Biometrika (2012), **99**, 3, *pp.* 755–762
doi: 10.1093/biomet/ass021

Printed in Great Britain

Quadratic inference function approach to merging longitudinal studies: validation and joint estimation

By FEI WANG, LU WANG and PETER X.-K. SONG

Department of Biostatistics, University of Michigan, Ann Arbor, Michigan 48109, U.S.A.

wafei@umich.edu luwang@umich.edu pxsong@umich.edu

Summary

Merging data from multiple studies has been widely adopted in biomedical research. In this paper, we consider two major issues related to merging longitudinal datasets. We first develop a rigorous hypothesis testing procedure to assess the validity of data merging, and then propose a flexible joint estimation procedure that enables us to analyse merged data and to account for different within-subject correlations and follow-up schedules in different studies. We establish large sample properties for the proposed procedures. We compare our method with meta analysis and generalized estimating equations and show that our test provides robust control of Type I error against both misspecification of working correlation structures and heterogeneous dispersion parameters. Our joint estimating procedure leads to an improvement in estimation efficiency on all regression coefficients after data merging is validated.

Some key words: Data merging; Estimation efficiency; Generalized estimating function; Heterogeneity; Meta analysis.

1. Introduction

Merging data from clinical trials or longitudinal cohort studies with identical or similar protocols can offer a powerful way to better understand effects of treatment and exposure on patient outcomes (e.g., Localio et al., 2001; Xie & Ahn, 2010). Appropriate data merging can increase statistical power. A well-known approach to this is meta analysis (e.g., Becker, 2007; Hartung et al., 2008) but this often utilizes summary statistics from individual analysis, with no or little justification provided on the validity of data merging. When the original datasets are fully available, a statistical model incorporating interaction terms between studies and covariates of interest may be used to characterize different effect sizes of covariates across studies. However, in such analysis most existing approaches use a common correlation structure and a common dispersion parameter for different studies. According to Crowder (1995), misspecification of working correlation structures, particularly for multiple longitudinal studies, may inflate Type I errors and distort power.

Another widely used approach is to model cross-study heterogeneity of regression coefficients through a mixed-effects model (e.g., Laird & Ware, 1982; Zhang et al., 2009). This requires a relatively large number of studies and correct distribution assumptions in order to adequately estimate the cross-study variability (Follmann & Proschan, 1999). In addition, the general theory regarding tests for nonzero variance components is difficult to apply (Stram & Lee, 1994; Crainiceanu & Ruppert, 2004), especially for nonnormal data (Zhang & Lin, 2008).

Breslow & Day's (1980) test for homogeneity of conditional odds ratios is a classical example of validation prior to the calculation of the common odds ratio for multiple strata. In this paper, we consider longitudinal studies that collect the same types of variables under similar protocols. We develop a novel quadratic inference function (Qu et al., 2000) strategy to validate longitudinal data merging by testing for the unbiasedness of the generalized estimating functions under a common set of regression coefficients but with possibly different covariance structures. The unbiasedness implies that study-specific estimating functions are compatible with a shared regression mean model, so that the resulting analysis of merged data would lead to consistent estimators of regression coefficients and robust control of Type I error against covariance misspecification.

2. FORMULATION

We consider $K \geqslant 2$ longitudinal studies that collect a common set of variables under similar study protocols. Let $Y_{k,ij}$ be the outcome for the jth observation of subject i in study k, and let $X_{k,ij}$ be the corresponding covariate vector for $i = 1, \ldots, n_k$, $j = 1, \ldots, m_k$ and $k = 1, \ldots, K$, where n_k and m_k are the numbers of subjects and the numbers of observations on each subject in study k, respectively. We assume a marginal model for outcome $Y_{k,ij}$, consisting of conditional mean model $E(Y_{k,ij} \mid X_{k,ij}) = \mu_{k,ij} = h(X_{k,ij}^{\mathrm{T}}\beta_k)$ and conditional variance $\mathrm{var}(Y_{k,ij} \mid X_{k,ij}) = \phi_k v(\mu_{k,ij})$, where $h(\cdot)$ is a known link function, β_k is a p-dimensional regression parameter, $v(\cdot)$ is the variance function, a known function of the mean, and ϕ_k is a dispersion parameter. The within-subject correlation is accommodated via a working correlation matrix $R_k(\alpha_k)$, as suggested by Zeger et al. (1988), where α_k is the correlation parameter of study k.

For study k, an estimator of β_k from generalized estimating equations solves

$$n_k^{-1} \sum_{i=1}^{n_k} D_{k,i}^{\mathrm{T}} A_{k,i}^{-1/2} R_k^{-1}(\alpha_k) A_{k,i}^{-1/2} (Y_{k,i} - \mu_{k,i}) = 0 \quad (k = 1, \ldots, K), \tag{1}$$

where $Y_{k,i} = (Y_{k,i1}, \ldots, Y_{k,im_k})^{\mathrm{T}}$, $\mu_{k,i} = (\mu_{k,i1}, \ldots, \mu_{k,im_k})^{\mathrm{T}}$, $D_{k,i} = \partial\mu_{k,i}/\partial\beta_k^{\mathrm{T}}$ and $A_{k,i} = \mathrm{diag}\{v(\mu_{k,i1}), \ldots, v(\mu_{k,im_k})\}$. To deal with merged data, we propose to use the quadratic inference function method (Qu et al., 2000) to join the study-specific estimating functions. A quadratic inference function is derived via approximating the inverse working correlation matrix by $R_k^{-1}(\alpha_k) \approx \sum_{s=1}^{s_k} a_{k,s} M_{k,s}$ for $k = 1, \ldots, K$, where $a_{k,1}, \ldots, a_{k,s_k}$ are constants possibly dependent on α_k and $M_{k,1}, \ldots, M_{k,s_k}$ are known basis matrices with elements 0 and 1, which are determined by a given correlation matrix $R_k(\alpha_k)$. Qu et al. (2000) give more details on the forms of basis matrices for some widely used correlation matrices. Inserting the expansion of $R_k^{-1}(\alpha_k)$ into (1) leads to

