

Kansas State University Libraries
New Prairie Press

Conference on Applied Statistics in Agriculture 2002 - 14th Annual Conference Proceedings

IMPACT OF VARIANCE COMPONENT ESTIMATES ON FIXED EFFECT INFERENCE IN UNBALANCED LINEAR MIXED MODELS

Walt Stroup

Ramon Littell

Follow this and additional works at: <https://newprairiepress.org/agstatconference>



Part of the [Agriculture Commons](#), and the [Applied Statistics Commons](#)



This work is licensed under a [Creative Commons Attribution-Noncommercial-No Derivative Works 4.0 License](#).

Recommended Citation

Stroup, Walt and Littell, Ramon (2002). "IMPACT OF VARIANCE COMPONENT ESTIMATES ON FIXED EFFECT INFERENCE IN UNBALANCED LINEAR MIXED MODELS," *Conference on Applied Statistics in Agriculture*. <https://doi.org/10.4148/2475-7772.1198>

This is brought to you for free and open access by the Conferences at New Prairie Press. It has been accepted for inclusion in Conference on Applied Statistics in Agriculture by an authorized administrator of New Prairie Press. For more information, please contact cads@k-state.edu.

IMPACT OF VARIANCE COMPONENT ESTIMATES ON FIXED EFFECT INFERENCE IN
UNBALANCED LINEAR MIXED MODELS

Walt Stroup, Department of Biometry, University of Nebraska-Lincoln 68583-0712
Ramon Littell, Department of Statistics, University of Florida, Gainesville, FL 32611

Abstract: Inference on fixed effects in mixed models depends on standard errors or test statistics which in turn depend on estimates of variance and covariance components. For unbalanced mixed models, even relatively simple models such as two-way cross-classification models with interaction where one factor is fixed and the other is random, dilemmas arise that have received inadequate attention to date. For example, if one uses SAS PROC MIXED, one can estimate variance components using expected means squares from Type I, II, or III sums of squares, or one can use likelihood-based algorithms such as the default restricted maximum likelihood. If there is a negative variance component estimate, one can set the estimate to zero and proceed with fixed effects inference, or one can allow the variance estimate to remain negative. These decisions affect inference on fixed effects in ways that are not generally well-understood. The purposes of this presentation are to 1) clarify what the main issues are and 2) present some guidelines data analysts can use.

1. Introduction

In mixed models, standard errors used to construct confidence intervals and statistics used to test hypotheses depend on estimated variance components. The resulting distribution theory upon which confidence interval and hypothesis test procedures are based can be affected by variance estimation. For linear mixed models with balanced data and uncorrelated random effects, distribution theory underlying well-known procedures based on the F and t distributions holds exactly. However, for unbalanced data, as well as models with correlated errors, inference must be based on asymptotic theory.

Kacker and Harville (1984) and Kenward and Roger (1997) addressed the impact on fixed effect inference in linear mixed models when estimated variance components are used. They found that for unbalanced data (as well as models with correlated errors), standard errors based on estimated variance components are biased downward and hence test statistics are biased upward. The most recent release of SAS[®] (Release 8) made Kenward and Roger's bias correction algorithm available as an option in *PROC MIXED*. However, the impact of *negative* variance component estimates on fixed effect inference has not received much attention. Furthermore, previous work has tended to focus on the behavior of likelihood-based variance estimation, e.g. restricted maximum likelihood (REML). However, other variance component estimation procedures exist, notably those based on the expected mean squares in analysis of variance. For unbalanced mixed models, there are several types of sums of squares, each leading to a different set of expected mean squares and hence different variance estimates and different test statistics. Swallow and Monahan (1984) compared the precision of various variance component estimation procedures, but their study was restricted to variance component estimation *per se* and did not address their impact on fixed effect inference.

The purpose of this paper is to present an example of a simple mixed model with unbalanced data where the variance component estimate is negative and to examine the consequences of various strategies to deal with negative estimates.

