Published in final edited form as:
Stat Med. 2008 December 20; 27(29): 6137–6157. doi:10.1002/sim.3429.

An R^2 Statistic for Fixed Effects in the Linear Mixed Model

Lloyd J. Edwards[1,*], **Keith E. Muller**[2], **Russell D. Wolfinger**[3], **Bahjat F. Qaqish**[1], and **Oliver Schabenberger**[3]

[1] Department of Biostatistics, School of Public Health CB# 7420, The University of North Carolina, Chapel Hill, NC 27599, U.S.A

[2] Division of Biostatistics, Department of Epidemiology and Health Policy Research, University of Florida, Gainesville, FL 32610, U.S.A

[3] SAS Institute, Inc., Cary, NC 27513, U.S.A

SUMMARY

Statisticians most often use the linear mixed model to analyze Gaussian longitudinal data. The value and familiarity of the R^2 statistic in the linear univariate model naturally creates great interest in extending it to the linear mixed model. We define and describe how to compute a model R^2 statistic for the linear mixed model by using only a single model. The proposed R^2 statistic measures multivariate association between the repeated outcomes and the fixed effects in the linear mixed model. The R^2 statistic arises as a 1–1 function of an appropriate F statistic for testing all fixed effects (except typically the intercept) in a full model. The statistic compares the full model to a null model with all fixed effects deleted (except typically the intercept) while retaining exactly the same covariance structure. Furthermore, the R^2 statistic leads immediately to a natural definition of a partial R^2 statistic. A mixed model in which ethnicity gives a very small p-value as a longitudinal predictor of blood pressure compellingly illustrates the value of the statistic. In sharp contrast to the extreme p-value, a very small R^2, a measure of statistical and scientific importance, indicates that ethnicity has an almost negligible association with the repeated blood pressure outcomes for the study.

Keywords

Goodness-of-fit; Longitudinal data; Model selection; Multiple correlation; Restricted maximum likelihood

1. Introduction

In the linear univariate model, the sample squared multiple correlation coefficient, R^2, measures the maximum overall linear association of a single dependent variable with several independent variables. In the univariate model, R^2 corresponds to comparing two models [1, Chapter 6, Sections 6.9–6.11]: 1. a *full* model that consists of $p-1$ independent predictors and an intercept; 2. a null model that has only the intercept. It also measures the overall *linear association* of one (dependent) variable $\boldsymbol{Y}$ with several other (independent) variables $X_1, X_2, \ldots, X_{p-1}$, which corresponds to adding $p-1$ predictors to an intercept-only model. Most linear regression and ANOVA software packages provide the model (overall) R^2. However, little attention has been given to developing an R^2 statistic for the linear mixed model, the most widely used statistical tool for analyzing Gaussian longitudinal data.

*email: Lloyd_Edwards@unc.edu.

NIH-PA Author Manuscript
NIH-PA Author Manuscript

The linear mixed model explicitly specifies not only the mean structure, but also the covariance structure. Hence three types of model comparisons can occur. I) Compare mean models with the same covariance structure. Nested mean models are the most common. II) Compare covariance models with the same mean structure. Two linear mixed models may be nested or nonnested in the covariance models. III) Compare linear mixed models with different mean and different covariance structures. Consequently any definition of an R^2 statistic for the linear mixed model must account for the distinction between the proportion of variation in the response explained by the fixed effects (in the mean model) and the proportion explained by the random effects (in the covariance model). The same distinction arises in measuring the degree of association between the repeated outcomes and the fixed effects. However, measuring association for random effects seems to require a distinct treatment. Here we describe an R^2 only for item I, i.e., comparing nested mean models with the same covariance structure.

Our focused approach aligns with earlier work in the area. The differences between fixed (mean) and random (covariance) effects led Snijders and Bosker [2] to propose distinct R^2 statistics for fixed and random effects in terms of the corresponding proportions of modeled variances. Kramer [3] concluded that “Different problems necessarily emphasize the importance of different parts of a model – this is a fundamental component to modeling a process and cannot be resolved mathematically. Thus, there can be no general definition of R^2 for mixed models that will cover every model, which is problematic for software developers.”

We use an approximate F statistic for a Wald test of fixed effects to define an R^2 statistic for fixed effects in the linear mixed model. The R^2 statistic measures multivariate association between the repeated outcomes and the fixed effects in the linear mixed model. Defining R^2 in terms of an F statistic for fixed effects allows computing it with results from fitting only a single model, i.e., there is no need to explicitly fit a null model. The approach necessarily assumes the covariance structure holds for both the model of interest and the implied null model.

2. The Linear Mixed Model

With N independent sampling units (often *persons* in practice), the linear mixed model for person i may be written [4, notation in 5, ch.5]

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

Here, $\boldsymbol{y}_i$ is a $p_i \times 1$ vector of observations on person i; $\boldsymbol{X}_i$ is a $p_i \times q$ known, constant design matrix for person i, with full column rank q while $\boldsymbol{\beta}$ is a $q \times 1$ vector of unknown, constant, population parameters. Also $\boldsymbol{Z}_i$ is a $p_i \times m$ known, constant design matrix with rank m for person i corresponding to the $m \times 1$ vector of unknown random effects $\boldsymbol{d}_i$, while $\boldsymbol{e}_i$ is a $p_i \times 1$ vector of unknown random errors. Gaussian $\boldsymbol{d}_i$ and $\boldsymbol{e}_i$ are independent with mean $\mathbf{0}$ and

$$\mathcal{V}\left(\begin{bmatrix} \boldsymbol{d}_i \\ \boldsymbol{e}_i \end{bmatrix}\right) = \begin{bmatrix} \sum_{di}(\boldsymbol{\tau}_d) & \mathbf{0} \\ \mathbf{0} & \sum_{ei}(\boldsymbol{\tau}_e) \end{bmatrix}. \tag{2}$$

Here $\mathcal{V}(\cdot)$ is the covariance operator, while both $\boldsymbol{\Sigma}_{di}(\boldsymbol{\tau}_d)$ and $\boldsymbol{\Sigma}_{ei}(\boldsymbol{\tau}_e)$ are positive-definite, symmetric covariance matrices. Therefore $\mathcal{V}(\boldsymbol{y}_i)$ may be written

$\sum_i = Z_i \sum_{di}(\tau_d) Z_i' + \sum_{ei}(\tau_e)$. We assume that $\boldsymbol{\Sigma}_i$ can be characterized by a finite set of parameters represented by an r×1 vector $\boldsymbol{\tau}$ which consists of the unique parameters in $\boldsymbol{\tau}_d$ and $\boldsymbol{\tau}_e$. Throughout $n = \sum_{i=1}^{N} p_i$.

We will also need to refer to a stacked data formulation of model (1) given by

$$\boldsymbol{y}_s = \boldsymbol{X}_s \beta + \boldsymbol{Z}_s \boldsymbol{d}_s + \boldsymbol{e}_s, \tag{3}$$

with $\boldsymbol{y}_s = [\boldsymbol{y}_1' \cdots \boldsymbol{y}_N']'$, $\boldsymbol{X}_s = [\boldsymbol{X}_1' \cdots \boldsymbol{X}_N']'$, $\boldsymbol{Z}_s = \text{diag}(\boldsymbol{Z}_1, \cdots, \boldsymbol{Z}_N)$, $\boldsymbol{d}_s = [\boldsymbol{d}_1' \cdots \boldsymbol{d}_N']'$, and $\boldsymbol{e}_s = [\boldsymbol{e}_1' \cdots \boldsymbol{e}_N']'$. Here $\boldsymbol{d}_s \sim \mathcal{N}_{Nq}[\boldsymbol{0}, \boldsymbol{\Sigma}_{di}(\boldsymbol{\tau}_d) \otimes \boldsymbol{I}_N]$ and $\boldsymbol{e}_s \sim \mathcal{N}_n(\boldsymbol{0}, \boldsymbol{\Sigma}_{es})$ for $\boldsymbol{\Sigma}_{es} = \text{diag}[\boldsymbol{\Sigma}_{e1}(\boldsymbol{\tau}_e), \cdots, \boldsymbol{\Sigma}_{eN}(\boldsymbol{\tau}_e)]$. In turn $\boldsymbol{y}_s \sim \mathcal{N}_n(\boldsymbol{X}_s\boldsymbol{\beta}, \boldsymbol{\Sigma}_s)$ with $\boldsymbol{\Sigma}_s = \mathcal{V}(\boldsymbol{y}_s) = \text{diag}(\boldsymbol{\Sigma}_1, \cdots, \boldsymbol{\Sigma}_N)$.

The advantage of reducing bias in covariance estimation has made restricted maximum likelihood (REML) estimation very popular for the linear mixed model. Our focus on variance estimates means that all parameter estimates in this paper are done using REML. However, the formulations also apply to computations based on maximum likelihood estimates.

3. A Review of R^2 Statistics for the Linear Mixed Model

The popularity of the R^2 statistic in the linear univariate model has led to direct analogs being proposed for the linear mixed model that measure the proportionate reduction in residual variation explained by the model. We provide a brief review of some of the R^2 statistics.

The choice of a null model plays an essential role in all definitions of R^2 statistics for the linear mixed model because it affects the interpretation and properties of R^2. Two null models have been discussed most often: 1) an intercept-only model in the fixed effects, i.e., for all persons in model (1) mentioned earlier, $\boldsymbol{\beta} = \beta_0$, $\boldsymbol{X}_i = \boldsymbol{1}_{pi}$, a p_i×1 vector of 1's, and $\boldsymbol{d}_i = \boldsymbol{0}$; 2) an intercept-only model in both the fixed and random effects, i.e., $\boldsymbol{\beta} = \beta_0$, $\boldsymbol{X}_i = \boldsymbol{1}_{pi}$, $\boldsymbol{d}_i = d_{0i}$, and $\boldsymbol{Z}_i = \boldsymbol{1}_{pi}$. However, other choices for the null model are possible, depending on the application, interpretation, and scientific question of interest. Guidelines for null model selection do not exist. As described in Section 4, evaluating fixed effects alone corresponds to comparing nested mean models and leads us to prefer a different null model.

Considering the linear mixed model as a random effects hierarchical two-level model, Snijders and Bosker [2] proposed an R^2 statistic for the fixed and random effects. The two-level model may be expressed in scalar form as

$$y_{ij} = \boldsymbol{X}_{ij}\beta + \sum_{h=0}^{m} Z_{hij} d_{hi} + e_{ij}, \tag{4}$$

where h ranges from 0 to m and $Z_{0ij} = 1$ for all i, j. Level 1 variables describe within person variation while level 2 variables describe between person variation. For longitudinal data, level-1 variables are indexed by j (within-person observations) and level-2 variables are indexed by i (between-person observations). Snijders and Bosker argued that the proportional reduction in variance components does not represent the joint importance of the predictor variables. The same authors interpreted their proposed R^2 statistics as the proportion of "modeled variance", as opposed to "explained variance". They argued that the principle of proportional reduction of prediction error gives the best way to define measures of modeled or explained variance.

