Research Article

Statistics in Medicine

Received 26 March 2010, Accepted 13 April 2011 Published online in Wiley Online Library

(wileyonlinelibrary.com) DOI: 10.1002/sim.4293

Avoiding bias in mixed model inference for fixed effects

Matthew J. Gurka,[a*†] Lloyd J. Edwards[b] and Keith E. Muller[c]

Analysis of a large longitudinal study of children motivated our work. The results illustrate how accurate inference for fixed effects in a general linear mixed model depends on the covariance model selected for the data. Simulation studies have revealed biased inference for the fixed effects with an underspecified covariance structure, at least in small samples. One underspecification common for longitudinal data assumes a simple random intercept and conditional independence of the within-subject errors (i.e., compound symmetry). We prove that the underspecification creates bias *in both small and large samples*, indicating that recruiting more participants will not alleviate inflation of the Type I error rate associated with fixed effect inference. Enumerations and simulations help quantify the bias and evaluate strategies for avoiding it. When practical, backwards selection of the covariance model, starting with an unstructured pattern, provides the best protection. Tutorial papers can guide the reader in minimizing the chances of falling into the often spurious software trap of nonconvergence. In some cases, the logic of the study design and the scientific context may support a structured pattern, such as an autoregressive structure. The sandwich estimator provides a valid alternative in sufficiently large samples. Authors reporting mixed-model analyses should note possible biases in fixed effects inference because of the following: (i) the covariance model selection process; (ii) the specific covariance model chosen; or (iii) the test approximation.

Keywords: compound symmetry; longitudinal data; random effects; type I error

1. Introduction

1.1. Motivation

The Context. The general linear mixed model has become a standard tool for modeling correlated continuous data from longitudinal and clustered sampling. The stacked-data structure creates parallels with univariate linear regression. The arrangement allows simple interpretation of mean values, which has helped make it popular with researchers with limited statistical background, as well as statisticians working with them.

The Problem. The flexibility of the general linear mixed model has created a false sense of security about robustness to covariance misspecification and assumption diagnostics, particularly in large samples. We demonstrate that underfitting the covariance structure for the responses can optimistically bias inference about fixed effects. Furthermore, the amount of bias remains unchanged as the number of subjects increases to infinity. Analytic, simulation, and enumeration results quantify the amount of bias in small and large samples.

The Solution. Good statistical practice can avoid or solve the problem. Conscientious application of existing tools gives a dependable strategy for selecting, verifying, and reporting a covariance model giving accurate inference about fixed effects in small and large samples. Analysis of a large longitudinal study of child care highlights the importance of the strategy, even in large samples.

[a]*Department of Community Medicine, School of Medicine, West Virginia University, PO Box 9190, Morgantown, WV 26506-9190, USA*

[b]*Department of Biostatistics CB #7420, University of North Carolina, 3105H McGavran-Greenberg, Chapel Hill, NC 27599-7420, USA*

[c]*Department of Health Outcomes and Policy, University of Florida College of Medicine, PO Box 100177, Gainesville, FL 32610-0177, USA*

**Correspondence to: Matthew J. Gurka, Department of Community Medicine, School of Medicine, West Virginia University, PO Box 9190, Morgantown, WV 26506-9190, USA.*

[†]*E-mail: mgurka@hsc.wvu.edu*

1.2. Asthma in childhood

Asthma affects more than nine million American children 0–17 years old [1]. In a cross-sectional study, Blackman and Gurka [2] found that children with severe asthma were more than four times more likely to have chronic developmental and behavioral problems of various sorts, including depression or anxiety and conduct problems. However, the cross-sectional sampling plan limits the value of the results.

In contrast, a longitudinal design lies at the heart of the Study of Early Child Care and Youth Development (SECCYD) [3]. The SECCYD followed up children from 10 US sites from birth through age 15 years. Families were recruited during hospital visits following the birth of the child in 1991. A total of 1364 families enrolled from the 8986 eligible births. By phase IV (2005–2008; age 14–15 years), 1056 families remained enrolled. The repeated collection of standard psychometric measures throughout childhood provided a unique opportunity to examine the role of chronic asthma in the development of a child. Interest lay solely in fixed effects inference: does the average longitudinal profile for children with asthma differ from the profile for those without? Although selection of the covariance model was secondary, the choice was found to be vital in making correct decisions for the primary aim.

2. Previous research and current views

The linear mixed model is composed of fixed effects, random effects, and a residual error. In medical research, the first component is typically of interest, whereas the latter two components comprise the covariance portion of the model. There are no fully reliable ways to identify the best covariance model, at least in small samples [4]. Verbeke and Molenberghs [5, pp. 125–127] noted that it is common to include random intercepts and additional random effects only for time-varying covariates. They then presented a set of guidelines for the inclusion of additional random effects, noting that they 'favor the inclusion of too many random effects rather than omitting some' (p. 127). However, they also noted that inclusion of too many random effects may lead to nonconvergence.

In practice, limitations of current software and user behaviors, magnified in small sample settings or in the presence of missing data, make nonconvergence common. Many analysts interpret a model failing to converge as empowering them to simplify the model, whether it be through the fixed effects or the covariance model. A more parsimonious fixed effect structure may be an effective remedy for nonconvergence, particularly in the presence of a small number of subjects and/or a limited number of repeated measures (in the case of time-varying covariates). In the context of covariance model selection, the press of time and a belief in the robustness of the mixed model with respect to fixed effects inference encourage this approach.

Some authors have considered covariance misspecification for tests about means, that is, fixed effects in mixed models. Liang and Zeger [6] proposed the 'sandwich' estimator for $\mathcal{V}(\widehat{\boldsymbol{\beta}})$ in seeking inference about the fixed effects robust to misspecified covariance. Verbeke and Molenberghs [5, p. 62] noted the following: (i) the sandwich estimator is less efficient than the one using the correct covariance model; and (ii) valid inference requires additional assumptions about the missing data. In addition, the sandwich estimator has not been fully evaluated in small samples.

Another approach to covariance misspecification has been to ask whether underfitting can allow valid inference. Jacqmin-Gadda *et al.* [7] demonstrated an inflated type I error rate with covariance misspecification in simulations in the general linear mixed model with Gaussian errors. For some special cases with complete and balanced Gaussian data, Lange and Laird [8] proved that 'fitting a parsimonious covariance structure need not give inappropriate variance estimates, even if the parsimonious structure is inadequate.' They demonstrated in the special cases considered that inclusion of both a random intercept and slope is conservative in that the variance estimates of the fixed intercept and slope will not be biased downward. The theory and simulations in the present paper expand on their discussion, specifically providing evidence against the use of the compound symmetry assumption in general longitudinal settings.

Scientists in a variety of health and social science settings often assume compound symmetry for the covariance model of longitudinal responses. Some do so by fitting a random-intercept-only mixed model and assuming residual error covariance is diagonal with homogeneous variances [9, 10], whereas others use the uncorrected univariate approach to repeated measures (UNIREP) test [11, 12]. The unadjusted UNIREP test assumes compound symmetry. Muller and Stewart [13], among many others, have provided further discussion of the UNIREP approach. In addition, convergence failure leads some data analysts to remove additional random effects and retain only a random intercept (and implicitly assume compound

symmetry of the responses). Scientists often assume compound symmetry to simplify a power analysis for the design of a longitudinal study. Some free software for planning repeated measures studies use compound symmetry exclusively [14].

Despite the widespread use of a compound symmetric (CS) covariance model for longitudinal responses, little evidence can be marshaled to defend the approach. Equally important, many authors studying the UNIREP approach, dating back to Box [15, 16], have demonstrated inflated type I error when falsely assuming compound symmetry. We present analytic and numerical evidence to demonstrate that underspecification of the covariance model can lead to very inaccurate fixed effect inference in a mixed model because of an inflated type I error rate and suggest ways to ensure accurate inference.

