix evaluated at the stage 1 estimates $\tilde{\alpha}$ and $\tilde{\rho}_j$, $j = 1, \ldots, g$.

SQLS then alternates between solving the GEE estimating equation (6) and the stage 2 estimating equations (10), in order to obtain the final estimates for β, α and ρ_j, $j = 1, \ldots, g$. Using (8), the stage 2 estimating equations for α and ρ_j, $j = 1, \ldots, g$ are given by

$$\sum_{i,j} \text{tr}\left(\tilde{W}_{ij}\left[E'_{ij}\left(R_j(\tilde{\rho}_j) \otimes \frac{\partial R(\tilde{\alpha})}{\partial \tilde{\alpha}}\right) E_{ij}\right] \tilde{W}_{ij} R_i(\alpha)\right) = 0 \tag{11}$$

and

$$\sum_{i=1}^{n_j} \text{tr}\left(\tilde{W}_{ij}\left[E'_{ij}\left(\frac{\partial R_j(\tilde{\rho}_j)}{\partial \tilde{\rho}_j} \otimes R(\tilde{\alpha})\right) E_{ij}\right] \tilde{W}_{ij} R_{ij}(\rho_j)\right) = 0, \tag{12}$$

respectively.

3.4 *Estimates of the Covariance Matrix in SQLS*

As in GEE, SQLS also implements the so-called *sandwich* estimate of the covariance matrix which uses the empirical evidence from the data to adjust the standard errors. The *sandwich* estimator is popular due to the consistency of the regression parameter $\hat{\beta}$ even under a misspecification of the true correlation structure under mild regularity assumptions (Liang and Zeger, 1986). The sandwich estimate of the covariance matrix in SQLS is given by

$$\widehat{\text{Cov}}(\hat{\beta}) = V^{-1}\left\{\sum_{i,j} X'_{ij} A_i^{1/2} \hat{W}_{ij}^{-1} \hat{z}_{ij} \hat{z}'_{ij} \hat{W}_{ij}^{-1} A_i^{1/2} X_{ij}\right\} V^{-1}$$

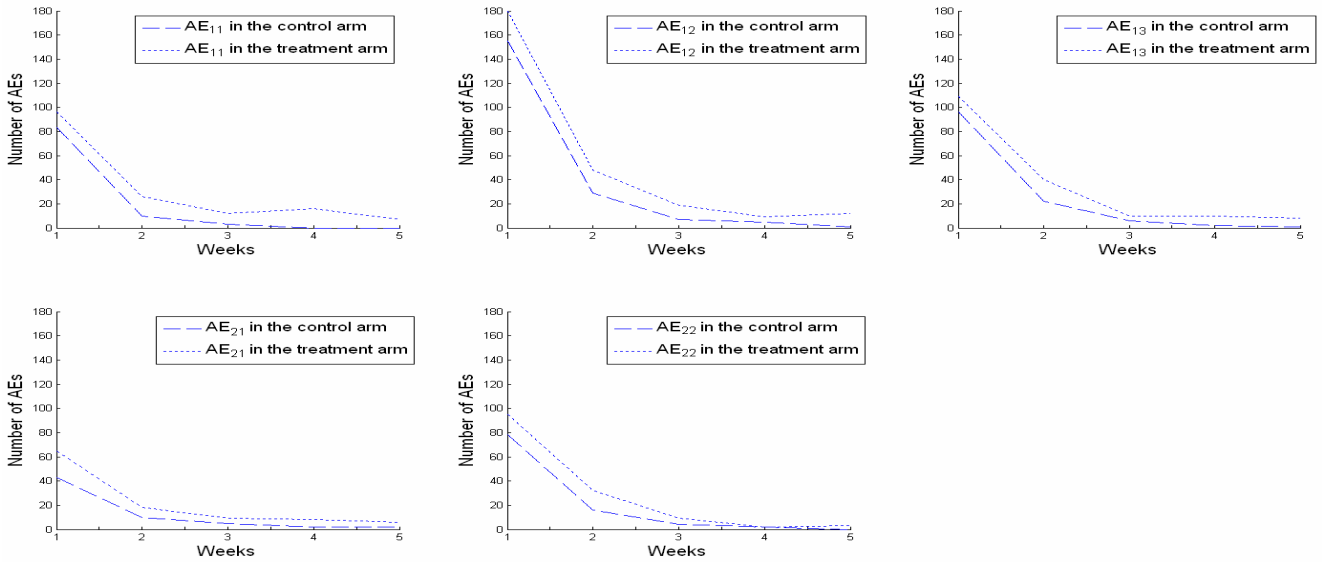

Figure 1. Number of AEs with rates greater than 10% in Table 2. AEs in the treatment arm were simulated based on the control arm using a mixture of Poisson and Bernoulli random variables.