Snijders and Bosker [2] defined R^2 for fixed effects (level-2 variables) in the linear mixed model as

$$R^2_{SB}=1-\frac{\widehat{\mathcal{V}}(\bar{y}_{i\cdot}-\overline{\boldsymbol{X}}_{i\cdot}\beta)}{\widehat{\mathcal{V}}(\overline{\boldsymbol{y}}_{i\cdot})}, \tag{5}$$

where $\bar{y}_{i\cdot}$ is the mean for person i across j and $\bar{\boldsymbol{X}}_{i\cdot}$ is the mean vector of predictor variables across $j \in \{1, 2, \cdots, p_i\}$. The quantities $\hat{\mathcal{V}}(\bar{y}_{i\cdot} - \bar{\boldsymbol{X}}_{i\cdot}\boldsymbol{\beta})$ and $\hat{\mathcal{V}}(\bar{y}_{i\cdot})$ are the (estimated) mean squared prediction error for the model of interest and the null model, respectively. As for several of the R^2 statistics discussed earlier, a very important assumption is that $\mathcal{V}(y_{ij})$ is the same for all i, j (homogeneity of variance) and that $\boldsymbol{\Sigma}_{ei}(\boldsymbol{\tau}_e) = \mathcal{V}(\boldsymbol{e}_i) = \sigma^2\boldsymbol{I}_{p_i}$ (within-person covariance) for all i. The null model contains only an intercept in the fixed and the random effects. The assumption restricts the null model covariance matrix within a person, $\mathcal{V}(\boldsymbol{y}_i)$ for model (1), to compound symmetry, which clearly may not be appropriate for a wide array of longitudinal and other types of data. Here $\widehat{\mathcal{V}}(\bar{y}_{i\cdot})=n_*^{-1}\widehat{\sigma}^2+\widehat{\sigma}^2_d$, where $\widehat{\sigma}^2_d$ is the estimated variance of the random intercepts. For balanced data, $p_i = n_*$ for all i. For unbalanced data, Snijders and Bosker [2] suggested choosing n_* as "either a value deemed a priori as being representative, or the harmonic mean, defined by $\left(N^{-1}\sum_1^N p_i^{-1}\right)^{-1}$." The estimated variance $\hat{\mathcal{V}}(\bar{y}_{i\cdot} - \bar{\boldsymbol{X}}_{i\cdot}\boldsymbol{\beta})$, assuming predictor variables are random variables over some population, is given by

$$\widehat{\mathcal{V}}(\bar{y}_{i\cdot}-\overline{\boldsymbol{X}}_{i\cdot}\beta)=\sum_{h,k}^{m}\left[\mathrm{E}\left(\overline{\boldsymbol{Z}}_{hi\cdot}\overline{\boldsymbol{Z}}'_{ki\cdot}\right)\right]\widehat{\sigma}_{hk}+n_*^{-1}\widehat{\sigma}^2, \tag{6}$$

with $\hat{\sigma}_{hk}$ the hk-th element of the $m\times m$ estimated random effects covariance matrix $\hat{\boldsymbol{\Sigma}}_{di}(\boldsymbol{\tau}_d)$.

The R^2_{SB} statistic has characteristics that help to highlight how an R^2 statistic for the linear mixed model can differ from R^2 for the linear univariate model. First, R^2_{SB} may decrease when more predictors are added. As Snijders and Bosker [2] explain, this is a property of the sample estimate R^2 (R^2_{SB}). The true population R^2, under suitable conditions, should not decrease. Secondly, the statistic can be negative. In comparison, both the true population and estimated R^2 for the linear univariate model do not decrease as more predictors are added and are nonnegative [6,7]. Snijders and Bosker [2] explained that the characteristics may reflect incorrect specification of the fixed-effect portion of the model. Hence they argued that the characteristics can help diagnose model misspecification.

Vonesh and Chinchilli [8, Chap. 8, pp 420–421] described a goodness-of-fit measure for generalized linear mixed models and interpreted it as the proportionate reduction in residual variation explained by the model. For their statistic, $\hat{\boldsymbol{y}}_i$ and $\hat{\boldsymbol{y}}_{i0}$ estimate $\boldsymbol{y}_i$ under the full model of interest and null model, respectively. In turn $g_i^2(\boldsymbol{V}_i)=(\boldsymbol{y}_i-\widehat{\boldsymbol{y}}_i)'\boldsymbol{V}_i^{-1}(\boldsymbol{y}_i-\widehat{\boldsymbol{y}}_i)$ and $g_{i0}^2(\boldsymbol{V}_i)=(\boldsymbol{y}_i-\widehat{\boldsymbol{y}}_{i0})'\boldsymbol{V}_i^{-1}(\boldsymbol{y}_i-\widehat{\boldsymbol{y}}_{i0})$ give weighted distance values in the spirit of Mahalanobis distances for the full and null model respectively. The distances depend on a matrix $\boldsymbol{V}_i$, which must be positive definite but does not necessarily equal the covariance matrix for person i (which would make $g_i^2()$ and $g_{i0}^2()$ Mahalanobis distances). Relative to the null, the full model reduces residual variation in the response of person i by the proportion

$$R_i^2(\boldsymbol{V}_i)=1-\frac{g_i^2(\boldsymbol{V}_i)}{g_{i0}^2(\boldsymbol{V}_i)}. \tag{7}$$

Overall the model reduces residual variation by the weighted average

$$R_{VC}^2(\boldsymbol{V})=1-\frac{\sum_{i=1}^{N} g_i^2(\boldsymbol{V}_i)}{\sum_{i=1}^{N} g_{i0}^2(\boldsymbol{V}_i)}=\sum_{i=1}^{N} w_i R_i^2(\boldsymbol{V}_i)/\sum_{i=1}^{N} w_i, \tag{8}$$

where $w_i=g_{i0}^2(\boldsymbol{V}_i)$ and $\boldsymbol{V}=\{\boldsymbol{V}_1,\cdots,\boldsymbol{V}_N\}$. Choices of an appropriate null model and $\boldsymbol{V}_i$ crucially affect the definition of $R_{VC}^2(\boldsymbol{V})$. Vonesh and Chinchilli [8] suggested using $\boldsymbol{\Sigma}_{ei}(\boldsymbol{\tau}_e) = \mathcal{V}(\boldsymbol{e}_i)$ or $\boldsymbol{\Sigma}_{ei0}(\boldsymbol{\tau}_e) = \mathcal{V}(\boldsymbol{e}_{i0})$ for $\boldsymbol{V}_i$. Either definition implicitly ignores a major component of the variance of $\boldsymbol{y}_i$, namely $\mathcal{V}(\boldsymbol{Z}_i\boldsymbol{d}_i)=\boldsymbol{Z}_i\sum_{di}(\boldsymbol{\tau}_d)\boldsymbol{Z}_i'$. Also, by definition, if the random effects covariance matrix model differs between the null and the full model, then the structure of $\mathcal{V}(\boldsymbol{y}_i)$and $\mathcal{V}(\boldsymbol{y}_{i0})$ differ. Vonesh and Chinchilli [8, Chap. 8, p 421] concluded that choosing $\boldsymbol{V}_i = \boldsymbol{\Sigma}_{ei0}(\boldsymbol{\tau}_e)$ meets the goal of having a goodness-of-fit measure that can be compared across different hypothesized models since the null model and $\boldsymbol{\Sigma}_{ei0}(\boldsymbol{\tau}_e)$ remain fixed. However, it should be noted that $R_{VC}^2(\boldsymbol{V})$ can be defined using $\boldsymbol{V}_i=\sum_i=\boldsymbol{Z}_i\sum_{di}(\boldsymbol{\tau}_d)\boldsymbol{Z}_i'+\sum_{ei}(\boldsymbol{\tau}_e)$, thereby extending $R_{VC}^2(\boldsymbol{V})$ to the case where the model of interest and ull model have the same covariance structure.

Vonesh and Chinchilli [8, Chap. 8, pp 422–424] discussed the interpretations associated with choosing either one of the null models. If the null model contains only an intercept in the fixed effects and no random effects, then $R_{VC}^2(\boldsymbol{V})$ measures the proportionate reduction in residual ariation explained by a set of fixed effects. However, using such a null model for longitudinal data ignores the correlation between observations and essentially uses a misspecified model with $\boldsymbol{V}_i = \boldsymbol{\Sigma}_{ei}(\boldsymbol{\tau}_e) = \sigma^2\boldsymbol{I}_{p_i}$, with σ^2 an unknown constant. Consequently

$$R_{VC}^2=1-\frac{\sum_{i=1}^{N}(\boldsymbol{y}_i-\widehat{\boldsymbol{y}}_i)'(\boldsymbol{y}_i-\widehat{\boldsymbol{y}}_i)}{\sum_{i=1}^{N}(\boldsymbol{y}_i-\overline{y}\mathbf{1}_{p_i})'(\boldsymbol{y}_i-\overline{y}\mathbf{1}_{p_i})}, \tag{9}$$

where $\boldsymbol{y}_i$ is the observed response vector for person i, $\bar{y}$ is the (scalar) grand mean of the y_{ij}, $\hat{\boldsymbol{y}}_i$ is the person's predicted response vector, and $\mathbf{1}_{p_i}$ is an $p_i\times1$ vector of 1's. A marginal view implies $\hat{\boldsymbol{y}}_i = \boldsymbol{X}_i\hat{\boldsymbol{\beta}}$ and leads to interpreting R^2_{VC} as the proportionate reduction in residual variation explained by the modeled response of the average person. A conditional view implies $\hat{\boldsymbol{y}}_i = \boldsymbol{X}_i\hat{\boldsymbol{\beta}} + \boldsymbol{Z}_i\,\hat{\boldsymbol{d}}_i$ and leads to interpreting R^2_{VC} as the proportionate reduction in residual variation explained by each person's modeled response. The fact that the marginal view models the average person ($\hat{\boldsymbol{y}}_i = \boldsymbol{X}_i\hat{\boldsymbol{\beta}}$) leads to the term *average model* R^2_{VC}. Vonesh and Chinchilli [8, Chap. 8, pp 422] described it as the "usual coefficient of determination one would get if one were to ignore the repeated measures aspect of the data." Since we focus on R^2 for fixed effects, in computing R^2_{VC}, we consider only the marginal view and hence assume $\hat{\boldsymbol{y}}_i = \boldsymbol{X}_i\hat{\boldsymbol{\beta}}$.