3. Analytic results

3.1. Notation

We use the notation in Chapters 3–5 of Muller and Stewart [13]. Subscripts have been added as needed to distinguish distinct expressions, which play parallel roles in distinct model formulations, with M for multivariate and m for mixed models. All results assume Gaussian errors and testable hypotheses (estimable secondary parameters with full-rank contrast matrices). The linear mixed model, that in the form which Laird and Ware [17] described, may be written in the notation of Muller and Stewart [13, Chapter 5] as follows:

$$\boldsymbol{y}_i = \boldsymbol{X}_i\boldsymbol{\beta} + \boldsymbol{Z}_i\boldsymbol{d}_i + \boldsymbol{e}_i\,. \tag{1}$$

Here, $i \in \{1,\ldots,N\}$, for N the number of independent sampling units (usually subjects) and $\boldsymbol{y}_i$ is a $p_i \times 1$ vector of observations on person i; $\boldsymbol{X}_i$ is a $p_i \times q$ known design matrix for person i, whereas $\boldsymbol{\beta}$ is a $q \times 1$ vector of unknown population parameters. Also, $\boldsymbol{Z}_i$ is a $p_i \times m$ known design matrix for person i corresponding to the $m \times 1$ vector of unknown random effects $\boldsymbol{d}_i$, whereas $\boldsymbol{e}_i$ is a $p_i \times 1$ vector of unknown residual errors. The vectors $\boldsymbol{d}_i$ and $\boldsymbol{e}_i$ are Gaussian and independent with mean $\boldsymbol{0}$ and covariance matrices $\boldsymbol{\Sigma}_{di}(\boldsymbol{\tau}_d)$ and $\boldsymbol{\Sigma}_{ei}(\boldsymbol{\tau}_e)$, respectively. In turn, $\boldsymbol{\mathcal{V}}(\boldsymbol{y}_i) = \boldsymbol{\Sigma}_i = \boldsymbol{Z}_i\boldsymbol{\Sigma}_{di}(\boldsymbol{\tau}_d)\boldsymbol{Z}_i' + \boldsymbol{\Sigma}_{ei}(\boldsymbol{\tau}_e)$ is characterized by the finite set of parameters in the $r \times 1$ vector $\boldsymbol{\tau}$ containing the unique parameters in $\boldsymbol{\tau}_d$ and $\boldsymbol{\tau}_e$.

A focus on inference about 'fixed' effects, the mean response values, has led others to consider the 'population average' model,

$$\boldsymbol{y}_i = \boldsymbol{X}_i\boldsymbol{\beta} + \boldsymbol{e}_{+i} \tag{2}$$

where $\boldsymbol{e}_{+i} = \boldsymbol{Z}_i\boldsymbol{d}_i + \boldsymbol{e}_i$. Formulation (1) allows conveniently specifying latent random intercepts, slopes, and others, which implies a covariance model for the responses. In contrast, formulation (2) allows directly specifying the response covariance model, which in some cases implies the latent random components. A functional mapping from random effects to a covariance structure always exists. However, the mapping need not be unique.

Clustered data arising from schools, hospitals, and others usually generate *CS* data. Such observations have a common variance and a common correlation for any pair. Cluster i has p_i observations. The same covariance structure arises by specifying the following: (i) a random intercept as the only random effect, so $\boldsymbol{Z}_i = \boldsymbol{1}_{p_i}$ and d_i is 1×1, a scalar; and (ii) $\boldsymbol{\Sigma}_{ei}(\boldsymbol{\tau}_e) = \sigma_e^2\boldsymbol{I}_{p_i}$ (a conditional independence assumption for within-subject residual). The responses have

$$\boldsymbol{\mathcal{V}}(\boldsymbol{y}_i) = \boldsymbol{\Sigma}_i = \boldsymbol{1}_{p_i}\sigma_d^2\boldsymbol{1}_{p_i}' + \sigma_e^2\boldsymbol{I}_{p_i}\,. \tag{3}$$

For intraclass correlation coefficient ρ between observation j and j', the model implies $\mathcal{V}(y_{ij}) = \sigma_d^2 + \sigma_e^2 = \sigma^2$ and $\sigma_d^2 = \sigma^2\rho$. Having $\sigma_d^2 \geqslant 0$ requires $\rho \geqslant 0$, whereas $-1/(p_i - 1) < \rho$ suffices for compound symmetry. The matrix $\boldsymbol{\mathcal{V}}(\boldsymbol{y}_i)$ has a largest eigenvalue of $\lambda_{1i} = \sigma^2[1 + (p_i - 1)\rho]$ with eigenvector $p_i^{-1/2}\boldsymbol{1}_{p_i}$, whereas all other $p_i - 1$ eigenvalues are $\lambda_2 = \sigma^2(1 - \rho)$. Typically, cluster designs are discussed in terms of $\{\sigma^2, \rho\}$ [18] rather than the equivalent pair $\{\lambda_{1i}, \lambda_2\}$.

3.2. A useful class of mixed models

We restrict attention to tests of fixed effects in the general linear mixed model with Gaussian errors. In this section and the next, we will prove the following statement:

Proposition 1

Inference for fixed effects in the general linear mixed model is ***not*** *robust to covariance misspecification.*

Specifically, we prove that falsely assuming compound symmetry, which ignores heterogeneity within subject, creates biased tests and confidence intervals for data with no missing values (complete and balanced). The result *holds in small and also in asymptotically large samples*. The often-used assumption of compound symmetry serves as a logical counterexample that allows us to make the previously mentioned claim.

To prove and enumerate the impact of under fitting the covariance model on tests of fixed effects, we focus on a restricted class of models, which allows finding closed-form and small-sample (exact) answers. Hence, our proof centers on the following: (i) mixed models, which can be stated as a multivariate model; and (ii) hypotheses, which can be stated as a general linear multivariate hypothesis. The multivariate model corresponds to the 'population average' model formulation in Equation (2), which has no explicitly defined random effects. This multivariate model, as Muller and Stewart [13, Chapter 3] defined, has restrictions that demand a design with complete balance within subjects, with no missing or mistimed data and no repeated covariates. An unstructured covariance matrix will be assumed to be the correct model for the responses.

Muller and Stewart [13, Table 6.5, Table 6.6, Section 12.1, and especially Equation 12.5] described how to state any multivariate model as a general linear mixed model. As traditionally written, the multivariate model $\boldsymbol{Y} = \boldsymbol{X}_M \boldsymbol{B} + \boldsymbol{E}$ describes N rows of independent sampling units (subjects) and p columns of repeated measures. Here, $\boldsymbol{B}$ is a $q \times p$ matrix of parameters with rows corresponding to between-subject effects and columns (time indicator) to within-subject (time) effects. By describing only the data for subject i and therefore row i in the multivariate model gives $1 \times p$ matrix $\boldsymbol{Y}_i = \boldsymbol{X}_{Mi} \boldsymbol{B} + \boldsymbol{E}_i$, with $\boldsymbol{Y}_i = \text{row}_i(\boldsymbol{Y})$ and $\boldsymbol{X}_{Mi} = \text{row}_i(\boldsymbol{X}_M)$, whereas $\boldsymbol{E}_i = \text{row}_i(\boldsymbol{E})$. The equivalent mixed model form arises as

$$\begin{aligned} \text{vec}(\boldsymbol{Y}_i) &= \text{vec}\left[(\boldsymbol{X}_{Mi}\boldsymbol{B})'\right] + \text{vec}(\boldsymbol{E}_i) \\ &= \left(\boldsymbol{X}_{Mi} \otimes \boldsymbol{I}_p\right) \text{vec}\left(\boldsymbol{B}'\right) + \text{vec}(\boldsymbol{E}_i) \\ \boldsymbol{y}_i &= \boldsymbol{X}_{mi}\boldsymbol{\beta} + \boldsymbol{e}_{+i}, \end{aligned} \tag{4}$$

which is the population average model with $n = N \cdot p$ observations, and $\otimes$ is the Kronecker product. The error term can be interpreted as $\boldsymbol{e}_{+i} = \boldsymbol{Z}_i \boldsymbol{d}_i + \boldsymbol{e}_i$. For example, if $\boldsymbol{Z}_i = \boldsymbol{1}_{p_i}$ and elements of the residual $\boldsymbol{e}_i$ are independent and identically distributed (i.i.d.), then $\boldsymbol{\mathcal{V}}(\boldsymbol{y}_i)$ will be CS.

For inference about population means (i.e., fixed effects), the parameter matrix of mean differences for the multivariate model is $\boldsymbol{\Theta} = \boldsymbol{C}_M \boldsymbol{B} \boldsymbol{U}$ with hypothesis $\text{H}_0 : \boldsymbol{C}_M \boldsymbol{B} \boldsymbol{U} = \boldsymbol{\Theta}_0$. Transforming to the mixed model creates vector $\boldsymbol{\theta} = \boldsymbol{C}_m \boldsymbol{\beta} = \text{vec}(\boldsymbol{\Theta}) = \left(\boldsymbol{C}_M \otimes \boldsymbol{U}'\right) \text{vec}\left(\boldsymbol{B}'\right)$. The multivariate matrix $\boldsymbol{C}_M$ defines contrasts *between groups*. The multivariate matrix $\boldsymbol{U}$ defines contrasts *within* an independent sampling unit (such as person) across level of response (such as time). The mixed model contrast matrix $\boldsymbol{C}_m$ does the work of both $\boldsymbol{C}_M$ and $\boldsymbol{U}$.

