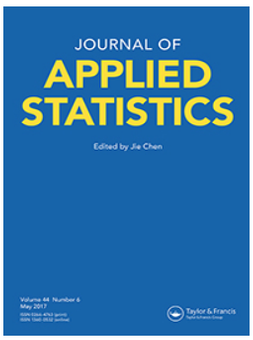

Taylor & Francis
Taylor & Francis Group

Journal of Applied Statistics

ISSN: 0266-4763 (Print) 1360-0532 (Online) Journal homepage: http://www.tandfonline.com/loi/cjas20

An R^2 statistic for fixed effects in the generalized linear mixed model

Byron C. Jaeger, Lloyd J. Edwards, Kalyan Das & Pranab K. Sen

To cite this article: Byron C. Jaeger, Lloyd J. Edwards, Kalyan Das & Pranab K. Sen (2017) An R^2 statistic for fixed effects in the generalized linear mixed model, Journal of Applied Statistics, 44:6, 1086-1105, DOI: 10.1080/02664763.2016.1193725

To link to this article: https://doi.org/10.1080/02664763.2016.1193725

JOURNAL OF APPLIED STATISTICS, 2017
VOL. 44, NO. 6, 1086–1105
http://dx.doi.org/10.1080/02664763.2016.1193725

Taylor & Francis
Taylor & Francis Group

An R^2 statistic for fixed effects in the generalized linear mixed model

Byron C. Jaeger[a], Lloyd J. Edwards[a], Kalyan Das[b] and Pranab K. Sen[a]

[a]Department of Biostatistics, University of North Carolina, Chapel Hill, NC, USA; [b]Department of Statistics, University of Calcutta, Kolkata, India

ABSTRACT

Measuring the proportion of variance explained (R^2) by a statistical model and the relative importance of specific predictors (semi-partial R^2) can be essential considerations when building a parsimonious statistical model. The R^2 statistic is a familiar summary of goodness-of-fit for normal linear models and has been extended in various ways to more general models. In particular, the generalized linear mixed model (GLMM) extends the normal linear model and is used to analyze correlated (hierarchical), non-normal data structures. Although various R^2 statistics have been proposed, there is no consensus in statistical literature for the most sensible definition of R^2 in this context. This research aims to build upon existing knowledge and definitions of R^2 and to concisely define the statistic for the GLMM. Here, we derive a model and semi-partial R^2 statistic for fixed (population) effects in the GLMM by utilizing the penalized quasi-likelihood estimation method based on linearization. We show that our proposed R^2 statistic generalizes the widely used marginal R^2 statistic introduced by Nakagawa and Schielzeth, demonstrate our statistics capability in model selection, show the utility of semi-partial R^2 statistics in longitudinal data analysis, and provide software that computes the proposed R^2 statistic along with semi-partial R^2 for individual fixed effects. The software provided is adapted for both SAS and R programming languages.

ARTICLE HISTORY
Received 1 December 2015
Accepted 21 May 2016

KEYWORDS
Blood pressure; CD4$^+$ cell counts; generalized linear mixed model; clustered data; R-squared; statistical software

AMS SUBJECT CLASSIFICATION
62-07; 62Fxx; 62Hxx; 62Pxx

1. Introduction

The generalized linear mixed model (GLMM) has proved to be a powerful tool for the analysis of data with clustered structure. Such data structures are known to occur in medical sciences [5], ecology and evolution [3], epidemiology [9], social and behavioral sciences [14], and numerous other fields of study. The GLMM incorporates random effects into the generalized linear model (GLM), and includes the linear mixed model (LMM) as a special case. Random effects may account for sources of heterogeneity and dependence in the data, thereby aiding in statistical inference. With the use of GLMMs increasing in multiple scientific categories, measuring overall goodness-of-fit and selecting the best model have become very important tasks. In the normal linear model (LM), these tasks are easily

CONTACT Byron C. Jaeger bcjaeger@live.unc.edu

handled using the coefficient of determination, denoted as the model R^2 statistic. Several desirable qualities underpin the R^2 statistic, making it an excellent goodness-of-fit measure. First, the statistic is interpreted as the proportion of variance explained by the proposed model, which is easily linked to goodness-of-fit. Second, the statistic is unit-less and ranges from 0 to 1, which means investigators can objectively evaluate the fit of models. Additionally, one may compare R^2 values across studies in a similar manner as standardized effect size statistics under certain circumstances (e.g. models with the same responses and similar set of predictors). Thus, the R^2 statistic can be utilized for meta-analyses [25]. Finally, partial and semi-partial R^2 statistics share the desirable properties of the model R^2 statistic, and provide a means of assessing the relative importance of predictor variables in determining the outcome.

Generalization of the R^2 statistic to GLMMs has proved to be a difficult task. In fact, generalization of the R^2 statistic from the LM to the LMM remains an unresolved problem. There are many proposed methods to compute R^2 for the LMM in the statistical literature [10,15,25,34,41]; however, fundamentally different approaches are taken in defining the R^2 for the LMM, and the structure of clustered data leads to some undesirable properties (see below) in R^2 estimates. Additionally, few authors provide software to compute their proposed statistic. As a result, consensus for the definition of R^2 in the LMM is yet to be obtained.

Recently, Nakagawa and Schielzeth [25] introduced conditional and marginal R^2 statistics for the LMM with random intercepts only (no random slopes) and later generalized their measures to the GLMM. The marginal statistic estimates the variance explained by the population parameters (fixed effects) in the model, while the conditional statistic estimates the variance explained by the entire model (fixed effects and random effects). These measures are constructed by dividing the variance of the predicted scores by the variance of the predicted scores plus within- and between-subject variance components. Johnson [20] later extended Nakagawa and Shielzeth's work so that their R^2 statistics could be used to estimate explained variance in models that contain random slopes. The approach taken by Nakagawa and Schielzeth is the first of its kind to be generalized to the GLMM. In this paper, we are pleased to introduce a novel method of calculating R^2 for the GLMM that builds upon the pioneering work of Nakagawa and Schielzeth.

The structure of this article is as follows. In Section 2, we discuss some of the more common issues when generalizing R^2. In Section 3, we summarize the salient features of a GLMM. In Section 4, we derive our proposed R^2 statistic and highlight some important characteristics. In Section 5, we briefly summarize Nakagawa and Schielzeth's marginal and conditional R^2 statistics. In Section 6, we compare our proposed R^2 measure to the marginal and conditional estimates. First, we show that the marginal R^2 statistic proposed by Nakagawa and Schielzeth is equivalent to a simplified form of the R^2 statistic, we introduce in the LMM. Next, we conduct simulation studies to empirically assess our claim and illustrate differences between the R^2 estimates. Lastly, We apply the R^2 estimates to longitudinal blood pressure and CD4 data. (Supplementary material can be viewed online at https://github.com/bcjaeger/R2FixedEffectsGLMM/) provides a SAS

macro [1] and R package (r2glmm) [18,19] to compute the proposed model and partial R^2 statistics along with the code used to conduct the two simulation studies in Section 6.

2. Difficulties of generalizing R^2

One of the first R^2 estimates for the LMM was introduced by Raudenbush and Bryk [29], who calculate R^2 measures for each variance component in a mixed model. That is, a separate R^2 statistic is computed for each random effect and the residual variance. Snijders and Bosker [34] later observed that addition of fixed effects may reduce the variance estimate for one component while simultaneously increasing variance for another. This feature causes some variance component R^2 estimates to be negative. As a remedy, Snijders and Bosker suggest calculating one R^2 statistic for each level of an LMM (e.g. the observation level, individual level, grouping level, etc.). However, Nakagawa and Schielzeth [25] point out that R^2 statistics corresponding to individual levels in a model may decrease with the addition of a fixed predictor in larger models. Snijders and Bosker [37] suggest that some decreases in R^2 estimates (i.e. changes in the 'wrong' direction) indicate mis-specification of the statistical model, and can therefore be a diagnostic tool for model selection. However, mis-specification is not always the cause for a decrease in R^2 values, which complicates the application of decreasing R^2 as a diagnostic tool. Gelman and Pardoe [15] introduced an extension of the multilevel R^2 proposed by Snijders and Bosker so that an arbitrary number of levels could be modeled, but the technical detail involved in applying the method acts as a deterrent for investigators that do not have the necessary statistical background. This short review covers only a select subset of R^2 estimates for the LMM (see [10,24,25] for more comprehensive summaries), but the difficulties presented (i.e. negative values, decreasing with additional fixed effects, implementation) represent some of the most common obstacles for generalizing R^2 to the LMM and the GLMM.