Xu [9] also presented an R^2 statistic as the proportionate reduction in residual variation explained by the model. Under model (1), Xu [9] restricts the definition of the residual variance to the variation not explained by the covariates for $\mathcal{V}(y_{ij}|\boldsymbol{X}, \boldsymbol{d}) = \mathcal{V}(e_{ij}) = \sigma^2$, with σ^2 an unknown constant. However, as observed for R^2_{VC}, the assumption ignores a major component of the variance of $\boldsymbol{y}_i$, namely $\mathcal{V}(\boldsymbol{Z}_i\boldsymbol{d}_i)=\boldsymbol{Z}_i\sum_{di}(\boldsymbol{\tau}_d)\boldsymbol{Z}_i^{\prime}$. Xu [9] defined the proportionate reduction in residual variation explained by the fixed effects only as

$$\Omega^2=1-\frac{\mathcal{V}(y_{ij}|\boldsymbol{X},\boldsymbol{d})}{\mathcal{V}(y_{ij})}=1-\frac{\sigma^2}{\sigma^2_{00}}, \tag{10}$$

where $\sigma^2_{00}=\mathcal{V}(y_{ij})$ for a null model containing only a fixed-effect intercept and no random effects, and all observations assumed to be independent. As observed for R^2_{VC} in relation to equation (9), using such a null model for longitudinal data ignores the correlation between observations and essentially uses a misspecified model. Also, by Xu's definition, if the random effects covariance matrix for the model of interest differs from that of the null model, then $\mathcal{V}(\boldsymbol{y}_i)$ will differ from $\mathcal{V}(\boldsymbol{y}_{i0})$.

Alternatively, Xu [9] proposed the proportionate reduction defined by each person's modeled response as

$$\Omega^2=1-\frac{\mathcal{V}(y_{ij}|\boldsymbol{X},\boldsymbol{d})}{\mathcal{V}(y_{ij}|d_{0i})}=1-\frac{\sigma^2}{\sigma^2_0}, \tag{11}$$

where $\sigma^2_0=\mathcal{V}(y_{ij}|d_{0i})$ for a null model with an intercept-only model in both the fixed and random effects. Effectively Xu [9] defined the R^2 statistic to be the proportionate reduction in σ^2_0, the null model within-unit error variance for homogeneous y_{ij}, due to the fixed-effect predictor variables $\boldsymbol{X}$ and random effects $\boldsymbol{Z}$ from the model of interest. However, Snijders and Bosker [2] suggested that R^2 should estimate the proportionate reduction in $\sigma^2_0+\sigma^2_{d0}$, the null model variance for homogeneous y_{ij} (σ^2_{d0} indicates the variance of random intercepts from the null model). Xu's statistic for the linear mixed model, denoted here as R^2_x, corresponds to the parameter in (10):

$$R^2_X = 1 - \frac{\widehat{\sigma}^2}{\widehat{\sigma}^2_{00}}. \tag{12}$$

Simply replacing $\widehat{\sigma}^2_{00}$ with $\widehat{\sigma}^2_0$ in (12) gives Xu's statistic corresponding to (11). Due to the scientific question of interest and resulting interpretation, Xu [9] chose an intercept-only model in both the fixed and random effects. As mentioned earlier, Vonesh and Chinchilli [8] suggested a null model with only an intercept in the fixed effects and no random effects gives one reasonable choice in equation (8). The difference helps illustrates the idea that the choice of null models may differ depending on the scientific question of interest.

Zheng [10] defined estimates of the proportionate reduction in deviance for generalized linear models for longitudinal data. For linear mixed models, Zheng's measures reduce to Vonesh and Chinchilli's statistics. Also, Gelman and Pardoe [11] presented Bayesian R^2 statistics that are equivalent in form to Xu's.

In the context of maximum likelihood estimation, Kramer [3] proposed the use of two existing R^2 statistics, including a modified R^2 by Buse [12],

$$R^2_W = 1 - \frac{\sum_{i=1}^{N} (\boldsymbol{y}_i - \widehat{\boldsymbol{y}}_i)' [\widehat{\mathcal{V}}(\boldsymbol{e}_i)]^{-1} (\boldsymbol{y}_i - \widehat{\boldsymbol{y}}_i)}{\sum_{i=1}^{N} (\boldsymbol{y}_i - \bar{y}\mathbf{1}_{p_i})' [\widehat{\mathcal{V}}(\boldsymbol{e}_i)]^{-1} (\boldsymbol{y}_i - \bar{y}\mathbf{1}_{p_i})}, \tag{13}$$

where $\hat{\boldsymbol{y}}_i = \boldsymbol{X}_i\hat{\boldsymbol{\beta}} + \boldsymbol{Z}_i \hat{\boldsymbol{d}}_i$ and $\widehat{\mathcal{V}}(\boldsymbol{e}_i) = \hat{\boldsymbol{\Sigma}}_{ei}(\boldsymbol{\tau}_e)$. Choosing $\boldsymbol{V}_i = \widehat{\mathcal{V}}(\boldsymbol{e}_i)$ in the Vonesh and Chinchilli form R^2_{VC} gives the same result. Kramer [3] also recommended a measure based on a likelihood ratio test proposed by Magee [13],

$$R^2_{LR} = 1 - \exp\left\{-2n^{-1}[L(\theta;\boldsymbol{y}) - L(\theta_0;\boldsymbol{y})]\right\}, \tag{14}$$

for $L(\boldsymbol{\theta}; \boldsymbol{y})$ the log-likelihood of the model of interest and $L(\boldsymbol{\theta}_0; \boldsymbol{y})$ the log-likelihood of the null model with only an intercept in the fixed effects. If $\boldsymbol{\theta}$ and $\boldsymbol{\theta}_0$ differ only in fixed effects, then two models have the same covariance structure. The two statistics R^2_W and R^2_{LR} give the same results for ordinary regression but may differ for the linear mixed model. The individual and relative performances of the R^2 statistics require further research [3]. In contrast to Kramer [3], we focus on REML estimation. Verbeke and Molenberghs [14, p 63) illustrated how computing a difference in log-likelihoods for REML, in contrast to the same calculation for ML, does not provide a valid test statistic for fixed effects. By extension, basing R^2_{LR} on REML estimates would not be valid.

Orelien and Edwards [15] investigated the performance of selected R^2 statistics for the linear mixed model, including R^2_{VC}, R^2_X, and Zheng's [10] statistic. The authors conducted simulations to assess the ability of the R^2 statistics to discriminate between the correct model and one with important fixed-effects covariates deleted. Statistics involving the residuals were unable to discriminate adequately if the computation of the predicted values for the

NIH-PA Author Manuscript

NIH-PA Author Manuscript

residuals included the random effects (conditional R^2). However, if the random effects were excluded from the computation of the predicted values that lead to the residuals, marginal R^2 statistics were able to select the correct model. The R^2 statistic proposed by Xu [9] performed poorly due to giving little variation in value between a full model and a reduced model.

4. An R^2 Statistic for Fixed Effects in the Linear Mixed Model

4.1 Model R²

As discussed in section 1, in a linear univariate model R^2 corresponds to comparing a full model to a null model which contains only the intercept. Similarly for our R^2 statistic intended to evaluate fixed effects (mean differences), we specify a null model with only the intercept in the fixed effects. To compare nested mean models, we differ from other authors in requiring the *same* covariance structure for both the null model and the model of interest.

We believe an R^2 statistic for the linear mixed model should generalize results from the univariate model, $\boldsymbol{y} = \boldsymbol{X\beta}+\boldsymbol{e}$. In linear univariate regression the R^2 statistic for measuring the overall *linear association* of one (dependent) variable Y with several other (independent) variables $\{X_1, X_2, \ldots, X_{q-1}\}$ reflects introducing $q - 1$ predictors to an intercept-only model:

$$R_q^2=\frac{SST_c - SSE(\beta_0,\beta_1,\cdots,\beta_{q-1})}{SST_c}. \tag{15}$$

Here SST_c indicates the corrected total sum of squares, and $SSE\ (\beta_0, \beta_1, \cdots, \beta_{q-1}\}$ the error sum of squares of the model of interest [1, Chapter 11, Section 11.4]. In turn, we can express R_q^2 in terms of an F statistic for the full model as follows:

$$\begin{aligned} F_q &= \frac{[SST_c-SSE(\beta_0,\beta_1,\cdots,\beta_{q-1})]/(q-1)}{SSE(\beta_0,\beta_1,\cdots,\beta_{q-1})/\nu} \\ &= \frac{R_q^2/(q-1)}{(1-R_q^2)/\nu}, \end{aligned} \tag{16}$$

for F_q the test statistic of corrected overall regression. Here N is the number of independent sampling units, while $\nu = N - \text{rank}(\boldsymbol{X})$ gives the denominator degrees of freedom. As mentioned in the introduction, R^2 (and hence F_q) corresponds to comparing two models: 1. a full model containing $q - 1$ independent predictors and an intercept; 2. a null model that has only the intercept. Both models have the same covariance structure, $\mathcal{V}(\boldsymbol{y}) = \mathcal{V}(\boldsymbol{e}) = \sigma^2\boldsymbol{I}_N$.

We extend (16) to the linear mixed model by interpreting F_q as an approximate F statistic for a Wald test of the appropriate set of model coefficients. The most common situation involves a model including an intercept and a hypothesis excluding the intercept, giving H_0: $\boldsymbol{C\beta} = \boldsymbol{0}$ for $\boldsymbol{C} = [\boldsymbol{0}_{(q-1)\times 1}\boldsymbol{I}_{q-1}]$ of rank $q - 1$. If we let $\boldsymbol{\theta} = \boldsymbol{C\beta}$, then

$$F(\widehat{\beta},\widehat{\sum})=\frac{(\boldsymbol{C}\widehat{\beta})'\left[\boldsymbol{C}\left(\boldsymbol{X}'\widehat{\sum}^{-1}\boldsymbol{X}\right)^{-1}\boldsymbol{C}'\right]^{-1}(\boldsymbol{C}\widehat{\beta})}{\text{rank}(\boldsymbol{C})}. \tag{17}$$

Approximations for denominator degrees of freedom include the Kenward-Roger, Satterthwaithe, and residual methods [16, Chapter 46]. For any choice we have

$$F(\widehat{\beta},\widehat{\sum})=\frac{R_\beta^2/(q-1)}{(1-R_\beta^2)/\nu}, \tag{18}$$

where R_β^2 denotes our proposed R^2 statistic. Solving for R_β^2 yields

$$R_\beta^2=\frac{(q-1)\nu^{-1}F(\widehat{\beta},\widehat{\sum})}{1+(q-1)\nu^{-1}F(\widehat{\boldsymbol{\beta}},\widehat{\sum})}. \tag{19}$$

The $F(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\Sigma}})$ statistic used for R_β^2 corresponds to a test of the null hypothesis H_0: $\beta_1 = \beta_2 = \cdots = \beta_{q-1} = 0$. Using the $F(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\Sigma}})$ statistic allows computing R_β^2 using a single model fit for the model of interest, rather than needing to fit a full model and a null model.

As stated previously, different choices exist for the approximate denominator degrees of freedom (d.f.), ν, in the $F(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\Sigma}})$ statistic for fixed effects. The natural extension of the univariate linear model uses the residual degrees of freedom, with $\nu = n - \text{rank}(\boldsymbol{X})$, where n is the total number of observations. Under REML estimation, the Kenward-Roger F [19] apparently provides the most accurate inference in small samples, while the Satterthwaite method also does well [20].