3.3. *Incorrectly assuming compound symmetry; ignoring heterogeneity within subject*

A test of all time trends from linear through $p-1$ for p repeated measures will illustrate the analytic and simulation results, which actually cover a wider class of situations. Stacking all of the data together as $\boldsymbol{y}_s = \left[\boldsymbol{y}_1' \ \cdots \ \boldsymbol{y}_i' \ \cdots \ \boldsymbol{y}_N'\right]'$ and $\boldsymbol{X}_s = \left[\boldsymbol{X}_1' \ \cdots \ \boldsymbol{X}_i' \ \cdots \ \boldsymbol{X}_N'\right]'$ allows writing $\boldsymbol{y}_s = \boldsymbol{X}_s \boldsymbol{\beta} + \boldsymbol{e}_{+s}$ and provides a convenient form for the Wald statistic in the mixed model:

$$F_m = (\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)'[\boldsymbol{C}_m (\boldsymbol{X}_s' \widehat{\boldsymbol{\Sigma}}_s^{-1} \boldsymbol{X}_s)^{-1} \boldsymbol{C}_m']^{-1} (\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)/a_m, \tag{5}$$

with $a_m = \text{rank}(\boldsymbol{C}_m)$ and $\widehat{\boldsymbol{\Sigma}}_s = \widehat{\boldsymbol{\mathcal{V}}}(\boldsymbol{y}_s)$. A mixed model corresponding to a multivariate model has $\boldsymbol{C}_m = \left(\boldsymbol{U}' \otimes \boldsymbol{C}_M\right)$ and $\widehat{\boldsymbol{\theta}} = \text{vec}(\widehat{\boldsymbol{\Theta}}) = \left(\boldsymbol{U}' \otimes \boldsymbol{C}_M\right) \text{vec}(\widehat{\boldsymbol{B}})$. Kenward and Roger [16] mentioned that in such models, F_m is a 1–1 function of the Hotelling–Lawley statistic in the multivariate model. The exact distribution of F_m when falsely assuming compound symmetry for a multivariate equivalent model may be derived as outlined in Appendix A. Assuming compound symmetry in the mixed model further reduces the statistic to the original UNIREP statistic. The following theorem and corollary summarize the most important conclusions that can be deduced.

Theorem 1
An interesting class of general linear mixed models have the following: (i) complete and balanced Gaussian data; (ii) a common (finite and full rank) population covariance matrix for each independent sampling unit (e.g., subject); and (iii) no time-varying covariates. Assuming CS covariance of the responses for any such model leads to a simple closed-form expression for the Wald statistic from a test of fixed effects. Algorithms for computing the distribution function of quadratic forms allows exact calculation of test size and power.

Proof
Appendix A contains an explicit derivation of easily computed expressions for the distribution function of the test statistic in small and asymptotic samples. □

Corollary 1
Incorrectly assuming compound symmetry among responses inflates type I error rate for Wald tests of fixed effects in a general linear mixed model for small or large samples.

Proof
Finite and infinite large-sample examples in the exact calculations in the following text (Section 3.4) prove the corollary by giving counterexamples to the implicit claim of guaranteed correct test size. □

An abundance of results [15, 16, 19, 20] make it clear that the uncorrected UNIREP test can badly inflate type I error rate in small samples. The same test arises in the general linear mixed model whenever compound symmetry has been assumed but does not hold. The between–within approach for mixed models (SAS (SAS Institute Inc., Cary, NC, USA) PROC MIXED) [21] provides one direct path. However, all other mixed model test choices (such as containment, residual, Satterthwaite, and Kenward–Roger in SAS) will also inflate type I error rate because of the assumption error. The proof contains results that show that type I error inflation remains just as high as $N \to \infty$.

3.4. *Exact calculations*

Exact calculations illustrate the bias described in the corollary for each combination of $N \in \{50, 100, \infty\}$ and $p \in \{5, 10\}$, based on models that assumed two groups with a mean model including only intercept, time (linear), group, and group×time (linear), so $\boldsymbol{X}_i$ had rank 4. The p_i time points were equally spaced from 0 to 1.

Using the same mean model structure, four covariance models were considered true. They may be summarized as *covariance model: random effects and within-unit residual error*, namely

(1) 1,IID: random intercept and i.i.d. within-unit residual (i.e., compound symmetry);
(2) 1,AR: random intercept and a first-order AR(1), autoregressive within-subject residual;
(3) 2,IID: random intercept and slope, with an i.i.d. within-unit residual; and
(4) 2,AR: random intercept and slope (correlated = 0.25) with an AR(1) within-unit residual.

Population variances of the random intercept and slope (when included) were 2 and 1, respectively. The within-unit residual variance was 1 with an AR(1) correlation of 0.25 when included.

Enumerations used the _QPROB module from the free software POWERLIB [22]. Kim *et al.* [23] described the _QPROB algorithm (because of Davies [24]). Exact asymptotic calculations used the SAS chi-square function. We calculated the type I error rate for testing the fixed effect of group×time (linear) interaction (i.e., the group difference in slopes) when assuming compound symmetry.

Correctly assuming compound symmetry (1,IID) gives the correct type I error rate of 5%. When $p = 5$, incorrectly assuming compound symmetry gives a type I error rate of 9.0% when the true covariance was (1, AR). Type I error rate increases to 9.8% and 13.7% when the true covariance also contains a random slope: (2,IID) and (2,AR), respectively. When $p = 10$, the type I error rate rises to 14.5% and 19.7% when assuming compound symmetry, and the true covariance is (2,IID) and (2,AR), respectively. The exact Type I error rates for $p \in \{5, 10\}$ were equal (to three digits of accuracy) for $N \in \{50, 100, \infty\}$. The proof requires no missing data (to take advantage of simplifications in the theory for multivariate models). Obviously, complete and balanced data rarely occur in some applications, such as clinical trials. However, the simulations (Section 4) support the general proposition that *with or without missing data, incorrectly assuming compound symmetry inflates type I error rate in the general linear mixed model.*

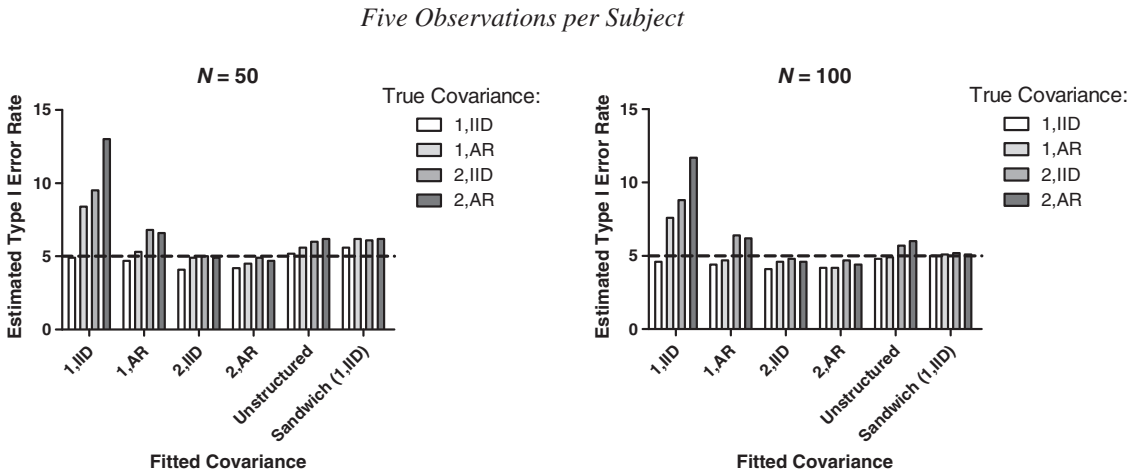
For the sake of brevity, parallel results about ignoring heterogeneity between subjects, as distinct from ignoring heterogeneity within subjects, have been omitted. Key features are well-known from the study of univariate linear models: heterogeneity between subjects strongly interacts with unequal cell sizes to inflate type I error rate for three or more groups in ANOVA [25]. The mixed model inherits the same vulnerability. However, a knowledgeable and conscientious analyst can directly model such heterogeneity in a mixed model formulation.

4. Simulation study

We have proved that incorrectly assuming compound symmetry leads to inflated type I error rates, with complete and balanced data. Simulations extend the results to missing data. All simulations used SAS/IML, PROC MIXED, and the DATA step. The normal function generated pseudorandom Gaussian values. The exact conditions presented in Section 3.4 were replicated for the simulations. We generated 10,000 pseudorandom samples for each $\{N, p_i\} \in \{\{50, 5\}, \{50, 10\}, \{100, 5\}, \{100, 10\}\}$, with each subject having p_i time points equally spaced from 0 to 1. With $\alpha = 0.05$, the 10,000 pseudorandom samples give a confidence region for the estimated type I error rate of $\widehat{\alpha} \in [0.0498, 0.0502]$. In practice, the acceptability of a particular type I error rate will vary with the preferences of the reader. A total of 20% of the data were set to missing by pseudorandom deletion within person (uniform sampling without replacement).