$$n_k^{-1} \sum_{i=1}^{n_k} \sum_{j=1}^{s_k} a_{k,j} D_{k,i}^{\mathrm{T}} A_{k,i}^{-1/2} M_{k,j} A_{k,i}^{-1/2} (Y_{k,i} - \mu_{k,i}) = 0 \quad (k = 1, \ldots, K),$$

which may be regarded as a combination of elements of the extended score vector

$$\bar{g}_k(\beta_k) = n_k^{-1} \sum_{i=1}^{n_k} g_{k,i}(\beta_k) = n_k^{-1} \sum_{i=1}^{n_k} \begin{pmatrix} D_{k,i}^{\mathrm{T}} A_{k,i}^{-1/2} M_{k,1} A_{k,i}^{-1/2} (Y_{k,i} - \mu_{k,i}) \\ \vdots \\ D_{k,i}^{\mathrm{T}} A_{k,i}^{-1/2} M_{k,s_k} A_{k,i}^{-1/2} (Y_{k,i} - \mu_{k,i}) \end{pmatrix} \quad (k = 1, \ldots, K).$$

Unlike generalized estimating equations, a quadratic inference function needs no estimates of nuisance coefficients $a_{k,1}, \ldots, a_{k,s_k}$ in order to estimate parameters $\beta = (\beta_1^{\mathrm{T}}, \ldots, \beta_K^{\mathrm{T}})^{\mathrm{T}}$ of interest.

Define the study indicator $\delta_i(k)$, with 1 indicating that subject i belongs to study k and 0 otherwise. For the merged longitudinal data, β can be estimated by $\hat{\beta} = \arg\min_{\beta} Q(\beta)$, where

$$Q(\beta) = n\bar{g}(\beta)^{\mathrm{T}} C^{-}(\beta) \bar{g}(\beta), \tag{2}$$

with $n = \sum_{k=1}^{K} n_k$, and

$$\bar{g}(\beta) = n^{-1} \sum_{i=1}^{n} \{\delta_i(1) g_{1,i}(\beta_1)^{\mathrm{T}}, \ldots, \delta_i(K) g_{K,i}(\beta_K)^{\mathrm{T}}\}^{\mathrm{T}} = n^{-1} \sum_{i=1}^{n} g_i(\beta),$$

$$C(\beta) = n^{-1} \sum_{i=1}^{n} \mathrm{diag}\{\delta_i(1) g_{1,i}(\beta_1) g_{1,i}(\beta_1)^{\mathrm{T}}, \ldots, \delta_i(K) g_{K,i}(\beta_K) g_{K,i}(\beta_K)^{\mathrm{T}}\}.$$

Here $C(\beta)$ is a block-diagonal matrix under the assumption of mutually independent study cohorts, which however may be relaxed in the case of related cohorts. We adopt the unique Moore–Penrose generalized inverse in equation (2) to enhance numerical stability, as the matrix $C(\beta)$ may become singular in some cases (Hu & Song, 2011). See Lemma 9.2.6 in (Harville, 2008) for the construction of such an inverse operation.

3. Homogeneity test

We develop methods of hypothesis test for global and partial homogeneity of regression parameters across multiple studies. By homogeneity, we mean the equality of regression parameters across all studies, including global homogeneity $\beta_1 = \cdots = \beta_K$ or equality on a subset of coefficients for partial homogeneity. To derive asymptotic distributions of the proposed test statistics, we assume the study-specific mean models are correctly specified, so β_k can be consistently estimated in the corresponding individual study k.

Let $\mathcal{M} \subset \{1, \ldots, p\}$ denote an index set, and then $|\mathcal{M}|$ denotes the number of elements in $\mathcal{M}$. Accordingly, $\beta_k(\mathcal{M})$ and $\beta_k(\mathcal{M}^c)$ are subsets of parameters indexed by $\mathcal{M}$ and its complementary set $\mathcal{M}^c$, respectively. Clearly, set $\mathcal{M} = \{1, \ldots, p\}$ leads to global homogeneity, while partial homogeneity is given by $\mathcal{M}$ being a certain subset of $\{1, \ldots, p\}$.

To test the hypothesis $H_0 : \beta_1(\mathcal{M}) = \cdots = \beta_K(\mathcal{M})$ against $H_a : \beta_i(\mathcal{M}) \neq \beta_j(\mathcal{M})$ for some $i \neq j$ and $i, j \in \{1, \ldots, K\}$, let $\Omega_0(\mathcal{M}) = \{(\beta_1^{\mathrm{T}}, \ldots, \beta_K^{\mathrm{T}})^{\mathrm{T}} : \beta_1(\mathcal{M}) = \cdots = \beta_K(\mathcal{M}), \beta_k \in R^p, k = 1, \ldots, K\}$ be the null parameter space under H_0, and let Ω be the whole parameter space. Estimators of β under $\Omega_0(\mathcal{M})$ and under Ω are, respectively

$$\hat{\beta}_{\Omega_0(\mathcal{M})} = \underset{\beta \in \Omega_0(\mathcal{M})}{\arg\min}\, Q(\beta), \quad \hat{\beta}_{\Omega} = \underset{\beta \in \Omega}{\arg\min}\, Q(\beta),$$

where $Q(\beta)$ is given by (2) with the corresponding parameterization. Under H_0, Theorem 1 establishes the asymptotic distribution of $Q(\hat{\beta}_{\Omega_0(\mathcal{M})})$.