Not surprisingly, analytic results in the remainder of the present section and numerical results in section 5 support using the Kenward-Roger or Satterthwaite F to define R_β^2. Just as for inference, the choice of residual degrees of freedom can substantially affect the value of R_β^2.

4.2 Choice of null model

As for the R^2 statistics reviewed in section 3, the choice of the null model plays a central role in defining R_β^2. We consider a model that includes an intercept in $\boldsymbol{X}_i$, the fixed effects design, and may or may not include an intercept in $\boldsymbol{Z}_i$ (the random effects design). The $F(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\Sigma}})$ statistic uses a single model fit to compare two models differing by the presence or absence of all fixed effect predictors except the intercept. With $\boldsymbol{X}_i = [1_i\ \boldsymbol{x}_{1,i} \cdots \boldsymbol{x}_{q-1,i}]$ and $\boldsymbol{\beta} = [\beta_0\ \beta_1 \cdots \beta_{q-1}]'$,

1. Model of Interest $\boldsymbol{y}_i = \boldsymbol{X}_i\boldsymbol{\beta} + \boldsymbol{Z}_i\boldsymbol{d}_i + \boldsymbol{e}_i$
2. Null Model $\boldsymbol{y}_i = \boldsymbol{1}_i\beta_0 + \boldsymbol{Z}_i\boldsymbol{d}_i + \boldsymbol{e}_i$.

The null model we choose has only an intercept in the fixed effects while the covariance structure, $\mathcal{V}(\boldsymbol{y}_i)$, coincides with the structure of the model of interest. Therefore the two models assume the same error covariance model $\mathcal{V}(\boldsymbol{e}_i)$, random effects design $\boldsymbol{Z}_i$, and random effects covariance $\mathcal{V}(\boldsymbol{d}_i)$. The R_β^2 statistic reflects our belief that the covariance model for $\boldsymbol{y}_i$ should remain the same when the comparison centers on fixed effects. A model of interest with $\mathcal{V}(\boldsymbol{y}_i)$ different from the null model gives a measure of association between the repeated outcomes and the fixed effects due to changing both the fixed effects *and* covariance structures. The approach applies in the context of any mixed model covariance structure. In contrast, all of the R^2 statistics discussed in section 3, with the exception of R_{VC}^2, only allow using $\boldsymbol{\Sigma}_{ei}(\boldsymbol{\tau}_e) = \sigma^2 \boldsymbol{I}_{p_i}$ for the model of interest and $\sum_{ei0}(\boldsymbol{\tau}_e)=\sigma_0^2 \boldsymbol{I}_{n_i}$ for the null model. For longitudinal data one may want to define $\boldsymbol{\Sigma}_{ei}(\boldsymbol{\tau}_e)$ to have heterogeneous variances, such as an unstructured covariance or a heterogeneous AR(1). The ability of R_β^2 to accommodate the full range of null model random effects and within-person error covariances provides a substantial advantage over all previously proposed R^2 statistics.

Although we have defined R_β^2 in terms of comparing a null model with only an intercept in the fixed effects to a larger model also containing an intercept, R_β^2 generalizes naturally in two other settings. Some models, such as those based on cell-mean coding, do not include an intercept but still spans an intercept in the sense that a q×1 vector $\boldsymbol{t}$ exists such that $\boldsymbol{Xt} = \boldsymbol{1}_q$. Computing R_β^2 (with an appropriate choice of contrast matrix, $\boldsymbol{C}$, which depends on the specific coding) for such models implicitly uses the intercept only model as the null model, and therefore requires no additional work. In other cases the model neither includes nor spans an intercept. Models that do not span (and therefore also do not include) an intercept would use $\boldsymbol{y}_i = \boldsymbol{Z}_i\boldsymbol{d}_i + \boldsymbol{e}_i$ as the null model and hence $\boldsymbol{C} = \boldsymbol{I}_q$. Again $\mathcal{V}(\boldsymbol{e}_i)$, $\boldsymbol{Z}_i$, and $\mathcal{V}(\boldsymbol{d}_i)$ coincide with the structures of the full model, which insures $\mathcal{V}(\boldsymbol{y}_i)$ coincides. Muller and Fetterman [1, Chapters 4–6] provided detailed introductions to the role of the intercept in model definition, interpretation, testing, and correlation in the special case context of the univariate linear model.

4.3 Interpretation of R^2 and what it estimates; special case connections to multivariate models

What does R_β^2 estimate and how do we to interpret it? Naturally R_β^2 reduces to the standard R^2 statistic for the special case of a linear univariate regression model since its definition generalizes the univariate definition. The linear univariate model R^2 measures the amount of overall linear association of one response variable with $q-1$ predictor variables. The description corresponds to interpreting $R=\sqrt{R^2}$ in terms of the geometry of the q dimensional scattergram for the data standardized to means of zero and standard deviations of one, while accounting for correlations among predictors. In turn, the linear mixed model uses a univariate approach to a multivariate linear model. The Wald statistic defining R_β^2 standardizes the response data by using (approximate) weighted least squares based on the estimated covariance among the repeated measures, while also accounting for correlations among predictors. By extension then, R_β^2 measures multivariate association between the response variable (repeated measures of a single outcome) and the predictor variables (fixed effects) in the linear mixed model.

Hence to understand what R_β^2 estimates we must also turn to the multivariate linear model,

$$\boldsymbol{Y}=\boldsymbol{X}\boldsymbol{B}+\boldsymbol{E}. \tag{20}$$

Rows of $\boldsymbol{Y}$ (N×p), $\boldsymbol{X}$ (N×q), and $\boldsymbol{E}$ correspond to N persons (independent sampling units), columns of $\boldsymbol{Y}$, $\boldsymbol{B}$ (q×p) and $\boldsymbol{E}$ correspond to time, while columns of design matrix $\boldsymbol{X}$ and rows of fixed, unknown $\boldsymbol{B}$ correspond to predictors. We assume $N > \text{rank}(\boldsymbol{X})$ independent rows of $\boldsymbol{E}$, with $\text{row}_i(\boldsymbol{E})' \sim \mathcal{N}_p(\boldsymbol{0}, \boldsymbol{\Sigma})$. The general linear hypothesis, H_0: $\boldsymbol{\theta} = \boldsymbol{\theta}_0$, tests $\boldsymbol{\theta} = \boldsymbol{CBU}$ (a×b), for fixed and known $\boldsymbol{C}$, $\boldsymbol{U}$, and $\boldsymbol{\theta}_0$. Here $\boldsymbol{C}$ (a×q) gives contrasts between person and $\boldsymbol{U}$ (p×b) gives contrasts within person. Muller and Stewart [5] provided details, including explicit representations of any such model as a general linear mixed model. The following discussion of measures of multivariate association draws from the same source.

In contrast to the univariate setting, no single multivariate test statistic satisfies all of the standard optimality criteria in the most complex conditions ($s = \min(a, b) > 1$). The three principles of union-intersection, likelihood, and substitution give four distinct and commonly used multivariate test statistics. Each distinct multivariate test statistic leads to a distinct measure of association. Measures of *multivariate association* generalize the concept of a squared multiple correlation corresponding to the hypothesis H_0: $\boldsymbol{\theta} = \boldsymbol{\theta}_0$.

One particular measure of multivariate association has special relevance for R^2_β. The Hotelling-Lawley-Trace (HLT) test statistic, sometimes called the ANOVA analog statistic, corresponds to a measure of multivariate association given by

$$\widehat{\eta} = \frac{\mathrm{HLT}/s}{1+\mathrm{HLT}/s}, \tag{21}$$

where $s = \min(a, b)$. Also $0 \le \hat{\eta} \le 1$, with 0 reflecting no association and 1 perfect association. For the multivariate general linear hypothesis H_0: $\boldsymbol{\theta} = \boldsymbol{\theta}_0$

$$\begin{aligned} \mathrm{HLT} &= \mathrm{trace}\left\{\left(\widehat{\boldsymbol{\Theta}} - \boldsymbol{\Theta}_0\right)' \boldsymbol{M}^{-1} \left(\widehat{\boldsymbol{\Theta}} - \boldsymbol{\Theta}_0\right) \left[(N - r)\boldsymbol{U}' \widehat{\Sigma} \boldsymbol{U}\right]^{-1}\right\} \\ &= \sum_{k=1}^{s} \frac{\widehat{\rho}_k^2}{(1-\widehat{\rho}_k^2)}, \end{aligned} \tag{22}$$

with $\left\{\widehat{\rho}_k^2/(1 - \widehat{\rho}_k^2)\right\}$ the eigenvalues of $(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta}_0)'\boldsymbol{M}^{-1}(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta}_0)[(N-r)\boldsymbol{U}'\hat{\boldsymbol{\Sigma}}\boldsymbol{U}]^{-1}$ while $\boldsymbol{M} = \boldsymbol{C}(\boldsymbol{X}'\boldsymbol{X})^{-}\boldsymbol{C}'$. Here $\{\widehat{\rho}_k^2\}$ estimate the (generalized) squared canonical correlations, ρ_k^2. Like the Wald statistic for the general linear mixed model, HLT was defined by the substitution principle. The distribution of the HLT statistic (times a constant) is approximated by that of an F with denominator degrees of freedom given by

$$\frac{[(N-r)^2 - (N-r)(2b+3)+b(b+3)](ab+2)}{(N-r)(a+b+1) - (a+2b+b^2-1)}+4. \tag{23}$$

Special cases give Hotelling T^2 statistics and correspond to exact noncentral F distributions, while general cases allow accurate approximations by noncentral F distributions.

Multivariate and mixed linear model theory overlap in many ways. Vonesh and Chinchilli [8, Chapter 6, p 236] showed that for complete and balanced data, the random coefficient growth curve model, a particular form of the linear mixed model, can be written in terms of the generalized multivariate analysis of variance (GMANOVA) model. In practice, GMANOVA can be implemented with careful use of the theory associated with the multivariate model in (20), [5, Chapters 4, 13, 17, 22].

In defining R^2_β for fixed effects in the linear mixed model, we see that the Wald $F(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\Sigma}})$ approximation plays the role of the F approximation for the HLT statistic given by (20). Both assess the Mahalanobis distance of the parameter vector from the alternative location

NIH-PA Author Manuscript

NIH-PA Author Manuscript

to the null. The fact that $F(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\Sigma}}) \geq 0$ insures $0 \leq R^2_\beta \leq 1$, with $R^2_\beta=0$ indicating no multivariate association between $\boldsymbol{y}$ and $\boldsymbol{X}$. On the other hand, as R^2_β nears 1, then the multivariate association between $\boldsymbol{y}$ and $\boldsymbol{X}$ becomes perfect. Hence R^2_β estimates the association between $\boldsymbol{y}$ and $\boldsymbol{X}$ controlled by the linear mixed model hypothesis H_0: $\boldsymbol{\theta} = \boldsymbol{0}$, i.e., H_0: $\beta_1 = \beta_2 = \cdots = \beta_{q-1} = 0$.