where

$$V = \sum_{i,j} X'_{ij} A_{ij}^{1/2} \hat{W}_{ij}^{-1} A_{ij}^{1/2} X_{ij},$$

$\hat{W}_{ij} = W(\hat{\rho}_j, \hat{\alpha})$, $\hat{z}_{ij} = z_{ij}(\hat{\beta})$, and X_{ij} is the design matrix for the ith cluster in the jth stratum.

4. Analysis of the Wyeth Safety Data

Table 2 summarizes the total proportion of subjects who experienced AEs in body systems 1 and 2. However, a more informative way to present the Wyeth safety data might be to summarize the total number of AEs per subject, during each week of follow up. Figure 1 displays information on the AEs with rates greater than 10% from Table 2, i.e. AE_{11}, AE_{12}, AE_{13}, AE_{21}, and AE_{22}, and for the simulated treatment data. The total number of AEs were calculated based on the total number of days for which a subject experienced AE_{ij} during

each week. For example, if a subject experienced AE_{11} for two consecutive days during the first week, we considered the total number of AEs for AE_{11} on this subject to be 2.

The number of AEs for the treatment data were simulated based on the control data using a mixture of Poisson and Bernoulli random variables. Let y_{ijkl} be the kth AE in the jth body system at the lth visit for the ith subject, where $i = 1, \ldots, n_j$, $j = 1, \ldots, g$, $k = 1, \ldots, n_{ij}$, and $l = 1, \ldots, n_{ijk}$. For the treatment arm, define

$$y_{ijkl} = \begin{cases} y_{ijkl} + xI_{0.2} & \text{for the on-therapy period, i.e. } l \leqslant 2 \\ y_{ijkl} + xI_{0.1} & \text{for the follow-up period, i.e. } l > 2 \end{cases}$$

where x is the Poisson random variable with mean 0.5, and $I_p = 1$ with probability p; otherwise 0 with probability $1 - p$, $0 < p < 1$. We therefore added the Poisson random variable with mean 0.5 to a randomly selected portion of the control data. For the on-therapy period, 20% of the control data were randomly selected, and for the follow-up period, only 10% of the control data were randomly selected.

From Figure 1, it is readily seen that the number of AEs for the treatment arm is greater than for the control arm for all AE types. However, for AE_{21} and AE_{22}, subjects in the treatment arm experienced more AEs during the on-therapy period than during the follow-up period. On the other hand, the opposite may be true for AE_{11}, i.e. subjects in the treatment arm experienced more AEs during the follow-up period than during the on-therapy period.

Table 3 summarizes the Fisher's exact p-values for comparing the proportions of each AE_{ij} for the control and the treatment arms. None of the AEs were significant at a nominal 5% significance level, which suggests that there was no strong signal for a difference in the proportion of subjects who had adverse event AE_{ij} between the two treatment groups.

However, for AE_{11} and AE_{21}, the p-values ≈ 0.06, which might warrant some additional investigation.

Table 3

Two-sided Fisher's exact p-values for AE_{11}, AE_{12}, AE_{13}, AE_{21}, and AE_{22}. Rates are calculated by the total number of subjects who experienced AE_{ij} over 412, the total number of subjects at the start of the trial for both arms. Y_{ij} and X_{ij} are the total number of subjects who experienced AE_{ij} for the control and the treatment arms respectively.

	Control		Treatment		Fisher's
AE_{ij}	$Y_{ij}^{\dagger}$	Rate	$X_{ij}^{\ddagger}$	Rate	exact p-value
AE_{11}	81	0.1966	104	0.2524	0.0661
AE_{12}	112	0.2718	125	0.3034	0.3557
AE_{13}	84	0.2039	100	0.2427	0.2095
AE_{21}	49	0.1189	69	0.1675	0.0585
AE_{22}	80	0.1942	86	0.2087	0.6642

A more sophisticated analysis which leads to a better understanding of the AE data may be accomplished via the method of SQLS using the Exchangeable(ρ_j)$\otimes$AR1(α) structure for which we allow different parameters ρ_j for multiple AEs within each body system j, and a common parameter α for the correlation over time.