Theorem 1. *Let $\hat{\beta}_{\Omega_0(\mathcal{M})}$ be a root-n consistent estimator of the true parameter β_0 under H_0. Suppose the following regularity conditions hold:* (a) *β_0 lies in the interior of $\mathcal{B} \subset \mathcal{R}^{Kp-(K-1)|\mathcal{M}|}$, and $\mathcal{B}$ is compact;* (b) *$g_i(\beta)$ is continuously differentiable in a neighbourhood $\mathcal{N}$ of β_0;* (c) *$E\{g_i(\beta)\} = 0$ if and only if $\beta = \beta_0$ and $E\{||g_i(\beta_0)||^2\}$ is finite;* (d) *$E\{\sup_{\beta \in \mathcal{N}} ||\partial g_i(\beta)/\partial \beta^{\mathrm{T}}||\} < \infty$;* (e) *$n^{1/2}\bar{g}(\beta_0)$ converges to $N(0, \Sigma)$ in distribution, where Σ is a block-diagonal matrix, $\Sigma = \mathrm{diag}(\rho_1^{-1}\Sigma_1, \ldots, \rho_K^{-1}\Sigma_K)$, with $\Sigma_k = \mathrm{cov}\{g_{k,i}(\beta_0)\}$ and $\rho_k = \lim_{n\to\infty} n_k/n$ for $k = 1, \ldots, K$;* (f) *$G^{\mathrm{T}}\Sigma^{-}G$ is nonsingular, where $G = E\{\partial g_i(\beta_0)/\partial \beta^{\mathrm{T}}\}$; and* (g) *$\Sigma\Sigma^{-}G = G$. Then $Q(\hat{\beta}_{\Omega_0(\mathcal{M})})$ converges in distribution to $\chi^2_{\mathrm{rank}(\Sigma)-Kp+(K-1)|\mathcal{M}|}$.*

Here $||\cdot||$ denotes the Euclidean norm. The proof of Theorem 1 is given in the Appendix. Since Σ may not be of full rank, the degrees of freedom of $Q(\hat{\beta}_{\Omega_0(\mathcal{M})})$ take the form of $\mathrm{rank}(\Sigma) - Kp + (K-1)|\mathcal{M}|$, where $\mathrm{rank}(\Sigma)$ can be estimated from orthogonal triangularization of an estimated Σ.

When all study-specific mean models are correctly specified, $\hat{\beta}_{\Omega}$ is a root-n consistent estimator of β_0. Under the null hypothesis, $\bar{g}(\beta)$ is an unbiased estimating function for $\beta \in \Omega_0(\mathcal{M})$, so we can obtain another root-n consistent estimator $\hat{\beta}_{\Omega_0(\mathcal{M})}$ of β_0. Therefore, we propose two test statistics. The first is $Q(\hat{\beta}_{\Omega_0(\mathcal{M})})$, mimicking a score test statistic, denoted as $\hat{Q}_S$. Its asymptotic chi-square distribution under H_0 is shown in Theorem 1. The second is $Q(\hat{\beta}_{\Omega_0(\mathcal{M})}) - Q(\hat{\beta}_{\Omega})$, which resembles a likelihood ratio test statistic, denoted as $\hat{Q}_{LR}$. The asymptotic distribution of $\hat{Q}_{LR}$ is given as follows.

Corollary 1. *Under the regularity conditions in Theorem 1, $\hat{Q}_{LR}$ converges in distribution to $\chi^2_{(K-1)|\mathcal{M}|}$.*

4. Joint estimation with merged data

When either global or partial homogeneity is established, the merged data will lead to efficiency improvement in estimation for both common and study-specific regression coefficients. To elucidate this without loss of generality we consider partial homogeneity. Let ζ denote a vector of common coefficients for covariates $X_{k,ij}$ shared by the studies and let γ_k denote study-specific parameters associated with covariates $Z_{k,ij}$. Then $\beta = (\zeta^{\mathrm{T}}, \gamma_1^{\mathrm{T}}, \ldots, \gamma_K^{\mathrm{T}})^{\mathrm{T}}$ represents all the parameters and $\beta_k = (\zeta^{\mathrm{T}}, \gamma_k^{\mathrm{T}})^{\mathrm{T}}$ contains those in study k only. Accordingly, the mean model is rewritten as $E(Y_{k,ij} \mid X_{k,ij}, Z_{k,ij}) = \mu_{k,ij} = h(X_{k,ij}^{\mathrm{T}}\zeta + Z_{k,ij}^{\mathrm{T}}\gamma_k)$ $(k = 1, \ldots, K)$. Consequently, we obtain a consistent estimate $\hat{\beta}$ by minimizing the function in (2) with the merged data. Under Assumptions (a)–(f) of Theorem 1, as shown in the Appendix,

$n^{1/2}(\hat{\beta} - \beta_0)$ converges in distribution to $N\{0, (G^{\mathrm{T}}\Sigma^{-}G)^{-1}\}$ with Σ defined in Theorem 1 and

$$G = E\left\{\frac{\partial g_i(\beta_0)}{\partial \beta^{\mathrm{T}}}\right\} = \begin{pmatrix} G_1 \\ \vdots \\ G_K \end{pmatrix}$$

$$= \begin{bmatrix} E\left\{\frac{\partial g_{1,i}(\beta_0)}{\partial \zeta^{\mathrm{T}}}\right\} & E\left\{\frac{\partial g_{1,i}(\beta_0)}{\partial \gamma_1^{\mathrm{T}}}\right\} & \cdots & \cdots & 0 \\ E\left\{\frac{\partial g_{2,i}(\beta_0)}{\partial \zeta^{\mathrm{T}}}\right\} & 0 & E\left\{\frac{\partial g_{2,i}(\beta_0)}{\partial \gamma_2^{\mathrm{T}}}\right\} & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ E\left\{\frac{\partial g_{K,i}(\beta_0)}{\partial \zeta^{\mathrm{T}}}\right\} & 0 & \cdots & \cdots & E\left\{\frac{\partial g_{K,i}(\beta_0)}{\partial \gamma_K^{\mathrm{T}}}\right\} \end{bmatrix}.$$

To find the efficiency gain in the merged data analysis, we focus on β_k. Let $\tilde{\beta}_k$ be an estimator obtained by minimizing the function in (2) using only the kth study data and let $\hat{\beta}_k$ be the subvector of $\hat{\beta}$, obtained with the merged data. The asymptotic variance for $\tilde{\beta}_k$ is $\{\rho_k(G_k^{\mathrm{T}}\Sigma_k^{-}G_k)_{[\beta_k,\beta_k]}\}^{-1}$, where G_k is the kth block-row of matrix G above, Σ_k is defined in Theorem 1, and $B_{[\beta_k,\beta_k]}$ denotes the sub-block matrix of B with rows and columns selected by those elements corresponding to β_k. The asymptotic variance for $\hat{\beta}_k$ is $\{(G^{\mathrm{T}}\Sigma^{-}G)^{-1}\}_{[\beta_k,\beta_k]}$. Theorem 2 below establishes the efficiency improvement achieved through the joint estimation with merged data.