2. Example

Table 1 describes a multi-location experiment conducted to compare two treatments. Specifically, the table gives the number of observations on each treatment at each of the nine locations. The data appear in Output 6.22 in Littell, et al (2002). A linear mixed model, assuming random locations, is

$$y_{ijk} = \mu + L_i + \tau_j + (\tau L)_{ij} + e_{ijk},$$

where y_{ijk} is the k^{th} observation on the i^{th} location and j^{th} treatment, μ is the intercept, τ_j is the j^{th} treatment effect, L_i is the i^{th} location effect, assumed independent and identically distributed (i.i.d.) $N(0, \sigma_L^2)$, $(\tau L)_{ij}$ is the ij^{th} treatment \times location effect, assumed i.i.d. $N(0, \sigma_{TL}^2)$, and e_{ijk} is random error, assumed i.i.d. $N(0, \sigma^2)$. The resulting analysis of variance table has the following general form:

Source of Variation	degrees of freedom	Expected Mean Square
location	8	
treatment	1	$\sigma^2 + k_1\sigma_{TL}^2 + k_2\sigma_L^2 + Q_{TRT}$
trt \times loc	8	$\sigma^2 + k_3\sigma_{TL}^2$
error	114	σ^2

The term Q_{TRT} is the quadratic form involving the sum of squared treatment effects. Note that because the data are unbalanced, the variance components have different coefficients for each source of variation. Moreover, the specific values of k_1 , k_2 , k_3 , and Q_{TRT} depend on which type of sum of squares one uses.

The default analysis using SAS PROC MIXED uses the following SAS statements:

```
proc mixed;
  class loc trt;
  model y=trt/ddfm=kr;
  random loc loc*trt;
```

The basic output appears in Table 2. The key results are

- the estimated location variance ($\hat{\sigma}_L^2$) is negative and has been set to 0
- the F -value for treatment is 1.83 with a p -value of 0.2075.

The PROC MIXED default that sets the negative variance component to zero is a typical convention in variance component estimation. This makes sense when one reports a variance estimate, *per se*, or a function of a variance, such as heritability. However, the “set to zero” convention upwardly biases the estimate of the variance component in question, has a ripple effect on estimates of the other variance components, and hence affects standard errors and test

statistics in ways that may not always be understood. Is the “set to zero” convention really appropriate for inference on fixed effects and, if not, what are the alternatives? We first examine the alternatives.

In this example, the negative variance component estimate is for “location,” i.e., the term immediately *before* “treatment” in the model and the ANOVA table. Data analysts using PROC MIXED have tended not to worry about this, but they *have* worried when the variance component for the *denominator* mean square of the test – in this case σ_{TL}^2 -- is negative. Despite a lack of systematic published documentation, there is something of an “oral tradition” among PROC MIXED users that the “set to zero” default results in an overly conservative test of treatment effect. Data analysts have been advised either to use the standard ANOVA F -test, i.e. $F = MS(trt) / MS(loc \times trt)$, or the NOBOUND option, which leaves the estimated variance component equal to the negative solution from the REML estimating equation rather than re-setting the estimate to zero.

One obtains the NOBOUND option by modifying the first line of the above SAS code as follows:

```
proc mixed nobound;
```

The rest of the program remains the same. Table 3 shows the resulting SAS output. The key results are:

- for the location variance, $\hat{\sigma}_L^2 = -16.06$
- $F_{TRT}=1.64$ with a p -value of 0.2610

In this case, the conclusions about treatment effect would probably not be affected, but it is easy to see that with certain data sets, they would.

One can also obtain tests based on the ANOVA table. However, SAS computes several different types of sums of squares. With unbalanced data, the resulting expected mean squares, variance component estimates, test statistics, and denominator degrees of freedom (which use Satterthwaite’s approximation based on the denominator term’s linear function of mean squares) are different for the various types of sums of squares. Table 4 shows the expected means squares and resulting F_{TRT} statistics as computed by SAS PROC GLM using the following statements:

```
proc glm;
  class loc trt;
  model flush=trt loc trt*loc / e1 e2 e3;
  random loc trt*loc / test;
```

Note that Version 8 of SAS PROC MIXED can also compute these ANOVA statistics using the option METHOD=TYPE1 (or TYPE2 or TYPE3) in place of NOBOUND in the PROC statement.