3. The generalized linear mixed model

Consider n independent sampling units with m_i repeated measurements on the ith subject and $N = \sum_{i=1}^{n} m_i$ total observations. Denote this vector of repeated outcomes as $\boldsymbol{y}_i = (y_{i,1}, \ldots, y_{i,m_i})$. Conditionally on the subject-specific random effects, assume that the observed outcomes are independent with a density function that is a member of the exponential family. That is, the observation corresponding to the ith subject at the jth measurement has probability density function

$$f(y_{ij}|\boldsymbol{b}_i) = \exp\left\{\frac{y_{ij}\theta_{ij} - b(\theta_{ij})}{a(\phi)} + c(\phi, y_{ij})\right\}, \tag{1}$$

where $\theta_{ij}(\mu_{ij}) = \theta_{ij}$ is the *canonical parameter*, ϕ is the *scale parameter*, and $a(\cdot), b(\cdot)$, and $c(\cdot)$ are known functions. We specify the observation-specific (scalar notation) GLMM [4] as

$$g(E[y_{ij}|\,\boldsymbol{b}_i]) = \eta_{ij}^{b} \quad \text{and} \quad \mathcal{V}(y_{ij}|\,\boldsymbol{b}_i) = \phi \cdot a_{ij} v(\mu_{ij}^{b}), \tag{2}$$

where $\eta_{ij}^b = \boldsymbol{x}_{ij}'\boldsymbol{\beta} + \boldsymbol{z}_{ij}'\boldsymbol{b}_i$ is the *linear predictor*, $g(\cdot)$ is the *link function*, $\mu_{ij}^b = g^{-1}(\eta_{ij}^b)$ is the *conditional expected value* of y_{ij}, $\boldsymbol{\beta}$ is a $p \times 1$ vector of unknown population level coefficients, $\boldsymbol{b}_i$ is the subject-specific $q \times 1$ random effect vector, $\boldsymbol{x}_{ij}$ and $\boldsymbol{z}_{ij}$ are $p \times 1$, $q \times 1$ vectors of known covariates for fixed effects and random effects, respectively, $\mathcal{V}(\cdot)$ is the *covariance operator*, ϕ is a dispersion parameter that may or may not be known, a_{ij} is a known constant (e.g. the reciprocal of a binomial denominator), and $v(\cdot)$ is a specified *variance function*. We will also utilize a subject-specific (matrix notation) model specification,

$$g(E[\boldsymbol{y}_i|\,\boldsymbol{b}_i]) = \boldsymbol{\eta}_i^b \quad \text{and} \quad \mathcal{V}(\boldsymbol{y}_i|\,\boldsymbol{b}_i) = \boldsymbol{V}_i, \tag{3}$$

where $\boldsymbol{\eta}_i^b = \boldsymbol{X}_i\boldsymbol{\beta} + \boldsymbol{Z}_i\boldsymbol{b}_i$, $\boldsymbol{X}_i = (\boldsymbol{x}_{i,1}, \ldots, \boldsymbol{x}_{i,m_i})'$ is the $m_i \times p$ fixed-effects design matrix for subject i, $\boldsymbol{Z}_i = (\boldsymbol{z}_{i,1}, \ldots, \boldsymbol{z}_{i,m_i})'$ is the $m_i \times q$ random effects design matrix for subject i, and $\boldsymbol{V}_i = \phi \cdot \text{diag}\{a_{i,1}^{-1}v(\mu_{i,1}^b), \ldots, a_{i,m_i}^{-1}v(\mu_{i,m_i}^b)\}$. Finally, it is also convenient to summarize the complete model (stacked matrix notation),

$$g(E[\boldsymbol{y}|\,\boldsymbol{b}]) = \boldsymbol{\eta}^b \quad \text{and} \quad \mathcal{V}(\boldsymbol{y}|\,\boldsymbol{b}) = \boldsymbol{V}, \tag{4}$$

where $\boldsymbol{\eta}^b = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{Z}\boldsymbol{b}$, $\boldsymbol{X} = (\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n)'$, $\boldsymbol{Z} = \text{diag}\{\boldsymbol{Z}_1, \ldots, \boldsymbol{Z}_n\}$, $\boldsymbol{b} = (\boldsymbol{b}_1, \ldots, \boldsymbol{b}_n)'$, and $\boldsymbol{V} = \text{diag}\{\boldsymbol{V}_1, \ldots, \boldsymbol{V}_n\}$. The random effects $\boldsymbol{b}_1, \ldots, \boldsymbol{b}_n$ are each assumed to follow a multivariate normal distribution with mean $\boldsymbol{0}_q$ and $q \times q$ covariance matrix $\boldsymbol{\Sigma}_b$. We have thus described a GLMM with a multivariate normal mixing distribution for the random effects. In a systematic review of statistical inference in the GLMM, Tuerlinckx *et al.* [36] note that the model described above has the structure applied most often in practice. The authors go on to describe the three most common tests used to draw inference from GLMMs: Likelihood ratio tests (LRTs), Wald tests, and score tests. Our proposed R^2 statistic is based on a Wald test of the appropriate fixed effects. In general, the Wald test is the most flexible and reliable when testing fixed effects for large samples, but it can be unreliable for small data sets. Thus, it is important to exercise caution if the proposed R^2 statistic is applied to a small sample.

4. Proposed R^2 for fixed effects in the GLMM

The choice of the null model plays a central role in defining a model R^2 statistic. To define the model R^2 statistics for fixed effects in the GLMM, we compare the following linear predictor models:

$$\text{Full Model} : \boldsymbol{\eta}^b = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{Z}\boldsymbol{b} \quad \text{Null Model} : \boldsymbol{\eta}^b = \beta_0\boldsymbol{1} + \boldsymbol{Z}\boldsymbol{b}, \tag{5}$$

where $\boldsymbol{1}$ is an $N \times 1$ vector of 1's. We consider models that include an intercept in $\boldsymbol{X}$ and may or may not include an intercept in $\boldsymbol{Z}$. We also require both models to have the same random effects and covariance structure. Comparing the full model to the null model in Equation (5) is equivalent to a test of the hypothesis $H_0 : \boldsymbol{C}\boldsymbol{\beta} = \boldsymbol{0}$ for $\boldsymbol{C} = [\boldsymbol{0}_{(p-1)\times 1}\boldsymbol{I}_{p-1}]$ of rank $p-1$.

4.1. Penalized quasi-likelihood estimation

The exact likelihood function for the GLMM involves an intractable high-dimensional integration. Several approximations to the likelihood function and approximate maximum

likelihood estimators (MLEs) have been proposed in the literature [31,40]. Penalized quasi-likelihood (PQL) [4], the default estimation technique used by Proc Glimmix in SAS, is one of the most widely used approaches. This technique applies Laplace's method for integral approximation and the resulting likelihood function is approximately Gaussian. Thus, the PQL approximation transforms the GLMM into a pseudo LMM. R^2 estimation methods for the LMM may be applied to the pseudo model, with estimation and inference under restricted maximum likelihood (REML), thereby producing the model R^2 corresponding to the original GLMM. It is well known [4] that this method produces biased estimates of variance components in logistic models when the number of observations occurring within subjects is small. This is a crucial point for which many researchers prefer the hierarchical likelihood approach, first introduced by Lee and Nelder [23]. Another popular alternative is using the marginal (integrated) likelihood approach, which resorts to various numerical integration methods (Laplace, Gaussian quadrature, etc). Some notable software packages implement this approach, including the `lme4` package [2] in R. Simulations in Section 6.3 show that the proposed R^2 statistic is biased in the 'worst-case' scenario of small cluster sizes with binary outcomes. Therefore, we recommend caution when applying our proposed R^2 statistic in logistic models.

Breslow and Clayton assume that the conditional mean $E[\boldsymbol{y}|\,\boldsymbol{b}] = g^{-1}(\boldsymbol{\eta}^b) = \boldsymbol{\mu}^b$ can be derived from a first-order approximation to the hierarchical model that is valid in the limit as the components of dispersion approach zero. Here, we will use the observation specific (scalar) model notation from Equation (2) and denote the *inverse link function* as $h(\cdot)$. We consider $y_{ij} \mid \boldsymbol{b}_i = h(\eta^b_{ij}) + \epsilon_{ij}$ with $\mathcal{V}(\epsilon_{ij}) = \phi \cdot a_{ij}\nu(\mu^b_{ij})$ and $\boldsymbol{b}_i \sim \mathcal{N}_q(\mathbf{0}, \boldsymbol{\Sigma}_b)$. A first-order Taylor series approximation of $h(\eta^b_{ij})$ about the estimated linear predictor $\hat{\eta}_{ij} = \boldsymbol{x}'_{ij}\hat{\boldsymbol{\beta}} + \boldsymbol{z}'_{ij}\hat{\boldsymbol{b}}_i$ yields

$$y_{ij} \approx h(\hat{\eta}_{ij}) + h'(\hat{\eta}_{ij})(\eta^b_{ij} - \hat{\eta}_{ij}) + \epsilon_{ij}, \tag{6}$$

where $h'(\hat{\eta}_{ij}) = \partial h(\hat{\eta}_{ij})/\partial \eta_{ij}$ is the first derivative of the conditional mean evaluated with the estimates $\hat{\boldsymbol{\beta}}$ and $\hat{\boldsymbol{b}}_i$. Equation (6) is equivalently written using the expanded forms of η^b_{ij} and $\hat{\eta}_{ij}$ as

$$y_{ij} \approx h(\hat{\eta}_{ij}) + h'(\hat{\eta}_{ij})\boldsymbol{x}'_{ij}(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}) + h'(\hat{\eta}_{ij})\boldsymbol{z}'_{ij}(\boldsymbol{b}_i - \hat{\boldsymbol{b}}_i) + \epsilon_{ij}, \tag{7}$$

Using the facts that $h(\hat{\eta}_{ij}) = \hat{\mu}_{ij}$ and $h'(\hat{\eta}_{ij}) = 1/g'(\hat{\mu}_{ij})$ [40, Section 2] along with some algebraic manipulation, Equation (7) is more conveniently written as

$$g'(\hat{\mu}_{ij}) \cdot (y_{ij} - \hat{\mu}_{ij}) + \boldsymbol{x}'_{ij}\hat{\boldsymbol{\beta}} + \boldsymbol{z}'_{ij}\hat{\boldsymbol{b}}_i \approx \boldsymbol{x}'_{ij}\boldsymbol{\beta} + \boldsymbol{z}'_{ij}\boldsymbol{b}_i + g'(\hat{\mu}_{ij}) \cdot \epsilon_{ij}. \tag{8}$$