THEOREM 2. *Suppose that $(G_k^{T}\Sigma_k^{-}G_k)_{[\beta_k,\beta_k]}$ is positive definite for study k, $k = 1, \ldots, K$. Then the asymptotic variances of $\hat{\beta}_k$ and $\tilde{\beta}_k$ satisfy*

$$\{(G^{\mathrm{T}}\Sigma^{-}G)^{-1}\}_{[\beta_k,\beta_k]} \preceq \frac{1}{\rho_k}\{(G_k^{\mathrm{T}}\Sigma_k^{-}G_k)_{[\beta_k,\beta_k]}\}^{-1} \quad (k = 1, \ldots, K),$$

where $\preceq$ is in the sense of Löwner's partial ordering in the space of nonnegative definite matrices.

The proof of Theorem 2 is given in the Appendix. Theorem 2 implies that the asymptotic variance of $\hat{\beta}_k$ is not larger than that of $\tilde{\beta}_k$. This suggests that the estimation with the merged data is not only flexible enough to accommodate different study-specific correlations and follow-up schedules but also leads to estimation efficiency gain on the regression coefficients. This efficiency benefit is not easily achieved by meta analysis, in which the effective sample size is not really increased from combining individual analyses. Moreover, when additional nuisance parameters are introduced into the joint estimation procedure in generalized estimating equations to account for study-specific covariance parameters, the efficiency gain is not guaranteed for the estimation of parameters of interest. This is because even though the merged data have more samples, the number of nuisance parameters increases too, which can offset the benefit from the increased sample size.

5. SIMULATION STUDY

Two simulation studies were conducted to investigate the finite sample performance of our proposed tests and to compare them with Wald-type tests using the method of generalized estimating equations. We consider several Wald-type test statistics, denoted by W_{zla}, W_{p}, W_{md} and W_{wl}. They are computed by using different sandwich variance estimators, proposed by Zeger et al. (1988), Pan (2001), Mancl & DeRouen's (2001) and Wang & Long (2011), respectively. Wald-type tests are applied to test for no interactions between study dummy covariates and covariates of interest under a common correlation structure for multiple studies. Technically speaking, these approaches may be modified to accommodate study-specific covariance matrices, but the resulting methods require iteratively updating regression parameters and study-specific covariance nuisance parameters, so their performances will be affected by the estimation of nuisance parameters. In this paper we do not implement such extended estimating equation approaches but focus on using robust sandwich variance estimators to account for covariance heterogeneity across multiple studies.

For meta analysis, we adopt Cochran's test for partial homogeneity. According to Hartung et al. (2008), this is approximately distributed as χ^2_{K-1} under the null hypothesis of homogeneous coefficients across all K studies. Similarly as for the Wald-type tests, we use T_{zla}, T_{p}, T_{md} and T_{wl} to denote Cochran's test statistics with robust sandwich variance estimators.

We also consider a homogeneity test using mixed-effects models by testing zero variance components of random slopes. In the linear mixed-effects model, the asymptotic distribution of a likelihood ratio test for one zero variance component is $0{\cdot}5\chi^2_0 + 0{\cdot}5\chi^2_1$ (Stram & Lee, 1994), while in a generalized linear mixed model with nonidentity link functions, such mixtures of chi-squares for likelihood ratio tests are hard to obtain (Fitzmaurice et al., 2007; Sinha, 2009). In our simulation studies, because data are generated by the population-average model with some prefixed correlation structures, tests for zero variance components cannot control Type I error at all. Thus, we do not include results from the mixed-effects models in the comparison.

The first simulation study is generated by a population-average linear model $Y_{k,ij} = \beta_{k,0} + \beta_{k,1}X_{k,ij} + \beta_{k,2}Z_{k,ij} + \varepsilon_{k,ij}$ for $j = 1, \ldots, m_k$, $i = 1, \ldots, n_k$ with $m_k = 8$, $n_k = 100$ and $k = 1, \ldots, K$. The covariate $Z_{k,i} = (Z_{k,i1}, \ldots, Z_{k,im_k})^{\mathrm{T}}$ is a time-dependent variable simulated from a multivariate normal distribution with mean vector $(1, \ldots, m_k)^{\mathrm{T}}$ and the identity covariance matrix I_{m_k}. The covariate $X_{k,i} = (X_{k,i1}, \ldots, X_{k,im_k})^{\mathrm{T}}$ is a time-independent, baseline covariate generated from an exponential distribution with rate parameter 4. The error terms $\varepsilon_{k,i} = (\varepsilon_{k,i1}, \ldots, \varepsilon_{k,im_k})^{\mathrm{T}}$ follow $N\{0, \phi_k R_k(\alpha_k)\}$ with correlation matrix $R_k(\alpha_k)$. Denote all correlation parameters and dispersion parameters by $\alpha = (\alpha_1, \ldots, \alpha_K)^{\mathrm{T}}$ and $\phi = (\phi_1, \ldots, \phi_K)^{\mathrm{T}}$, respectively, and denote the order-1 autoregressive and compound symmetric correlations by R_{AR} and R_{CS} respectively. We consider three cases: (i) $K = 4$, $\phi = (50, 10, 4, 1)^{\mathrm{T}}$, $\alpha = (0{\cdot}7, 0{\cdot}4, 0{\cdot}2, 0{\cdot}1)^{\mathrm{T}}$, and $\{R_1(\cdot), R_2(\cdot), R_3(\cdot), R_4(\cdot)\} = \{R_{\text{AR}}, R_{\text{CS}}, R_{\text{CS}}, R_{\text{AR}}\}$; (ii) $K = 3$, $\phi = (10, 4, 1)^{\mathrm{T}}$, $\alpha = (0{\cdot}7, 0{\cdot}2, 0{\cdot}1)^{\mathrm{T}}$, $\{R_1(\cdot), R_2(\cdot), R_3(\cdot)\} = \{R_{\text{AR}}, R_{\text{CS}}, R_{\text{AR}}\}$; and (iii) $K = 2$, $\phi = (10, 1)^{\mathrm{T}}$, $\alpha = (0{\cdot}7, 0{\cdot}2)^{\mathrm{T}}$, and $\{R_1(\cdot), R_2(\cdot)\} = \{R_{\text{AR}}, R_{\text{CS}}\}$. Let $\beta_k = (\beta_{k,0}, \beta_{k,1}, \beta_{k,2})^{\mathrm{T}}$ for $k = 1, \ldots, K$. We are interested in a global test $H_0 : \beta_1 = \cdots = \beta_K$ and a partial test concerning only the coefficients of $X_{k,ij}$, $H_0 : \beta_{1,1} = \cdots = \beta_{K,1}$. Type I errors are computed with $\beta_k = (-1, -2, 3)^{\mathrm{T}}$ for $k = 1, \ldots, K$, while power is calculated under $\beta_1 = \beta_3 = \beta_4 = (-1, -2, 3)^{\mathrm{T}}$ and $\beta_2 = (-1, -1{\cdot}85, 3)^{\mathrm{T}}$. In the use of Wald-type tests for zero interaction effects between covariates and study indicators, only coefficients of interaction terms will be involved in the test.