Table 4 shows the type I, II, and III SS results. For these data, type III and type IV SS are identical, so type IV results are not shown. The key results are:

- for type I SS, $F_{TRT}=0.91$, with a p -value of 0.3742
- for type II SS, $F_{TRT}=1.10$, with a p -value of 0.3307
- for type III SS, $F_{TRT}=2.17$, with a p -value of 0.1674

Obviously, these F -values are quite different. Note that the “traditional” ANOVA-based estimator follows from Henderson’s Method 3 (1953). This is SAS PROC VARCOMP’s METHOD=TYPE1, which is identical to GLM or MIXED using type I SS. Type

II and III SS are not as widely used in variance component estimation, but there is no theoretical reason why they cannot be used. How should the data analyst decide which result to report?

3. Comparison of various approaches

Comparison can be based on the following criteria: 1) for hypothesis tests, power and control of type I error; 2) for confidence intervals, percent coverage. This section first considers some issues related to power that can be evaluated analytically, then addresses more comprehensive power, type I error, and coverage issues via simulation.

For the various types of sums of squares, the power of the test of treatment effect is a function of the non-centrality parameter, $\phi = Q_{TRT}/EMS = Q_{TRT}/(\sigma^2 + k_1\sigma_{TL}^2 + k_2\sigma_L^2)$. Noting that k_2 is either zero or negligible, the non-centrality parameter is essentially a function of Q_{TRT}/k_1 .

One can compute the values of Q_{TRT} for each type of SS by adding the option Q to the RANDOM statement in the above SAS PROC GLM code, i.e.

```
random loc trt*loc /  $\alpha$  test;
```

The resulting output appears in Table 5. Using these values, the non-centrality parameters of the three types of SS are:

- type I, $\phi = 32.99/9.15 = 3.61$
- type II, $\phi = 32.81/9.14 = 3.59$
- type III, $\phi = 20.98/4.66 = 4.50$

By this criterion, the test of treatment effect based on the type III SS should have the greatest power. Assuming none of the ANOVA-based tests have inflated type I error rate, the type III is the test of choice for this particular design and pattern of unbalanced data.

If one runs the type III test in PROC MIXED, one encounters a final complication. Table 6 shows the PROC MIXED output for the METHOD=TYPE3 option. Note that in addition to the ANOVA F -statistic, 2.17, MIXED computes its approximate Wald-type F -statistic from variance component estimates obtained from type III EMS. The resulting approximate F is equal to 1.71 with a p -value of 0.2481. Again, the question arises: which F should a data analyst use?

To help provide guidance to data analysts, power characteristics of the various methods of obtaining F were compared. Power was compared via simulation because the small sample characteristics of REML default, REML-nobound, and approximate Wald-based F tests using ANOVA variance component estimates cannot be assessed analytically. The following tests were compared

- REML default-based approximate F
- REML nobound-based approximate F
- type I MS ratio F -test
- type II MS ratio F -test
- type III MS ratio F -test
- approximate F using variance component estimates from type I EMS

- approximate F using variance component estimates from type II EMS
- approximate F using variance component estimates from type III EMS

The simulation used a design identical to the set up of the example in the previous section. Two scenarios were investigated: one where negative estimates of σ_L^2 are likely and one where negative estimates of σ_{TL}^2 are likely.

The first simulation set the variance components to $\sigma^2 = 1$, $\sigma_L^2 = 0.01$, and $\sigma_{TL}^2 = 0.25$. Power was simulated by generating 1000 simulated experiments for treatment mean differences of 0, 0.2, 0.4, 0.6, 0.8, and 1.0, and determining the number of times the null hypothesis of no treatment effect, i.e. $H_0: \tau_j=0$ for all j , was rejected at $\alpha=0.05$. For the simulated data with treatment mean difference equal to zero, the rejection rate was also obtained for approximate F tests using maximum likelihood (ML), as opposed to REML, estimators of the variance components.