The description just given leads to a very appealing result in the special case with a multivariate model that is expressed as linear mixed model and a linear mixed model hypothesis that coincides exactly with a multivariate linear hypothesis. In Lemma 1 in the Appendix we prove that in such special cases the R^2_β statistic reduces exactly to $\hat{\eta}$ corresponding to HLT (as in equation 21). If the models and hypotheses do not coincide, then R^2_β and $\hat{\eta}$ need not coincide.

4.4 Partial R^2

Testing H_0: $\beta_j = 0$ for $j \in \{1, \cdots, q-1\}$, i.e., testing a particular fixed effect regression coefficient, gives a partial F statistic. It measures the marginal contribution of X_k when all the other predictors have already been included in the model. Just as for overall regression, the partial F in the linear mixed model defines a partial R^2_β by way of equation (19). It should be noted that none of the R^2 statistics discussed in section 3 lead to defining a partial R^2 statistic.

More generally, for fixed effects, *any* test of a single variable, a group of variables, or a general linear contrast, corresponds to an F statistic and hence provides a corresponding R^2_β. The concept applies to both added-in-order and added-last tests. The validity of the claim for the general linear contrast rests on the proof of Lemma 2 in the Appendix.

Muller and Fetterman [1, Chapters 5–6] provided detailed introductions to the variety of partial correlations and associated tests in the special case context of the univariate linear model. The special case results and the nature of the tests involved make it easy to deduce that the partial R^2_β described here can more precisely be called a semi-partial, as distinct from a full partial. As in the univariate model, the process directly gives semi-partial correlations which assess the importance of the parameters tested. A semi-partial correlation assesses the relationship of a subset of predictors with the response for the subset of predictors (but not the response) adjusted for other predictors in the model. In contrast, a full partial correlation assesses the relationship of a subset of predictors with the response for the subset of predictors and the response adjusted for other predictors in the model. As for univariate models, computing full partial correlations for the linear mixed model would require more than one step.

4.5 Analytic Results Recommend Kenward-Roger F and Degrees of Freedom in Defining R^2_β

As results in section 4 illustrate in many ways, the approximate denominator degrees of freedom play an important role in the interpretation of the Wald $F(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\Sigma}})$ statistic for fixed effects. Although it may seem natural to use the residual degrees of freedom, at the present time we recommend the Kenward-Roger approach or perhaps the Satterthwaite approach. Their superior accuracy in inference [20] leads to accurately mapping R^2_β into the multivariate measure of association when the latter exists. Using Kenward-Roger best maintains the highly desirable correspondence between inference about regression coefficients, F statistics, and correlations seen in univariate and multivariate models.

5. Example Computations and Interpretations

5.1 Example 1, Dental Data Comparison of Mixed Models with R_β^2

We use a well-known example from Potthoff and Roy [21] to demonstrate results for R_β^2 proposed in section 4. The data come from an orthodontic study with 27 children, 16 boys and 11 girls. For each child, the distance (mm) from the center of the pituitary to the pterygomaxillary fissure was measured at ages 8, 10, 12, and 14 years. The objectives were to determine whether, on the average over time, distances are larger for boys than for girls and whether, on the average over time, the rate of change of the distance is similar for boys and girls. The interested reader may refer to Zheng [10] and Orelien and Edwards [15] for the performance of R_{VC}^2, R_x^2 with the dental data.

For the example, we fitted linear mixed models with three different fixed-effect structures (in addition to a fixed intercept) using REML estimation: (I) a model with continuous age effect only; (II) a model with linear Age and Gender effect; (III) a model with linear Age, Gender, and their interaction. The null model for expected values contains only an intercept, while the covariance structure remains the same as for the model of interest. We considered three different covariance structures: (1) random intercepts, $\sum_{di}(\boldsymbol{\tau}_d)=\sigma_d^2$ (scalar), and $\boldsymbol{\Sigma}_{ei}(\boldsymbol{\tau}_e) = \sigma^2\boldsymbol{I}_{n_i}$; (2) random intercepts and slopes with unstructured covariance, $\boldsymbol{\Sigma}_{di}(\boldsymbol{\tau}_d)$ (2×2), and $\boldsymbol{\Sigma}_{ei}(\boldsymbol{\tau}_e) = \sigma^2\boldsymbol{I}_{n_i}$; (3) random intercepts and slopes with unstructured covariance, $\boldsymbol{\Sigma}_{di}(\boldsymbol{\tau}_d)$ (2×2), and $\sum_{ei}(\boldsymbol{\tau}_e)=\sigma_1^2\boldsymbol{I}_{n_i}$ for boys and $\sum_{ei}(\boldsymbol{\tau}_e)=\sigma_2^2\boldsymbol{I}_{n_i}$ for girls.

Table 1 provides the estimates, standard errors, and p-values for fixed effects in Models I, II, and III described above, as well as covariance parameter estimates. For covariance structures 2 and 3, the estimated variances of the random intercepts and slopes are $\widehat{\sigma}_{d1}^2$ and $\widehat{\sigma}_{d2}^2$, respectively, and the correlation is given by $\hat{\rho}$. REML estimates were used, with fixed effect standard errors computed via a method in Kackar and Harville [17] and Prasad and Rao [18]. In turn p-values used the Kenward-Roger F and associated denominator degrees of freedom.

Recall that different choices exist for the approximate denominator degrees of freedom, $\boldsymbol{\nu}$, for the $F(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\Sigma}})$ statistic for fixed effects. Table 2 provides R_β^2 as well as F statistics and denominator degrees of freedom using the residual, Kenward-Roger, Contain, and Satterthwaite approximations [16, Chapter 46]. In Table 2 the Kenward-Roger results vary from R_β^2=0.59 to 0.84. In comparison, the Residual F and denominator degrees of freedom values range from R_β^2=0.45 to 0.56, which is in most cases much lower than R_β^2 using Kenward-Roger. The denominator degrees of freedom for the Residual F treats the observations as if they are independent and therefore the degrees of freedom are much larger, which deflates R_β^2.

Results in Table 2 demonstrate that R_β^2 using Kenward-Roger (or Contain or Satterhwaite) can decrease when adding a fixed effect. For linear mixed models, Snijders and Bosker (1994) argued that most fixed effect predictors have both within-group and between-group variability. As such, adding a predictor may increase estimated residual variance and hence cause the estimated R^2 for the linear mixed model to decrease. Snijders and Bosker (1994) also stated that “We would like to stress that the possibility, discussed above, of an increase of residual variance estimates when predictor variables are added is not a consequence of misspecification. The model without and the model with the extra predictor variable could both be valid statistical models for the observations at hand (although the latter model would be better in the sense of having a greater explanatory power).” For example, Table 1 shows that adding a fixed effect for any of the covariance structures impacts the estimate of the random effects variance. The most dramatic affect can be seen for covariance model 2 where adding gender inflates the variance of the random intercept from 5.4 to nearly 8 and as a result increases the variance of the response. It also affects the correlation between random intercept and slope, increasing in absolute value from 0.61 to 0.77. To a lesser extent this is also true for covariance model 3. In turn, this affects the value of the R_β^2 measures using the Kenward-Roger, Satterwaithe, and Contain methods which compute the denominator degrees of freedom using estimates of parameters. In the linear univariate and multivariate models, adding a fixed effect can either explain more of the variance or add no additional explaination. Hence, for the linear univariate and mulitvariate models, the true population R^2 and estimated R^2 increase or remain the same. However, unlike the linear univariate and multivariate models, adding a fixed effect in the linear mixed model may actually increase the overall variance estimate as illustrated above. In such cases, R_β^2 is interpreted as indicating a decrease in measure of association possibly due to either misspecification of the “full” model and/or of sampling variation resulting in changes to estimates of variance components.

5.2 Example 1, Dental Data Comparison of Mixed and Multivariate Models with R_β^2

The nature of the dental data allows defining a multivariate model which corresponds to a mixed model with an unstructured covariance matrix. Using the approach referenced in the proof of Lemma 1 in the appendix, a growth curve analysis can be formulated in terms of the GMANOVA model [21] as

$$\boldsymbol{Y}=\boldsymbol{X}_{\mathrm{M}}\boldsymbol{B}\boldsymbol{T}+\boldsymbol{E} \tag{24}$$

$$\begin{bmatrix} y_{1,1} & \cdots & y_{1,4} \\ \vdots & & \vdots \\ y_{27,1} & \cdots & y_{27,4} \end{bmatrix} = \begin{bmatrix} \mathbf{1}_{16} & \mathbf{0}_{16} \\ \mathbf{0}_{11} & \mathbf{1}_{11} \end{bmatrix} \begin{bmatrix} \beta_{b0} & \beta_{b1} & \beta_{b1} & \beta_{b3} \\ \beta_{g0} & \beta_{g1} & \beta_{g2} & \beta_{g3} \end{bmatrix} \begin{bmatrix} 0.50 & 0.50 & 0.50 & 0.50 \\ -0.67 & -0.22 & 0.22 & 0.67 \\ 0.50 & -0.50 & -0.50 & 0.50 \\ -0.22 & 0.67 & -0.67 & 0.22 \end{bmatrix} + \boldsymbol{E},$$

where $\boldsymbol{Y}$ and $\boldsymbol{E}$ are 27×4. The between subject cell-mean design matrix $\boldsymbol{X}_{\mathrm{M}}$ (27×2) represents the two groups, boys and girls. The matrix $\boldsymbol{B}$ (2×4) contains unknown parameters for intercepts, linear, quadratic, and cubic orthogonal polynomials for both boys and girls. The within-subject design matrix $\boldsymbol{T}$ (4×4) consists of zero order, linear, quadratic, and cubic orthonormal polynomials generated from the natural cubic polynomial matrix $\boldsymbol{P}$ (4×4) with row k + 1 being $\boldsymbol{p}_k = [8^k\ 10^k\ 12^k\ 14^k]$. Using a square $\boldsymbol{T}$ allows converting the GMANOVA model into a MANOVA model by multiplying both sides of (24) by $\boldsymbol{T}^{-1} = \boldsymbol{T}'$. If $\boldsymbol{Y}_s = \boldsymbol{YT}'$ and $\boldsymbol{E}_s = \boldsymbol{ET}'$, the transformation gives the equivalent MANOVA model as

$$\boldsymbol{Y}_s=\boldsymbol{X}_{\mathrm{M}}\boldsymbol{B}+\boldsymbol{E}_s. \tag{25}$$

Using equation 12.6 in Muller and Stewart [5, p 245] as in the proof of Lemma 1 gives explicit expressions for the general linear multivariate model stacked by person into a corresponding mixed model with no random effects and unstructured covariance within independent sampling unit. Mixed model authors often refer to the stacked model as a population-average model:

$$\boldsymbol{y}_s=\boldsymbol{X}_s\beta+\boldsymbol{e}_s. \tag{26}$$

Here $\boldsymbol{y}_s=\text{vec}(\boldsymbol{Y}_s^{\prime})$ (108×1), $\boldsymbol{X}_s = \boldsymbol{X}_\text{M}\otimes\boldsymbol{I}_4$ (108×8), $\boldsymbol{\beta} = [\beta_{b0}\ \beta_{b1}\ \beta_{b2}\ \beta_{b3}\ \beta_{g0}\ \beta_{g1}\ \beta_{g2}\ \beta_{g3}]'$ (8×1), and $\boldsymbol{e}_s = \text{vec}(\boldsymbol{E}'_s)$ (108×1). For comparing $\hat{\eta}$ using HLT as in equation 21 and R_β^2 using Kenward-Roger, we consider tests of gender (in the intercepts), age (joint linear, quadratic, cubic effects), and gender×age (linear, quadratic, cubic). LINMOD 3.3 (freely available at web site http://ehpr.ufl.edu/muller) was used to compute $\hat{\eta}$.

