Published in final edited form as:

Commun Stat Theory Methods. 2012 January 1; 41(6): 953–969. doi:10.1080/03610926.2010.535631.

Random effects coefficient of determination for mixed and meta-analysis models

Eugene Demidenko^{1,*}, James Sargent², and Tracy Onega^{1,2}

¹Dartmouth Medical School, Hanover, NH 03755

²Norris Cotton Cancer Center, Lebanon, NH 03756

Abstract

The key feature of a mixed model is the presence of random effects. We have developed a coefficient, called the random effects coefficient of determination, R_r^2 , that estimates the proportion of the conditional variance of the dependent variable explained by random effects. This coefficient takes values from 0 to 1 and indicates how strong the random effects are. The difference from the earlier suggested fixed effects coefficient of determination is emphasized. If R_r^2 is close to 0, there is weak support for random effects in the model because the reduction of the variance of the dependent variable due to random effects is small; consequently, random effects may be ignored and the model simplifies to standard linear regression. The value of R_r^2 apart from 0 indicates the evidence of the variance reduction in support of the mixed model. If random effects coefficient of determination is close to 1 the variance of random effects is very large and random effects turn into free fixed effects—the model can be estimated using the dummy variable approach. We derive explicit formulas for R_r^2 in three special cases: the random intercept model, the growth curve model, and meta-analysis model. Theoretical results are illustrated with three mixed model examples: (1) travel time to the nearest cancer center for women with breast cancer in the U.S., (2) cumulative time watching alcohol related scenes in movies among young U.S. teens, as a risk factor for early drinking onset, and (3) the classic example of the meta-analysis model for combination of 13 studies on tuberculosis vaccine.

Keywords

clustered data; growth curve; F -test; longitudinal data; random effect; random intercept

1 Introduction and motivation

The linear mixed effects model has become a major tool for the analysis of clustered/ correlated data. Several books have been published recently on the topic including Vonesh and Chin-chilli (1997), Brown and Prescott (1999), Demidenko (2004), Jiang (2007), and McCulloch et al. (2008), among others. In many applications, the mixed model is applied to the analysis of longitudinal data, e.g. Verbeke and Molenberghs (2000) and Diggle et al. (2002). Several books focus exclusively on software for mixed models, e.g. Pinheiro and Bates (2000) and Littell et al. (2006).

Despite a plethora of publications on this topic, a fundamental question remains difficult to answer, *Why fit a mixed model, or equivalently, why random effects?* Although employing

*Address for correspondence to Eugene Demidenko, Section of Biostatistics and Epidemiology, Dartmouth Medical School, Hanover, NH 03755, USA; eugened@dartmouth.edu.

random effects models has become trendy, these models have disadvantages and may add unnecessary complexity to the analysis. It would be very desirable to have a measure/coefficient by which one could judge whether fitting a mixed model was appropriate versus a traditional linear regression model. If that coefficient was close to zero, one might say that the data may be analyzed by standard linear regression because there are no random effects, and when that coefficient departs from 0, the presence of random effects is strong and mixed model would be valid. The goal of the present paper is to introduce such a coefficient, which will be called the random effects coefficient of determination. This coefficient reflects the reduction of the variance of the dependent variable due to the presence of random effects.

Several other measures of explained variance in the framework of linear mixed model have been developed previously in the literature including Snijders and Bosker (1994), Vonesh and Chinchilli (1997), Xu (2003), Gelman and Pardoe (2006). Those measures resemble the coefficient of determination as they reflect the goodness-of-fit with fixed effects, as emphasized in a recent paper by Edwards et al. (2008). Specifically, the *fixed* effects coefficient of determination in the framework of linear mixed effects model $\mathbf{y}_i = \mathbf{X}_i\boldsymbol{\beta} + \mathbf{Z}_i\mathbf{b}_i + \boldsymbol{\varepsilon}_i$ is a straightforward generalization of the classical coefficient of determination,

$$R_f^2 = 1 - \frac{\sum_{i=1}^N (\mathbf{y}_i - \mathbf{X}_i\hat{\boldsymbol{\beta}})' \hat{\mathbf{V}}_i^{-1} (\mathbf{y}_i - \mathbf{X}_i\hat{\boldsymbol{\beta}})}{\sum_{i=1}^N (\mathbf{y}_i - \bar{y}_i \mathbf{1}_i)' \hat{\mathbf{V}}_i^{-1} (\mathbf{y}_i - \bar{y}_i \mathbf{1}_i)}, \quad (1)$$

where N is the number of clusters, $\hat{\boldsymbol{\beta}}$ is the maximum likelihood estimator of fixed effects, and $\hat{\mathbf{V}}_i$ is an estimate of the intra cluster covariance matrix (it is assumed that fixed effects contain an intercept term), $\bar{y}_i$ is the average value of the dependent variable in cluster i , and $\mathbf{1}_i$ is the vector-column of ones. Similarly to the classical coefficient of determination, we interpret R_f^2 as the proportion of the variance of the dependent variable, $\mathbf{y}$ explained by predictors $\mathbf{X}$.

The goal of this paper is to develop a similar figure of merit for random effects, the *random* effects coefficient of determination, R_r^2 . Such coefficient should take larger values for larger values of the variance of the random effect but on a fixed scale, so that comparisons can be made across mixed models. While (1) shows the strength of fixed effects, R_r^2 characterizes the strength of random effects in a linear mixed model. The measures R_f^2 and R_r^2 reflect the strength of a linear mixed effects from different perspective and we believe they should be a part of the software output every time a linear mixed effects model is estimated.