The left-hand side of Equation (8) is analogous to the pseudo variable used in the estimating equations for the GLM [35, Section 4.3]. Denoting this pseudo outcome as y^*_{ij} and letting ϵ^*_{ij} denote $g'(\hat{\mu}_{ij}) \cdot \epsilon_{ij}$, Equation (8) can be expressed as

$$y^*_{ij} \approx \boldsymbol{x}'_{ij}\boldsymbol{\beta} + \boldsymbol{z}'_{ij}\boldsymbol{b}_i + \epsilon^*_{ij}. \tag{9}$$

Turning to the subject specific notation introduced in Equation (3), the working vector of pseudo responses $\boldsymbol{y}^*_i = (y^*_{i,1}, \ldots, y^*_{i,m_i})$ can be specified as an approximate LMM with

multivariate normal $\boldsymbol{b}_i$ and $\boldsymbol{\epsilon}_i^*$ independent with mean zero and

$$\mathcal{V}\begin{pmatrix}\boldsymbol{b}_i\\ \boldsymbol{\epsilon}_i^*\end{pmatrix} = \begin{pmatrix}\boldsymbol{\Sigma}_b & 0\\ 0 & \boldsymbol{W}_i^{-1}\end{pmatrix},$$

where $\boldsymbol{W}_i = \text{diag}\{w_1, \ldots, w_{m_i}\}$ with

$$w_j = \left\{\phi \cdot a_{ij} v(\mu_{ij}^b)\left[g'(\mu_{ij}^b)\right]^2\right\}^{-1}$$

can be recognized as the GLM iterated weights matrix [4, Section 2.1]. In the LMM with uncorrelated random error, $\boldsymbol{W}_i$ reduces to the familiar form $\sigma^2 \cdot \boldsymbol{I}_{m_i}$. Model fitting is accomplished by iterating between updating the working vector of pseudo-responses and obtaining REML estimates from the approximate linear mixed model in Equation (9) until convergence.

4.2. $R^2_{\beta^*}$, an Extension of R^2_β to the GLMM

Upon convergence of the PQL estimating equations, we calculate an R^2 statistic for the pseudo LMM that approximates the original GLMM using the F-statistic for a Wald test of fixed effects,

$$F(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\Sigma}}) = \frac{(\boldsymbol{C}\hat{\boldsymbol{\beta}})'[\boldsymbol{C}(\boldsymbol{X}'\hat{\boldsymbol{\Sigma}}^{-1}\boldsymbol{X})^{-1}\boldsymbol{C}']^{-1}(\boldsymbol{C}\hat{\boldsymbol{\beta}})}{\text{rank}(\boldsymbol{C})}, \tag{10}$$

where $\boldsymbol{C} = [\boldsymbol{0}_{p-1\times 1}\boldsymbol{I}_{p-1}]$ has rank $p-1$ and $\hat{\boldsymbol{\Sigma}}$ is the estimated covariance matrix for the pseudo outcomes having a block diagonal structure with subject-specific estimated covariance matrices $\hat{\boldsymbol{\Sigma}}_i = \boldsymbol{Z}_i\hat{\boldsymbol{\Sigma}}_b\boldsymbol{Z}_i' + \hat{\boldsymbol{W}}_i^{-1}$, $i = 1, \ldots, n$, constituting the blocks. Our proposed R^2 statistic is calculated using the one-to-one relationship between R^2 and the F -statistic,

$$R^2_{\beta^*} = \frac{\hat{\nu}^{-1}(p-1)\cdot F(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\Sigma}})}{1 + \hat{\nu}^{-1}(p-1)\cdot F(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\Sigma}})}, \tag{11}$$

where the subscript β^* represents a generalization of R^2_β, the R^2 statistic for the LMM proposed by Edwards *et al.* [10]. In other words, our proposed R^2 statistic extends R^2_β to the GLMM. The quantity $\hat{\nu}$ is the estimated denominator degrees of freedom for $F(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\Sigma}})$. In a study of the properties of R^2_β, Matuszewski and Edwards [24, chap. 3] show that the method of estimation used to determine $\hat{\nu}$ plays a significant role in determining the value of $R^2_{\beta^*}$. In particular, the authors concluded that R^2_β is more robust to covariance misspecification using the residual approximation for $\hat{\nu}$, and that the residual approximation often converged to a lower asymptotic value of R^2_β in comparison to other methods (e.g. Kenward–Roger, Satterthwaite, and containment) of approximation. On the other hand, the authors found that using the Kenward–Roger approximation for $\hat{\nu}$ leads to different asymptotic values of R^2_β when different covariance models are specified. Specifically, the authors found that using the Kenward–Roger estimate for denominator degrees of freedom leads to a lower asymptotic limit of R^2_β when the covariance model is underspecified. In other words, more complex covariance models tend to increase the estimate

of R^2_β when using the Kenward–Roger approximation for $\hat{\nu}$. Since under-specification of the covariance model is known to bias inference for fixed effects [16], and as the original authors of R^2_β recommend [10, Sec. 5], we explore the properties of $R^2_{\beta*}$ using the Kenward–Roger approximation for $\hat{\nu}$ throughout this article. In future research, other methods of approximation will be explored.

$R^2_{\beta*}$ has many desirable qualities in addition to the general characteristics of R^2 statistics. First, $R^2_{\beta*}$ is the most generalized definition of R^2, with R^2 statistics corresponding to the LM, LMM, and GLM being special cases of $R^2_{\beta*}$. This result follows from the fact that the estimating equations used in LMs, LMMs, and GLMs are special cases of the PQL estimating equations [35, Section 4.5]. Thus, $R^2_{\beta*}$ reduces to an R^2 for any of the corresponding models above by using the corresponding form of the PQL estimating equations.

Second, semi-partial estimates of correlation with the outcome of interest may also be computed for any combination of fixed effects using this technique, allowing $R^2_{\beta*}$ to be used as a measure of effect size. To our knowledge, this is a unique feature of $R^2_{\beta*}$ that no other measure in its class provides. The semi-partial R^2 statistic corresponds to a Wald test of the desired subset of fixed effects, and represents the strength of association between a subset of predictors and the outcome, adjusted for other predictors in the full model. Applied statistical practice focuses on two broad questions: (1) Is a predictor statistically significant (i.e. non-zero regression coefficients with high probability), (2) Is the predictor *relatively* important? For example, a predictor may be nonzero, yet still provide a negligible contribution to the outcome. Semi-partial R^2 statistics address question (2), and are denoted here as $R^2_{\beta_j*}$, where j represents the index of the parameter considered in the full model.

Third, as Matuszewski and Edwards [24, chap. 3] demonstrate, the Kenward–Roger approximation leads to a lower asymptotic value of R^2_β when covariance is under-specified, thereby making the $R^2_{\beta*}$ capable of both mean model selection and covariance model selection. This is a pressing topic of future research and not the focus of this article, but some results presented here do indicate this potential.

5. R^2_{NSJ}, an R^2 statistic for marginal and conditional explained variance in the GLMM

Nakagawa and Schielzeth's marginal and conditional R^2 statistics (with the extension proposed by Johnson [20]) for the LMM are written as

$$R^2_{\mathrm{NSJ}(m)} = \frac{\hat{\sigma}^2_f}{\hat{\sigma}^2_f + \hat{\sigma}^2_\varepsilon + \bar{\sigma}^2_l} \quad \text{and} \quad R^2_{\mathrm{NSJ}(c)} = \frac{\hat{\sigma}^2_f + \bar{\sigma}^2_l}{\hat{\sigma}^2_f + \hat{\sigma}^2_\varepsilon + \bar{\sigma}^2_l}, \tag{12}$$

where $\bar{\sigma}^2_l = N^{-1}\sum_{i=1}^{n} \mathrm{Tr}\{\boldsymbol{Z}_i \hat{\boldsymbol{\Sigma}}_b \boldsymbol{Z}_i'\}$ is the average variance of the random effects, N is the total number of observations, $\hat{\sigma}^2_\varepsilon$ is the estimated random error, and $\hat{\sigma}^2_f$ is the fixed-effects variance component introduced by Snijders and Bosker [33, pg. 114]. The (m) in parentheses indicates that the statistic is calculated marginally, which is to say that $R^2_{\mathrm{NSJ}(m)}$ is formulated without conditional knowledge of random effects. The (c) in parentheses indicates that $R^2_{\mathrm{NSJ}(c)}$ calculates explained variance using fixed and random effects. Equation (12) can be extended to non-normal outcomes by decomposing the residual variance σ^2_ε into three components: (i) multiplicative dispersion, (ii) additive dispersion and

(iii) distribution-specific variance. We refer the reader to [25, pg. 137] for further details on this generalization.

6. Comparison of $R^2_{\beta^*}$ and R^2_{NSJ}

Here, we implement the R^2 statistics from Section 4 and 5 in four examples of longitudinal data analysis. The analysis of longitudinal data is a special case of clustered data analysis where clustering typically occurs in time and sampling units are persons. In Section 6.1, we show that $R^2_{\text{NSJ}(m)}$ is a special case of $R^2_{\beta^*}$ under a normal link function. In Section 6.2, we demonstrate this correspondence through simulation and evaluate the efficacy of all R^2 measures summarized above in the LMM. In Section 6.3, we move from the LMM to the GLMM and consider a simulation study with continuous, binary, and count outcomes. Finally, in Section 6.4 and 6.5, we consider two real datasets: a retrospective CD4 Cell study with count outcomes and a blood pressure study with binary outcomes.