We calculated the type I error rate for testing the fixed effect of Group×Time (Linear) interaction (i.e., the group difference in slopes). For each of the four true population covariance structures (listed in Section 3.4), we tabulated type I error rates when fitting the same four distinct covariance models, in addition to fitting an unstructured covariance. We employed restricted maximum likelihood (REML) estimation with the Kenward–Roger approach [26] for all five cases. We also tabulated the type I error rate when assuming compound symmetry with the sandwich estimator, which required the containment method [21] because the Kenward–Roger approach does not apply.

Figure 1 presents empirical type I error rates for the simulations with $p_i = 5$ observations per subject. Results from the simulations with $\{N, p_i\} \in \{\{50, 10\}, \{100, 10\}\}$ were similar and are not shown. The results illustrate the generality of the theorem and the substantial bias that can occur. Even with a correctly specified covariance model, observed type I error rate has some modest bias for smaller sample sizes, especially when the true models include, an AR(1) within-unit error covariance. The (true) (1,AR) model has a 5.3% type I error rate when correct, and the (true) (2,AR) model has a 4.7% type I error



Covariance Model Notation:

1,IID: Random intercept, i. i. d. within-unit residual (compound symmetry)
1,AR: Random intercept, AR(1) within-unit residual
2,IID: Random intercept and random slope, i. i. d. within-unit residual
2,AR: Random intercept and random slope, AR(1) within-unit residual

*Except in the sandwich cases, we performed inference using the Kenward-Roger approach.

Figure 1. Observed type I error rate (×100) of fixed effect interaction for $\alpha = 0.05$. Four true covariance models, 20% missing completely at random, 10,000 replications per condition*.

rate when correct, for N = 50. As expected, moderate to severe bias in type I error rate can occur when incorrectly assuming compound symmetry, with type I error rates no smaller than 7.6% for any of the other examined true covariance models. The bias remains substantial even when the number of subjects increases from 50 to 100.

It is important to note what works in the absence of reliable covariance model identification techniques. Not surprisingly, if the covariance model chosen contained the true model as a special case, or differed only modestly from the true model, then type I error rate was roughly correct. For example, the random intercept and slope (i.i.d. error) covariance model that Lange and Laird [8] considered controlled type I error rate for the four simple models considered, with a small conservative bias when the true model is indeed compound symmetry (4.1% and 4.2% when N = 50 and N = 100, respectively). Our results imply that it would be a simple exercise to find other covariance patterns for which the choice fails to control type I error rate. Fitting an unstructured covariance model always gives a safe choice when focused on inference about fixed effects, if one is willing to accept a type I error rate no greater than about 6%. It is clear that the unstructured covariance model is over-fitting data generated from a model with at most two random effects. As many practitioners know, estimating an unstructured covariance can create convergence problems, especially as the number of observations per subject increases or the number of independent observations decreases. For example, with $N = 100$ and $p = 10$ (results not shown), convergence for the unstructured model was achieved only 61% of the time. Among the models that converged, type I error rates were inflated, no matter what the underlying true model was. For the sample sizes considered, using the sandwich estimator improved inference no matter what covariance model was assumed, without the convergence problems. However, bias similar to what was observed for the unstructured model remains, particularly for small samples.

5. Application to the Study of Early Child Care and Youth Development asthma data

Modeling behavioral outcomes over time in the SECCYD data further illustrate the analytical and numerical results. Specifically, assuming CS errors or equivalently including only a random intercept with conditional independence of within-subject error leads to different inferences than for more general covariance models. Both look plausible, although both cannot be correct because they disagree. The analytic and simulation results suggest that the results for the complex covariance model are the correct ones.

Developmental measures collected in the SECCYD allow comparing children with and without various phenotypes of asthma. The primary outcome was the Child Behavior Checklist (CBCL) [27]. The CBCL measures social competence as well as externalizing and internalizing behavior problems. The parent completed age-specific versions of the CBCL, standardized as a 'T-score' (mean = 50, SD = 10), at age 36 and 54 months, as well as during grades 1, 3, 4, 5, and 6, and at age 15 years. Higher scores indicate more problematic behavior, with T-scores greater than 70 considered as clinically relevant.

We compared internalizing behavior trends between those who developed asthma early (by age 4 years) and those who never developed asthma. Asthma status could be classified for 795 children: 58 children had asthma, whereas 737 did not. Linear, through cubic orthogonal polynomial trends for time, were initially included as predictors, as well as all corresponding interactions, with the interactions of age with asthma status of particular interest. Covariates included sex, maternal age at birth, the mean maternal depression score during years 0–3, the mean home environment score (higher indicating better) during years 0–3, and the mean child care quality score during years 0–3 (higher indicating better). Initially, compound symmetry was assumed. A backward stepwise selection procedure was performed for the fixed effects by removing the least-significant higher-order interactions one by one until only significant interactions (and/or main effects) remained. Table I displays the estimates and the p-values for the final fixed effects model of internalizing CBCL T-scores when assuming compound symmetry. In addition, estimates were calculated for the final (fixed effect) model using two additional covariance models: (i) unstructured; and (ii) random intercept and slope with an AR(1) within-unit error. The sandwich estimator (with containment) assuming compound symmetry was also applied. With respect to the recommendation of Verbeke and Molenberghs [5, Section 2.3], we fitted additional models not shown here. Given the consistency of these other models, we did not display results from these others, including one with random effects through cubic age. Table I also displays the corresponding fixed effects estimates and the p-values for these other covariance models.

Table I. Study of Early Child Care and Youth Development final model results of child behavior checklist internalizing *T*-scores; 795 children (58 with asthma by age 4 years) followed up from 3–15 years old

	Covariance model							
	Random intercept (CS)		Unstructured		Random Int + Slope		Sandwich estimator	
Fixed predictor*	Estimate	*p*-value	Estimate	*p*-value	Estimate	*p*-value	Estimate	*p*-value
Intercept	143.6	< 0.01	143.4	< 0.01	143.7	< 0.01	143.6	< 0.01
Years	−1.63	< 0.01	−1.77	< 0.01	−1.72	< 0.01	−1.63	< 0.01
Years^2	0.73	0.01	0.84	< 0.01	1.19	< 0.01	0.73	< 0.01
Years^3	−2.22	< 0.01	−2.08	< 0.01	−2.31	< 0.01	−2.22	< 0.01
Differences in intercept								
Male	−1.02	0.47	−1.29	0.35	−1.19	0.40	n	0.47
Maternal age	−0.37	< 0.01	−0.39	< 0.01	−0.38	< 0.01	−0.37	< 0.01
Maternal depression	1.25	< 0.01	1.29	< 0.01	1.28	< 0.01	1.25	< 0.01
Home environment	0.41	0.02	0.44	0.01	0.42	0.02	0.41	0.02
Asthma	0.39	0.88	−0.28	0.92	0.25	0.93	0.39	0.88
Interaction with years								
Male	−1.10	**0.02**	−1.10	0.08	−1.01	0.12	−1.10	0.09
Home environment	0.18	< 0.01	0.18	0.01	0.18	0.01	0.18	0.02
Asthma	2.10	**0.03**	2.08	0.11	2.10	0.11	2.10	0.10
Interaction with years^2								
Home environment	−0.12	0.01	−0.04	< 0.05	−0.15	< 0.01	−0.12	0.08
Interaction with years^3								
Home environment	0.20	< 0.01	0.15	< 0.01	0.19	< 0.01	0.20	< 0.01

*We used orthogonal polynomial coefficients for the time variables; we centered maternal age, depression, and home environment scores around their means.
**P-values in bold indicate significance for a fixed predictor in the compound symmetry model that was not significant in the other models.

Comparisons of the fixed effect estimates and the *p*-values for the various covariance models reveal agreement for most values (Table I). However, asthma status was found to significantly interact with age ($p = 0.03$) in a linear fashion when assuming compound symmetry; there was no such significant interaction when fitting any of the other covariance models ($p \geq 0.10$ in all cases). If one were to have fit a random-intercept-only model to these data, one would have concluded, apparently erroneously, that those with asthma exhibited higher internalizing behavior scores over time. A similar contrast is also observed in considering gender and behavior over time, with the compound symmetry model giving a significant linear change in behavior scores between boys and girls as they mature. We emphasize that it is safer to select the fixed-effect model based on a sufficiently complex covariance model. We omit the details simply for the sake of brevity and simplicity.

6. Discussion

A common choice for a covariance model of the responses in a general linear mixed model arises from including only a random intercept and assuming homogeneous and independent errors within subject. The choice requires the responses to have CS covariance, an assumption that arises in a variety of ways.