The organization of the paper is as follows. In the next section we introduce the notation and lay out basic facts about linear mixed model. In Section 3, we introduce the F -test for the presence of random effects, which will be used in the next section to define the random effects coefficient of determination. We derive explicit formulae for three special cases of mixed model: random intercept, growth curve, and meta-analysis models. Finally, in Section 5, we illustrate the computation of R_r^2 with three examples: In the first example, we use mixed model with random intercepts that represent the state-specific travel time of breast cancer patients to cancer centers in the U.S. In the second example, we demonstrate the validity of the growth curve model applied to the longitudinal data on alcohol movie watching among young adolescents and its association with drinking initiation. In the third example, we show the application of R_r^2 to the meta-analysis model using the classic data set on 13 studies with the overall effect of a tuberculosis vaccine.

2 Preliminaries

In this paper, we use the linear mixed effects (LME) model

$$y_i = X_i \beta + Z_i b_i + \varepsilon_i, \quad (2)$$

where i indicates the cluster (subject or unit); y_i is the $n_i \times 1$ vector of observations of the dependent variable; X_i is the $n_i \times m$ matrix of fixed effects variables/covariates; β is the $m \times 1$ vector of unknown fixed effects coefficients; Z_i is the $n_i \times k$ design matrix of random effects; b_i is the $k \times 1$ vector of random effects with zero mean and unknown covariance matrix $\sigma^2 \mathbf{D}$; ε_i is the error term with zero mean and covariance matrix $\sigma^2 \mathbf{I}$, $i = 1, 2, \dots, N$ (number of clusters). We assume that the observations between clusters do not correlate, but do correlate within clusters due to the presence of random effects. Also, it is assumed that random vectors are normally distributed, $b_i \sim N(\mathbf{0}, \sigma^2 \mathbf{D})$ and $\varepsilon_i \sim N(\mathbf{0}, \sigma^2 \mathbf{I})$. To avoid

degeneration, we shall assume that matrix $\sum_{i=1}^N X_i' X_i$ is nonsingular (boldface is used for vectors and matrices).

If matrix $\mathbf{D}$ is known, the generalized linear least squares (GLS) estimator of β takes the form

$$\hat{\beta}_{GLS} = \left[\sum_{i=1}^N X_i' (I + Z_i \mathbf{D} Z_i')^{-1} X_i \right]^{-1} \left[\sum_{i=1}^N X_i' (I + Z_i \mathbf{D} Z_i')^{-1} y_i \right], \quad (3)$$

where $V_i = I + Z_i \mathbf{D} Z_i'$ is the intra cluster covariance matrix. If random effects are not present (the variances of random effects are zero), the LME model (2) reduces to a linear regression

$$y_i = X_i \beta + \varepsilon_i. \quad (4)$$

Then, since errors do not correlate, we can combine data over i and apply ordinary least squares, which produces an efficient estimator of the fixed effects coefficients.

Alternatively, we can derive this estimator by letting $\mathbf{D} = \mathbf{0}$ in (3), $\hat{\beta}_{LS}$, which may be called the zero random effects estimator. The least squares (LS) estimator yields the residual sum of squares

$$S_{LS} = \sum_{i=1}^N \|y_i - X_i \hat{\beta}_{LS}\|^2. \quad (5)$$

A more interesting extreme situation is when the variances of random effects become large. Specifically, we are interested in the limit of $\hat{\beta}_{GLS}$ when matrix $\mathbf{D}$ goes to infinity. That is, we set $\mathbf{D} = d\mathbf{I}$, where $\mathbf{I}$ is the $k \times k$ identity matrix, and let $d \rightarrow \infty$. Omitting some algebra, we arrive at the estimator

$$\hat{\beta}_{\infty} = \left[\sum_{i=1}^N X_i' (I - Z_i Z_i') X_i \right]^+ \left[\sum_{i=1}^N X_i' (I - Z_i Z_i') y_i \right], \quad (6)$$

where $^+$ means the Moore-Penrose generalized matrix inverse. Note that it is necessary to use the generalized matrix inverse operation because matrix $\mathbf{Z}_i$ may be not of full rank. We call this estimator the *fixed effects* estimator because (6) can be viewed as the least squares estimate from the linear regression

$$y_i = X_i\beta + Z_i b_i + \varepsilon_i, \quad (7)$$

treating $\{b_i\}$ as fixed and unknown, similarly to β , with the minimal sum of squares

$$S_{\min} = \min_{\beta, b_1, \dots, b_N} \sum_{i=1}^N \|y_i - X_i\beta - Z_i b_i\|^2. \quad (8)$$

Clearly, $S_{\min} = S_{LS}$. Practically, we obtain this residual sum of squares by running linear regression

$$y = W\gamma + \varepsilon$$

with combined data

$$y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_N \end{bmatrix}, W = \begin{bmatrix} X_1 & Z_1 & \mathbf{0} & \cdots & \mathbf{0} \\ X_2 & \mathbf{0} & Z_2 & \mathbf{0} & \mathbf{0} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ X_N & \cdots & \mathbf{0} & \mathbf{0} & Z_N \end{bmatrix}, \gamma = \begin{bmatrix} \beta \\ b_1 \\ \vdots \\ b_N \end{bmatrix}, \quad (9)$$

assuming that

$$\text{rank}(W) > m. \quad (10)$$

This is the reason we call (6) the fixed effects estimator. Alternatively, one can compute $S_{\min}$ as the residual sum of squares in the OLS regression after regressing the residuals from regression of y_i and X_i on Z_i in each cluster on X .

For finite matrix $\mathbf{D}$, the GLS for β and random effects, $\{b_i, i = 1, 2, \dots, N\}$ are found from the penalized least squares

$$\min_{\beta, b_1, \dots, b_N} \sum_{i=1}^N \left[\|y_i - X_i\beta - Z_i b_i\|^2 + b_i' D^{-1} b_i \right].$$

Some algebra shows that the solution to this optimization problem is (6) and the estimate of the i th random effect is

$$\widehat{b}_i = (Z_i' Z_i + D^{-1})^{-1} Z_i' r_i, \quad (11)$$

where $r_i = y_i - X_i \hat{\beta}_{GLS}$ is the residual vector for the i th cluster.