6.1. $R^2_{\beta^*}$ generalizes $R^2_{NSJ(m)}$ in the LMM

While $R^2_{\beta^*}$ incorporates the estimated covariance matrix $\hat{\boldsymbol{\Sigma}}$ for clustered data, $R^2_{\text{NSJ}(m)}$ and $R^2_{\text{NSJ}(c)}$ use a scalar representation. Recall that

$$R^2_{\text{NSJ}(m)} = \frac{\hat{\sigma}^2_f}{\hat{\sigma}^2_f + \hat{\sigma}^2_\varepsilon + N^{-1}\sum_{i=1}^n \text{Tr}\left\{\boldsymbol{Z}_i\hat{\boldsymbol{\Sigma}}_b\boldsymbol{Z}_i'\right\}}.$$

Noting that

$$\hat{\sigma}^2_\varepsilon + N^{-1}\sum_{i=1}^n \text{Tr}\left\{\boldsymbol{Z}_i\hat{\boldsymbol{\Sigma}}_b\boldsymbol{Z}_i'\right\} = N^{-1}\left(N\cdot\hat{\sigma}^2_\varepsilon + \sum_{i=1}^n \text{Tr}\{\boldsymbol{Z}_i\hat{\boldsymbol{\Sigma}}_b\boldsymbol{Z}_i'\}\right) = N^{-1}\cdot\text{Tr}(\hat{\boldsymbol{\Sigma}})$$

and denoting $N^{-1}\cdot\text{Tr}(\hat{\boldsymbol{\Sigma}})$ as $\hat{\sigma}^2_{nsj}$, the marginal statistic is equivalently written as

$$R^2_{\text{NSJ}(m)} = \frac{\hat{\sigma}^2_f}{\hat{\sigma}^2_f + \hat{\sigma}^2_{nsj}}.$$

Here, we show in the context of an LMM that simplifying $R^2_{\beta^*}$ by substituting $\hat{\sigma}^2_{nsj}$ for $\hat{\boldsymbol{\Sigma}}$ and using $\hat{\nu} = N - 1$ induces an equivalence between $R^2_{\text{NSJ}(m)}$ and the simplified $R^2_{\beta^*}$, which is defined as

$$R^2_{\text{NSJ}(\beta^*)} = \frac{(N-1)^{-1}\cdot(p-1)F(\hat{\boldsymbol{\beta}},\hat{\sigma}^2_{nsj})}{1 + (N-1)^{-1}\cdot(p-1)F(\hat{\boldsymbol{\beta}},\hat{\sigma}^2_{nsj})},$$

where $F(\hat{\boldsymbol{\beta}},\hat{\sigma}^2_{nsj})$ is the Wald F statistic used to test $H_0 : \beta_1 = \ldots = \beta_{p-1} = 0$ using $\hat{\sigma}^2_{nsj}$ as the estimated covariance. Since $\hat{\sigma}^2_{nsj}$ is scalar, the Wald statistic reduces to the sum of squares for the contrast defined by the estimable function $\boldsymbol{C}\hat{\boldsymbol{\beta}}$ [35, pg. 161–162], which

Table 1. Simulation of dental data using 10,000 datasets.

		R^2 Statistic			
Cov	Model	$R^2_{\beta^*}$	$R^2_{NSJ(\beta^*)}$	$R^2_{\text{NSJ}(c)}$	$R^2_{\text{NSJ}(m)}$
1	Full	0.67 (0.05)	0.42 (0.08)	0.78 (0.05)	0.42 (0.08)
	Reduced	0.59 (0.06)	0.25 (0.05)	0.76 (0.05)	0.25 (0.05)
	Noise	0.04 (0.03)	0.03 (0.03)	0.45 (0.1)	0.03 (0.03)
2	**Full**	0.8 (0.05)	0.42 (0.08)	0.81 (0.04)	0.42 (0.08)
	Reduced	0.76 (0.06)	0.25 (0.05)	0.81 (0.05)	0.25 (0.05)
	Noise	0.05 (0.04)	0.02 (0.02)	0.83 (0.05)	0.02 (0.02)

Notes: Interpretation of covariance specifications:
(1): Random intercepts with $\boldsymbol{W}_i = \sigma^2 \boldsymbol{I}_{m_i}$.
(2): Random intercepts and random slopes with $\boldsymbol{W}_i = \sigma^2 \boldsymbol{I}_{m_i}$.
Values are mean (standard deviation). The bold text indicates the covariance specification and fixed-effects structure used to generate the data (i.e. the correct model).

may be written as $\hat{\boldsymbol{\beta}}'(\boldsymbol{X}'\boldsymbol{X})\hat{\boldsymbol{\beta}} - N \cdot \bar{y}^2$ [7, pg. 125]. Therefore,

$$\begin{aligned}
\frac{p-1}{N-1} F(\hat{\boldsymbol{\beta}}, \hat{\sigma}^2_{nsj}) &= \frac{p-1}{N-1} \cdot \frac{(\boldsymbol{C}\hat{\boldsymbol{\beta}})'[\boldsymbol{C}(\tilde{\boldsymbol{X}}'[\hat{\sigma}^2_{nsj}]^{-1}\tilde{\boldsymbol{X}})^{-1}\boldsymbol{C}']^{-1}(\boldsymbol{C}\hat{\boldsymbol{\beta}})}{p-1} \\
&= [\hat{\sigma}^2_{nsj}(N-1)]^{-1}[\hat{\boldsymbol{\beta}}'(\boldsymbol{X}'\boldsymbol{X})\hat{\boldsymbol{\beta}} - N \cdot \bar{y}^2] \\
&= [\hat{\sigma}^2_{nsj}(N-1)]^{-1}(N-1) \cdot \hat{\sigma}^2_f \\
&= \frac{\hat{\sigma}^2_f}{\hat{\sigma}^2_{nsj}},
\end{aligned}$$

where we have used the fact that

$$\hat{\boldsymbol{\beta}}'(\boldsymbol{X}'\boldsymbol{X})\hat{\boldsymbol{\beta}} - N \cdot \bar{y}^2 = \sum_{i=1}^{N}(\hat{y}_i^2) - N \cdot \bar{y}^2 = \sum_{i=1}^{N}(\hat{y}_i - \bar{y})^2 = (N-1)\hat{\sigma}^2_f.$$

Thus,

$$R^2_{NSJ(\beta^*)} = \frac{\hat{\sigma}^2_f/\hat{\sigma}^2_{nsj}}{1 + \hat{\sigma}^2_f/\hat{\sigma}^2_{nsj}} = \frac{\hat{\sigma}^2_f}{\hat{\sigma}^2_f + \hat{\sigma}^2_{nsj}} = R^2_{NSJ(m)}.$$

The equivalence is demonstrated with simulated data in Table 1.

6.2. Simulated normal longitudinal data

We base this simulation study off the well-known dental data from Potthoff and Roy [27]. 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. We conducted a simulation study in R using these data as a template. All models in the simulation employed REML estimation in R version 3.2.0 [28]. First, an initial LMM was fit to the dental data using the `lme4` [2] package, including fixed effects for age, gender, and their interaction as

well as correlated random intercepts and slopes. Next, we simulated 10,000 sets of outcomes from the fitted values of the initial LMM. For each vector of simulated outcomes, we fit six candidate LMMs using three different fixed-effect structures and two covariance specifications. The fixed-effect structures were: (*Full*) a model with age, gender, and their interaction; (*Reduced*) a model with just an effect for age; and (*Noise*) a model with three effects that had no relation to the outcomes. The covariance specifications were:

(1) Random intercept b_0 with $\boldsymbol{\Sigma}_b = \mathcal{V}(b_0) = \sigma^2_{b_0}$ and $\boldsymbol{W}_i = \sigma^2 \boldsymbol{I}_{m_i}$.
(2) Random intercept b_0 and random slope b_1 with $\boldsymbol{W}_i = \sigma^2 \boldsymbol{I}_{m_i}$ and

$$\boldsymbol{\Sigma}_b = \mathcal{V}(b_0, b_1) = \begin{pmatrix} \sigma^2_{b_0} & \sigma_{b_0 b_1} \\ \sigma_{b_0 b_1} & \sigma^2_{b_1} \end{pmatrix}.$$

For each model, we calculated four estimates of R^2. The first estimate is $R^2_{\beta^*}$ as described in Section 4. The second measure, denoted $R^2_{NSJ(\beta^*)}$ is the simplified version of $R^2_{\beta^*}$ defined above. The other measures are the marginal $R^2_{\mathrm{NSJ}(m)}$ and conditional $R^2_{\mathrm{NSJ}(c)}$ described in Section 5.