Table 1 summarizes Type I errors and power of all test statistics at a significance level $0{\cdot}05$ over 4000 replications. For a fair comparison, we compute Wald-type tests, meta analyses and our proposed tests under a common correlation structure, R_{AR} or R_{CS}. The results clearly show that no matter which working correlation structure is used, our proposed tests, $\hat{Q}_{LR}$ and $\hat{Q}_S$, can properly control Type I error rates. In contrast, Wald-type tests and meta analyses cannot, particularly for global homogeneity tests and for $K > 2$. Wald-type tests have inflated Type I error rates, mostly because the modified robust variance estimators underestimate variances of regression coefficients and cannot fully account for differences among covariance structures across studies. Meta analyses appear to have fewer inflated Type I errors than Wald-type tests, but since meta analyses cannot sufficiently utilize all data information, they tend to have lower power.

The second simulation study concerns a binary outcome $Y_{k,ij}$, which follows a population-average logistic model $\text{logit}\{E(Y_{k,ij} \mid Z_{k,ij})\} = \beta_{k,0} + \beta_{k,1}Z_{k,ij}$ for $j = 1, \ldots, 8$, $i = 1, \ldots, 100$ and $k = 1, \ldots, K$, with $Z_{k,ij}$ generated from $\text{Un}(0, 1)$ distribution. We consider the same global and partial homogeneity hypotheses as in the first simulation study. Within-subject correlations are specified in three cases: (i) $K = 4$, $\alpha = (0{\cdot}7, 0{\cdot}4, 0{\cdot}2, 0{\cdot}1)^{\mathrm{T}}$, and $\{R_1(\cdot), R_2(\cdot), R_3(\cdot), R_4(\cdot)\} = \{R_{\text{AR}}, R_{\text{CS}}, R_{\text{CS}}, R_{\text{AR}}\}$; (ii) $K = 3$, $\alpha = (0{\cdot}7, 0{\cdot}2, 0{\cdot}1)^{\mathrm{T}}$, and $\{R_1(\cdot), R_2(\cdot), R_3(\cdot)\} = \{R_{\text{AR}}, R_{\text{CS}}, R_{\text{AR}}\}$; and (iii) $K = 2$, $\alpha = (0{\cdot}7, 0{\cdot}2)^{\mathrm{T}}$, and $\{R_1(\cdot), R_2(\cdot)\} = \{R_{\text{AR}}, R_{\text{CS}}\}$. Type I errors are computed with $\beta_k = (-0{\cdot}2, 1{\cdot}5)^{\mathrm{T}}$ for all $k = 1, \ldots, K$, while the power is calculated under $\beta_2 = (-0{\cdot}2, 2{\cdot}5)^{\mathrm{T}}$ and $\beta_k = (-0{\cdot}2, 1{\cdot}5)^{\mathrm{T}}$ for $k \neq 2$.

Table 2 presents results summarized over 4000 replications at significance level $0{\cdot}05$. Similar conclusions are drawn to those obtained in the case of continuous outcomes. Wald-type tests and meta analyses both produce inflated Type I errors. For instance, the Type I error of W_{p} appears to be above 7% when $K = 4$ studies are considered, regardless of the working correlation structure used. Among all Wald-type

Table 1. *Average Type I error rates and power of test statistics of the proposed* (Qs)*, Wald-type* (Ws) *and meta-based* (Ts) *versions over* 4000 *replications for continuous outcomes from* K *studies. Upper and lower panels correspond to the global and partial homogeneity tests respectively. Two correlations are used: order-*1 *autoregression,* R_{AR}*, and compound symmetry,* R_{CS}