It is easy to see that we come to the same limit sum of squares (8) assuming that $\mathbf{D} = d\mathbf{D}_0$, where $\mathbf{D}_0$ is any positive definite fixed matrix. Indeed, we can express $\mathbf{D}_0 = \mathbf{T}\mathbf{T}'$, and letting $\mathbf{U}_j = \mathbf{Z}_j\mathbf{T}$, reduce the problem to the previous case with $\mathbf{D} = d\mathbf{I}$. Since $\mathbf{U}_j\mathbf{b}_j = \mathbf{Z}_j(\mathbf{T}\mathbf{b}_j)$ and $\mathbf{b}_j$ is unrestricted, the minimal sum of squares remains the same (8) as in the case $\mathbf{D}_0 = \mathbf{I}$.

Now we are ready to introduce the F -test which is the basis of our random effects coefficient of determination. While the fixed effects coefficient of determination (1) is a straightforward generalization of the classical regression model, the random effects coefficient of determination requires more insight into the nature of the mixed model.

3 The F -test for the presence of random effects

In order to prove that the LME model does not reduce to standard linear regression, we need to test the hypothesis

$$H_0: \mathbf{D} = \mathbf{0}. \quad (12)$$

The justification of our test below is as follow: The two sums of squares (5) and (8) represent two extreme case with $\mathbf{D} = \mathbf{0}$ and $\mathbf{D} = \infty$. If $\mathbf{D} = \mathbf{0}$, it is plausible to expect that the latter sum of squares is much smaller than the former one. In order to construct the test for (12), we recall the F -test for linear regression when testing the hypothesis on a linear combination of parameters $H_0: \mathbf{C}\boldsymbol{\beta} = \mathbf{0}$,

$$F = \frac{(S_0 - S_{\min})/q}{S_{\min}/(n - m)}, \quad (13)$$

where S_0 and $S_{\min}$ are residual sums of squares under restrictions and with no restriction, q is the number of restrictions, n is the number of observations, and m is the dimension of vector $\boldsymbol{\beta}$. In effect, we apply the same F -test to regression (7) to test whether the coefficients $\{\mathbf{b}_i, i = 1, 2, \dots, N\}$ are zero. Specifically, the following fact holds (Demidenko, 2004, p. 137).

Theorem 1

For linear mixed model (2) under the null hypothesis (12), we have

$$\frac{(S_{LS} - S_{\min})/(r - m)}{S_{\min}/(N_T - r)} \sim F(r - m, N_T - r), \quad (14)$$

where $r = \text{rank}(\mathbf{W})$ and $N_T = \sum_{i=1}^N n_i$.

This test, applied to single random effect variance, has deep roots in ANOVA and the variance components literature: Seely and El-Bassiouni (1983), Searle et al. (1992), Mathew and Sinha (1992), Khuri et al. (1998), starting from a pioneering work of Wald (1947). A comprehensive comparison study is found in a recent paper Scheipl et al. (2008). Note that (a) test (14) is exact, it yields the exact significance level, and (b) it applies to multivariate random effects. To the best of our knowledge, this test is not a part of existing software on mixed models, including SAS and S-Plus. This test is the basis for our random effects coefficient of determination.

4 Random effects coefficient of determination

As follows from the previous section, S_{LS} and $S_{\min}$ are the two extreme values of the residual sum of squares (8); the former corresponds to when random effects, are zero ($\mathbf{b}_i = \mathbf{0}$), and the latter corresponds to when there is no restriction on random effects, or in other words, random effects are fixed effects, $S_{\min} = S_{LS}$. When the mixed model is estimated we obtain estimates of random effects, $\hat{\mathbf{b}}_i$ with the sum of squares

$$S_{MM} = \sum_{i=1}^N \|y_i - \mathbf{X}_i \hat{\boldsymbol{\beta}}_{GLS} - \mathbf{Z}_i \hat{\mathbf{b}}_i\|^2.$$

We use subscript MM to indicate that the sum of squares is computed from the mixed model. Apparently, $S_{\min}$ and S_{LS} are the lower and upper bounds of sum of squares in the estimated mixed model, $S_{\min} \leq S_{MM} \leq S_{LS}$.

We associate these three residual sum of squares with conditional variance similarly to regression analysis. Apparently, S_{MM}/N_T can be viewed as the variance of y conditional on the estimated random effects. Then S_{LS}/N_T is the highest and $S_{\min}/N_T$ is the lowest variance of y with $(S_{MM} - S_{\min})/N_T$ as the reduction of the conditional variance due to the presence of the random effects. Finally, the random effects coefficient of determination for mixed model is defined as the reduction of the conditional variance of y with respect to the maximum reduction, $(S_{LS} - S_{\min})/N_T$, namely,

$$R_r^2 = 1 - \frac{S_{MM} - S_{\min}}{S_{LS} - S_{\min}}, \quad (15)$$

or alternatively,

$$R_r^2 = \frac{S_{LS} - S_{MM}}{S_{LS} - S_{\min}}. \quad (16)$$

This coefficient lies in the interval from 0 to 1 under assumption (10), see Appendix for a rigorous proof. If $R_r^2 = 1$ then $S_{MM} = S_{\min}$ so that random and fixed effects coincide ($\mathbf{D} = \infty$). If $R_r^2 = 0$ then $S_{LS} = S_{MM}$ and random effects are not present (matrix $\mathbf{D} = \mathbf{0}$). We interpret R_r^2 as the proportion of the reduction of the conditional variance of the dependent variable on the scale of maximum reduction due to random effects.

Below, we illustrate the computation of the random effects coefficient of determination with three popular special cases of the LME model and provide closed-form solutions.