Recall that the primary goal of the analysis is to detect any signal that there is a difference in the two arms with respect to the AEs. We will consider several regression models in a sequential fashion to accomplish our analytic goal. First, we consider a model to compare the overall differences in the number of AEs per subject, given by

$$\log(\mu_{ijkl}) = \beta_0 + \beta_1 \text{trt}_i \tag{13}$$

where μ_{ijkl} is the mean of the Poisson random variable, and $\text{trt}_i = 1$ if the ith subject is in the treatment group; and $= 0$ otherwise. After fitting Model (13), we may consider a model with two-way interaction terms between each type of AE and trt_i, for testing a difference in the number of AE_{ij} between the control and the treatment arms. This model is given by

Table 4
Summary of the model fit for models (13), (14), and (15) via the method of SQLS using Exchangeable(ρ_j)$\otimes$AR1(α) working correlation structure for correlation within body system j, $j = 1, 2$ and for common correlation over time.

Model	Parameter	Estimate	Standard Error	z-value	p-value	95% Confidence Interval Lower	Upper
	Intercept	-1.683	0.045	-37.08	< 0.001	-1.772	-1.594
	Treatment	0.358	0.060	6.01	< 0.001	0.241	0.474
(13)	Exchangeable(ρ_1)	0.244					
	Exchangeable(ρ_2)	0.200					
	AR1(α)	0.117					
	Intercept	-1.562	0.089	-17.54	< 0.001	-1.737	-1.388
	AE_{11}	-0.439	0.128	-3.43	0.001	-0.690	-0.188
	AE_{12}	0.259	0.129	2.01	0.045	0.006	0.512
	AE_{13}	-0.174	0.138	-1.27	0.206	-0.446	0.096
	AE_{21}	-0.514	0.179	-2.87	0.004	-0.865	-0.163
	Treatment	0.298	0.124	2.41	0.016	0.056	0.541
(14)	AE_{11} × Treatment	0.165	0.176	0.94	0.349	-0.180	0.510
	AE_{12} × Treatment	-0.001	0.174	-0.01	0.993	-0.342	0.339
	AE_{13} × Treatment	0.016	0.186	0.09	0.930	-0.349	0.381
	AE_{21} × Treatment	0.234	0.225	1.04	0.298	-0.206	0.674
	Exchangeable(ρ_1)	0.255					
	Exchangeable(ρ_2)	0.215					
	AR1(α)	0.100					
	Intercept	-3.886	0.455	-8.54	< 0.001	-4.778	-2.994
	AE_{11}	-1.088	0.722	-1.51	0.132	-2.504	0.327
	AE_{12}	0.367	0.536	0.68	0.494	-0.683	1.416
	AE_{13}	0.001	0.617	0.00	0.999	-1.209	1.210
	AE_{21}	0.410	0.458	0.89	0.371	-0.489	1.308
	Treatment	0.841	0.537	1.57	0.118	-0.212	1.893
	Ontherapy	3.068	0.479	6.41	< 0.001	2.129	4.007
	AE_{11} × Treatment	1.596	0.800	1.99	0.046	0.027	3.164
	AE_{12} × Treatment	0.279	0.635	0.44	0.660	-0.965	1.524
	AE_{13} × Treatment	0.289	0.721	0.40	0.689	-1.125	1.703
	AE_{21} × Treatment	0.085	0.571	0.15	0.882	-1.033	1.203
(15)	AE_{11} × Ontherapy	0.658	0.747	0.88	0.379	-0.806	2.122
	AE_{12} × Ontherapy	-0.116	0.555	-0.21	0.834	-1.204	0.971
	AE_{13} × Ontherapy	-0.196	0.646	-0.30	0.762	-1.462	1.071
	AE_{21} × Ontherapy	-1.021	0.495	-2.06	0.039	-1.990	-0.051
	Ontherapy × Treatment	-0.580	0.568	-1.02	0.308	-1.693	0.534
	AE_{11} × Ontherapy × Treatment	-1.584	0.834	-1.90	0.058	-3.219	0.051
	AE_{12} × Ontherapy × Treatment	-0.326	0.661	-0.49	0.622	-1.622	0.970
	AE_{13} × Ontherapy × Treatment	-0.318	0.759	-0.42	0.676	-1.805	1.169
	AE_{21} × Ontherapy × Treatment	0.109	0.615	0.18	0.860	-1.097	1.135
	Exchangeable(ρ_1)	0.127					
	Exchangeable(ρ_2)	0.112					
	AR1(α)	0.012					