	$K=4$				$K=3$				$K=2$			
	Size %		Power %		Size %		Power %		Size %		Power %	
Test	R_{AR}	R_{CS}	R_{AR}	R_{CS}	R_{AR}	R_{CS}	R_{AR}	R_{CS}	R_{AR}	R_{CS}	R_{AR}	R_{CS}
$\hat{Q}_S$	4·1	4·1	27·0	38·7	4·8	4·6	85·4	95·2	4·5	4·0	50·5	65·7
$\hat{Q}_{LR}$	4·0	4·2	42·4	43·4	4·4	4·1	95·9	96·6	4·0	4·3	73·6	72·3
W_{zla}	9·6	9·6	79·0	80·2	8·8	8·3	99·9	100	7·1	7·4	91·3	90·1
W_{p}	11·4	12·1	42·5	40·0	8·5	9·2	95·2	94·8	5·8	5·9	90·8	89·4
W_{md}	6·7	7·0	71·2	73·3	6·6	6·6	99·8	99·9	5·7	6·0	88·0	87·0
W_{wl}	10·2	11·0	39·4	36·8	7·5	8·2	94·6	94·0	5·4	5·2	89·4	88·8
$\hat{Q}_{LR}$	4·7	4·6	71·1	72·2	5·0	4·7	98·8	98·8	5·3	5·3	87·4	87·0
W_{zla}	8·4	8·2	88·8	90·0	7·4	7·3	100	100	7·2	7·1	96·5	95·5
W_{p}	9·0	8·8	62·1	60·4	7·2	7·6	99·4	99·5	6·1	6·0	96·4	95·6
W_{md}	6·2	6·4	85·0	86·6	5·4	5·6	100	100	5·5	5·8	95·0	94·1
W_{wl}	8·5	8·4	60·5	58·7	6·9	7·2	99·3	99·5	5·5	5·4	96·1	95·4
T_{zla}	9·5	9·8	65·0	65·7	8·2	8·2	99·2	99·1	7·6	7·6	78·7	76·9
T_{p}	5·7	5·4	60·2	60·6	5·7	5·8	99·2	99·2	5·4	5·3	77·0	74·9
T_{md}	6·2	6·5	56·4	57·0	5·6	6·0	98·6	98·6	5·7	5·8	72·7	71·8
T_{wl}	4·9	4·8	57·9	58·6	5·1	5·0	99·0	99·0	4·8	4·7	75·2	73·5

Table 2. *Average Type I error rates and power of test statistics of the proposed* (Qs)*, Wald-type* (Ws) *and meta-based* (Ts) *versions over* 4000 *replications for binary outcomes from* K *studies*

	$K=4$				$K=3$				$K=2$			
	Size %		Power %		Size %		Power %		Size %		Power %	
Test	R_{AR}	R_{CS}	R_{AR}	R_{CS}	R_{AR}	R_{CS}	R_{AR}	R_{CS}	R_{AR}	R_{CS}	R_{AR}	R_{CS}
$\hat{Q}_S$	4·8	4·9	65·2	66·9	5·0	4·9	79·8	79·6	5·1	5·4	57·0	51·0
$\hat{Q}_{LR}$	5·2	5·1	86·0	82·6	4·8	4·9	92·0	91·3	5·1	5·5	78·7	71·0
W_{zla}	5·2	5·5	73·6	76·6	5·5	4·9	89·3	90·5	6·0	5·9	78·6	76·5
W_{p}	7·8	8·1	81·6	81·6	7·6	6·1	90·1	89·4	6·3	6·6	79·5	77·5
W_{md}	4·5	4·8	71·6	75·2	5·0	4·6	88·6	89·7	5·5	5·4	77·6	75·7
W_{wl}	7·4	7·8	80·5	80·7	7·1	5·9	89·5	88·9	6·0	6·3	79·1	77·0
$\hat{Q}_{LR}$	5·3	5·0	88·2	89·4	5·2	5·3	95·6	95·4	5·0	4·9	86·0	80·8
W_{zla}	5·1	4·9	82·5	85·9	5·4	5·2	94·4	95·2	5·5	4·8	85·2	83·4
W_{p}	6·4	6·8	87·9	89·6	6·6	6·6	94·6	95·2	5·6	5·8	85·9	84·3
W_{md}	4·7	4·6	81·5	85·0	5·2	4·9	94·1	94·8	5·2	4·3	84·8	82·5
W_{wl}	6·1	6·7	87·5	89·2	6·2	6·4	94·4	95·1	5·4	5·5	85·7	84·0
T_{zla}	5·1	5·4	58·8	65·4	5·4	5·4	70·8	69·3	5·0	5·0	63·2	52·2
T_{p}	10·4	10·2	58·7	64·6	5·7	4·8	70·4	69·2	4·8	4·7	63·2	52·0
T_{md}	4·8	4·8	56·8	63·8	4·9	5·0	69·5	68·3	4·6	4·7	62·4	51·1
T_{wl}	10·3	10·0	58·2	64·4	5·5	4·7	70·1	68·8	4·7	4·6	62·9	51·8

tests, the one based on Mancl & DeRouen's (2001) sandwich variance estimator, W_{md}, seems to have a reasonable control of Type I error. To deal with the violation of a common correlation structure, their method strives to reduce the bias in estimation of the covariance matrix while the other methods focus on improving correlation matrix estimation. To compare the power of Mancl & DeRouen's test to our test, a ratio, (Power of W_{md})/(Power of $\hat{Q}_{LR}$), decreases as the number of studies increases, dropping from 98·6% in the case of two studies to 88·8% in the case of four studies for the global homogeneity test under R_{AR} working correlation. This implies that although Mancl & DeRouen's method can correct for the bias in the covariance estimation, it is inferior to $\hat{Q}_{LR}$ in terms of power. Since meta analysis does not utilize data from individual studies efficiently, it has lower power than our method even when its Type I error is properly controlled. Finally, since the degrees of freedom of a chi-square test statistic increase when the number of studies increases, our test statistics may lose power in the setting of many studies. In this scenario, we recommend using mixed-effects models to handle merged data if distribution assumptions for multiple studies can be properly prespecified.

Acknowledgement

The authors thank the editor, the associate editor and three anonymous referees for their constructive comments that led to an improvement of this paper. The third author's research is supported in part by the U.S. National Science Foundation.