4.1 Random intercept model

This model is a special case of the linear mixed model with one random effect: It is assumed that the intercepts are random and deviations of the cluster-specific intercepts from the common intercept are independent and identically distributed with zero mean and variance $\sigma^2 d$, where d is the relative variance of the random intercept. Mathematically, the random intercept model is expressed as a system of linear regression with a cluster-specific intercept,

$$y_i = X_i \beta + b_i \mathbf{1}_i + \varepsilon_i,$$

where $b_i \sim N(0, \sigma^2 d)$, assuming that all elements in the first column of matrix X_i are 1 (the common intercept is the first element of vector β), and $\mathbf{1}_i$ is the $n_i \times 1$ vector of ones, $i = 1, 2, \dots, N$. Omitting some algebra, we derive

$$S_{\min} = \sum_{i=1}^N \|\tilde{y}_i - \tilde{X}_i \tilde{\beta}\|^2, \quad (17)$$

where ‘tilde’ means deviation from the mean (centerized data), $\tilde{y}_i = y_i - \mathbf{1}_i \bar{y}_i$ and $\tilde{X}_i = X_i - \mathbf{1}_i \bar{x}_i'$ and

$$\tilde{\beta} = \left(\sum_{i=1}^N \tilde{X}_i' \tilde{X}_i \right)^{-1} \left(\sum_{i=1}^N \tilde{X}_i' \tilde{y}_i \right). \quad (18)$$

Note that $\tilde{X}_i$ is a $n_i \times (m-1)$ matrix, so only slope coefficients are computed in (18) because intercepts are eliminated. Alternatively, $S_{\min}$ could be computed using standard linear regression with dummy intercepts. Finally, for the random intercept model,

$$R_r^2 = \frac{\sum_{i=1}^N \|y_i - X_i \hat{\beta}\|^2 - S_{MM}}{\sum_{i=1}^N \|y_i - X_i \hat{\beta}\|^2 - \sum_{i=1}^N \|\tilde{y}_i - \tilde{X}_i \tilde{\beta}\|^2},$$

where S_{MM} is computed after the mixed model is estimated by restricted or standard maximum likelihood (procedure mixed or lme in S-Plus/R). It is easy to see that the rank of matrix W for the random intercept model is $r = N + m$.

4.2 Growth curve model

The growth curve model is another popular mixed model, which can be reduced to (2) letting $X_i = Z_i A_i$, where A_i is a $k \times m$ fixed matrix. We assume here that matrices $\{Z_i, i = 1, 2, \dots, N\}$ have full rank. In this case, $\beta_i = A_i \beta - b_i$ and b_i are not restricted and therefore

$$S_{\min} = \sum_{i=1}^N \|y_i - Z_i \hat{\beta}_i\|^2,$$

where

$$\hat{\beta}_i = (Z_i' Z_i)^+ Z_i' y_i$$

is the cluster-specific LS estimator. Thus, finally,

$$R_r^2 = \frac{\sum_{i=1}^N \|y_i - X_i \widehat{\beta}\|^2 - S_{MM}}{\sum_{i=1}^N \|y_i - X_i \widehat{\beta}\|^2 - \sum_{i=1}^N \|y_i - X_i \tilde{\beta}_i\|^2},$$

where $\tilde{\beta}_i$ is the LS estimate from cluster i . For this model $r = Nm$.

4.3 Meta-analysis model

The meta-analysis model became a very popular statistical model for combination of studies to improve statistical power for detection of the overall treatment effect. Usually the metaanalysis model is formulated as a simple random-effect model

$$y_i = \mu + b_i + \varepsilon_i, \quad i = 1, 2, \dots, n, \quad (19)$$

where y_i the variable of interest, say, log odds ratio of the treatment effect from the i th study (we shall use this interpretation hereafter), μ is the overall treatment effect, b_i is the random effect that reflects the studies' heterogeneity, and ε_i is the error term due to uncertainty of y_i estimation from study i (Demidenko, 2004, chapter 5; Borenstein et al., 2009). The estimates y_i and their variances $\text{var}(\varepsilon_i) = \sigma_i^2$ are known but the variance of the random effect, $\text{var}(b_i) = \tau^2$ are not and is subject of estimation along with μ . We associate τ^2 with heterogeneity of the studies: if $\tau^2 = 0$ the only variation factor is due to imprecision within the study; if τ^2 is significantly apart from 0 studies are heterogeneous. The difference with the general mixed model described in Section 2 is that the variances of ε_i are not equal and given so we have to consider the weighted least squares instead of ordinary least squares. If the heterogeneity variance τ^2 is fixed the estimates of the common treatment and random effects are found from minimization of the penalized weighted least squares

$$S(\tau^2) = \sum_{i=1}^n \frac{(y_i - \mu - b_i)^2}{\sigma_i^2} + \frac{1}{\tau^2} \sum_{i=1}^n b_i^2,$$

that yields

$$\widehat{\mu}(\tau^2) = \frac{\sum_{i=1}^n y_i / (\tau^2 + \sigma_i^2)}{\sum_{i=1}^n 1 / (\tau^2 + \sigma_i^2)}, \quad \widehat{b}_i(\tau^2) = \frac{\tau^2}{\tau^2 + \sigma_i^2} (y_i - \widehat{\mu}) \quad (20)$$

with

$$S_{\min}(\tau^2) = \sum_{i=1}^n \frac{(y_i - \widehat{\mu} - \widehat{b}_i(\tau^2))^2}{\tau^2 + \sigma_i^2}.$$