Table 3 gives results for corresponding multivariate and mixed models and hypotheses, contrast matrices, and measures of association for the dental data. The multivariate model in equation (25) has zero, linear, quadratic, and cubic orthonormal polynomial trend scores. The multivariate and mixed models assume unstructured covariance. Both $\hat{\eta}$ and R_β^2 yield values of 0.27, 0.97, and 0.07, respectively, for the three hypotheses listed. We also computed R_β^2 defined by the Satterwaithe, Contain, and residual methods (not shown in table). The Satterwaithe results were identical to the Kenward-Roger. The Contain and residual methods were identical and yielded values of 0.09, 0.88, and 0.02, respectively, for the three hypotheses listed.

5.3 Numerical Results Recommend Kenward-Roger Degrees of Freedom in Defining R_β^2

The theoretical properties of R_β^2 and related distribution theory in sections 4.3–4.5 agree with the numerical comparisons of mixed models in section 4.1, and also with the numerical comparisons of mixed and corresponding multivariate models in section 4.2. We conclude that the linear mixed model statistic R_β^2 for fixed effects should be defined using the Kenward-Roger F and the associated denominator degrees of freedom. Doing so insures coincidence with an appropriate corresponding measure of multivariate association, when possible, based on the Hotelling-Lawley-Trace statistic. We find it felicitous and not accidental that both the Wald mixed model statistic and the Hotelling-Lawley-Trace were originally defined by applying the “substitution principle” to the error covariance matrix in the respective models. If the models and hypotheses do not coincide, then R_β^2 and $\hat{\eta}$ need not coincide.

5.4 Example 1, Dental Data Partial R_β^2

Table 4 presents partial R_β^2 values for model II. For model II, the partial R_β^2 measures the partial multivariate association between the repeated outcomes and Gender or Age after controlling for the effect of the other predictor. The results indicate that the partial multivariate association between the response and Age is much larger than response and Gender. Comparing results between two such variable added-last tests in the same model involve nonnested models. In contrast, Table 2 illustrates comparisons among a sequence of nested models, which correspond to added-in-order tests and allow simple and appealing comparisons and interpretations among R_β^2 values. As noted earlier, the potential exchangeability of expected value and covariance parameters in imperfect mixed models may give mild disordinality in R_β^2 values for nested model sequences.

5.5 Example 2, Blood Pressure and Race Strength of Relationship

Statisticians have long criticized reporting p-values without a corresponding measure of importance such as a confidence interval for a mean or a correlation. The lack of a credible measure of importance for mixed models has led scientists to rely too much on p-values for interpreting a predictor of particular interest, such as race or treatment. Our concern arises from the fact that the predictor could be statistically significant yet explain very little of the variation in the outcome. Data from a retrospective longitudinal cohort study of 459 adults with hypertension [22] illustrate the problem and the utility of our statistic. Longitudinal blood pressure (BP) level were taken on patients making at least four visits to the Family Practice Center at UNC during a two year period, 1999–2001. Predictor variables in the linear mixed model include indicators for Continuity of Care, and Race, Gender, Insurance status, Provider type, Marital status, as well as continuous linear Age at first measurement and linear Time. The random effects include intercept and linear Time, with unstructured covariance among them, and a within-person error covariance of $\boldsymbol{\Sigma_{ei}}(\boldsymbol{\tau_e}) = \sigma^2 \boldsymbol{I}_{n_i}$ for person i. Both systolic and diastolic BPfell over the two years (systolic 2.2 mmHg/yr and diastolic 2.8 mmHg/yr). Many studies have confirmed that blacks, on average, have higher BP than whites. In this study higher BP was associated with blacks versus whites with p-value = 0.0042 for systolic BP and 0.0053 for diastolic BP. The race effect remained the same across time (there was no Race × Time interaction).

The overall model R_β^2 statistic was computed for both systolic BP ($R_\beta^2=0.09$) and diastolic BP ($R_\beta^2=0.27$; predictors were Continuity of Care, Gender, Insurance status, Provider type, Marital status, linear Age, linear Time, Race). The partial R_β^2 for Race when modeling either systolic or diastolic BP was 0.02 (corresponding to the added-last test for Race). Thus the combined predictors have only a very small association with repeated measures of systolic and diastolic BP in this study. Furthermore, despite being statistically significant, Race (blacks versus whites) has a nearly negligible association with the responses, after controlling for the covariates available. Other factors influence BP that were not available, such as diet, weight, sleep status, alcohol and coffee consumption, smoking, family history, emotional factors, medication, socioeconomic status, and seasonal factors. The significance of racial health disparities in BP have been emphasized in many reports [23]. However, for this longitudinal study the R_β^2 statistic suggests that race has a very weak association with BP when other appropriate status variables have been accounted for. As many statisticians have said before us, we believe R_β^2 and corresponding measures of importance should always accompany p-value measures of “significance”. Reviewing the health disparities in BP findings with the purpose of assessing strength of association would seem meritorious.

5.6 The Possible Role of Changing Covariance Matrices

Although comparing fixed effects across increasing complexity of covariance models seems appealing, any such comparison of association only makes sense in the uninteresting case in which the simplest covariance model actually holds true and all models are well estimated. Otherwise differences in association can occur due to the fixed effects specified by the underlying hypothesis or by some unrecognized combination of fixed and random effects, i.e., some unrecognized combination of model differences in means and covariances. We leave discussion of possible analogs of R_β^2 allowing different covariance matrices, which corresponds to evaluating random effects, to future research.

6. Discussion and Conclusions

6.1 Positive Results

The value and familiarity of the R^2 statistic in the linear univariate model naturally generates great interest in extending it to the linear mixed model. However, the development of R^2 statistics for the linear mixed model has received comparatively little attention. We were motivated to develop a statistic by the substantial limitations of existing R^2 statistics.

We proposed a new R^2 statistic, R_β^2, for assessing fixed effects in the linear mixed model. We interpret R_β^2 as a measure of multivariate association between the response variable and the fixed effects in the linear mixed model. The analytic properties of R_β^2 overcome the limitations of previous proposals, while longitudinal data examples demonstrated the impact the proposed R^2 statistic can have on future practice. The same statistic generalizes to define a partial R^2 statistic for marginal (fixed) effects of all sorts. None of the other R^2 statistics reviewed appear to have the same important property. Given the sound principles underlying R_β^2, the ease of computation due to using a single model, and generalization to a partial R^2statistic, we believe that R_β^2 should be used to measure association for fixed effects in the linear mixed model.

6.2 Limitations and Opportunities

We have concentrated on the fixed effects or “marginal” portion of the linear mixed model and related residuals, while not considering the deviations within individuals, which correspond to a different type of residual. Basing R_β^2 on a marginal statistics means it cannot be used to determine person-specific goodness-of-fit. We believe a different and purpose-specific measure must be developed to answer the very different questions about variation within persons.

The coincidence of R_β^2 with a measure of multivariate association for multivariate model hypotheses has interesting implications. The HLT provides 1 of 4 widely used test statistics with corresponding measures of association for multivariate models. The variety of tests reflect the lack of a uniformly most powerful multivariate test (among similarly invariant and size α tests). The special case relationship to mixed models immediately implies the same lack of a single optimal test for mixed models. Hence other combinations of test and measure of association likely merit consideration as competitors to R_β^2, a proposition we leave to future consideration. For example, a reviewer postulated that it can be argued that all of the R^2 measures based on any of the denominator degrees of freedom converge to well-defined population values defined by each test’s non-centrality parameter as $N \to \infty$. We note the following: 1) The choice of degrees of freedom clearly affects the rate of convergence. 2) We think, but are not sure, that the choice of degrees of freedom may change the parameter being estimated. The latter is a key question for subsequent work on asymptotics. However, the rates of convergence appear to be so disparate as to make the KR choice (due to its second order convergence, in contrast to first order for others) a clear winner, in our view.

As in the special case of the univariate model, R_β^2 was developed for comparing two models with nested fixed effects. Comparing models with non-nested fixed effects disallows applying the distribution theory of the F approximations. As in the univariate case, the scientific value and interpretation of comparing non-nested models rests with the user in a specific application. Readers familiar with the R^2 statistic for the linear univariate and multivariate models may at first be skeptical of a feature of R_β^2 that allows the measure to decrease when adding predictors. In the linear univariate and multivariate models, adding a fixed effect results in an increase (or no change) in the amount of variance explained by the predictors and hence the monotonic property of both the sample R^2 and true population R^2. However, unlike the linear univariate and multivariate models, in the linear mixed model adding a predictor in the fixed effects (between-subject effect) can increase the estimated variance of the random effects (within-subject effect) and hence increase the estimated variance of the response. In such cases, R_β^2 is interpreted as indicating a decrease in measure of association possibly due to either misspecification of the "full" model and/or of sampling variation resulting in changes to the variance components estimates. The true population R^2 that R_β^2 estimates, under suitable conditions, should not decrease when a predictor is added (Snijders and Bosker [2]).

We caution the reader that a long line of research supports the conclusion that R^2 measures have limited use in model building [1, Chapter 11, Section 11.1]. In building a univariate linear model, R^2 measures serve only as an adjunct to a suite of model diagnostic and selection tools (as outlined in [1, Chapter 11]). Unfortunately model diagnostic and selection tools have not been as well developed for the linear mixed model. Also, for a variety of reasons, the methods that do exist have not become widely known. Even worse in terms of practical effect, most popular software does not provide easy access. Hence the development, dissemination, and provision of easily accessible software for model diagnostics and model selection seems of the highest priority for linear mixed models.

Acknowledgments

This research was supported in part by the following grants: Keith Muller by NCI P01 CA47 982-04, NCI RO-1 CA095749-01A1 and NIAID 9P30 AI 50410.