We proved that incorrectly assuming compound symmetry inflates type I error for inference about the fixed effects in a general linear mixed model, in both small and large samples. Numerical results illustrate the magnitude of the problem. Simulations demonstrate that the problem occurs with missing data. Our focus here has been the compound symmetry assumption; the proof does not directly apply to other assumed covariance structures (e.g., random intercept plus slope, etc.). However, our results prove that robustness of fixed effects inference with an underspecified covariance model cannot be guaranteed in small or large samples. The generalization applies not only to mixed models that incorporate random effects but also to any multivariate model, particularly those with repeated measurements.

The results provide some guidance for achieving accurate inference for fixed effects. The first step involves considering general covariance models. Many analysts, including ourselves, would not consider compound symmetry for longitudinal data. Despite that, many recently published studies explicitly or implicitly make the assumption. The false hope that large samples allow ignoring insufficient covariance model complexity may motivate the approach. In contrast to longitudinal data, cluster-based sampling typically provides exchangeability of observations and hence CS correlation within cluster. A careful review of the sampling plan to ensure the validity of the exchangeability assumption seems necessary.

In addition to the use in data analysis, the compound symmetry assumption has been used often for power analysis. Although preferred for cluster samples, realistic correlation patterns for longitudinal data include a linear exponent autoregressive (LEAR) reliable or damped exponential model [28], which generalize AR(1) structure. In parallel to the results on type I error rate, a valid power analysis requires a covariance model aligned with the population. The example power values for repeated measures designs in [29] allow concluding that misspecification of the covariance matrix can lead to computing power either *too high or too low* because of misalignment.

If the data allow it, one should model an unstructured covariance. Unfortunately, convergence often becomes an issue with mixed models. Cheng *et al.* [30] provided practical advice on improving the chances for convergent models. The least appreciated strategies center on minimizing collinearity in both fixed and random predictors. The importance of centering and scaling all predictors (especially time) and the use of full-rank coding schemes, preferably orthogonal, cannot be overemphasized.

When a simple covariance model (compound symmetry, unstructured) can not be assumed, the mixed model provides advantages in that one can incorporate random effects to model the covariance structure. Our work extends the results of Lange and Laird [8] by demonstrating that modest expansion of the random effects (covariance structure) helps but cannot be guaranteed to be sufficient. We warn against underspecification of the covariance model, no matter whether it is a general structure or one composed of random effects. Fortunately, conscientious attention to good statistical practices in covariance model fitting can provide confidence in fixed effect inference in a mixed model.

The sandwich estimator appeals simply because it does not require an elaborate covariance model selection process and can simplify convergence. As far as we know, its performance in small-sample settings has been relatively unstudied and should be the focus of future work. Specifically, DOF approximations should be studied with the sandwich estimator. However, in large-sample settings, where we prove that assuming compound symmetry results in substantial bias, the sandwich estimator seems to be a valid alternative.

We have focused on settings with primary interest in inference about fixed effects and little interest in covariance structure itself. Even with a focus on fixed effects, accurate inference requires selecting an adequate approximation for the true covariance structure. Unfortunately, although many model selection criteria have been suggested, none has been found to be clearly superior. Part of the difficulty stems from the interaction between fixed effect and covariance model selection. *Overall, one principle seems clear: controlling type I error for tests of fixed effects demands avoiding an underfitted covariance model.*

Appendix A. Proof of the theorem

For the sake of brevity, we use the notation and many results in [13, Chapters 1, 3, 5, 12, 14, 16, 18] without specific references. *We restrict attention to mixed models that are valid multivariate models*. Stacking all of the data sorted by subject gives $\mathrm{vec}(\boldsymbol{Y}') = \mathrm{vec}\left[(\boldsymbol{X}_M\boldsymbol{B})'\right] + \mathrm{vec}\left(\boldsymbol{E}'\right) = \left(\boldsymbol{X}_M \otimes \boldsymbol{I}_p\right)\mathrm{vec}\left(\boldsymbol{B}'\right) + \mathrm{vec}\left(\boldsymbol{E}'\right)$, which may be stated as $\boldsymbol{y}_s = \boldsymbol{X}_s\boldsymbol{\beta} + \boldsymbol{e}_{+s}$, the population average model for all $n = N \cdot p$ observations. Here, $\boldsymbol{y}_s = \left[\boldsymbol{y}_1' \;\cdots\; \boldsymbol{y}_i' \;\cdots\; \boldsymbol{y}_N\right]'$ and $\boldsymbol{X}_s = \left[\boldsymbol{X}_1' \;\cdots\; \boldsymbol{X}_i' \;\cdots\; \boldsymbol{X}_N\right]'$.

With $\boldsymbol{\mathcal{V}}(\boldsymbol{y}_i) = \boldsymbol{\Sigma}_i$ for subject i, the multivariate model gives maximum-likelihood estimations noniteratively: $\widetilde{\boldsymbol{B}} = \left(\boldsymbol{X}_M'\boldsymbol{X}_M\right)^{-1}\boldsymbol{X}_M'\boldsymbol{Y}$ and $\widetilde{\boldsymbol{\Sigma}} = \left(\boldsymbol{Y} - \boldsymbol{X}_M\widetilde{\boldsymbol{B}}\right)'\left(\boldsymbol{Y} - \boldsymbol{X}_M\widetilde{\boldsymbol{B}}\right)/N$. Here $\widetilde{\boldsymbol{B}}$ is functionally independent of $\widetilde{\boldsymbol{\Sigma}}_i$, so $\widetilde{\boldsymbol{B}}$ is invariant to $\widetilde{\boldsymbol{\Sigma}}_i$, a property not guaranteed in the mixed model. REML estimates are $\widehat{\boldsymbol{B}} = \widetilde{\boldsymbol{B}}$ and $\widehat{\boldsymbol{\Sigma}} = \widetilde{\boldsymbol{\Sigma}}N/(N - r_M)$ for $r_M = \mathrm{rank}(\boldsymbol{X}_M)$. Mixed model estimates are $\widehat{\boldsymbol{\beta}} = \mathrm{vec}(\widehat{\boldsymbol{B}})$, $\widehat{\boldsymbol{\theta}} = \mathrm{vec}(\widehat{\boldsymbol{\Theta}})$, $\widehat{\boldsymbol{\Sigma}}_i$ for subject i and $\widehat{\boldsymbol{\Sigma}}_s = \boldsymbol{I}_N \otimes \widehat{\boldsymbol{\Sigma}}_i$. Hence, $\boldsymbol{X}_s'\widehat{\boldsymbol{\Sigma}}_s^{-1}\boldsymbol{X}_s = \left(\boldsymbol{X}_M \otimes \boldsymbol{I}_p\right)'\left(\boldsymbol{I}_N \otimes \widehat{\boldsymbol{\Sigma}}_i\right)^{-1}\left(\boldsymbol{X}_M \otimes \boldsymbol{I}_p\right) = \left(\boldsymbol{X}_M' \otimes \boldsymbol{I}_p\right)\left(\boldsymbol{I}_N \otimes \widehat{\boldsymbol{\Sigma}}_i^{-1}\right)\left(\boldsymbol{X}_M \otimes \boldsymbol{I}_p\right) = \boldsymbol{X}_M'\boldsymbol{X}_M \otimes \widehat{\boldsymbol{\Sigma}}_i^{-1}$ and $\boldsymbol{C}_m(\boldsymbol{X}_s'\widehat{\boldsymbol{\Sigma}}_s^{-1}\boldsymbol{X}_s)^{-1}\boldsymbol{C}_m' = \left(\boldsymbol{C}_M \otimes \boldsymbol{U}'\right)\left[\boldsymbol{X}_M'\boldsymbol{X}_M \otimes \widehat{\boldsymbol{\Sigma}}_i^{-1}\right]^{-1}\left(\boldsymbol{C}_M' \otimes \boldsymbol{U}\right) = \left[\boldsymbol{C}_M\left(\boldsymbol{X}_M'\boldsymbol{X}_M\right)^{-1}\boldsymbol{C}_M'\right] \otimes \left(\boldsymbol{U}'\widehat{\boldsymbol{\Sigma}}_i\boldsymbol{U}\right) = \boldsymbol{M}_M \otimes \widehat{\boldsymbol{\Sigma}}_*$. The Wald statistic reduces to $F_m = (\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)'[\boldsymbol{C}_m(\boldsymbol{X}_s'\widehat{\boldsymbol{\Sigma}}_s^{-1}\boldsymbol{X}_s)^{-1}\boldsymbol{C}_m']^{-1}(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)/a_m = [\mathrm{vec}(\widehat{\boldsymbol{\Theta}} - \boldsymbol{\Theta}_0)]'(\boldsymbol{M}_M \otimes \widehat{\boldsymbol{\Sigma}}_*)^{-1}[\mathrm{vec}(\widehat{\boldsymbol{\Theta}} - \boldsymbol{\Theta}_0)]/a_m = [\mathrm{vec}(\widehat{\boldsymbol{\Theta}} - \boldsymbol{\Theta}_0)]'(\boldsymbol{M}_M^{-1} \otimes \widehat{\boldsymbol{\Sigma}}_*^{-1})$