$$\begin{aligned}\log(\mu_{ijkl}) = \beta_0 &+ \beta_1 \mathrm{AE}_{i11l} + \beta_2 \mathrm{AE}_{i12l} \\ &+ \beta_3 \mathrm{AE}_{i13l} + \beta_4 \mathrm{AE}_{i21l} + \beta_5 \mathrm{trt}_i \\ &+ \beta_6(\mathrm{AE}_{i11l} \times \mathrm{trt}_i) + \beta_7(\mathrm{AE}_{i12l} \times \mathrm{trt}_i) \\ &+ \beta_8(\mathrm{AE}_{i13l} \times \mathrm{trt}_i) + \beta_9(\mathrm{AE}_{i21l} \times \mathrm{trt}_i)\end{aligned} \tag{14}$$

where $\mathrm{AE}_{ijkl} = 1$ if the ith subject experienced the kth AE in the jth body system at time l; and $= 0$ otherwise; and $\mathrm{AE}_{ijkl} \times \mathrm{trt}_i$ is the interaction term between AE_{ijkl} and trt_i. Lastly,

we may also investigate whether the number of AEs differ during the on-therapy period vs. the follow-up period by fitting the following model:

$$\begin{aligned}
\log(\mu_{ijkl}) = {} & \beta_0 + \beta_1 \mathrm{AE}_{i11l} + \beta_2 \mathrm{AE}_{i12l} + \beta_3 \mathrm{AE}_{i13l} \\
& + \beta_4 \mathrm{AE}_{i21l} + \beta_5 \mathrm{trt}_i + \beta_6 \mathrm{Ontherapy}_{ijk} \\
& + \beta_7 (\mathrm{AE}_{i11l} \times \mathrm{trt}_i) + \beta_8 (\mathrm{AE}_{i12l} \times \mathrm{trt}_i) \\
& + \beta_9 (\mathrm{AE}_{i13l} \times \mathrm{trt}_i) + \beta_{10} (\mathrm{AE}_{i21l} \times \mathrm{trt}_i) \\
& + \beta_{11} (\mathrm{AE}_{i11l} \times \mathrm{Ontherapy}_{ijk}) \\
& + \beta_{12} (\mathrm{AE}_{i12l} \times \mathrm{Ontherapy}_{ijk}) \\
& + \beta_{13} (\mathrm{AE}_{i13l} \times \mathrm{Ontherapy}_{ijk}) \\
& + \beta_{14} (\mathrm{AE}_{i21l} \times \mathrm{Ontherapy}_{ijk}) \\
& + \beta_{15} (\mathrm{Ontherapy}_{ijk} \times \mathrm{trt}_i) \\
& + \beta_{12} (\mathrm{AE}_{i11l} \times \mathrm{Ontherapy}_{ijk} \times \mathrm{trt}_i) \\
& + \beta_{13} (\mathrm{AE}_{i12l} \times \mathrm{Ontherapy}_{ijk} \times \mathrm{trt}_i) \\
& + \beta_{14} (\mathrm{AE}_{i13l} \times \mathrm{Ontherapy}_{ijk} \times \mathrm{trt}_i) \\
& + \beta_{15} (\mathrm{AE}_{i21l} \times \mathrm{Ontherapy}_{ijk} \times \mathrm{trt}_i)
\end{aligned} \tag{15}$$

where $\mathrm{Ontherapy}_{ijk} = 1$ if the ith subject experienced the kth AE in the jth body system during the on-therapy period; 0 otherwise.

As in a GEE analysis, misspecification of the working correlation structure may only affect the efficiency of estimation of the regression parameter β, but not the consistency of $\hat{\beta}$. Consequently, the estimates of the correlation parameters are mainly used to improve efficiency, rather than to make formal inferences regarding the association parameters.

Table 4 summarizes the model fit for models (13), (14), and (15) via SQLS using the KP working correlation structure Exchangeable(ρ_j)$\otimes$AR1(α). Model (13) shows a significant overall treatment effect (p-value < 0.001) which can be interpreted as follows: The overall expected number of AEs for the treatment arm is $\exp(0.358) = 1.43$ times higher than that for the control arm (95% CI: $1.27 \sim 1.61$).