References

1. Muller, KE.; Fetterman, BA. Regression and ANOVA: An Integrated Approach Using SAS®Software. SAS Institute: Cary, NC; 2002.
2. Snijders TAB, Bosker RJ. Modeled variance in two-level models. Sociological Methods and Research. 1994; 22:342–363.
3. Kramer, M. R^2 Statistics for Mixed Models. Presented at the 17th Annual Kansas State University Conference on Applied Statistics in Agriculture; 2005; April 24–26;
4. Laird NM, Ware JH. Random-effects models for longitudinal data. Biometrics. 1982; 38:963–974. [PubMed: 7168798]
5. Muller, KE.; Stewart, PW. Linear Model Theory: Univariate, Multivariate, and Mixed Models. Wiley; New York: 2006.
6. Kvålseth TO. Cautionary note about R^2. The American Statistician. 1985; 39:279–285.
7. Cameron C, Windmeijer FAG. R-squared measures of count data regression models with applications to health-care utilization. Journal of Business and Econonomic Statistics. 1996; 14:209–220.
8. Vonesh, EF.; Chinchilli, VM. Linear and Nonlinear Models for the Analysis of Repeated Measurements. Marcel Dekker; New York: 1997.
9. Xu RH. Measuring explained variation in linear mixed effects models. Statistics in Medicine. 2003; 22:3527–3541. [PubMed: 14601017]
10. Zheng BY. Summarizing the goodness of fit of generalized linear models for longitudinal data. Statistics in Medicine. 2000; 19:1265–1275. [PubMed: 10814976]

11. Gelman A, Pardoe I. Bayesian measures of explained variance and pooling in multilevel (hierarchical) models. Technometrics. 2006; 48:241–251.
12. Buse A. Goodness of fit in generalized least squares estimation. The American Statistician. 1973; 27:106–108.
13. Magee L. R^2 measures based on Wald and likelihood ratio joint significance tests. The American Statistician. 1990; 44:250–253.
14. Verbeke, G.; Molenberghs, G. Linear Mixed Models For Longitudinal Data. Springer-Verlag; New York: 2000.
15. Orelien JG, Edwards LJ. Fixed effect variable selection in linear mixed models using R^2 statistics. Computational Statistics and Data Analysis. 2008; 52:1896–1907.
16. SAS Institute Inc. SASeSTAT® 9.1 User's Guide, Chapter 46. SAS Institute; Cary, NC: 2004.
17. Kackar RN, Harville DA. Approximations for standard errors of estimators of fixed and random effects in mixed linear models. Journal of the Amercian Statistical Association. 1984; 79:853–862.
18. Prasad NGN, Rao JNK. The estimation of mean squared error of small-area estimators. Journal of the Amercian Statistical Association. 1990; 85:163–171.
19. Kenward MG, Roger JH. Small sample inference for fixed effects from restricted maximum likelihood. Biometrics. 1997; 53:983–997. [PubMed: 9333350]
20. Muller KE, Edwards LJ, Simpson SL, Taylor DJ. Statistical tests with accurate size and power for balanced linear mixed models. Statistics in Medicine. 2007; 26:3639–3660. [PubMed: 17394132]
21. Potthoff RF, Roy SN. A generalized multivariate analysis of variance model useful especially for growth curve problems. Biometrika. 1964; 51:313–326.
22. Fisher M, Sloane P, Edwards L, Gamble G. Continuity of care and hypertension control in a university-based practice. Ethnicity and Disease. 2007; 17:693–698. [PubMed: 18072381]
23. Cooper R, Rotimi C. Hypertension in blacks. American Journal Hypertension. 1997; 10:804–812.

Appendix: Some Analytic Properties of Mixed Models

Muller and Stewart [5, Section 12.1] describe how a multivariate model can be stated as a special case of a mixed model. In the mixed model formulation, the defining characteristics of the special case are "Kronecker design" and "Kronecker covariance."

Lemma 1

If a linear mixed model corresponds to a multivariate general linear model, and a hypothesis of interest corresponds to a multivariate general linear hypothesis, then the mixed model Wald statistic reduces to the Hotelling-Lawley multivariate statistic, except perhaps for a scaling constant.

Proof

Equation 12.6 in Muller and Stewart [5, p 245] gives explicit expressions for the general linear multivariate model $\boldsymbol{Y} = \boldsymbol{X}_{\mathrm{M}}\boldsymbol{B} + \boldsymbol{E}$ stacked by person into a corresponding mixed model with no random effects and unstructured covariance within independent sampling unit (often person). For the sake of brevity, we detail notation only when it varies from that in Muller and Stewart and omit most intermediate steps involving direct product ($\mathbf{A} \otimes \boldsymbol{L} = \{a_{ij}\boldsymbol{L}\}$) and vec (stack by columns) operators. Here $\boldsymbol{X}_s = \boldsymbol{X}_{\mathrm{M}} \otimes \boldsymbol{I}_p$, $\boldsymbol{\Sigma}_i \equiv \boldsymbol{\Sigma}_{i'}$, and $\boldsymbol{\Sigma}_{\mathbf{s}} = \boldsymbol{I}_p \otimes \boldsymbol{\Sigma}_{\mathrm{i}}$.

Multivariate hypothesis H_0: $\boldsymbol{C}_{\mathrm{M}}\boldsymbol{BU} = \theta_0$ gives $\boldsymbol{M}_{\mathrm{M}}=\boldsymbol{C}_{\mathrm{M}}(\boldsymbol{X}_{\mathrm{M}}'\boldsymbol{X}_{\mathrm{M}})^{-1}\boldsymbol{C}_{\mathrm{M}}'$. Without loss of generality, we assume $\theta_0 = \mathbf{0}$ and a contrast that retains all times, which implies $\boldsymbol{U} = \boldsymbol{I}_p$ and $\theta_{\mathrm{M}} = \boldsymbol{C}_{\mathrm{M}}\boldsymbol{B}\boldsymbol{I}_p = [\boldsymbol{C}_{\mathrm{M}}\boldsymbol{\beta}_1\ \boldsymbol{C}_{\mathrm{M}}\boldsymbol{\beta}_2 \cdots \boldsymbol{C}_{\mathrm{M}}\boldsymbol{\beta}_p]$.

The mixed model uses $\boldsymbol{C}_s=\boldsymbol{C}_{\mathrm{M}} \otimes \boldsymbol{I}_p, \theta=\mathrm{vec}(\boldsymbol{\Theta}_{\mathrm{M}}')$, and

$$\mathcal{V}(\widehat{\theta})=\boldsymbol{C}_s(\boldsymbol{X}_s'\textstyle\sum_s^{-1}\boldsymbol{X}_s)^{-1}\boldsymbol{C}_s'=\boldsymbol{C}_s[(\boldsymbol{X}_{\mathrm{M}} \otimes \boldsymbol{I}_p)'(\boldsymbol{I}_N \otimes \textstyle\sum_i)^{-1}(\boldsymbol{X}_{\mathrm{M}} \otimes \boldsymbol{I}_p)]^{-1}\boldsymbol{C}_s'=\boldsymbol{M}_{\mathrm{M}}\otimes\textstyle\sum_i.$$

Except for a scalar constant, the mixed model form may be written

$$\begin{aligned}\theta'[C_s(X_s'\textstyle\sum_s^{-1}X_s)^{-1}C_s']^{-1}\theta\\ &=\theta'(M_{\mathrm{M}}\otimes\textstyle\sum_i)^{-1}\theta\\ &=[\Theta_{\mathrm{M1}}\ \Theta_{\mathrm{M2}}\ \cdots\ \Theta_{\mathrm{M}a}](M_{\mathrm{M}}^{-1}\otimes\textstyle\sum_i^{-1})[\Theta_{\mathrm{M1}}\ \Theta_{\mathrm{M2}}\ \cdots\ \Theta_{\mathrm{M}a}]'\\ &=\mathrm{vec}(\Theta_{\mathrm{M}}')'\mathrm{vec}(\textstyle\sum_i^{-1}\Theta_{\mathrm{M}}'M_{\mathrm{M}}^{-1})\\ &=\mathrm{tr}(\Theta_{\mathrm{M}}\textstyle\sum_i^{-1}\Theta_{\mathrm{M}}'M_{\mathrm{M}}^{-1})\\ &=\textstyle\sum_{k=1}^{s}\rho_k^2\\ &/(1-\rho_k^2)=s\cdot\eta\end{aligned}$$

Lemma 2

For any testable general linear hypothesis for a general linear mixed model, a linearly equivalent model may be found in which $\boldsymbol{C} = [\mathbf{I}_a\ \mathbf{0}]$ provides the original test.

Proof

The proof generalizes Theorem 16.16 in Muller and Stewart [5]. For the sake of brevity, many intermediate results have been omitted, as have details of mixed model notation in Muller and Stewart [5, Chapters 5, 14, 18). Considering all observations, the model $\boldsymbol{y}_s = \boldsymbol{X}_s\boldsymbol{\beta} + \boldsymbol{Z}_s\boldsymbol{d}_s + \boldsymbol{e}_s$ has testable secondary parameter $\theta = \boldsymbol{C\beta}$ for H_0: $\boldsymbol{\theta} = \mathbf{0}$ and $a \times q$ $\boldsymbol{C}$ of rank a. Singular value decomposition gives $\boldsymbol{C} = \boldsymbol{L}[\mathrm{Dg}(\boldsymbol{s}_C)\mathbf{0}_{a\times(q-a)}][\boldsymbol{R}_1\ \boldsymbol{R}_0]'$ with $\boldsymbol{L}$ and $\mathrm{Dg}(\boldsymbol{s}_C)$ $a \times a$ and $\boldsymbol{R}$ $q \times q$. In turn $\boldsymbol{CC}'=\boldsymbol{L}\mathrm{Dg}(\boldsymbol{s}_C^2)\boldsymbol{L}'$, and $\boldsymbol{LL}' = \boldsymbol{L}'\boldsymbol{L} = \boldsymbol{I}_a$. With $\boldsymbol{R}_1$ $q\times a$ and $\boldsymbol{R}_0$ $q \times (q-a)$, $\boldsymbol{C}'\boldsymbol{C}=[\boldsymbol{R}_1\ \ \boldsymbol{R}_0]\begin{bmatrix}\mathrm{Dg}(\boldsymbol{s}_C^2) & \mathbf{0}\\ \mathbf{0} & \mathbf{0}\end{bmatrix}\begin{bmatrix}\boldsymbol{R}_1'\\ \boldsymbol{R}_0'\end{bmatrix}$ and $\boldsymbol{RR}'\ \boldsymbol{R}'\boldsymbol{R} = \boldsymbol{I}_q$. If $\boldsymbol{S}=\begin{bmatrix}\boldsymbol{C}\\ \boldsymbol{R}_0'\end{bmatrix}=\begin{bmatrix}\boldsymbol{L}\mathrm{Dg}(\boldsymbol{s}_C) & \boldsymbol{R}_1'\\ & \boldsymbol{R}_0'\end{bmatrix}$, then $\boldsymbol{S}^{-1} = [\boldsymbol{R}_1\ \mathrm{Dg}(\boldsymbol{s}_C)^{-1}\boldsymbol{L}'\ \boldsymbol{R}_0]=[C'(CC')^{-1}\boldsymbol{R}_0]$and

$$\begin{aligned}\boldsymbol{y}_s &=\boldsymbol{X}_s\boldsymbol{S}^{-1}\boldsymbol{S}\beta+\boldsymbol{Z}_s\boldsymbol{d}_s+\boldsymbol{e}_s\\ &=[\boldsymbol{X}_s\boldsymbol{C}'(\boldsymbol{CC}')^{-1}\ \ \boldsymbol{X}_s\boldsymbol{R}_0]\begin{bmatrix}\theta\\ \theta_\perp\end{bmatrix}+\boldsymbol{Z}_s\boldsymbol{d}_s+\boldsymbol{e}_s\\ &=[\boldsymbol{X}_{s,c}\ \ \boldsymbol{X}_{s,\perp}]\begin{bmatrix}\theta\\ \theta_\perp\end{bmatrix}+\boldsymbol{Z}_s\boldsymbol{D}_s+\boldsymbol{e}_s.\end{aligned}$$

The lemma has many implications. Here rank $(\boldsymbol{X}_{s,\mathrm{c}}) = a$. Testing H_0:$\boldsymbol{\theta} = \mathbf{0}$ compares full model $\boldsymbol{y}_s = \boldsymbol{X}_{s,\mathrm{c}}\boldsymbol{\theta} + \boldsymbol{X}_{s,\perp}\ \boldsymbol{\theta}_\perp + \boldsymbol{Z}_s\boldsymbol{d}_s + \boldsymbol{e}_s$ to the nested model $\boldsymbol{y}_s = \boldsymbol{X}_{s,\perp}\boldsymbol{\theta}_\perp + \boldsymbol{Z}_s\ \boldsymbol{d}_s + \boldsymbol{e}_s$.