$[\mathrm{vec}(\widehat{\boldsymbol{\Theta}}-\boldsymbol{\Theta}_0)]/a_m$. Here, F_m is invariant to full-rank transformation of columns of $\boldsymbol{U}$ and $\boldsymbol{\Theta}_0$: $(\boldsymbol{M}_M^{-1}\otimes\widehat{\boldsymbol{\Sigma}}_*^{-1})\mathrm{vec}(\widehat{\boldsymbol{\Theta}}-\boldsymbol{\Theta}_0)=(\boldsymbol{M}_M^{-1}\otimes\widehat{\boldsymbol{\Sigma}}_*^{-1})\left(\boldsymbol{T}^{-1}\otimes\boldsymbol{I}_a\right)(\boldsymbol{T}\otimes\boldsymbol{I}_a)\mathrm{vec}(\widehat{\boldsymbol{\Theta}}-\boldsymbol{\Theta}_0)=(\boldsymbol{M}_M^{-1}\otimes\widehat{\boldsymbol{\Sigma}}_*^{-1})\mathrm{vec}[(\boldsymbol{C}\widehat{\boldsymbol{B}}\boldsymbol{U}\boldsymbol{T}-\boldsymbol{\Theta}_0\boldsymbol{T})]$. Singular value decomposition gives $\boldsymbol{U}=\boldsymbol{L}_+\mathrm{Dg}(\boldsymbol{s})\,\boldsymbol{R}'$ with $\boldsymbol{L}_+$ and $\boldsymbol{R}$ orthonormal by column. Using $\boldsymbol{T}=\boldsymbol{R}\mathrm{Dg}(\boldsymbol{s})^{-1}$ allows assuming $\boldsymbol{U}=\boldsymbol{L}_+$, the eigenvectors of $\boldsymbol{U}\boldsymbol{U}'$ for nonzero eigenvalues.

A $p\times p$ CS covariance is $\boldsymbol{\Sigma}_{i\mathrm{CS}}=\sigma^2\left[\mathbf{1}_p\mathbf{1}_p'\rho+\boldsymbol{I}_p(1-\rho)\right]=\boldsymbol{V}\mathrm{Dg}(\boldsymbol{\lambda})\,\boldsymbol{V}'=\begin{bmatrix}\boldsymbol{v}_0 & \boldsymbol{V}_t\end{bmatrix}\begin{bmatrix}\lambda_1 & \mathbf{0}\\ \mathbf{0} & \lambda_2\boldsymbol{I}_{p-1}\end{bmatrix}\begin{bmatrix}\boldsymbol{v}_0'\\ \boldsymbol{V}_t'\end{bmatrix}$. Here, $\lambda_1=\sigma^2[1+(p-1)\rho]$ and $\lambda_2=\sigma^2(1-\rho)$, whereas $\boldsymbol{v}_0=\mathbf{1}_p/p^{1/2}$ and $\boldsymbol{V}_t$ may be taken to be any $p\times(p-1)$ orthonormal matrix with $\boldsymbol{V}_t'\boldsymbol{v}_0=\mathbf{0}$.

Three distinct classes of hypotheses occur: (i) pure between hypotheses have $p\times1$ $\boldsymbol{U}=w\cdot\boldsymbol{v}_0$ for $w\neq0$; (ii) pure within hypotheses have $\boldsymbol{U}'\boldsymbol{v}_0=\mathbf{0}$; and (iii) combinations have $\boldsymbol{U}'\boldsymbol{v}_0\neq\mathbf{0}$ and $\mathrm{rank}(\boldsymbol{U})>1$. For the sake of brevity, we provide an explicit form only for pure within hypotheses. The other cases have the same structure with different weights in the resulting quadratic forms. If $\boldsymbol{U}'\boldsymbol{v}_0=\mathbf{0}$ and $p\times b$ $\boldsymbol{U}=\boldsymbol{L}_+$, then without loss of generality $\boldsymbol{V}_t=\begin{bmatrix}\boldsymbol{L}_+ & \boldsymbol{L}_0\end{bmatrix}$ with $\boldsymbol{L}_0$ any $p-1-b\geqslant0$ eigenvectors of $\boldsymbol{U}\boldsymbol{U}'$ for *zero* eigenvalues.

Compound symmetric estimates may be computed in terms of the unstructured maximum likelihood ($\widetilde{\boldsymbol{\Sigma}}_i$) and unstructured REML ($\widehat{\boldsymbol{\Sigma}}_i$) covariance estimates [31]. Here, $\widehat{\sigma}^2=\mathrm{tr}(\widehat{\boldsymbol{\Sigma}}_i)/p=\widetilde{\sigma}^2N/(N-r_M)$ and $\widetilde{\rho}=\left[\mathbf{1}_p'\widetilde{\boldsymbol{\Sigma}}_i\mathbf{1}_p-\mathrm{tr}(\widetilde{\boldsymbol{\Sigma}}_i)\right]/\left[\mathrm{tr}(\widetilde{\boldsymbol{\Sigma}}_i)(p-1)\right]=\mathbf{1}_p'\widehat{\boldsymbol{\Sigma}}_i\mathbf{1}_p-\mathrm{tr}(\widehat{\boldsymbol{\Sigma}}_i)/\left[\mathrm{tr}(\widehat{\boldsymbol{\Sigma}}_i)(p-1)\right]=\widehat{\rho}$ give $\widehat{\lambda}_1=\widehat{\sigma}^2[1+(p-1)\widehat{\rho}]=\mathrm{tr}(\widehat{\boldsymbol{\Sigma}}_i)/p+[\mathbf{1}_p'\widehat{\boldsymbol{\Sigma}}_i\mathbf{1}_p-\mathrm{tr}(\widehat{\boldsymbol{\Sigma}}_i)]/p=\mathbf{1}_p'\widehat{\boldsymbol{\Sigma}}_i\mathbf{1}_p/p$ and $\widehat{\lambda}_2=\widehat{\sigma}^2(1-\widehat{\rho})=\mathrm{tr}(\widehat{\boldsymbol{\Sigma}}_i)/p-[\mathbf{1}_p'\widehat{\boldsymbol{\Sigma}}_i\mathbf{1}_p-\mathrm{tr}(\widehat{\boldsymbol{\Sigma}}_i)]/[p(p-1)]=\left[p\mathrm{tr}(\widehat{\boldsymbol{\Sigma}}_i)-\mathbf{1}_p'\widehat{\boldsymbol{\Sigma}}_i\mathbf{1}_p\right]/[p(p-1)]$. Hence, the REML estimate for a CS covariance matrix in a general linear mixed model corresponding to a multivariate model may be computed noniteratively as $\widehat{\boldsymbol{\Sigma}}_{i\mathrm{CS}}=\begin{bmatrix}\boldsymbol{v}_0 & \boldsymbol{V}_t\end{bmatrix}\begin{bmatrix}\widehat{\lambda}_1 & \mathbf{0}\\ \mathbf{0} & \widehat{\lambda}_2\boldsymbol{I}_{p-1}\end{bmatrix}\begin{bmatrix}\boldsymbol{v}_0'\\ \boldsymbol{V}_t'\end{bmatrix}$. Assuming compound symmetry in the mixed model gives $\widehat{\boldsymbol{\Sigma}}_i=\widehat{\boldsymbol{\Sigma}}_{i\mathrm{CS}}$ and $\widehat{\boldsymbol{\Sigma}}_s=(\boldsymbol{I}_N\otimes\widehat{\boldsymbol{\Sigma}}_{i\mathrm{CS}})$.