The structure used to generate the data is the full model using covariance specification 2, and $R^2_{\beta^*}$ is the only R^2 statistic favoring this model. The marginal $R^2_{\mathrm{NSJ}(m)}$ and simplified $R^2_{\mathrm{NSJ}(\beta^*)}$ correctly prefer the full model, but cannot distinguish the correct covariance structure. The conditional $R^2_{\mathrm{NSJ}(c)}$ prefers the correct covariance structure, but chooses the noise model. The behavior of $R^2_{\mathrm{NSJ}(c)}$ agrees with previous research by Orelien and Edwards [26], who show that conditional estimates of R^2 do not select the correct fixed-effects model. We believe that the tendency to select a model with noise parameters is not acceptable for an R^2 estimate. Thus, our main focus of comparison for the remaining examples will be $R^2_{\mathrm{NSJ}(m)}$ and $R^2_{\beta^*}$.

This example illustrates how the treatment of the estimated covariance matrix $\hat{\boldsymbol{\Sigma}}$ differentiates $R^2_{\beta^*}$ from $R^2_{\mathrm{NSJ}(m)}$ in the LMM. In the GLMM, these statistics are further differentiated by the use of PQL in the calculation of $R^2_{\beta^*}$ and the generalization of $\hat{\sigma}^2_e$ in the calculation of $R^2_{\mathrm{NSJ}(m)}$. The example also provides some evidence to favor $R^2_{\beta^*}$ over $R^2_{\mathrm{NSJ}(m)}$ and $R^2_{\mathrm{NSJ}(c)}$ when conducting model selection and reporting goodness-of-fit for LMMs, particularly when the covariance specification contains random slopes.

The reader may be understandably concerned by the large difference between the values of $R^2_{\beta^*}$ and $R^2_{NSJ(m)}$ for identical models. It is important to bear in mind that although both measures may be applied to the LMM and GLMM, these R^2 estimates do not necessarily measure the same quantity. More specifically, $R^2_{\beta^*}$ measures the multivariate association between the outcome and the fixed effects within the context of correlated observations, while $R^2_{NSJ(m)}$ measures the proportion of explained variance assuming that unexplained variance may be estimated with a scalar and correlation between observations may be ignored.

6.3. Simulated non-normal longitudinal data

Here, we extend the scope of simulated data to non-gaussian outcomes. All simulated datasets were generated in R version 3.2.0. For computational reasons (see supplementary materials at https://github.com/bcjaeger/R2FixedEffectsGLMM/), $R^2_{\beta^*}$ was calculated in SAS. All models in R were fit using `glmer` in the `lme4` package, while the models fit in SAS used proc GLIMMIX [30] with PQL estimation. The structure of our simulated data is based on the response profile analysis of longitudinal outcomes, as described by Fitzmaurice *et al.* [13, ch.5]. We conduct our study in two scenarios in order to assess the performance of $R^2_{\beta^*}$ in the 'worst-case ' scenario for PQL estimation (i.e. small cluster size with binary outcomes). For each scenario, we consider 1000 simulated datasets with $n = 100$ subjects. The timing of observations is common across subjects, but the number of observations within subject i, denoted m_i, varies with uniform probability from 10 to 20 for the first scenario and 3 to 6 observations in the second. Half of the subjects are allocated to the treatment group, while the rest are assigned to the control group. All outcomes are generated using $\boldsymbol{\beta} = (0, -1, 1/2, 1/2)$, the vector of known fixed effects for intercept, treatment group, time, and the interaction of treatment and time, respectively. Random effects are simulated using covariance (2) as defined in Section 6.2 with covariance matrix

$$\boldsymbol{\Sigma}_b = \mathcal{V}(b_0, b_1) = \begin{pmatrix} 6.25 & 0.15 \\ 0.15 & 0.04 \end{pmatrix}.$$

Outcomes are generated based on the subject-specific linear predictor $\boldsymbol{\eta}_i = \boldsymbol{X}_i\boldsymbol{\beta} + \boldsymbol{Z}_i\boldsymbol{b}_i$ (see Section 3). Continuous Gaussian outcomes, denoted $\boldsymbol{y}_i$, are generated by setting $\boldsymbol{y}_i = \boldsymbol{\eta}_i + \boldsymbol{\epsilon}_i$, where $\boldsymbol{\epsilon}_i \sim \mathcal{N}_{m_i}(0, (2.5)^2 \cdot \boldsymbol{I}_{m_i})$. The Gaussian outcomes are then used as an underlying continuous score to generate count and binary outcomes. For example, longitudinal binary outcomes are generated by dichotomizing the continuous outcome along the median value. Figure 1 displays one such set of simulated outcomes. Loess curves [8], drawn as thicker lines, show population trends over time, while individual trajectories (thinner lines) demonstrate the variation between subjects and correlation within subjects. We fit four candidate GLMMs using two different fixed-effect structures and two covariance specifications. The fixed-effect structures were: (*Full*) a model with treatment group, time, and their interaction; (*Reduced*) a model with just an effect for treatment group and time. The covariance specifications were (1) and (2), as defined in Section 6.2.

Tables 2 and 3 show results from the simulation study using large and small cluster sizes. The results demonstrate several important points. First, $R^2_{\beta^*}$ accurately estimates the decreasing trend in explained variance, but $R^2_{\text{NSJ}(m)}$ does not. This point must be explained in some detail. Since the count and binary outcomes were generated using the continuous outcomes as a latent variable, the underlying mechanism for data generation is invariant between outcomes, and in turn the true R^2 (assuming identical model specifications) is invariant between outcomes; however, loss of information occurs from categorization of continuous outcomes, and in turn we should expect to see lower estimates of explained variance. The loss of information is greatest from a dichotomy, so the R^2 estimates should be maximized for normal outcomes and minimized for binary outcomes. However, $R^2_{\text{NSJ}(m)}$ is maximized for binary outcomes and minimized for count outcomes in both simulation scenarios. Furthermore, in the small cluster scenario, $R^2_{\text{NSJ}(m)}$ shows no appreciable difference between the correct fixed effects model and the reduced model. On the other hand,

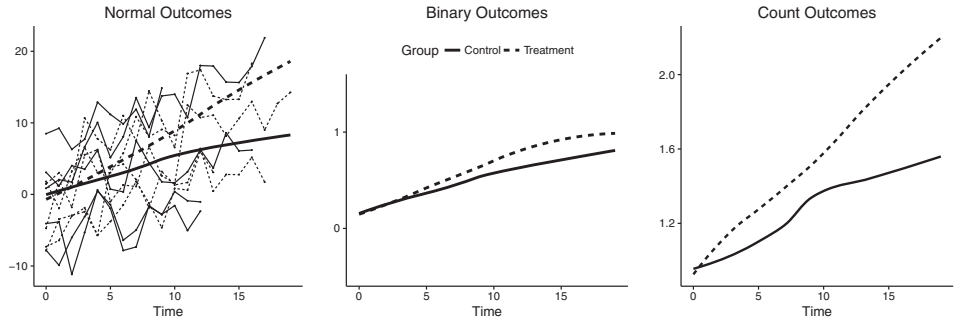

Figure 1. Loess curves (thicker lines) with sample of individual trajectories (thinner lines) from simulated data for normal, binary, and count outcomes. For visual clarity, individual trajectories are omitted from binary and count plots.

Table 2. Simulation of Longitudinal data using 1000 data sets and 10–20 observations within subjects.

			Statistic		
Cov	Model	Outcome	$R^2_{\text{NSJ}(m)}$	$R^2_{\text{NSJ}(c)}$	$R^2_{\beta^*}$
1	Full	Continuous	0.30 (0.04)	0.87 (0.01)	0.87 (0.01)
		Count	0.06 (0.01)	0.10 (0.02)	0.63 (0.03)
		Binary	0.31 (0.04)	0.87 (0.02)	0.56 (0.03)
	Reduced	Continuous	0.27 (0.04)	0.84 (0.01)	0.89 (0.01)
		Count	0.06 (0.01)	0.10 (0.02)	0.73 (0.03)
		Binary	0.27 (0.04)	0.85 (0.02)	0.67 (0.03)
2	**Full**	Continuous	0.30 (0.04)	0.88 (0.01)	0.90 (0.02)
		Count	0.06 (0.01)	0.10 (0.02)	0.80 (0.04)
		Binary	0.31 (0.04)	0.87 (0.02)	0.75 (0.04)
	Reduced	Continuous	0.24 (0.03)	0.89 (0.01)	0.77 (0.03)
		Count	0.06 (0.01)	0.10 (0.02)	0.70 (0.05)
		Binary	0.27 (0.05)	0.87 (0.02)	0.65 (0.04)

Notes: Interpretation of covariance specifications:
(1): Random intercepts with $\boldsymbol{W}_i = \sigma^2 \boldsymbol{I}_{m_i}$.
(2): Random intercepts and random slopes with $\boldsymbol{W}_i = \sigma^2 \boldsymbol{I}_{m_i}$.
Values are mean (standard deviation). Bold text indicates the model used to generate the data (i.e. the correct model).

$R^2_{\beta^*}$ successfully distinguishes this order in both covariance specifications when cluster size is sufficiently large, but mistakenly overestimates explained variance for binary outcomes when cluster size is small and covariance is mis-specified. Thus, results indicate that for data with sufficient cluster sizes, $R^2_{\beta^*}$ correctly estimates a loss in explained variance when continuous outcomes are transformed to count or binary outcomes.

Turning attention to patterned behavior under different covariance models, we note that under covariance (1), a mis-specification, $R^2_{\beta^*}$ *increases* when the interaction term is removed from the model for each outcome type. Under covariance (2), the correct specification, $R^2_{\beta^*}$ *decreases* with the removal of this effect. Naturally, this increase in estimated explained variance is not desirable. In this study, $R^2_{\beta^*}$ exhibits this property for each outcome type under covariance (1), while exhibiting the correct *decrease* in explained variance under covariance (2). These results indicate that generalization via PQL estimation appears to retain the properties of R^2_{β} (using the Kenward–Roger approximation) demonstrated in the LMM by Matuszewski and Edwards [24].