Table 1

Mixed Model Estimates, Standard Errors (SE) and p-values for the Dental Data

						Covariance Estimates	
Cov*	Model	Fixed Effect	Estimate	SE	P-value	Random effects	Error
1	I	Intercept	16.76	0.802	<0.0001	$\widehat{\sigma}_d^2=4.472$	$\hat{\sigma}^2 = 2.050$
		Age	0.66	0.062	<0.0001		
	II	Intercept	17.71	0.834	<0.0001	$\widehat{\sigma}_d^2=3.267$	$\hat{\sigma}^2 = 2.050$
		Age	0.66	0.062	<0.0001		
		Gender	−2.32	0.761	0.0054		
	III	Intercept	16.34	0.981	<0.0001	$\widehat{\sigma}_d^2=3.299$	$\hat{\sigma}^2 = 1.922$
		Age	0.78	0.078	<0.0001		
		Gender	1.03	1.537	0.5035		
		Age×Gender	−0.30	0.121	0.0141		
2	I	Intercept	16.76	0.775	<0.0001	$\widehat{\sigma}_{d1}^2=5.415$	$\hat{\sigma}^2 = 1.716$
		Age	0.66	0.071	<0.0001	$\widehat{\sigma}_{d2}^2=0.051$	
						$\hat{\rho} = -0.61$	
	II	Intercept	17.64	0.891	<0.0001	$\widehat{\sigma}_{d1}^2=7.823$	$\hat{\sigma}^2 = 1.716$
		Age	0.66	0.071	<0.0001	$\widehat{\sigma}_{d2}^2=0.051$	
		Gender	−2.15	0.792	0.0120	$\hat{\rho} = -0.77$	
	III	Intercept	16.34	1.019	<0.0001	$\widehat{\sigma}_{d1}^2=5.786$	$\hat{\sigma}^2 = 1.716$
		Age	0.78	0.086	<0.0001	$\widehat{\sigma}_{d2}^2=0.033$	
		Gender	1.03	1.596	0.5192	$\hat{\rho} = -0.67$	
		Age×Gender	−0.30	0.135	0.0257		

NIH-PA Author Manuscript

NIH-PA Author Manuscript

Cov*	Model	Fixed Effect	Estimate	SE	P-value	Covariance Estimates: Random effects	Covariance Estimates: Error
3	I	Intercept	17.30	0.656	<0.0001	$\widehat{\sigma}_{d1}^2=3.608$	$\widehat{\sigma}_1^2=0.442$
		Age	0.61	0.062	<0.0001	$\widehat{\sigma}_{d2}^2=0.039$	$\widehat{\sigma}_2^2=2.707$
						$\hat{\rho}=-0.45$	
	II	Intercept	18.44	0.818	<0.0001	$\widehat{\sigma}_{d1}^2=4.994$	$\widehat{\sigma}_2^2=0.440$
		Age	0.58	0.062	<0.0001	$\widehat{\sigma}_{d2}^2=0.037$	$\widehat{\sigma}_2^2=2.761$
		Gender	−1.94	0.777	0.0194	$\hat{\rho}=-0.63$	
	III	Intercept	16.34	1.135	<0.0001	$\widehat{\sigma}_{d1}^2=3.891$	$\widehat{\sigma}_1^2=0.444$
		Age	0.78	0.099	<0.0001	$\widehat{\sigma}_{d2}^2=0.025$	$\widehat{\sigma}_2^2=2.656$
		Gender	1.03	1.377	0.4586	$\hat{\rho}=-0.50$	
		Age×Gender	−0.30	0.119	0.0143		

* 1 ≡ Random intercept only and $\boldsymbol{\Sigma}_{ei}(\boldsymbol{\tau}_e)=\sigma^2\boldsymbol{I}_{n_i}$

2 ≡ Random intercept and slope with unstructured $\boldsymbol{\Sigma}_{di}(\boldsymbol{\tau}_d)$ and $\boldsymbol{\Sigma}_{ei}(\boldsymbol{\tau}_e)=\sigma^2\boldsymbol{I}_{n_i}$

3 ≡ Random intercept and slope with unstructured $\boldsymbol{\Sigma}_{di}(\boldsymbol{\tau}_d)$ and $\sum_{ei}(\boldsymbol{\tau}_e)=\sigma_1^2\boldsymbol{I}_{n_i}$ for boys and $\sum_{ei}(\boldsymbol{\tau}_e)=\sigma_2^2\boldsymbol{I}_{n_i}$ for girls

NIH-PA Author Manuscript
NIH-PA Author Manuscript

Table 2

Model R^2_β Values for the Dental Data

Cov*	Model	Kenward-Rodger	Contain	Satterthwaite	Residual
		F, e, R^2_β	F, e, R^2_β	F, e, R^2_β	F, e, R^2_β
1	I	114.8, 80.0, *0.59*	114.8, 80, *0.59*	114.8, 80.0, *0.59*	114.8, 106, *0.52*
	II	61.3, 50.1, *0.71*	62.1, 80, *0.61*	62.1, 37.5, *0.77*	62.1, 105, *0.54*
	III	45.4, 66.6, *0.67*	46.0, 79, *0.64*	46.0, 95.0, *0.59*	46.0, 104, *0.57*
2	I	85.9, 26.0, *0.77*	85.9, 26, *0.77*	85.9, 26.0, *0.77*	85.9, 106, *0.45*
	II	45.7, 33.3, *0.73*	46.9, 54, *0.63*	46.9, 25.5, *0.79*	46.9, 105, *0.47*
	III	36.2, 27.6, *0.80*	37.5, 54, *0.68*	37.5, 33.5, *0.77*	37.5, 104, *0.52*
3	I	96.4, 17.9, *0.84*	104.3, 26, *0.80*	104.3, 17.9, *0.85*	104.3, 106, *0.50*
	II	48.5, 23.1, *0.81*	54.0, 54, *0.67*	54.0, 19.6, *0.85*	54.0, 105, *0.51*
	III	41.8, 24.0, *0.84*	44.1, 54, *0.71*	44.1, 26.8, *0.83*	44.1, 104, *0.56*

*1 ≡ Random intercept only and $\mathbf{\Sigma}_{ei}(\boldsymbol{\tau}_e) = \sigma^2 \mathbf{I}_{n_i}$

2 ≡ Random intercept and slope with unstructured $\mathbf{\Sigma}_{di}(\boldsymbol{\tau}_d)$ and $\mathbf{\Sigma}_{ei}(\boldsymbol{\tau}_e) = \sigma^2 \mathbf{I}_{n_i}$

3 ≡ Random intercept and slope with unstructured $\mathbf{\Sigma}_{di}(\boldsymbol{\tau}_d)$ and $\sum_{ei}(\boldsymbol{\tau}_e)=\sigma_1^2 \boldsymbol{I}_{n_i}$ for boys and $\sum_{ei}(\boldsymbol{\tau}_e)=\sigma_2^2 \boldsymbol{I}_{n_i}$ for girls

Table 3

Multivariate and Mixed Model Hypotheses, Contrasts, and Measures of Association for the Dental Data With Unstructured Covariance

	Multivariate Model			Mixed Model		
Hypothesis	C_M	U	(ν_h, ν_e)	C	(ν_h, ν_e)	$\hat{\eta}=R_\beta^2$
Gender Effect	$[1\ -1]$	$\begin{bmatrix} 1 \\ \mathbf{0}_{3\times1} \end{bmatrix}$	1, 25	$[1\ 0\ 0\ 0 -1\ 0\ 0\ 0]$	1, 25	0.27
Joint Age	$[1\ 1]/2$	$\begin{bmatrix} \mathbf{0}_{1\times3} \\ \boldsymbol{I}_3 \end{bmatrix}$	3, 23	$[\mathbf{0}_{3\times1}\ \boldsymbol{I}_3\ \mathbf{0}_{3\times1}\ \boldsymbol{I}_3]/2$	3, 23	0.97
Gender×Age	$[1\ -1]$	$\begin{bmatrix} \mathbf{0}_{1\times3} \\ \boldsymbol{I}_3 \end{bmatrix}$	3, 23	$[\mathbf{0}_{3\times1}\ \boldsymbol{I}_3\ \mathbf{0}_{3\times1}\ -\boldsymbol{I}_3]$	3, 23	0.07

NIH-PA Author Manuscript
NIH-PA Author Manuscript

Table 4

Partial R_β^2 Results Using Kenward-Rodger *F* for the Dental Data with Model II

Cov*	Variable	F	ν	Partial R_β^2
1	Gender	9.29	25.0	0.27
	Age	114.84	80.0	0.59
2	Gender	7.34	25.0	0.23
	Age	85.85	26.0	0.77
3	Gender	6.24	25.2	0.20
	Age	87.38	16.9	0.84

*1 ≡ Random intercept only and $\mathbf{\Sigma}_{ei}(\boldsymbol{\tau}_e) = \sigma^2 \mathbf{I}_{n_i}$

2 ≡ Random intercept and slope with unstructured $\mathbf{\Sigma}_{di}(\boldsymbol{\tau}_d)$ and $\mathbf{\Sigma}_{ei}(\boldsymbol{\tau}_e) = \sigma^2 \mathbf{I}_{n_i}$

3 ≡ Random intercept and slope with unstructured $\mathbf{\Sigma}_{di}(\boldsymbol{\tau}_d)$ and $\sum_{ei}(\boldsymbol{\tau}_e)=\sigma_1^2\mathbf{I}_{n_i}$ for boys and $\sum_{ei}=(\boldsymbol{\tau}_e)=\sigma_2^2\mathbf{I}_{n_i}$ for girls