Figure 1 shows the power curves. Because σ_L^2 was close to zero, negative estimates ($\hat{\sigma}_L^2$) occurred in roughly 50% of the simulated data sets. Several points emerge:

- The ML variance component estimate yields an inflated F statistic and a type I error rate a little over twice the nominal $\alpha=0.05$ criterion. This is typical of ML variance component estimates, which are biased downward, and thus produce upwardly biased F statistics even without negative variance component estimates. Setting a negative $\hat{\sigma}_L^2$ to zero further deflates $\hat{\sigma}_{TL}^2$ thus upwardly biasing F even more. ML variance component estimates should never be used in conjunction with fixed effect inference in mixed models.
- REML-default variance component estimates produced inflated test statistics and hence type I error rates roughly twice the nominal $\alpha=0.05$ criterion. One can immediately conclude that the REML default is an unacceptable procedure when negative variance component estimates occur.
- All of the other procedures – the REML-nobound and all ANOVA-based methods – have observed rejection rates essentially equal to the nominal $\alpha=0.05$ criterion. Hence, any of these procedures would be acceptable from the viewpoint of controlling type I error.
- For the ANOVA-based procedures, the results are consistent with the evaluation of non-centrality parameters for the various types of SS discussed earlier. The F test using the MS ratio based on type III SS had a somewhat higher rejection rate, i.e. somewhat greater observed power, for all non-zero treatment mean differences.
- The ANOVA tests that use MS ratios have somewhat greater observed power than the approximate Wald-type F using ANOVA-based variance component estimates.
- The observed power of the REML-nobound test was similar to that of the type III MS ratio ANOVA test. *However*, the REML nobound procedure *was* prone to convergence failure due to infinite likelihood. Failures to converge varied between 1.2% and 2.9% of the simulated data sets at the various levels of treatment mean difference. Variation in convergence failure rate appeared to be strictly random: there was no apparent relationship between failure rate and size of mean difference.

Figure 1 also shows the theoretical power curve assuming known variance components, computed using a procedure described by Stroup (1999). The REML-default power curve exceeds the theoretical power, but only because its type I error rate is inflated and it is therefore not a valid procedure. Note that none of the *legitimate* tests of treatment effect obtain an observed power equal to the theoretical power. Such is the price of additional variability inherent in using estimated rather than known variance components.

The choice of variance component estimator also affects confidence intervals for estimable functions of the fixed effects. Table 7 shows the observed percent coverage for 95% confidence intervals for treatment mean difference computed from the REML default, REML nobound, and ANOVA type III SS procedures. The REML nobound and ANOVA procedures produce observed coverages of 0.9473 and 0.9450, respectively, close to the nominal 95%. On the other hand, because the REML default results in an underestimate of the standard error of the treatment difference, the observed coverage is only 0.9170.

Table 8a shows additional detail for the REML default, REML nobound, and ANOVA Type III F tests. The rejection rates for differences of 0 through 0.8 are divided into those for all simulated data sets and those for which $\hat{\sigma}_L^2$ was negative. Rejection rates were consistently lower when $\hat{\sigma}_L^2$ was negative. Also, the rejection rates for the REML default were relatively more inflated relative to the other two methods when $\hat{\sigma}_L^2$ was negative. This is expected, because setting $\hat{\sigma}_L^2$ to 0 tends to deflate $\hat{\sigma}_{TL}^2$ and thus inflate F . Note that the rejection rates for the REML default when $\hat{\sigma}_L^2$ was positive were actually lower than ANOVA type III. This suggests that a strategy using ANOVA Type III for negative $\hat{\sigma}_L^2$ and REML for positive $\hat{\sigma}_L^2$ might actually be somewhat conservative.

A second simulation was done with $\sigma^2 = 1$, $\sigma_L^2 = 0.5$, and $\sigma_{TL}^2 = 0.01$. This was done to check the “oral tradition” mentioned earlier regarding negative estimates of σ_{TL}^2 . The results appear in Table 8b. For the lower treatment mean differences (0, 0.2, and 0.4) the results tend to support conventional wisdom. Rejection rates for REML default do appear to be excessively low. However, the results also contain surprises. For larger treatment mean differences, the situation reverses: if anything, the ANOVA type III F -test appears to be overly conservative! Also, for data sets for which $\hat{\sigma}_{TL}^2$ was negative, rejection rates for the REML nobound-based F -test are much greater than ANOVA type III, yet the overall power characteristics of REML nobound are about right. Unfortunately, REML nobound also showed a 10.6-15.0% failure to converge rate with these simulated data sets.