When $\tau^2 = 0$ we have $\widehat{\mu}(0) = \bar{y}$, $\widehat{b}_i(0) = 0$ and the weighted least squares gives

$$S_{wLS} = \sum_{i=1}^n \frac{(y_i - \bar{y})^2}{\sigma_i^2},$$

where $\tilde{y} = (\sum y_i / \sigma_i^2) (\sum 1 / \sigma_i^2)^{-1}$ is the weighted average. When $\tau^2 \rightarrow \infty$ we have $\hat{\mu}(\infty) = \bar{y}$, $\hat{\sigma}_\mu^2(\infty) = \bar{y}_i - \bar{y}$ and $S_{\min} = 0$. Following formula (15) we obtain

$$R_r^2 = 1 - \frac{S_{MM}}{S_{WLS}}, \quad (21)$$

where $S_{MM} = S_{\min}(\hat{\tau}^2)$ and $\hat{\tau}^2$ is an estimate of τ^2 . This estimate is usually obtained using method of moments or maximum likelihood (DerSimonian and Laird, 1986; DerSimonian and Kacker, 2007). We prove in the Appendix that $0 \leq R_r^2 \leq 1$. We interpret R_r^2 for the metaanalysis model as the proportion of conditional variance of y explained by the heterogeneity of studies reflected in the random effect. When studies are homogeneous (the heterogeneity variance is zero) $R_r^2 = 0$. When studies are highly heterogeneous R_r^2 is close to 1. Note that the fixed effects coefficient of determination for the meta-analysis model does not make sense.

5 Examples

We illustrate the application of the random effects coefficient of determination with three examples. The random intercept model is used in the first example to account for variation in breast cancer patients travel time to the nearest National Cancer Institute (NCI) Cancer Center. In this example, we have a relatively small number of clusters (states) and a large number of individuals in each cluster. In the second example, we model the cumulative time of watching alcohol movie scenes as a predictor of drinking initiation among young adolescents. This is an example of longitudinal data analysis where there are a few observations in each cluster (4 surveys), but many clusters/individuals (about five thousand). In both examples, the validity of the mixed model is under question and we address it through computation of R_r^2 . The third example illustrates the random effects coefficient of determination with a meta-analysis of 13 studies on tuberculosis vaccine previously used in several publications.

5.1 Travel time to the nearest NCI Cancer Center

Access to specialized cancer care is an important issue of national concern given that cancer is the leading cause of death in people under age 85, Twombly (2005). Of particular importance is the analysis of patients' geographic access to cancer centers expressed in terms of travel time, Onega et al. (2008). Understanding the differences in travel time may help in efficient allocation of existing resources and in the development of new cancer centers. In this example, we analyze travel time (TT) in minutes for Medicare breast cancer patients (a comprehensive description of the data is found in the above reference); the total number of observations/patients $N_T = 47,383$ from 46 states. Special attention is paid to racial differences when it comes to cancer treatment accessibility. To identify the differences, five races were modeled via a dummy variable approach: Caucasian, African American, Asian, Hispanic, Native American, other/unknown (Caucasian is the reference). It is suspected that TT varies significantly across states, so the random (state-specific) intercept model seems appropriate here. Travel time is heavily skewed to the right, so log transformation is used to reduce the bias towards long travel. For these data there are $N = 46$ clusters and the number of variables in the model is $m = 8$ (without intercept). Besides racial factors, three other variables were chosen to explain variation in TT: age of woman, breast cancer stage expressed on the nominal scale from 0 to 4, and whether breast surgery occurred (1) or not (0).

Three methods of estimation are used: (a) ordinary least squares, (b) fixed effects least squares, (c) a linear mixed (random intercept) model; the results are presented in Table 1. The sum of squares for the linear mixed model was computed by

$$S_{MM} = \sum \|y_i - X_i \hat{\beta}_{ML} - \hat{b}_i\|^2,$$

where $\hat{b}_i$ is the estimated random effect, or deviation of the random intercept from its common value (function lme was used in S-Plus). As follows from Table 1, the estimates in methods (b) and (c) are very close; the estimates of the standard deviations of the error term and random intercept are $\hat{\sigma}_\varepsilon = 0.79$ and $\hat{\sigma}_a = 0.82$ which imply $\hat{d} = \sigma_\varepsilon^2 / \sigma_a^2$ and $\rho = 0.52$ indicating that the variation across states is comparable with the variation around the model.

We see considerable differences of TT across races. For example, using the LME model, Asian women spend $(1 - e^{-0.61376}) \times 100\% = 46\%$ less time on travel than White women. On the other hand, Native American women spend considerably more time. We attribute these differences to where the majority of women live: Asian women live in urban areas while Native Americans live in remote parts of the country. Neither breast cancer stage nor surgery contribute much to TT. We compute the national average of the travel time in hours for a White woman of age 80 as $e^{4.9035 - 0.00418 \times 80} / 60 = 1.6$ (1 hour 40 min) based on the estimate from LME model.

Now we turn our attention to testing of the validity of the LME model. The F -test (14) gives

$$F = \frac{(55278 - 29864) / (54 - 8)}{29864 / (47383 - 54)} = 876,$$

with the critical value 1.36—the random effect is highly significant. The random effects coefficient of determination is

$$R_r^2 = \frac{56278 - 29867}{56278 - 29864} = 0.999.$$

We conclude that this model can be estimated by the dummy variable approach where each state has its own intercept term. Indeed, as follows from Table 1, the difference of estimates between fixed effects and random effects approaches are practically equivalent.

We present the differences in TT across the U.S. states graphically in Figure 1. As we see, there are substantial differences across states. For example, a white woman of 80 years spends only 30 minutes on travel to an NCI Cancer Center in New York state. On the other hand, this woman would spend more than 9 hours living in Montana.

5.2 Alcohol drinking initiation in young adolescents

Early onset of alcohol use, when drinking begins in early adolescence, is a cause of many behavioral and life problems, including reckless driving, school dropout, and accidental death (DeWit, 2000; Hingson et al., 2000). Several studies have confirmed that early drinkers also have a higher risk of developing alcoholism in adulthood, Kraus et al. (2000). Recently researchers have drawn a parallel between early drinking and alcohol depiction in

movies, Sargent et al. (2006). In this section, we apply a linear mixed model to analyze timed exposure to movie alcohol depiction using data described in the above paper. More than five thousand U.S. teens were surveyed through telephone interviews approximately every 8 months over two years (four waves of interviews, kids of age from 10 to 14 at the baseline). Covariates include parents' education, family income, whether the teen owned branded merchandise and whether he/she had ever drunk alcohol. The teen were interviewed at home: to ensure privacy and honesty of answers to sensitive questions, teens were able to key in their answers using the telephone pad rather than answering by voice. Special attention was given to measuring exposure to alcohol depiction in movies. Children were asked what popular movies they have seen from a list of the most popular recent movies. A team of researchers at Dartmouth quantified the length of the alcohol scenes in those movies and the time of exposure to movie alcohol has been reconstructed for each adolescent over four waves.