Appendix

Proof of Theorem 1. Let $\hat\beta$ be a root-n consistent estimator for β_0. A Taylor expansion of $\bar g(\hat\beta)$ about β_0 gives $\bar g(\hat\beta) = \bar g(\beta_0) + \hat G(\beta^*)(\hat\beta - \beta_0)$, where β^* lies between $\hat\beta$ and β_0 and $\hat G(\beta^*) = \partial \bar g(\beta^*)/\partial\beta^{\mathrm{T}}$. Substituting this expression for $\bar g(\hat\beta)$ into $Q(\hat\beta)$, we have

$$Q(\hat\beta) = \|n^{1/2}\{C^-(\hat\beta)\}^{1/2}\{\bar g(\beta_0) + \hat G(\beta^*)(\hat\beta - \beta_0)\}\|^2, \tag{A1}$$

where $\|\cdot\|$ is the Euclidean norm. Another Taylor expansion of $\bar g(\hat\beta)$ about β_0 in the first-order condition of $\hat\beta$, $\partial Q(\hat\beta)/\partial\beta^{\mathrm{T}} = 0$, gives $\hat G(\hat\beta)^{\mathrm{T}} C^-(\hat\beta)\{\bar g(\beta_0) + \hat G(\beta^{**})(\hat\beta - \beta_0)\} + o_p(1) = 0$, where β^{**} is between β_0 and $\hat\beta$. Provided that $\hat G(\hat\beta)^{\mathrm{T}} C^-(\hat\beta)\hat G(\beta^{**})$ is nonsingular,

$$(\hat\beta - \beta_0) = -\{\hat G(\hat\beta)^{\mathrm{T}} C^-(\hat\beta)\hat G(\beta^{**})\}^{-1} a_n(\hat\beta)\bar g(\beta_0), \tag{A2}$$

where $a_n(\hat\beta) = \hat G(\hat\beta)^{\mathrm{T}} C^-(\hat\beta)$. Substituting (A2) for $\hat\beta - \beta_0$ into (A1) yields

$$Q(\hat\beta) = \|n^{1/2}\{C^-(\hat\beta)\}^{1/2}[I - \hat G(\beta^*)\{\hat G(\hat\beta)^{\mathrm{T}} C^-(\hat\beta)\hat G(\beta^{**})\}^{-1} a_n(\hat\beta)]\bar g(\beta_0)\|^2.$$

By assumptions (b) and (d) of Theorem 1 and Davidson (2001, Theorem 21.6), we obtain

$$\frac{\partial \bar g(\hat\beta)}{\partial\beta^{\mathrm{T}}} = G + o_p(1), \quad \frac{\partial \bar g(\beta^*)}{\partial\beta^{\mathrm{T}}} = G + o_p(1), \quad \frac{\partial \bar g(\beta^{**})}{\partial\beta^{\mathrm{T}}} = G + o_p(1), \quad C^-(\hat\beta) = \Sigma^- + o_p(1).$$

Assumptions (c) and (e) give $n^{1/2}\bar g(\beta_0) \to Y \sim N(0, \Sigma)$ in distribution, where Σ could be singular. The extended definition for multivariate normal distribution with singular covariance matrix is given by Definition 2.4.1 (Anderson, 2003). Then Slutsky's Theorem implies that

$$n^{1/2}\{C^-(\hat\beta)\}^{1/2}[I - \hat G(\beta^{**})\{\hat G(\hat\beta)^{\mathrm{T}} C^-(\hat\beta)\hat G(\beta^{**})\}^{-1} a_n(\hat\beta)]\bar g(\beta_0) \to (\Sigma^-)^{1/2}(I - P)Y$$

in distribution, where $P = G(G^{\mathrm{T}}\Sigma^- G)^{-1}G^{\mathrm{T}}\Sigma^-$ and $Y \sim N(0, \Sigma)$. Let $S = (\Sigma^-)^{1/2}(I - P)\Sigma(I - P)^{\mathrm{T}}(\Sigma^-)^{1/2}$. Since P is idempotent, so is S. Thus $Q(\hat\beta)$ converges in distribution to $Y^{\mathrm{T}}(I - P)^{\mathrm{T}}\Sigma^-(I - P)Y \sim \chi^2_{\mathrm{rank}(S)}$, where $\mathrm{rank}(S) = \mathrm{trace}(S) = \mathrm{rank}(\Sigma) - Kp + (K-1)|\mathcal{M}|$. □

Proof of Theorem 2. Note that $G^{\mathrm{T}} = (G_1^{\mathrm{T}}, G_2^{\mathrm{T}}, \ldots, G_K^{\mathrm{T}})$ and $\Sigma^- = \mathrm{diag}\{\rho_1\Sigma_1^-, \ldots, \rho_K\Sigma_K^-\}$ where G_k $(k = 1, \ldots, K)$ and Σ are defined in Theorem 1. We have $G^{\mathrm{T}}\Sigma^- G = \rho_1 G_1^{\mathrm{T}}\Sigma_1^- G_1 + \cdots + \rho_K G_K^{\mathrm{T}}\Sigma_K^- G_K$. Denote $B = G^{\mathrm{T}}\Sigma^- G$ and $B_k = \rho_k G_k^{\mathrm{T}}\Sigma_k^- G_k$ for $k = 1, \ldots, K$. Then B can be partitioned as

$$B = \begin{pmatrix} B_{[\beta_1,\beta_1]} & B_{[\beta_1,-\beta_1]} \\ B_{[-\beta_1,\beta_1]} & B_{[-\beta_1,-\beta_1]} \end{pmatrix} = \begin{pmatrix} \sum_{k=1}^K B_{k[\beta_1,\beta_1]} & \sum_{k=1}^K B_{k[\beta_1,-\beta_1]} \\ \sum_{k=1}^K B_{k[-\beta_1,\beta_1]} & \sum_{k=1}^K B_{k[-\beta_1,-\beta_1]} \end{pmatrix},$$

where $-\beta_1$ means not corresponding to β_1, the block-diagonal matrix $\sum_{k=1}^K B_{k[-\beta_1,-\beta_1]} = \mathrm{diag}\{B_{2[\gamma_2,\gamma_2]}, \ldots, B_{K[\gamma_K,\gamma_K]}\}$, $\sum_{k=1}^K B_{k[\beta_1,-\beta_1]} = \{B_{2[\beta_1,\gamma_2]}, \ldots, B_{K[\beta_1,\gamma_K]}\}$ and $\sum_{k=1}^K (B_{k[-\beta_1,\beta_1]})^{\mathrm{T}} =$

$\{(B_{2[\gamma_2,\beta_1]})^{\mathrm{T}}, \ldots, (B_{K[\gamma_K,\beta_1]})^{\mathrm{T}}\}$. Following Horn & Johnson (1990, page 18), one can easily derive the inverse of partitioned matrix B and

$$(B^{-1})_{[\beta_1,\beta_1]} = \left\{ \sum_{k=1}^{K} B_{k[\beta_1,\beta_1]} - \sum_{k=2}^{K} B_{k[\beta_1,\gamma_k]} (B_{k[\gamma_k,\gamma_k]})^{-1} B_{k[\gamma_k,\beta_1]} \right\}^{-1}.$$