4. Discussion and Conclusions

Negative variance component estimates can and do affect fixed effect inference in linear mixed models. This is true for *all* variance components in the model, not merely the variance component associated with the denominator error term. The simulation results suggest that there is much we still do not know about the behavior of these tests when negative variance estimates

are obtained. Nonetheless, despite the limited nature of the simulation and a few surprises, some general guidelines for the data analyst do emerge. These are:

- for negative variance estimates of the term *preceding* the fixed effect of interest (e.g. “location” in the example in this paper), setting the estimate to zero inflates the type I error rate and reduces coverage by confidence intervals
- for negative variance estimates of the term *following* the fixed effect of interest (e.g. “location \times treatment” in the example), setting the estimate to zero deflates the rejection rate, resulting in an excessively conservative test and loss of power

When one obtains a negative variance component, one should therefore override “set to zero” defaults such as the one used by SAS PROC MIXED.

The standard ANOVA F tests based on ratios of mean squares yield acceptable control of type I error, as does the REML nobound option. On the other hand, one should *never* use straight maximum likelihood variance components for inference on fixed effects because it does not control type I error.

Among the procedures that do control type I error and allow variance component estimates to go negative, there are no simple answers. REML nobound appears to have more consistent power characteristics, but there is no guarantee that it will converge. For unbalanced data, the different types of SS produce different results. All of them control type I error acceptably, but their power characteristics differ. Although the example in this paper showed the type III SS maximizing power, different designs and patterns of unbalance may result in a different type of SS having greater power. Therefore, one should compare the non-centrality parameters of each type of SS for a given data set before deciding which to use. The classical ANOVA tests based on ratios of linear functions of mean squares have greater, or at least equal, power than the approximate F using ANOVA-based variance component estimates computed by PROC MIXED. Unfortunately, the power characteristics of even the best of the ANOVA-based procedures (Type III in this paper’s example) were not entirely consistent, e.g. the excessively conservative results with large treatment differences and small σ_{TL}^2 .

This paper has considered only one configuration of unbalanced data from a simple variance-component-only mixed model. However, the implications of negative variance component estimates even in these examples make it obvious that this subject has not received adequate attention. Clearly, there is much to learn even about relatively simple models. For more complex models, several issues remain to be addressed. For example, how do negative variance component estimates affect inference with correlated error models, such as repeated measures or spatial data? How do negative variance component estimates affect degree of freedom or bias adjustment algorithms, e.g. Kenward and Roger’s procedure? Finally, how do all of these issues translate to *nonlinear* mixed models? All of these questions could be expanded to include inference on EBLUP’s involving random effects. These are among the questions that need to be addressed in future research. What *this* presentation makes clear is that negative variance components cannot be ignored and that the defaults that many data analysts have used without comment often result in misleading or outright incorrect inference.

5. References

- Henderson, C.R. 1953. The estimation of variance and covariance components. *Biometrics* **9**: 226-252.
- Kacker, R.N. and D.A. Harville. 1984. Approximations for standard errors and estimators of fixed and random effects in mixed linear models. *J. Amer. Statist. Assoc.* **79**: 853-861.
- Kenward, M.G. and J.H. Roger. 1997. Small sample inference for fixed effects from restricted maximum likelihood. *Biometrics* **53**: 983 -997.
- Littell, R.C., W.W. Stroup, and R.J. Freund. 2002. *SAS System for Linear Models, 4th edition*. Cary, NC: SAS Institute, Inc.
- Stroup, W.W. 1999. Mixed model procedures to assess power, precision, and sample size in the design of experiments. *American Statistical Association 1999 Proceedings of the Biopharmaceutical Section*. pp. 15-24.
- Swallow, W.H. and J.F. Monahan. 1984. Monte Carlo comparisons of ANOVA, MINQUE, REML, and ML estimators of variance components. *Technometrics* **20**: 265-272.