We hypothesize that the cumulative exposure to alcohol scenes in movies is one of the factors contributing to early drinking. As in the previous example, the dependent variable is heavily skewed, so the log transformation is used. The log cumulative movie alcohol time was modeled as a function of age adjusting for the following binary variables: drinking status (Drinker=1, otherwise=0), sex (Boy=1, Girl=0), race (White=1, otherwise=0), having any alcohol branded merchandise (Alcohol item=1, otherwise=0), parents' education (High school or higher education=1, otherwise=0), family income (income < \$20K =1, otherwise 0), and performance in school (below average=1, otherwise 0). The intercept and slope at age were assumed to be random and teen specific. The total number of observations is $N_T = 18,482$ with the number of clusters/subjects $N = 5,124$. The results of estimation by two methods are presented in Table 2 and graphically illustrated in Figure 2. As follows from this figure and Table 2, drinkers saw many hours of alcohol scenes in movies, and the cumulative exposure increased exponentially with age. White adolescents are more exposed to alcohol in movies, while other factors, such as parents' education or income had only a small effect. Since all independent variables/predictors, other than age, did not change with time, we obtained $S_{\min}$ simply by regressing each $\mathbf{y}_i$ on Age and taking the sum of squares of residuals, which yields $S_{\min} = 61$. The MLE of the linear mixed model gives $\hat{\sigma}_e = 0.089$ with SD intercept =1.92 and SD slope=0.134.

In Figure 2, we show the cumulative time of alcohol scene watching as function of age for two groups of adolescents, nondrinkers and drinkers. The two parallel lines represent nondrinkers (drinker=0, solid) and drinkers (drinker=1, dotted); the distance between these lines is 0.3223. The rank of matrix W in this example is $r = 2 \times 5124$, and we obtain

$$F = \frac{(6582 - 61)/(2 \times 5124 - 9)}{61/(18482 - 2 \times 5124)} = 86,$$

yielding the p -value < 0.001. The random effect coefficient of determination,

$$R_r^2 = \frac{6582 - 76}{6582 - 61} = 0.997,$$

which indicates a very strong support of random effects in this model. This means that variation between adolescents is much higher than change in habit of movie watching as they get older.

As in the previous example, the variation of the baseline alcohol movie watching (at age 10) is so large that the random effects model can be replaced with the fixed effects model estimated by the dummy variable approach.

5.3 Meta-analysis model for vaccine studies

To illustrate the random effects coefficient of determination for the meta-analysis model we use an example from Berkey et al. (1995). The log odds-ratio of the efficacy of the BCG vaccine for prevention of tuberculosis along with its variance was estimated in 13 studies with different sample sizes. The treatment effect with the 95% confidence intervals (horizontal segments) is shown in Figure 3. We start with testing the homogeneity of the studies using the Q -statistic, DerSimonian and Laird (1986), which in our notation is S_{LS} . Apparently, under the null, $H_0: \tau^2 = 0$, this statistic has a chi-square distribution with $n - 1 = 12$ degrees of freedom. In our case $Q = S_{WLS} = 152$ but the critical value is 22, so the p -value = 0; there is a strong indication of the presence of random effects.

Now we turn our attention to computation of the random effect coefficient of determination. The original data (log OR) for each study is depicted by a circle, the data adjusted by the random effect, depicted by a cross, is computed as $y_i - \hat{b}_i(\hat{\tau}^2)$, an “estimate” of ε_i using formula from (20). The meta-analysis model (19) is estimated by maximum likelihood which yielded an overall vaccine effect $\hat{\mu}_{ML} = -0.71$ (OR = $e^{-0.71} = 0.49$) and the heterogeneity variance $\hat{\tau}_{ML}^2 = 0.28$ (SE = 0.53). The 95% confidence band (shaded rectangular) for the overall vaccine effect is shown in Figure 3. The weighted least squares estimate, $\tilde{y} = -0.43$ with $S_{WLS} = 152$ yielding $R_r^2 = 0.91$. We conclude that studies are quite heterogeneous because 91% of the conditional variance of log OR can be explained by the random effect. The strong presence of the random effect is seen from the figure—the LOG OR adjusted for the random effect substantially “move” toward 0.

6 Conclusion

In this paper, we aimed at answering two fundamental questions in the framework of linear mixed model: “Why fit a linear mixed model and how to measure the strength of the presence of random effects?” We have developed such a measure, called the random effects determination coefficient by appealing to the proportion of the residual sum of squares on the scale from its minimum to maximum values. R_r^2 is originated from the F -test based on the fixed effects estimation. It estimates the proportion of the conditional variance of the dependent variable, with respect to the maximum reduction, explained by random effects. R_r^2 close to zero indicates that the data can be modeled using the standard linear model—the presence of random effects is weak. On the other hand, if R_r^2 is apart from 0, the random effects lead to a significant reduction in the variance due to the presence of random effect. If R_r^2 is close to 1 random effects lead to the maximum reduction so that the mixed model can be replaced with a linear model with fixed effects estimated by the dummy variable approach.

In the examples presented, R_r^2 was close to 1 indicating that random effects can be replaced with fixed effects using the dummy variable approach estimated by the standard linear regression which simplifies the analysis and the theory considerably (for example, using standard linear hypothesis vehicle one may test whether dummy coefficients/intercepts are equal in some specific clusters). This finding could be easily overlooked without computation of the random effects determination coefficient.