Model (14) is one of the most important models to answer the primary goal of the AE analysis. In this model, we are mainly interested in the two-way interaction terms since the

significance of an interaction term indicates that there is a difference in the number of AEs for the treatment and the control arms, given the adverse event AE_{ij}. For example, from Model (14), the expected difference in log of the number of AEs for the treatment vs. the control arms with regards to AE_{11} can by computed as

$$\log\left(\frac{\mu_{ijkl}|\text{AE}_{i11l}=1,\text{trt}_i=1}{\mu_{ijkl}|\text{AE}_{i11l}=1,\text{trt}_i=0}\right)=\beta_5\text{trt}_i+\beta_6(\text{AE}_{i11l}\times\text{trt}_i) \tag{16}$$

while holding other variables constant. From Table 4, none of the interaction terms are significant at a 5% significance level. Therefore, we may conclude that there is no statistically strong evidence (or signal) that the number of AEs are different for two groups with respect to each type of AE.

Lastly, Model (15) can be used to further explore the three-way interactions between each type of AE, treatment, and on-therapy indicator variables. The expected difference in log of the number of AEs for the treatment vs. the control arms during the on-therapy period with regards to AE_{11} can by computed as

$$\begin{aligned}\log\left(\frac{\mu_{ijkl}|\text{AE}_{i11l}=1,\text{Ontherapy}_i=1,\text{trt}_i=1}{\mu_{ijkl}|\text{AE}_{i11l}=1,\text{Ontherapy}_i=1,\text{trt}_i=0}\right)&\\ =\beta_5\text{trt}_i+\beta_7(\text{AE}_{i11l}\times\text{trt}_i)+\beta_{11}(\text{Ontherapy}_{ijk}\times\text{trt}_i)&\\ +\beta_{12}(\text{AE}_{i11l}\times\text{Ontherapy}_{ijk}\times\text{trt}_i)&\end{aligned} \tag{17}$$

while holding other variables constant. Similarly, for the follow-up period, the expected difference in log of the number of AEs is given by

$$\log\left(\frac{\mu_{ijkl}|\text{AE}_{i11l}=1,\text{Ontherapy}_i=0,\text{trt}_i=1}{\mu_{ijkl}|\text{AE}_{i11l}=1,\text{Ontherapy}_i=0,\text{trt}_i=0}\right)=\beta_5\text{trt}_i+\beta_7(\text{AE}_{i11l}\times\text{trt}_i) \tag{18}$$

while holding other variables constant.

From Table 4, only the three-way interaction with respect to AE_{11}, i.e. $\hat{\beta}_{12}$ in Model (15), is marginally significant at a 5% significance level (p-values = 0.058). Using (17) and

(18), this suggests that the expected number of AEs for AE_{11} in the treatment arms is $\exp(0.841 + 1.596 - 0.580 - 1.584) = 1.31$ times greater than that that in the control arm during the on-therapy period whiles it is $\exp(0.841 + 1.596) = 11.4$ times greater during the follow-up period. This can also be visually checked from Figure 1, where the number of AEs for the treatment arm is much greater than in the control arm during the follow-up period, i.e. after the 2nd week.

5. Discussion

We described various regression strategies to model multiple longitudinal AEs data via the method of SQLS using the Exchangeable(ρ_j)$\otimes$AR1(α) working correlation structure. In particular, we allow a different correlation parameter ρ_j for each body system j assuming the Exchangeable structure, and a common correlation parameter α for correlation over time assuming the AR1 structure for all subjects. Consequently, SQLS fits a marginal model by borrowing information across subjects, within each subject as well as across AEs within each body system. This is one of the main advantages over conventional univariate methods such as the χ^2 test of independence, and the Fisher's exact test. In addition, our approach takes the number and type of AEs per subject over time into account, while application of the Fisher's exact test reduces the longitudinal follow-up information on each subject to whether or not they had a particular type of AE during the follow-up period.

As shown in the analysis of the Wyeth safety data in §4, SQLS regression models (13), (14), and (15) lead to a better understanding of the data by exploring the overall treatment effect, the two-way, and three-way interactions among factors that are of interest to the investigator(s) and/or regulatory agencies. These models allow useful interpretations that

may better characterize the difference between the treatment vs. the control arms with respect to each type of AE. A regression model may also adjust for the number of doses if this information is available in the raw data, e.g. higher doses may lead to more AEs and vice versa.