Table 1. Description of design used in example by number of observations per location \times treatment combination

LOC	TRT		Total
	A	B	
42	2	1	3
43	14	14	28
44	5	6	11
45	11	11	22
46	6	7	13
47	6	7	13
48	7	8	15
49	10	10	20
50	4	3	7
Total	65	67	132

Table 2. Default output from SAS PROC MIXED

Covariance Parameter Estimates

Cov Parm	Estimate
LOC	0
LOC*TRT	75.3629
Residual	447.57

Type 3 Tests of Fixed Effects

Effect	Num DF	Den DF	F Value	Pr > F
TRT	1	9.3	1.83	0.2075

Differences of Least Squares Means

Effect	TRT	_TRT	Estimate	Standard Error	DF	t Value	Pr > t
	TRT	A B	-7.8198	5.7728	9.3	-1.35	0.2075

Table 3. Output from SAS PROC MIXED with REML nobound option

Covariance Parameter Estimates								
		Cov Parm	Estimate					
		LOC	-16.0628					
		LOC*TRT	93.4778					
		Residual	446.87					
Type 3 Tests of Fixed Effects								
		Num	Den					
Effect		DF	DF	F Value	Pr > F			
TRT		1	4.58	1.64	0.2610			
Differences of Least Squares Means								
Effect	TRT	_TRT	Estimate	Standard Error	DF	t Value	Pr > t	
TRT	A	B	-7.5903	6.1898	4.58	-1.23	0.2793	

Applied Statistics in Agriculture

Table 4. Output from SAS PROC GLM: EMS and F_{TRT} for various types of SS

Source	Type I Expected Mean Square
TRT	$\text{Var}(\text{Error}) + 9.1461 \text{Var}(\text{LOC}*\text{TRT}) + 0.04 \text{Var}(\text{LOC}) + Q(\text{TRT})$
LOC	$\text{Var}(\text{Error}) + 7.1543 \text{Var}(\text{LOC}*\text{TRT}) + 14.213 \text{Var}(\text{LOC})$
LOC*TRT	$\text{Var}(\text{Error}) + 7.0585 \text{Var}(\text{LOC}*\text{TRT})$

Tests of Hypotheses for Mixed Model Analysis of Variance

Source	DF	Type I SS	Mean Square	F Value	Pr > F
TRT	1	1134.560964	1134.560964	0.91	0.3742
Error	6.5931	8236.875597	1249.317056		
Error: $0.0028*MS(\text{LOC}) + 1.2929*MS(\text{LOC}*\text{TRT}) - 0.2957*MS(\text{Error})$					

Source	Type II Expected Mean Square
TRT	$\text{Var}(\text{Error}) + 9.1385 \text{Var}(\text{LOC}*\text{TRT}) + Q(\text{TRT})$
LOC	$\text{Var}(\text{Error}) + 7.1543 \text{Var}(\text{LOC}*\text{TRT}) + 14.213 \text{Var}(\text{LOC})$
LOC*TRT	$\text{Var}(\text{Error}) + 7.0585 \text{Var}(\text{LOC}*\text{TRT})$

Tests of Hypotheses for Mixed Model Analysis of Variance

Source	DF	Type II SS	Mean Square	F Value	Pr > F
TRT	1	1377.550724	1377.550724	1.10	0.3307
Error	6.5739	8212.186760	1249.204926		
Error: $1.2947*MS(\text{LOC}*\text{TRT}) - 0.2947*MS(\text{Error})$					

Source	Type III Expected Mean Square
TRT	$\text{Var}(\text{Error}) + 4.6613 \text{Var}(\text{LOC}*\text{TRT}) + Q(\text{TRT})$
LOC	$\text{Var}(\text{Error}) + 7.0585 \text{Var}(\text{LOC}*\text{TRT}) + 14.117 \text{Var}(\text{LOC})$
LOC*TRT	$\text{Var}(\text{Error}) + 7.0585 \text{Var}(\text{LOC}*\text{TRT})$

Tests of Hypotheses for Mixed Model Analysis of Variance

Dependent Variable: FLUSH

Source	DF	Type III SS	Mean Square	F Value	Pr > F
TRT	1	1843.572090	1843.572090	2.17	0.1674
Error	11.689	9943.710652	850.710718		
Error: $0.6604*MS(\text{LOC}*\text{TRT}) + 0.3396*MS(\text{Error})$					