Table 3. Simulation of Longitudinal data using 1000 data sets and 3–6 observations within subjects.

			Statistic		
Cov	Model	Outcome	$R^2_{\text{NSJ}(m)}$	$R^2_{\text{NSJ}(c)}$	$R^2_{\beta^*}$
1	Full	Continuous	0.05 (0.02)	0.81 (0.02)	0.23 (0.04)
		Count	0.01 (0.00)	0.01 (0.00)	0.08 (0.03)
		Binary	0.05 (0.02)	0.76 (0.05)	0.14 (0.05)
	Reduced	Continuous	0.04 (0.01)	0.81 (0.02)	0.26 (0.05)
		Count	0.01 (0.00)	0.01 (0.00)	0.10 (0.04)
		Binary	0.05 (0.02)	0.76 (0.05)	0.18 (0.06)
2	**Full**	Continuous	0.05 (0.02)	0.82 (0.02)	0.38 (0.09)
		Count	0.01 (0.00)	0.01 (0.00)	0.21 (0.11)
		Binary	0.05 (0.02)	0.78 (0.05)	0.2 (0.07)
	Reduced	Continuous	0.04 (0.02)	0.82 (0.02)	0.33 (0.07)
		Count	0.01 (0.00)	0.01 (0.00)	0.16 (0.08)
		Binary	0.05 (0.02)	0.78 (0.05)	0.15 (0.06)

Notes: Interpretation of covariance specifications:
(1): Random intercepts with $\boldsymbol{W}_i = \sigma^2 \boldsymbol{I}_{m_i}$.
(2): Random intercepts and random slopes with $\boldsymbol{W}_i = \sigma^2 \boldsymbol{I}_{m_i}$.
Values are mean (standard deviation). Bold text indicates the model used to generate the data (i.e. the correct model).

6.4. Longitudinal CD4$^+$ cell counts in an HIV study

Here, we consider count data from a retrospective longitudinal CD4 count study [38] of $N = 303$ subjects with HIV. Subjects in the analysis were divided into two age groups, younger ($N_y = 202$) and older ($N_o = 101$). The study was a retrospective case-control study with pairwise (1:2) matching. For each case and two matched controls, longitudinal CD4$^+$ cell counts were obtained at each visit. Reduction in sample size resulted from exclusions due to inadmissible dates. The analysis for this example has 94 older adults and 197 younger adults. We consider the CD4$^+$ cell counts using a Poisson distribution with log link for count data in the GLMM. We have as predictors group status (younger vs. older), time in months since clinic entry, and a group-by-time interaction. Figure 2 shows population loess smoothed curves for young and old age groups, with a small sample of individual response profiles. The loess curves suggest that little change occurs within the population over time, and that age groups do not exhibit different patterns over time.

Table 4 gives the calculated values of $R^2_{\beta^*}$, $R^2_{\text{NSJ}(m)}$, and $R^2_{\text{NSJ}(c)}$ for four candidate models, featuring two sets of fixed effects and two covariance specifications. Semi-partial $R^2_{\beta^*}$ statistics are calculated for each predictor in each model. Observing the semi-partials, one may note the interesting pattern of R^2 estimates for time in the models. For this predictor, we estimate semi-partial statistics of 0.010, 0.031, 0.074, and 0.212 for the full and reduced models using covariance structures 1 and 2, respectively. In the full and reduced models using covariance 1, the statistically significant p-value of $<.001$ for time is misleading given the model estimates predict an average subject in the study will gain about 1 CD4$^+$ cell per month. Here, we see the partial $R^2_{\beta^*}$ statistic is 0.010, indicating that time is the most important predictor in the model, but that it is relatively non-informative. Using covariance 2 with random slopes for time, estimates for the rate of change in CD4$^+$ cell counts are allowed to vary between subjects. In contrast to the random intercept model, random

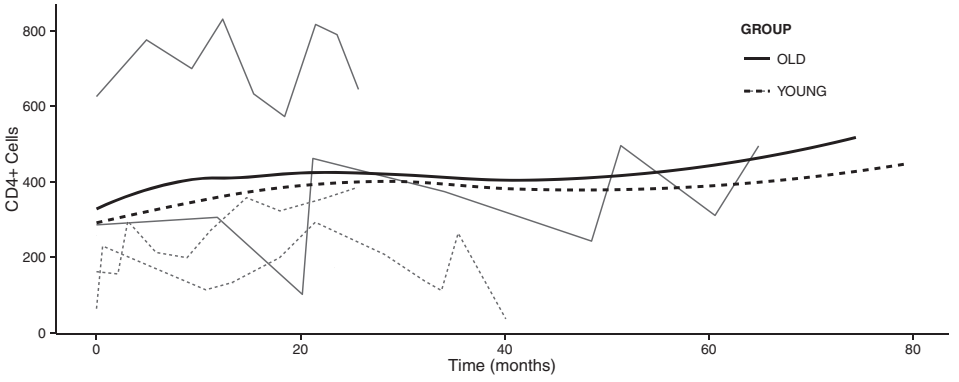

Figure 2. Loess curves (thicker lines) with sample of subject specific $CD4^+$ count trajectories (thinner lines).

Table 4. Estimates for R^2 and semi-partial R^2 with $CD4^+$ count data.

						R^2 Estimates	
Covariance	Model	Fixed effect	Estimate	SE	*p*-value	Semi-partial	Model
1	Full	Intercept	5.6269	0.0793	<.001	–	$R^2_{\beta^*} = 0.063$
		Group	−0.0736	0.0964	.445	0.001	$R^2_{\text{NSJ}(m)} = 0.013$
		Time	0.0036	0.0001	<.001	0.010	$R^2_{\text{NSJ}(c)} = 0.754$
		Group × Time	0.0004	0.0002	.023	<0.001	
	Reduced	Intercept	5.6233	0.0793	<.001	–	$R^2_{\beta^*} = 0.095$
		Group	−0.0682	0.0964	.479	<0.001	$R^2_{\text{NSJ}(m)} = 0.013$
		Time	0.0039	0.0001	<.001	0.031	$R^2_{\text{NSJ}(c)} = 0.754$
2	Full	Intercept	5.5475	0.0878	<.001	–	$R^2_{\beta^*} = 0.170$
		Group	−0.0688	0.1066	.518	0.001	$R^2_{\text{NSJ}(m)} = 0.066$
		Time	0.0148	0.0057	.009	0.074	$R^2_{\text{NSJ}(c)} = 0.900$
		Group × Time	0.0011	0.0069	.872	<0.001	
	Reduced	Intercept	5.5432	0.0838	<.001	–	$R^2_{\beta^*} = 0.141$
		Group	−0.0626	0.0992	.528	<0.001	$R^2_{\text{NSJ}(m)} = 0.066$
		Time	0.0156	0.0033	<.001	0.212	$R^2_{\text{NSJ}(c)} = 0.899$

Notes: Interpretation of covariance specifications:
(1): Random intercepts with $\boldsymbol{W}_i = \sigma^2 \boldsymbol{I}_{m_i}$.
(2): Random intercepts and random slopes with $\boldsymbol{W}_i = \sigma^2 \boldsymbol{I}_{m_i}$.

slopes reduce unexplained subject-specific variation over time. The addition of these random effects noticeably increases the effect of time at the population level, and this shift in variance can be gauged by the increased model and semi-partial $R^2_{\beta^*}$ estimates.

As noted in Section 6.3, $R^2_{\beta^*}$ may sometimes *decrease* with the addition of a fixed effect under an incorrectly specified covariance. On the other hand, the statistic correctly increases with the addition of a fixed effect under a correctly specified covariance. Table 4 shows the same pattern as the simulated longitudinal data in Section 6.3. Thus, we select the full model under covariance 2 for this data. The $R^2_{\text{NSJ}(m)}$ estimate is 0.066 for both the full model and the reduced model under covariance 2, so one may conclude that $R^2_{\text{NSJ}(m)}$

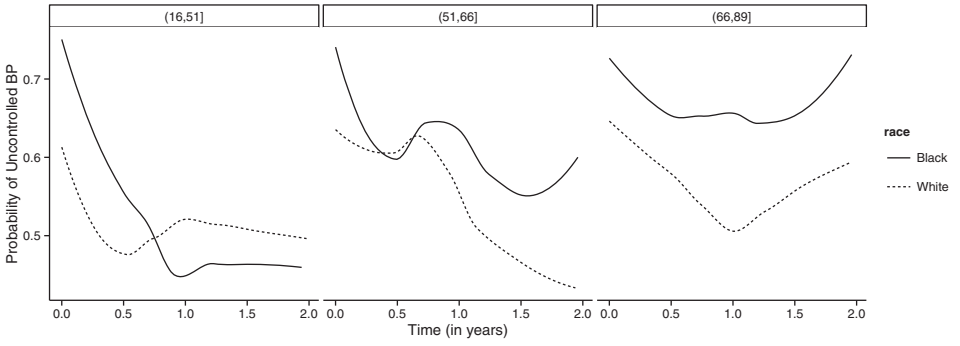

Figure 3. Loess curves of probability of uncontrolled blood pressure. Panels in the plot represent subgroups for age.

selects the reduced model for a more parsimonious fit. This example demonstrates the utility of semi-partial $R^2_{\beta*}$ statistics with real longitudinal data that has some similarity to the simulated data in Section 6.3 and agrees with conclusions drawn therein.