Fixed and random effects coefficients of determination represent the quality of a mixed linear model from two perspectives and are equally important. While the former reflects the strength of the fixed effects, the latter characterizes the presence of the random effects. It would be beneficial if these figures of merits were computed as a standard output software every time a mixed model or a meta-analysis models are estimated.

Acknowledgments

The work of Demidenko was supported by NIH/NCI grant CA130880 and the work of Sargent was supported by grant CA77026.

8 Appendix: Proofs

1. Proof that $0 \leq R_r^2 \leq 1$ for the general linear mixed effects model defined in equation (16). The denominator is positive if $\text{rank}(\mathbf{W}) > m$ and $S_{LS} > 0$. This follows from the following fact of linear regression analysis [24]: if $\mathbf{X}^{n \times m}$ and $\mathbf{H}^{n \times k}$ such that the space spanned by vector-columns of matrix $\mathbf{X}$ is a subspace of space spanned by vector-columns of matrix $\mathbf{H}$, and $\text{rank}(\mathbf{X}) < \text{rank}(\mathbf{H}) < n$ and $\min_{\beta} \|\mathbf{y} - \mathbf{X}\beta\|^2 > 0$ then $\min_{\gamma} \|\mathbf{y} - \mathbf{H}\gamma\|^2 < \min_{\beta} \|\mathbf{y} - \mathbf{X}\beta\|^2$. Now we prove that for any fixed $\mathbf{D}$ we have $S_{\min} \leq S_{MM} \leq S_{LS}$. Expressing $\mathbf{b}_i$ through β as in (11) and denoting $\mathbf{P}_i = \mathbf{Z}_i'(\mathbf{Z}_i'\mathbf{Z}_i + \mathbf{D}^{-1})^{-1}\mathbf{Z}_i'$ we can rewrite

$$S_{MM} = \min_{\beta} \sum_{i=1}^N (\mathbf{y}_i - \mathbf{X}_i\beta)'(\mathbf{I} - \mathbf{P}_i)^2(\mathbf{y}_i - \mathbf{X}_i\beta) \leq \min_{\beta} \sum_{i=1}^N (\mathbf{y}_i - \mathbf{X}_i\beta)'(\mathbf{y}_i - \mathbf{X}_i\beta) \leq S_{LS}$$

because $0 \leq \mathbf{I} - \mathbf{P}_i \leq \mathbf{I}$, the right inequality is established. To prove the left inequality we notice that

$$S_{\min} = \min_{\beta, b_1, \dots, b_N} \sum_{i=1}^N \|\mathbf{y}_i - \mathbf{X}_i\beta - \mathbf{Z}_i b_i\|^2 \leq \min_{\beta} \sum_{i=1}^N \|\mathbf{y}_i - \mathbf{X}_i\beta - \mathbf{Z}_i \widehat{b}_i(\beta)\|^2$$

for any function $\widehat{b}_i(\beta)$ because $S_{\min}$ is the unconstrained minimum. Particularly, this is true for (11), which proves $S_{\min} \leq S_{MM}$.

2. Proof that $0 \leq R_r^2 \leq 1$ for the meta-analysis model defined in equation (21). The inequality $R_r^2 \leq 1$ follows from the fact that S_{WLS} and $S_{\min}$ are nonnegative. Now we prove that $S_{\min}(\tau^2)$ is a decreasing function of τ^2 . Indeed, applying the chain rule the derivative is

$$\frac{dS_{\min}(\tau^2)}{d\tau^2} = - \sum_{i=1}^n \frac{(y_i - \widehat{\mu}(\tau^2))^2}{(\tau^2 + \sigma_i^2)^2} - \sum_{i=1}^n \frac{y_i - \widehat{\mu}(\tau^2)}{\tau^2 + \sigma_i^2} \times \frac{d\widehat{\mu}(\tau^2)}{d\tau^2}.$$

But the second term vanishes due to (20). This implies that the derivative is nonnegative and $S_{\min}(\tau^2) \leq S_{\min}(0) = S_{WLS}$ which proves $R_r^2 \geq 0$.