Table 5. Output from SAS PROC GLM: Q_{TRT} output for various types of SS

Quadratic Forms of Fixed Effects in the Expected Mean Squares

Source: Type I Mean Square for TRT

	TRT A	TRT B
TRT A	32.99242424	-32.99242424
TRT B	-32.99242424	32.99242424

Source: Type II Mean Square for TRT

	TRT A	TRT B
TRT A	32.80309690	-32.80309690
TRT B	-32.80309690	32.80309690

Source: Type III Mean Square for TRT

	TRT A	TRT B
TRT A	20.97586951	-20.97586951
TRT B	-20.97586951	20.97586951

Table 6. Output from SAS PROC GLM: Q_{TRT} output for various types of SS

Type 3 Analysis of Variance				
Source	DF	Sum of Squares	Mean Square	
TRT	1	1843.572090	1843.572090	
LOC	8	7081.377266	885.172158	
LOC*TRT	8	8512.586561	1064.073320	
Residual	114	49684	435.825358	

Source	Expected Mean Square
TRT	Var(Residual) + 4.6613 Var(LOC*TRT) + Q(TRT)
LOC	Var(Residual) + 7.0585 Var(LOC*TRT) + 14.117 Var(LOC)
LOC*TRT	Var(Residual) + 7.0585 Var(LOC*TRT)
Residual	Var(Residual)

Source	Error Term	Error		
		DF	F Value	Pr > F
TRT	0.6604 MS(LOC*TRT) + 0.3396 MS(Residual)	11.689	2.17	0.1674
LOC	MS(LOC*TRT)	8	0.83	0.5995
LOC*TRT	MS(Residual)	114	2.44	0.0178
Residual

Covariance Parameter Estimates	
Cov Parm	Estimate
LOC	-12.6728
LOC*TRT	89.0063
Residual	435.83

Type 3 Tests of Fixed Effects				
Effect	Num DF	Den DF	F Value	Pr > F
TRT	1	4.97	1.71	0.2481

Differences of Least Squares Means							
Effect	TRT	_TRT	Estimate	Standard Error	DF	t Value	Pr > t
TRT	A	B	-7.9297	6.0625	4.97	-1.31	0.2481

Table 7. Percent coverage of Confidence Interval for estimate of treatment difference

<i>Method</i>	<i>s.e.(diff)</i>	<i>% Conf Lim Coverage</i>
theory	0.3037	0.9500
REML default	0.2783	0.9170
NoBound	0.2961	0.9473
Type 3	0.2948	0.9450

Table 8. Simulation Results: Observed Rejection Rates for H_0 : no treatment effect from tests based on REML-default, REML-nobound, and Type 3 ANOVA

a. $\sigma_L^2 = 0.01, \sigma_{TL}^2 = 0.25$

Trt Diff	Power from Theory	All Simulated Data Sets			Data Sets where $\hat{\sigma}_L^2 < 0$			
		REML Default	REML nobound*	Type 3 ANOVA	Pct Neg	REML Default	REML nobound	Type 3 ANOVA
0	0.050	0.083	0.052	0.054	47.8%	0.063	0.030	0.032
0.2	0.090	0.111	0.092	0.079	47.5%	0.118	0.056	0.062
0.4	0.214	0.240	0.188	0.199	46.4%	0.209	0.114	0.125
0.6	0.413	0.495	0.423	0.413	45.2%	0.477	0.312	0.307
0.8	0.638	0.727	0.634	0.643	45.6%	0.684	0.525	0.527