For pure within hypotheses (such as a test of the linear trend), $\boldsymbol{U}'\widehat{\boldsymbol{\Sigma}}_{i\mathrm{CS}}\boldsymbol{U}=\boldsymbol{L}_+'\left(\begin{bmatrix}\boldsymbol{v}_0 & \boldsymbol{L}_+ & \boldsymbol{L}_0\end{bmatrix}\begin{bmatrix}\widehat{\lambda}_1 & \mathbf{0}\\ \mathbf{0} & \widehat{\lambda}_2\boldsymbol{I}_{p-1}\end{bmatrix}\begin{bmatrix}\boldsymbol{v}_0 & \boldsymbol{L}_+ & \boldsymbol{L}_0\end{bmatrix}'\right)\boldsymbol{L}_+=\widehat{\lambda}_2\boldsymbol{I}_b$. With $\widehat{\boldsymbol{\theta}}_k$ indicating column k of $\widehat{\boldsymbol{\Theta}}$,

$$\begin{aligned}a_mF_m&=[\mathrm{vec}(\widehat{\boldsymbol{\Theta}}-\boldsymbol{\Theta}_0)]'[(\boldsymbol{M}_M\otimes\widehat{\lambda}_2\boldsymbol{I}_b)^{-1}][\mathrm{vec}(\widehat{\boldsymbol{\Theta}}-\boldsymbol{\Theta}_0)]\\&=[\mathrm{vec}(\widehat{\boldsymbol{\Theta}}-\boldsymbol{\Theta}_0)]'\left(\boldsymbol{M}_M^{-1}\otimes\boldsymbol{I}_b\right)\left[(\widehat{\boldsymbol{\theta}}_1-\boldsymbol{\theta}_{0,1})'\quad\cdots\quad(\widehat{\boldsymbol{\theta}}_1-\boldsymbol{\theta}_{0,1})'\right]'/\widehat{\lambda}_2\\&=\widehat{\lambda}_2^{-1}\sum_{k=1}^{b}(\widehat{\boldsymbol{\theta}}_k-\boldsymbol{\theta}_{0,k})'\boldsymbol{M}_M^{-1}(\widehat{\boldsymbol{\theta}}_k-\boldsymbol{\theta}_{0,k})=\mathrm{tr}[(\widehat{\boldsymbol{\Theta}}-\boldsymbol{\Theta}_0)'\boldsymbol{M}_M^{-1}(\widehat{\boldsymbol{\Theta}}-\boldsymbol{\Theta}_0)]/\widehat{\lambda}_2\,.\end{aligned}\tag{6}$$

Writing $\boldsymbol{\Delta}=(\boldsymbol{\Theta}-\boldsymbol{\Theta}_0)'\,\boldsymbol{M}_M^{-1}\,(\boldsymbol{\Theta}-\boldsymbol{\Theta}_0)$ implies $a_mF_m=\mathrm{tr}(\widehat{\boldsymbol{\Delta}})/\widehat{\lambda}_2$ with $a_m=a_Mb$ for $\boldsymbol{\Theta}$ $a_M\times b$. Multivariate model results give $\widehat{\boldsymbol{\Delta}}=(\widehat{\boldsymbol{\Theta}}-\boldsymbol{\Theta}_0)'\boldsymbol{M}_M^{-1}(\widehat{\boldsymbol{\Theta}}-\boldsymbol{\Theta}_0)\sim\mathcal{W}_b\left(a_M,\boldsymbol{U}'\boldsymbol{\Sigma}_i\boldsymbol{U},\boldsymbol{\Delta}\right)$ and $\widehat{\boldsymbol{\Delta}}$ independent of $\widehat{\boldsymbol{\Sigma}}_i$, which insures $\widehat{\boldsymbol{\Delta}}$ independent of $\widehat{\lambda}_2=[p\mathrm{tr}(\widehat{\boldsymbol{\Sigma}}_i)-\mathbf{1}_p'\widehat{\boldsymbol{\Sigma}}_i\mathbf{1}_p]/[p(p-1)]$.

Under H_0, $\mathrm{tr}(\widehat{\boldsymbol{\Delta}})=\sum_{k=1}^{b}\lambda_{hk}X_{hk}$ for i.i.d. $X_{hk}\sim\chi^2(a_M)$ and λ_{hk} an eigenvalue of $\boldsymbol{\Sigma}_h=\boldsymbol{U}'\boldsymbol{\Sigma}_i\boldsymbol{U}$. If $\widehat{\boldsymbol{E}}=\boldsymbol{Y}-\boldsymbol{X}_M\widehat{\boldsymbol{B}}$, then $p(p-1)\nu_e\widehat{\lambda}_2=p\mathrm{tr}(\widehat{\boldsymbol{E}}'\widehat{\boldsymbol{E}})-\mathrm{tr}(\mathbf{1}_p'\widehat{\boldsymbol{E}}'\widehat{\boldsymbol{E}}\mathbf{1}_p)=p\mathrm{tr}(\widehat{\boldsymbol{E}}\widehat{\boldsymbol{E}}')-\mathrm{tr}(\widehat{\boldsymbol{E}}\mathbf{1}_p\mathbf{1}_p'\widehat{\boldsymbol{E}}')$. In turn, $\nu_e(p-1)\widehat{\lambda}_2=\mathrm{tr}[\widehat{\boldsymbol{E}}\left(\boldsymbol{I}_p-\mathbf{1}_p\mathbf{1}_p'/p\right)\widehat{\boldsymbol{E}}']$. Here, $\left(\boldsymbol{I}_p-\mathbf{1}_p\mathbf{1}_p'/p\right)=\begin{bmatrix}\boldsymbol{v}_0 & \boldsymbol{V}_t\end{bmatrix}\begin{bmatrix}0 & \mathbf{0}\\ \mathbf{0} & \boldsymbol{I}_{p-1}\end{bmatrix}\begin{bmatrix}\boldsymbol{v}_0'\\ \boldsymbol{V}_t'\end{bmatrix}=\boldsymbol{V}_t\boldsymbol{V}_t'$ so $\nu_e(p-1)\widehat{\lambda}_2=\mathrm{tr}(\widehat{\boldsymbol{E}}\boldsymbol{V}_t\boldsymbol{V}_t'\widehat{\boldsymbol{E}}')=\mathrm{tr}(\boldsymbol{V}_t'\widehat{\boldsymbol{E}}'\widehat{\boldsymbol{E}}\boldsymbol{V}_t)$ and $\widehat{\lambda}_2=\mathrm{tr}(\boldsymbol{V}_t'\widehat{\boldsymbol{\Sigma}}_i\boldsymbol{V}_t)/(p-1)$ with $\nu_e\boldsymbol{V}_t'\widehat{\boldsymbol{\Sigma}}_i\boldsymbol{V}_t\sim\mathcal{W}_{p-1}\left(\nu_e,\boldsymbol{V}_t'\boldsymbol{\Sigma}\boldsymbol{V}_t,\mathbf{0}\right)$. Hence, $\nu_e(p-1)\widehat{\lambda}_2=\sum_{k=1}^{p-1}\lambda_{ek}X_{ek}$ for i.i.d. $X_{ek}\sim\chi^2(\nu_e,0)$ and λ_{ek} an eigenvalue of $\boldsymbol{\Sigma}_e=\boldsymbol{V}_t'\boldsymbol{\Sigma}\boldsymbol{V}_t$. Finally, $\Pr\{F_m\leqslant f\}=\Pr\{\mathrm{tr}(\widehat{\boldsymbol{\Delta}})/(a_m\widehat{\lambda}_2)\leqslant f\}=\Pr\{\mathrm{tr}(\widehat{\boldsymbol{\Delta}})-a_mf\widehat{\lambda}_2\leqslant0\}=\Pr\{Q\leqslant0\}$, for $Q=\sum_{k=1}^{b}\lambda_{hk}X_{hk}-a_mf\sum_{k=1}^{p-1}\lambda_{ek}X_{ek}$ a quadratic form with positive and negative weights.

The fact that $\Pr\{F_m\leqslant f\}=\Pr\{Q\leqslant0\}$ allows computing the exact distribution function of the Wald statistic *in finite samples* with the Kenward–Roger approximation. Results in [12] allows using modules from POWERLIB [28] that implements exact methods described in [29].

In large samples, under H_0, the numerator distribution of F_m does not change as $N \to \infty$ because $\text{tr}(\widehat{\boldsymbol{\Delta}})$ does not depend on N for i.i.d. $X_{hk} \sim \chi^2(a_M)$. The denominator converges to a constant, $\widehat{\lambda}_2 = \text{tr}(\boldsymbol{V}_t' \widehat{\boldsymbol{\Sigma}}_i \boldsymbol{V}_t)/(p-1) \to \text{tr}(\boldsymbol{V}_t' \boldsymbol{\Sigma}_i \boldsymbol{V}_t)/(p-1)$ and thereby becomes known. Hence, under the null $\lim_{N\to\infty} F_m = \sum_{k=1}^{b} \lambda_{hk} X_{hk} / \left[a_M \text{tr}\left(\boldsymbol{U}' \boldsymbol{\Sigma}_i \boldsymbol{U}\right)\right]$, a quadratic form, which allows exact probability calculations. Satterthwaite's method gives an accurate approximation.