6.5. Longitudinal blood pressure study with binary outcome

Data from a retrospective longitudinal cohort blood pressure study [12] of $N = 459$ adults with hypertension illustrates application to a real-world problem and the utility of our statistic for multiple logistic regression in the GLMM. Longitudinal blood pressure (BP) levels were taken on patients making at least four visits to the Family Practice Center at UNC during a two-year period, 1999–2001. We constructed a repeated dichotomous outcome, controlled or uncontrolled BP, as a composite measure from separate systolic and diastolic BP. Using the JNC 7 classification of BP [6], BP is considered controlled if systolic BP is less than 140 mmHg and diastolic BP is less than 90 mmHg. We created a binary outcome that indicates whether a person's BP was controlled or uncontrolled at the time of measurement. Figure 3 displays loess curves for the binary data. The estimated curves suggest older subjects have a higher likelihood of uncontrolled blood pressure.

Fixed effects for the analysis include Age, Race, Sex, Continuity of Care (COC), and continuous linear time in years. We fit a full model and reduced model under the two covariance structures introduced in Section 6.2. Table 5 shows $R^2_{\mathrm{NSJ}(m)}$, $R^2_{\mathrm{NSJ}(c)}$, and $R^2_{\beta*}$ with semi-partial statistics. Both $R^2_{\mathrm{NSJ}(m)}$ and $R^2_{\beta*}$ prefer the full model using covariance structure 1. Partial $R^2_{\beta*}$ statistics favor race, age, and time, indicating that these variables are most strongly associated with the outcome when accounting for all other covariates. This example demonstrates application of our proposed R^2 statistic to binary data, where PQL is known to be biased.

7. Discussion

In a review of R^2 statistics for the LMM, Lahuis *et al.* [22] discuss several major issues that researchers may encounter when reporting explained variance in mixed models. The concerns addressed by the authors are a lack of consensus regarding the measure to report, how to deal with random slopes, and how to compute and interpret the change in explained

Table 5. Estimates for R^2 and semi-partial R^2 with blood pressure data.

						R^2 Estimates	
Covariance	Model	Fixed effect	Estimate	SE	*p*-value	Semi-partial	Model
1	Full	Intercept	0.481	0.322	.135	–	$R^2_{\beta^*} = 0.089$
		Age	0.012	0.005	.017	0.015	$R^2_{\text{NSJ}(m)} = 0.024$
		Race	−0.130	0.150	.385	0.015	$R^2_{\text{NSJ}(c)} = 0.346$
		Sex	−0.260	0.260	.318	0.002	
		COC	−0.401	0.065	<.001	0.004	
		Time	−0.393	0.147	.008	0.009	
	Reduced	Intercept	0.318	0.300	.288	–	$R^2_{\beta^*} = 0.071$
		Age	0.012	0.005	.018	0.015	$R^2_{\text{NSJ}(m)} = 0.022$
		Race	−0.398	0.065	<.001	0.015	$R^2_{\text{NSJ}(c)} = 0.346$
		Time	−0.420	0.146	.004	0.009	
2	Full	Intercept	0.446	0.326	.171	–	$R^2_{\beta^*} = 0.080$
		Age	0.012	0.005	.016	0.016	$R^2_{\text{NSJ}(m)} = 0.023$
		Race	−0.108	0.152	.476	0.018	$R^2_{\text{NSJ}(c)} = 0.368$
		Sex	−0.222	0.266	.405	0.001	
		COC	−0.403	0.076	<.001	0.002	
		Time	−0.410	0.149	.006	0.070	
	Reduced	Intercept	0.309	0.302	.306	–	$R^2_{\beta^*} = 0.074$
		Age	0.012	0.005	.017	0.015	$R^2_{\text{NSJ}(m)} = 0.022$
		Race	−0.398	0.076	<.001	0.021	$R^2_{\text{NSJ}(c)} = 0.369$
		Time	−0.4330.148	0.003	.069		

Notes: Interpretation of covariance specifications:
(1): Random intercepts with $\boldsymbol{W}_i = \sigma^2 \boldsymbol{I}_{m_i}$.
(2): Random intercepts and random slopes with $\boldsymbol{W}_i = \sigma^2 \boldsymbol{I}_{m_i}$.

variance between two nested models. These concerns underpin the motivation for this discussion. As we have shown in the preceding examples, $R^2_{\beta^*}$ fully addresses these points. First, there is no confusion regarding which measure to report when using $R^2_{\beta^*}$ because we summarize goodness-of-fit for the proposed fixed effects with a single measure. Simulations have demonstrated that this measure consistently selects the correct model with more efficacy than competing statistics.

In response to the second concern of Lahuis *et al.* [22], Nakagawa and Schielzeth agree with the claim made by Snijders and Bosker, who state that explained variance does not vary based on inclusion of random slopes. However, Gurka *et al.* [16] show that falsely assuming a random intercept model in an LMM creates biased tests and confidence intervals for fixed effects, even in asymptotically large samples. Thus, we believe incorporation of random slopes is desirable and easily accomplished using $R^2_{\beta^*}$.

With regard to the third concern, the Wald test for fixed effects (and hence $R^2_{\beta^*}$) compares two models with nested fixed effects. When computing and interpreting $R^2_{\beta^*}$ for two nested models, we calculate the model $R^2_{\beta^*}$ for each model under an identical covariance structure. The difference between $R^2_{\beta^*}$ measures can then be interpreted as the difference in explained variance between the two models. This type of comparison may be complicated by the possible decrease in $R^2_{\beta^*}$ with the addition of fixed effects (e.g. a negative difference in explained variance). Characterizing the nature of this occurrence is an aim of future research (see below).

To summarize the principle differences between the R^2 statistics contained in this article, $R^2_{\beta*}$ estimates the variance explained by fixed effects in the context of random effects, which induce clustering on the observed outcomes. The marginal R^2 statistic proposed by Nakagawa and Schielzeth estimates explained variance under the assumptions that the covariance between observations may be ignored and the variation arising from random effects may be treated as random error. On the other hand, their conditional statistic treats variation from random effects as explained variance, and this strategy leads to incorrect population model selection. The limited results from this article indicate that $R^2_{\beta*}$ is the superior tool for population model selection in the GLMM, particularly when the covariance model is correct.

7.1. Limitations and future research

R^2_β has received criticism for the potential to decrease with the addition of fixed effects. Since $R^2_{\beta*}$ is a direct extension of R^2_β, the same potential to decrease exists. However, further research may show how to judiciously use this characteristic as a tool for covariance model selection. In the simulation studies from Section 6.3, the decrease occurs only in the context of an under-specified covariance model. No decrease in R^2 was observed under the correct covariance specification in either simulation. Thus, negative values of explained variance between nested models may be indicative of a mis-specified covariance structure. Characterizing and understanding this behavior is an interesting topic with a great deal of potential utility. Incorrect specification of the covariance is known to cause biased estimation of standard errors for fixed effects [13, chap. 7] as well as biased estimates for random effects [11,21]. Additionally, under-specification of the covariance model inflates type I error for tests of fixed effects [16]. Each of these findings present interesting potential sources of explaining the cause of $R^2_{\beta*}$ decreasing with the addition of fixed effects in a mis-specified covariance model. Although this possible decrease is not a generally desired property of R^2 statistics, it may be an invaluable tool for covariance selection. Matuszewski and Edwards also found that R^2_β, when calculated using residual degrees of freedom, was more robust to covariance mis-specification than using the Kenward–Roger approach. Future research is warranted to explore whether this approximation technique may be used in the GLMM to build parsimonious population models even when the covariance model is incorrect.

8. Conclusion

Here, we have proposed a new R^2 statistic, $R^2_{\beta*}$, for assessing fixed effects in the GLMM based on PQL estimation, the most widely implemented estimation technique for the GLMM. We interpret $R^2_{\beta*}$ as a measure of multivariate association between the outcome and the fixed effects in the GLMM. The proposed statistic possesses several desirable qualities: first, it is a truly generalized definition of R^2 from the linear univariate model, and also generalizes the marginal R^2 statistic proposed by Nakagawa and Schielzeth; second, it offers semi-partial R^2 statistics to assess the relative importance of individual fixed effects; and third, the estimated covariance matrix, $\hat{\mathbf{\Sigma}}$, is fully incorporated in the formulation of $R^2_{\beta*}$. We have used real and simulated longitudinal data to demonstrate application of the statistic for assessing goodness-of-fit and conducting model selection. The benefit of semi-partial $R^2_{\beta*}$ is evident in both of the real data sets considered. In summary, we believe

that $R^2_{\beta^*}$ is the most sensible definition of R^2 for GLMMs, and that the proposed R^2 statistic can have a dramatic impact on the applied practice of clustered data analysis.

Acknowledgments

Lloyd Edwards and Pranab Sen were supported by the National Center for Research Resources and the National Center for Advancing Translational Sciences, National Institutes of Health, through Grant Award Number UL1TR000083; Kalyan Das was supported by the Pranab K. Sen Distinguished Visiting Professorship, January 1–June 30, 2013.

Disclosure statement

No potential conflict of interest was reported by the author(s).