We have also developed a user-written SAS macro `%QLS version 2` for fitting SQLS regression models for continuous, binary, and count multivariate (unbalanced) longitudinal data using the Exchangeable(ρ_j)$\otimes$AR1(α) working correlation structure. Our macro is available for download from `http://www.cceb.upenn.edu/~sratclif/QLSproject.html`. (See Kim and Shults (2008) for a detailed description and demonstration of the software).

6. Acknowledgement

We are grateful to a statistical programmer Rudram Das, MS at Wyeth Research for providing us with the Wyeth safety Data.

References

Berry, S. and Berry, D. (2004) Accounting for Multiplicities in Assessing Drug Safety; A Three-Level Hierarchical Mixture Model. *Biometrics* **60**: 418–426.

Chaganty, N. (1997) An alternative approach to the analysis of longitudinal data via generalized estimating equations. *J. Statist. Plan. Inf.* **63**: 39–54.

Chaganty, N. and Naik, D. (2002) Analysis of multivariate longitudinal data using quasi-least squares. *J. Statist. Plan. Inf.* **103**: 421-436.

Chaganty, N. and Shults, J. (1999) On eliminating the asymptotic bias in the quasi-least squares estimate of the correlation parameter. *J. Statist. Plan. Inf.* **76**: 127-144.

Chow S. and Liu J. (2003) Design and Analysis of Clinical Trials - Revised and Expanded, Second Edition. John Wiley and Sons, New York, New York.

Galecki, A. (1994). General class of covariance structures for two or more repeated factors in longitudinal data analysis. *Communications in Statistics- Theory and Methods* **23**: 3105-3120.

Goldberg-Alberts, R. and Page, S. (2006) Multivariate Analysis of Adverse Events. *Drug Information Journal* **40**: 99–110.

Kim, H. and Shults, J. (2008) Analysis of unbalanced multi-outcome longitudinal data in SAS *UPenn Biostatistics Working Papers*. `http://biostats.bepress.com/upennbiostat/papers`

Liang, K. and Zeger, S. (1986). Longitudinal data analysis using generalized linear models. *Biometrika* **73**: 13–22.

Lu, N. and Zimmerman, D. (2005). The likelihood ratio test for a seperable covariance matrix. *Statistics & Probability Letters* **73**: 449–457.

Mehrotra, D. and Heyse, J. (2004). Multiplicity considerations in clinical safety analyses. *Statistical Methods in Medical Research* **13**: 227–238.

Naik, D. and Rao, S. (2001). Analysis of multivariate repeated measures data with a Kronecker product structured covariance matrix, *Journal of Applied Statistics* **28**: 91-105.

MedDRA: Medical Dictionary for Regulatory Activities / Introduction. Available at `http://www.ich.org`

Roy, A. and Khattree, R. (2005). Testing the hypothesis of a Kroneckar product covariance matrix in multivariate repeated measures data. *SAS Users Group International,*

Proceedings of the Statistics and Data Analysis Section. **Paper 199-30**: 1–11.

Roy, A. (2006). A new classification rule for incomplete doubly multivariate data using mixed effects model with performance Comparisons on the Imputed Data. *Statistics in Medicine* **25(10)**: 1715–1728.

Roy, A. and Leiva, R. (2007). Discrimination with jointly equicorrelated multi-level multivariate data. *Advances in Data Analysis and Classification* **1(3)**: 175–199.

Schildcrout, J., Jenkins, C., Ostroff, J., Gillen., D., Harrel, F. and Trost, D. (2008) Analysis of longitudinal laboratory data in the presence of common selection mechanisms: A view toward greater emphasis on pre-marketing pharmaceutical safety. *Statistics in Medicine* **27**: 2248–2266.

Shults, J. and Morrow, A. (2002) Use of quasi-least squares to adjust for two levels of correlation. *Biometrics* **58**, 521–530.

Shults, J., Whitt, C. and Kumanyika, S. (2004) Analysis of data with multiple sources of correlation in the framework of generalized estimating equations. *Statistics in Medicine* **23**(20), 3209–3226.

Shults, J. and Radcliffe, S. (2007) Analysis of multi-level correlated data in the framework of generalized estimating equations via xtmultcorr procedures in Stata and qls functions in Matlab. *UPenn Biostatistics Working Papers* Working Paper 15. `http://biostats.bepress.com/upennbiostat/papers/art15` (This paper is also in press at *Statistics and Its Interface.*)