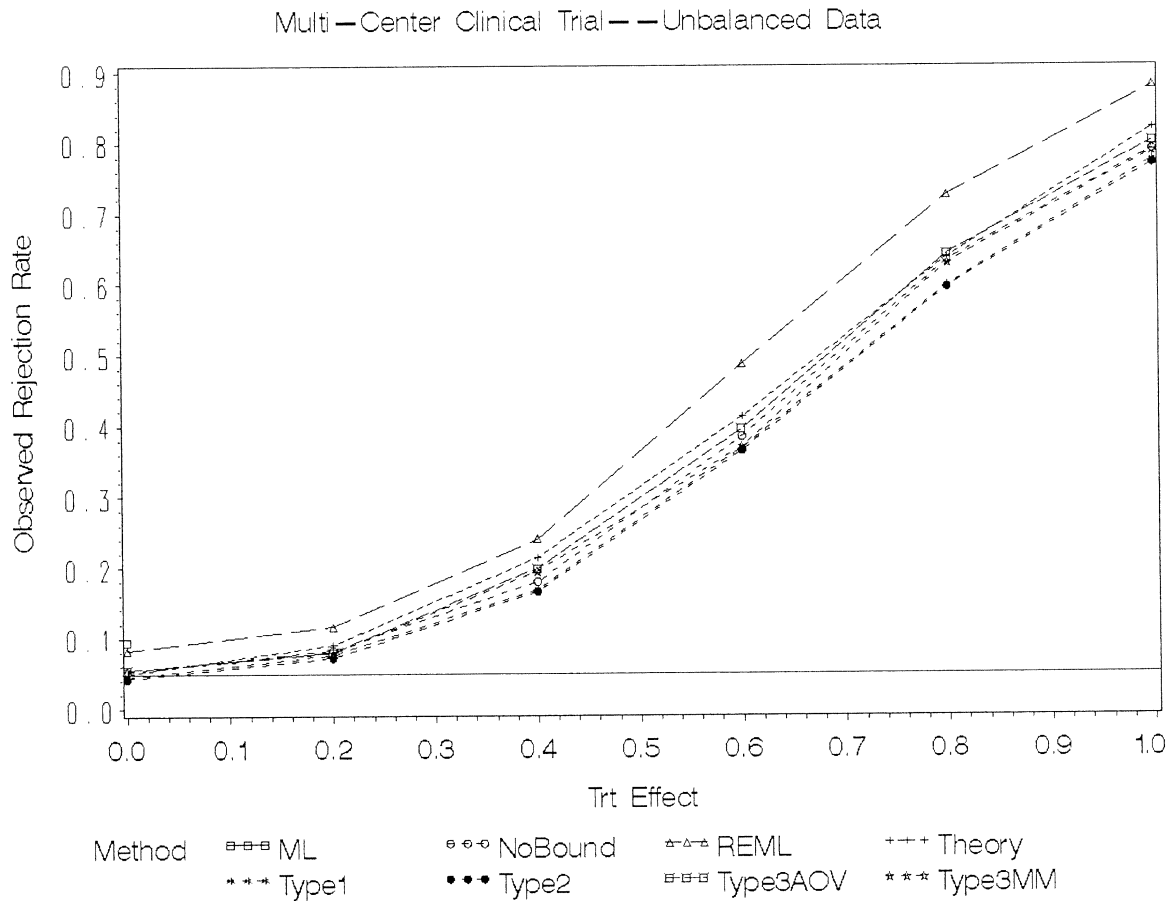
* Nobound failed to converge in 1.2-2.9% of simulated data sets

b. $\sigma_L^2 = 0.5, \sigma_{TL}^2 = 0.01$

Trt Diff	Power from Theory	All Simulated Data Sets			Data Sets where $\hat{\sigma}_{TL}^2 < 0$			
		REML Default	REML nobound*	Type 3 ANOVA	Pct Neg	REML Default	REML nobound	Type 3 ANOVA
0	0.050	0.026	0.052	0.044	52.2%	0.044	0.108	0.055
0.2	0.163	0.080	0.134	0.102	50.4%	0.119	0.255	0.139
0.4	0.489	0.396	0.455	0.408	52.2%	0.483	0.624	0.473
0.6	0.822	0.764	0.784	0.728	52.8%	0.830	0.897	0.822
0.8	0.969	0.952	0.959	0.898	53.0%	0.977	0.995	0.961

* Nobound failed to converge in 10.6-15.0% of simulated data sets

Figure 1. Power curve for tests of treatment effect using various variance component estimators



- NoBound failed to converge in 1.2 – 2.9% of simulated data sets
- estimate of σ_L^2 negative $\approx 48\%$ of data sets
- ML estimate shown only for Trt Effect=0.0