References

[1] N. Ballarini and B. Jaeger, *A macro to compute r^2 for the glmm in sas*. Available at https://github.com/nicoballarini/SAS-Macros.
[2] B.B.D. Bates, M. Maechler, and W. Steven, 2014, http://CRAN.R-project.org/package = lme4, R package version 1.1-7.
[3] B.M. Bolker, M.E. Brooks, C.J. Clark, S.W. Geange, J.R. Poulsen, M.H.H. Stevens, and J.S.S. White, *Generalized linear mixed models: A practical guide for ecology and evolution*, Trends Ecol. Evolut. 24 (2009), pp. 127–135.
[4] N.E. Breslow and D.G. Clayton, *Approximate Inference in Generalized Linear Mixed Models*, J. Amer. Statist. Assoc. 88 (1993), pp. 9–25, http://dx.doi.org/10.2307/2290687.
[5] M. Casals, M. Girabent-Farrs, and J.L. Carrasco, *Methodological quality and reporting of generalized linear mixed models in clinical medicine (2000–2012): A systematic review*, PLoS ONE 9 (2014), p. e112653.
[6] A.V. Chobanian, G.L. Bakris, H.R. Black, W.C. Cushman, L.A. Green, J.L. Izzo Jr, D.W. Jones, B.J. Materson, S. Oparil, and J.T. Wright Jr, et al., *The seventh report of the joint national committee on prevention, detection, evaluation, and treatment of high blood pressure: The jnc 7 report*, Jama 289 (2003), pp. 2560–2571.
[7] R. Christensen, *Plane answers to Complex Questions: The Theory of Linear Models*, Springer Science & Business Media, New York, NY, 2011.
[8] W.S. Cleveland, *Lowess: A program for smoothing scatterplots by robust locally weighted regression*, Amer. Statist. 35 (1981), pp. 54–54, http://www.jstor.org/stable/2683591.
[9] C.B. Dean and J.D. Nielsen, *Generalized linear mixed models: A review and some extensions*, Lifetime Data Anal. 13 (2007), pp. 497–512.
[10] L.J. Edwards, K.E. Muller, R.D. Wolfinger, B.F. Qaqish, and O. Schabenberger, *An R^2 statistic for fixed effects in the linear mixed model*, Stat. Med. 27 (2008), pp. 6137–6157, http://dx.doi.org/10.1002/sim.3429.
[11] J. Ferron, K.Y. Hogarty, M. Hess, J. Romano, J.D. Kromrey, J.D. Niles, G. Dawkins and C. Sentovich, *A macro for examining the consequences of error structure mis-specifications*, Proceedings of the Twenty-Seventh Annual SAS Users Group International Conference (2002), Orlando, Florida, pp. 201–227.
[12] M. Fisher, P. Sloane, L. Edwards and G. Gamble, *Continuity of care and hypertension control in a university-based practice*, Ethnicity Disease 17 (2007), pp. 693–698.
[13] G.M. Fitzmaurice, N.M. Laird and J.H. Ware, *Applied Longitudinal Analysis, Vol. 998*, John Wiley & Sons, Boston, MA, 2012.
[14] A. Gelman and J. Hill, *Data Analysis Using Regression and Multilevel/Hierarchical Models*, Cambridge University Press, New York, NY, 2006.
[15] A. Gelman and I. Pardoe, *Bayesian measures of explained variance and pooling in multilevel (hierarchical) models*, Technometrics 48 (2006), pp. 241–251.

[16] M.J. Gurka, L.J. Edwards and K.E. Muller, *Avoiding bias in mixed model inference for fixed effects*, Stat. Med. 30 (2011), Available at http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3396027/.
[17] U. Halekoh and S. Højsgaard, *A Kenward–Roger approximation and parametric bootstrap methods for tests in linear mixed models–the R package pbkrtest*, J. Statist. Softw. 59 (2014), pp. 1–30, http://www.jstatsoft.org/v59/i09/.
[18] B. Jaeger, *Software in SAS and R to compute R^2 for fixed effects in the linear and generalized linear mixed model*. Available at https://github.com/bcjaeger/R2FixedEffectsGLMM/.
[19] B. Jaeger, *r2glmm: Computes R squared for mixed (multilevel) models (LMMs and GLMMs)*, (2016). R package version 0.1.0. Available at https://github.com/bcjaeger/r2glmm.
[20] P.C. Johnson, *Extension of nakagawa & schielzeth's r2glmm to random slopes models*, Methods Ecol. Evolut. 5 (2014), pp. 944–946, http://dx.doi.org/10.1111/2041-210X.12225.
[21] O.M. Kwok, S.G. West and S.B. Green, *The impact of misspecifying the within-subject covariance structure in multiwave longitudinal multilevel models: A monte carlo study*, Multivariate Behav. Res. 42 (2007), pp. 557–592.
[22] D.M. LaHuis, M.J. Hartman, S. Hakoyama and P.C. Clark, *Explained variance measures for multilevel models*, Organizational Res. Methods 17 (2014), pp. 433–451.
[23] Y. Lee and J.A. Nelder, *Hierarchical generalized linear models*, J. R. Statist. Soc. Ser. B (Methodol.) (1996), pp. 619–678.
[24] J. Matuszewski and L.J. Edwards, *Properties of an r2 statistic for fixed effects in the linear mixed model for longitudinal data*, 2013. Available at https://cdr.lib.unc.edu/record/uuid:053ce978-e62b-4021-bd67-d 1c5db669aad.
[25] S. Nakagawa and H. Schielzeth, *A general and simple method for obtaining r2 from generalized linear mixed-effects models*, Methods Ecol. Evolut. 4 (2013), pp. 133–142, http://dx.doi.org/10.1111/j.2041-210x.2012.00261.x.
[26] J.G. Orelien and L.J. Edwards, *Fixed-effect variable selection in linear mixed models using R^2 statistics*, Comput. Stat. Data Anal. 52 (2008), pp. 1896–1907.
[27] R.F. Potthoff and S. Roy, *A generalized multivariate analysis of variance model useful especially for growth curve problems*, Biometrika (1964), pp. 313–326.
[28] R Core Team, *R Foundation for Statistical Computing*, Vienna, Austria, 2015. Available at http://www.R-project.org/.
[29] S. Raudenbush and A.S. Bryk, *A hierarchical model for studying school effects*, Sociol. Ed. (1986), pp. 1–17.
[30] SAS Institute Inc., Cary, NC, 2003. Available at http://www.sas.com/.
[31] R. Schall, *Estimation in generalize linear models with random effects*, Biometrika 78 (1991), pp. 719–727.
[32] H. Singmann, B. Bolker and J. Westfall, 2015, http://CRAN.R-project.org/package = afex, R package version 0.13-145.
[33] T. Snijders and R. Bosker, *Multilevel Analysis: An Introduction to Basic and Advanced Multilevel Modeling*, SAGE Publications, 2011, Available at http://books.google.com/books?id = N1BQvcomDdQC.
[34] T.A.B. Snijders and R.J. Bosker, *Modeled variance in two-level models*, Sociol. Methods Res. 22 (1994), pp. 342–363.
[35] W.W. Stroup, *Generalized Linear Mixed Models: Modern Concepts, Methods and Applications*, CRC Press, Sheffield, UK, 2012.
[36] F. Tuerlinckx, F. Rijmen, G. Verbeke, and P. De Boeck, *Statistical inference in generalized linear mixed models: A review*, British J. Math. Statist. Psychol. 59 (2006), pp. 225–255, http://dx.doi.org/10.1348/000711005X79857.
[37] A.B. Tom, Tom AB Snijders, Roel J. Bosker and Roel J. Bosker. Multilevel analysis: An introduction to basic and advanced multilevel modeling. Sage, 1999.
[38] M.F. Wellons, L. Sanders, L.J. Edwards, J.A. Bartlett, A.E. Heald, and K.E. Schmader, *Hiv infection: treatment outcomes in older and younger adults*, J. Amer. Geriatrics Soc. 50 (2002), pp. 603–607.
[39] H. Wickham and R. Francois, 2015, http://CRAN.R-project.org/package = dplyr, R package version 0.4.1.

[40] R. Wolfinger, *Laplace's approximation for nonlinear mixed models*, Biometrika 80 (1993), pp. 791–795, http://www.jstor.org/stable/2336870.
[41] R. Xu, *Measuring explained variation in linear mixed effects models*, Stat. Med. 22 (2003), pp. 3527–3541.

Computation of $R^2_{\beta^*}$ in SAS and R

Proc Glimmix in SAS [30] fits GLMMs using PQL as the default estimation technique. The procedure is based on a more general macro, %Glimmix, which applies the PQL algorithm by performing repeated calls to Proc Mixed. Our procedure to calculate $R^2_{\beta^*}$ in SAS utilizes this macro, saving output parameters from a full model contrast statement upon the convergent iteration of the PQL algorithm. The approximate F-statistic and denominator degrees of freedom from the pseudo linear mixed model are then used to compute $R^2_{\beta^*}$. The method outlined here is implemented in a SAS macro available in [18].

The r2glmm package computes $R^2_{\beta^*}$ for the LMM using the Kenward-Roger approach as well as the method proposed by Nakagawa and Schielzeth [17]. In addition to semi-partial R squared values, the package supplies confidence limits. As a note of caution, the authors of the `pbkrtest` package cannot guarantee that the results agree with other implementations of the Kenward–Roger approach. Therefore, it is highly encouraged to compare $R^2_{\beta^*}$ values calculated with R to corresponding calculations in SAS when performing inference. We recommend favoring results from SAS.