Since $B_{k[\beta_k,\beta_k]}$ is positive definite, so is $B_{k[\zeta,\zeta]} - B_{k[\zeta,\gamma_k]}(B_{k[\gamma_k,\gamma_k]})^{-1}B_{k[\gamma_k,\zeta]}$. By the fact that $B_{k[\beta_1,\beta_1]} - B_{k[\beta_1,\gamma_k]}(B_{k[\gamma_k,\gamma_k]})^{-1}B_{k[\gamma_k,\beta_1]}$ is a block-diagonal matrix with diagonal components $B_{k[\zeta,\zeta]} - B_{k[\zeta,\gamma_k]}(B_{k[\gamma_k,\gamma_k]})^{-1}B_{k[\gamma_k,\zeta]}$ and a zero matrix, we show that $\sum_{k=2}^{K} B_{k[\beta_1,\beta_1]} - \sum_{k=2}^{K} B_{k[\beta_1,\gamma_k]}(B_{k[\gamma_k,\gamma_k]})^{-1} B_{k[\gamma_k,\beta_1]}$ is nonnegative definite. Applying Horn & Johnson (1990, Theorem 7.7.4), we obtain $(B^{-1})_{[\beta_1,\beta_1]} \preceq (B_{1[\beta_1,\beta_1]})^{-1}$, where $(B^{-1})_{[\beta_1,\beta_1]}$ and $(B_{1[\beta_1,\beta_1]})^{-1}$ are root-n asymptotic variances for $\hat{\beta}_1$ and $\tilde{\beta}_1$, respectively. Rearranging the order of $\gamma_1, \ldots, \gamma_K$ in parameter β, we can prove $(B^{-1})_{[\beta_k,\beta_k]} \preceq (B_{k[\beta_k,\beta_k]})^{-1}$ for all $k = 1, \ldots, K$. $\square$

REFERENCES

ANDERSON, T. (2003). *An Introduction to Multivariate Statistical Analysis*, 3rd ed. New Jersey: Wiley.

BECKER, B. J. (2007). Multivariate meta-analysis: Contributions of Ingram Olkin. *Statist. Sci.* **22**, 401–6.

BRESLOW, N. & DAY, N. (1980). *Statistical Methods in Cancer Research: Volume I: The Analysis of Case-control Studies*. Lyon: IARC Scientific Publications.

CRAINICEANU, C. M. & RUPPERT, D. (2004). Likelihood ratio tests in linear mixed models with one variance component. *J. R. Statist. Soc.* B **66**, 165–85.

CROWDER, M. (1995). On the use of a working correlation matrix in using generalised linear models for repeated measures. *Biometrika* **82**, 407–10.

DAVIDSON, J. (2001). *Stochastic Limit Theory*. New York: Oxford University Press.

FITZMAURICE, G. M., LIPSITZ, S. R. & IBRAHIM, J. G. (2007). A note on permutation tests for variance components in multilevel generalized linear mixed models. *Biometrics* **63**, 942–6.

FOLLMANN, D. A. & PROSCHAN, M. A. (1999). Valid inference in random effects meta-analysis. *Biometrics* **55**, 732–7.

HARTUNG, J., KNAPP, G. & SINHA, B. K. (2008). *Statistical Meta-Analysis with Applications*. New Jersey: Wiley.

HARVILLE, D. A. (2008). *Matrix Algebra from a Statistician's Perspective, Corrected Edition*. New York: Springer.

HORN, R. A. & JOHNSON, C. R. (1990). *Matrix Analysis*. New York: Cambridge University Press.

HU, Y. & SONG, P. X.-K. (2011). Sample size determination for quadratic inference functions method in longitudinal design with dichotomous outcomes. *Statist. Med.* **31**, 787–800.

LAIRD, N. M. & WARE, J. H. (1982). Random-effects models for longitudinal data. *Biometrics* **38**, 963–74.

LOCALIO, A. R., BERLIN, J. A., TEN HAVE, T. R. & KIMMEL, S. E. (2001). Adjustments for center in multicenter studies: an overview. *Ann. Internal Med.* **135**, 112–23.

MANCL, L. A. & DEROUEN, T. A. (2001). A covariance estimator for generalized estimating equations with improved small-sample properties. *Biometrics* **57**, 126–34.

PAN, W. (2001). On the robust variance estimator in generalised estimating equations. *Biometrika* **88**, 901–6.

QU, A., LINDSAY, B. G. & LI, B. (2000). Improving generalised estimating equations using quadratic inference functions. *Biometrika* **87**, 823–76.

SINHA, S. K. (2009). Bootstrap tests for variance components in generalized linear mixed models. *Can. J. Statist.* **37**, 219–34.

STRAM, D. O. & LEE, J. W. (1994). Variance components testing in the longitudinal mixed effects model. *Biometrics* **50**, 1171–7.

WANG, M. & LONG, Q. (2011). Modified robust variance estimator for generalized estimating equations with improved small-sample performance. *Statist. Med.* **30**, 1278–91.

XIE, Y. & AHN, C. (2010). Statistical methods for integrating multiple types of high-throughput data. *Meth. Molec. Biol.* **620**, 511–29.

ZEGER, S. L., LIANG, K.-Y. & ALBERT, P. S. (1988). Models for longitudinal data: a generalized estimating equation approach. *Biometrics* **44**, 1049–60.

ZHANG, D. & LIN, X. (2008). Variance component testing in generalized linear mixed models for longitudinal/clustered data and other related topics. In *Random Effect and Latent Variable Model Selection*, Ed. D. B. Dunson. New York: Springer.

ZHANG, W., FAN, J. & SUN, Y. (2009). A semiparametric model for cluster data. *Ann. Statist.* **37**, 2377–408.

[*Received February* 2011. *Revised February* 2012]