Under H_A, as in [15], $\boldsymbol{U}' \boldsymbol{\Sigma}_i \boldsymbol{U} = \boldsymbol{\Upsilon} \text{Dg}(\boldsymbol{\lambda}_*) \boldsymbol{\Upsilon}'$ defines noncentrality $\omega_{*k} = \boldsymbol{v}_k' (\boldsymbol{\Theta} - \boldsymbol{\Theta}_0)' \boldsymbol{M}_M^{-1} (\boldsymbol{\Theta} - \boldsymbol{\Theta}_0) \boldsymbol{v}_k / \lambda_{*k}$, a diagonal element of $\boldsymbol{\Omega}_* = \boldsymbol{\Upsilon}' \boldsymbol{\Delta} \boldsymbol{\Upsilon} \text{Dg}(\boldsymbol{\lambda}_h)^{-1}$. Here, $\text{tr}(\widehat{\boldsymbol{\Delta}}) = \sum_{k=1}^{b} \lambda_{hk} X_{hk}$ for i.i.d. $X_{hk} \sim \chi^2(a_M, \omega_{*k})$ and λ_{hk} an eigenvalue of $\boldsymbol{\Sigma}_h = \boldsymbol{U}' \boldsymbol{\Sigma}_i \boldsymbol{U}$. Methods in [29] allow exact computation and accurate approximation of the cumulative distribution function of F_m for finite samples. In large samples, as under H_0, $\widehat{\lambda}_2 \to \text{tr}(\boldsymbol{V}_t' \boldsymbol{\Sigma}_i \boldsymbol{V}_t)/(p-1)$. With nonlocal alternatives $\lim_{N\to\infty} F_m \to \infty$ and power $\to 1$. For local alternatives $\boldsymbol{\beta} \cdot N^{-1/2}$ replaces $\boldsymbol{\beta}$ and $\omega_{*k} = \boldsymbol{v}_k' (\boldsymbol{\Theta} - \boldsymbol{\Theta}_0)' \boldsymbol{C}_M \left[\text{Es}(\boldsymbol{X}_M)' \text{Es}(\boldsymbol{X}_M)\right]^{-1} \boldsymbol{C}_M' (\boldsymbol{\Theta} - \boldsymbol{\Theta}_0) \boldsymbol{v}_k / \lambda_{*k}$. The *essence matrix* $\text{Es}(\boldsymbol{X}_M)$ contains one copy of each unique row of $\boldsymbol{X}_M$, and $\boldsymbol{M}_M = N^{-1} \boldsymbol{C}_M \left[\text{Es}(\boldsymbol{X}_M)' \text{Es}(\boldsymbol{X}_M)\right]^{-1} \boldsymbol{C}_M'$. Here, ω_{*k} does not depend on N and F_m converges to a noncentral quadratic form, distinct from the finite sample form.

Acknowledgements

NIH/NICHD grant R03-HD055298 and NIH/NIDDK grant R21-DK085363-01A1 partly supported Dr Gurka. NIH/NIDCR grants U54-DE019261 and R01DE020832-01A1, NIH/NCRR grant K30-RR022258, NIH/NHLBI grant R01-HL091005, and NIH/NIAAA grant R01-AA013458-01 partly supported Dr Muller. Use of the SECCYD data was classified as exempt by the University of Virginia and the West Virginia University Institutional Review Boards.

References

1. National Center for Health Statistics. Asthma prevalence, health care use and mortality, 2002. http://www.cdc.gov/nchs/products/pubs/pubd/hestats/asthma/asthma.htm, Accessed: [February 2010].
2. Blackman JA, Gurka MJ. Developmental and behavioral co-morbidities of asthma in children. *Journal of Developmental and Behavioral Pediatrics* 2007; **28**:92–99.
3. Gurka MJ, Blackman JA, Heymann PW. Risk of childhood asthma in relation to the timing of early child care exposures. *Journal of Pediatrics* 2009; **155**:781–787.
4. Gurka MJ. Selecting the best linear mixed model under REML. *The American Statistician* 2006; **60**:19–26.
5. Verbeke G, Molenberghs G. *Linear Mixed Models for Longitudinal Data*. Springer: New York, 2000.
6. Liang KY, Zeger SL. Longitudinal data analysis using generalized linear models. *Biometrika* 1986; **73**:13–22.
7. Jacqmin-Gadda H, Sibillot S, Proust C, et al. Robustness of the linear mixed model to misspecified error distribution. *Computational Statistics and Data Analysis* 2007; **51**:5142–5154.
8. Lange N, Laird NM. The effect of covariance structure on variance estimation in balanced growth-curve models with random parameters. *Journal of the American Statistical Association* 1989; **84**:241–247.
9. Bainbridge KE, Sowers MF, Crutchfield M, et al. Natural history of bone loss over 6 years among premenopausal and early postmenopausal women. *American Journal of Epidemiology* 2002; **156**:410–417.
10. Kotch JB, Lewis T, Hussey JM, et al. Importance of early neglect for childhood aggression. *Pediatrics* 2008; **121**:725–731.
11. Conner-Spady BL, Cumming C, Nabholtz JM, et al. A longitudinal prospective study of health-related quality of life in breast cancer patients following high-dose chemotherapy with autologous blood stem cell transplantation. *Bone Marrow Transplantation* 2005; **36**:251–259.
12. Murphy MM, Molloy AM, Ueland PM, et al. Longitudinal study of the effect of pregnancy on maternal and fetal cobalamin status in healthy women and their offspring. *Journal of Nutrition* 2007; **137**:1863–1867.
13. Muller KE, Stewart PW. *Linear Model Theory: Univariate, Multivariate, and Mixed Models*. John Wiley and Sons, Inc: Hoboken, New Jersey, 2006.
14. Spybrook J, Raudenbush SW, Liu XF, et al. *Optimal design for longitudinal and multilevel research: documentation for the "Optimal Design" software* (V 1.7.6), 2008. Available at http://sitemaker.umich.edu/group-based/optimal_design_software, Accessed: [February 2010].
15. Box GEP. Some theorems on quadratic forms applied in the study of analysis of variance problems: I. effects of inequality of variance in the one-way classification. *Annals of Mathematical Statistics* 1954a; **25**:290–302.
16. Box GEP. Some theorems on quadratic forms applied in the study of analysis of variance problems: II. effects of inequality of variance and of correlation between errors in the two-way classification. *Annals of Mathematical Statistics* 1954b; **25**:484–498.
17. Laird NM, Ware JH. Random-effects models for longitudinal data. *Biometrics* 1982; **38**:963–974.
18. Murray DM. *Design and Analysis of Group-Randomized Trials*. Oxford University Press: New York, 1998.

19. Muller KE, Edwards LJ, Simpson S, Taylor DT. Statistical tests with accurate size and power for balanced linear mixed models. *Statistics in Medicine* 2007; **26**:3639–3660.
20. Muller KE, Barton CN. Approximate power for repeated-measures ANOVA lacking sphericity. *Journal of the American Statistical Association* 1989; **84**:549–555. (corrigendum 1991, **86**, 255–256).
21. SAS Institute Inc. *SAS/STAT User's Guide,* Version 9.1. SAS Institute Inc: Cary, NC, 2003.
22. Johnson JL, Muller KE, Slaughter JC, et al. POWERLIB: SAS/IML software for computing power in multivariate linear models. *Journal of Statistical Software* 2009; **30**(5):1–27.
23. Kim H, Gribbin MJ, Muller KE, Taylor DJ. Analytic and computational forms for the ratio of a noncentral chi square and a Gaussian quadratic form. *Journal of Computational and Graphical Statistics* 2006; **15**:443–459.
24. Davies R B. Algorithm AS 155: the distribution of a linear combination of χ^2 random variables. *Applied Statistics* 1980; **29**:323–333.
25. Greenhouse SW, Geisser S. On methods in the analysis of profile data. *Psychometrika* 1959; **24**:95–112.
26. Kenward MG, Roger JH. Small sample inference for fixed effects from restricted maximum likelihood. *Biometrics* 1997; **53**:983–997.
27. Achenbach T M. *Manual for the Child Behavior Checklist/4-18 and 1991 Profile*. University of Vermont, Department of Psychiatry: Burlington, VT, 1991.
28. Simpson SL, Edwards LJ, Muller KE, et al. A linear exponent AR (1) family of correlation structures. *Statistics in Medicine* 2010; **29**:1825–1838.
29. Muller KE, LaVange LM, Ramey SL, Ramey CT. Power calculations for general linear multivariate models including repeated measures applications. *Journal of the American Statistical Association* 1992; **87**:1209–1226.
30. Cheng J, Edwards LJ, Maldonado-Molina MM, et al. Real longitudinal data analysis for real people: building a good enough mixed model. *Statistics in Medicine* 2010; **29**:504–520.
31. Kistner EO, Muller KE. Exact distributions of intraclass correlation and Cronbach's alpha with Gaussian data and general covariance. *Psychometrika* 2004; **69**:459–474.