References

- Berkey CS, Hoaglin DC, Mosteller F, Colditz GA. A random-effects regression model for meta-analysis. *Statistics in Medicine*. 1995; 14:395–411. [PubMed: 7746979]
- Borenstein, M.; Hedges, LV.; Higgins, JPT.; Rothstein, HR. *Introduction to Meta-Analysis*. New York: Wiley; 2009.
- Brown, H.; Prescott, R. *Applied Mixed Models in Medicine*. Chichester: Wiley; 1999.
- Demidenko, E. *Mixed Models: Theory and Applications*. New York: Wiley; 2004.
- DerSimonian R, Laird NM. Meta-analysis in clinical trials. *Controlled Clinical Trials*. 1986; 7:177–188. [PubMed: 3802833]
- DerSimonian R, Kackerb R. Random-effects model for meta-analysis of clinical trials: An update. *Contemporary Clinical Trials*. 2007; 28:105–114. [PubMed: 16807131]
- DeWit DJ, Adlaf EM, Offord DR, Ogborne AC. Age at first alcohol use: A risk factor for the development of alcohol disorders. *American Journal of Psychiatry*. 2000; 157:745–750. [PubMed: 10784467]
- Diggle, PJ.; Kung-Yee, L.; Zeger, SL. *Analysis of Longitudinal Data*. Oxford: Oxford University Press; 2002.
- Edwards LJ, Muller KE, Wolfinger RD, Qaqish BF, Schabenberger O. An R^2 statistic for fixed effects in the linear mixed model. *Statistics in Medicine*. 2008; 27:6137–6157. [PubMed: 18816511]
- Gelman A, Pardoe I. Bayesian measures of explained variance and pooling in multilevel (hierarchical) model. *Technometrics*. 2006; 48:241–251.
- Hingson RW, Heeren T, Jamanka A, Howland J. Age of drinking onset and unintentional injury involvement after drinking. *Journal of the American Medical Association*. 2000; 284:1527–1533. [PubMed: 11000646]
- Jiang, J. *Linear and Generalized Linear Mixed Models and Their Applications*. New York: Springer; 2007.
- Khuri, AI.; Mathew, T.; Sinha, BK. *Statistical Tests for Mixed Linear Models*. New York: Wiley; 1988.
- Kraus L, Bloomfield K, Augustin R, Reese A. Prevalence of alcohol use and the association between onset of use and alcohol related problems in a general population sample in Germany. *Addiction*. 2000; 95:1389–1401. [PubMed: 11048357]
- Littell, RC.; Milliken, GA.; Stroup, WW.; Wolfinger, RD.; Schabenberger, O. *SAS for Mixed Models*. Cary: SAS Institute; 2006.
- Mathew T, Sinha BK. Exact and optimum tests in unbalanced split-plot designs under mixed and random models. *Journal of the American Statistical Association*. 1992; 87:192–200.
- McCulloch, CE.; Searle, SR.; Neuhaus, JM. *Generalized, Linear, and Mixed Models*. New York: Wiley; 2008.
- Onega T, Duel EJ, Shi X, Wang D, Demidenko E, Goodman D. Geographic access to cancer care in the U.S. *Cancer*. 2008; 138:1919–1933.
- Pinheiro, JC.; Bates, DM. *Mixed-Effects Models in S and S-PLUS*. New York: Springer; 2000.
- Sargent JD, Wills TA, Stoolmiller M, Gibson J, Gibbons FX. Alcohol use in motion pictures and its relation with early-onset teen drinking. *Journal of Studies on Alcohol and Drugs*. 2006; 67:54–65.
- Scheipl F, Greven S, Küchenhoff H. Size and power of tests for a zero random effect variance or polynomial regression in additive and linear mixed models. *Computational Statistics & Data Analysis*. 2008; 52:3283–3299.
- Searle, SR.; Casella, G.; McCulloch, CM. *Variance Components*. New York: Wiley; 1992.
- Seely JF, El-Bassiouni Y. Applying Wald's variance component test. *Annals of Statistics*. 1983; 11:197–201.
- Searle, SR. *Linear Models*. New York: Wiley; 1971.
- Snijders TAB, Bosker RJ. Modeled variance in two-level models. *Sociological Methods & Research*. 1994; 22:342–363.
- Twombly R. Cancer surpasses heart disease as leading cause of death for all but the very elderly. *Journal of National Cancer Institute*. 2005; 97:330–331.

- Verbeke, G.; Molenberghs, G. Linear Mixed Models for Longitudinal Data. New York: Springer; 2000.
- Vonesh, EF.; Chinchilli, VM. Linear and Nonlinear Models for the Analysis of Repeated Measurements. New York: Marcel Dekker; 1997.
- Xu R. Measuring explained variation in linear mixed effects models. *Statistics in Medicine*. 2003; 22:3527–3541. [PubMed: 14601017]
- Wald A. A note on regression analysis. *Annals of Mathematical Statistics*. 1947; 18:586–589.

Figure 1.

State-specific averages estimated from the random intercept model. There are substantial differences in TT across states, $R^2_r=0.999$. The mixed model reduces to the dummy variable approach.

Figure 2.

Cumulative length of alcohol scene watched in movies for the two groups of adolescents as a function of age. The solid and dotted lines represent the cumulative growth in nondrinker and drinker groups, respectively.

Figure 3.

Log odds ratios from 13 studies of the tuberculosis vaccine effect with estimation of the overall effect using the meta-analysis model. The studies are highly heterogeneous because $R^2=0.91$.

Table 1

Estimation of TT model by three methods

Variable	(a) OLS, $\hat{\beta}$		(b) Fixed Effects, $\hat{\beta}$		(c) LME, $\hat{\beta}_{ML}$	
	Coefficient	SE	Coefficient	SE	Coefficient	SE
Intercept	4.3291	0.0624			4.8851	0.1334
Age	-0.0058	0.0007	-0.00418	0.0005	-0.00418	0.0005
African American	-0.2657	0.0194	-0.36309	0.0146	-0.36307	0.0146
Asian	-0.4990	0.0518	-0.34313	0.0379	-0.34321	0.0379
Hispanic	-0.8740	0.0445	-0.61353	0.0327	-0.61376	0.0327
Native American	-0.2863	0.0448	-0.09938	0.0329	-0.09947	0.0329
Race unknown	0.6248	0.1364	0.56905	0.0099	0.56898	0.0099
BC stage	0.0098	0.0023	0.00886	0.0017	0.00883	0.0017
Surgery	0.0564	0.0221	0.03700	0.0162	0.03695	0.0161
Sum of Squares	$S_{OLS} = 56278$		$S_{min} = 29864$		$S_{ML} = 29867$	

Table 2

Estimation results of cumulative exposure to alcohol scenes in movies among teens.

Variable	OLS, $\hat{\beta}$		LME, $\hat{\beta}_{ML}$	
	Coefficient	SE	Coefficient	SE
Intercept	-0.5272	0.0396	-2.3264	0.0390
Drinker	0.3223	0.0102	0.1783	0.0199
Age	0.1779	0.0030	0.3250	0.0022
Boy	0.0129	0.0089	0.0100	0.0181
White	0.2396	0.0156	0.2463	0.0307
Alcohol item	0.2202	0.0146	0.1787	0.0294
Parents' education	-0.0289	0.0100	-0.0334	0.0196
Family income	0.0630	0.0131	0.0429	0.0262
School performance	-0.0784	0.0096	-0.0525	0.0196
Sum of Squares	$S_{OLS} = 6582$		$S_{MM} = 76$	