



**HAL**  
open science

# A COMPARISON OF ESTIMATION METHODS ON THE COVERAGE PROBABILITY OF SATTERTHWAITE CONFIDENCE INTERVALS FOR ASSAY PRECISION WITH UNBALANCED DATA

Edwin van den Heuvel

► **To cite this version:**

Edwin van den Heuvel. A COMPARISON OF ESTIMATION METHODS ON THE COVERAGE PROBABILITY OF SATTERTHWAITE CONFIDENCE INTERVALS FOR ASSAY PRECISION WITH UNBALANCED DATA. *Communications in Statistics - Simulation and Computation*, 2010, 39 (04), pp.777-794. 10.1080/03610911003646373 . hal-00583551

**HAL Id: hal-00583551**

**<https://hal.science/hal-00583551>**

Submitted on 6 Apr 2011

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



**A COMPARISON OF ESTIMATION METHODS ON THE COVERAGE PROBABILITY OF SATTERTHWAITE CONFIDENCE INTERVALS FOR ASSAY PRECISION WITH UNBALANCED DATA**

Journal:	<i>Communications in Statistics - Simulation and Computation</i>
Manuscript ID:	LSSP-2009-0221.R1
Manuscript Type:	Original Paper
Date Submitted by the Author:	20-Jan-2010
Complete List of Authors:	Van den Heuvel, Edwin; University of Groningen, Faculty of Mathematics & Natural Sciences
Keywords:	Henderson III Moment Estimators, (Restricted) Maximum Likelihood Estimators, Variance Components, Unbalanced Data, Satterthwaite Confidence Intervals, Assay Precision
Abstract:	Construction of closed-form confidence intervals on linear combinations of variance components were developed generically for balanced data and studied for simple analysis of variance models. Satterthwaite approach is generalized to unbalanced data and modified to increase its coverage probability. They are applied on measures of precision in combination with (restricted) maximum likelihood and Henderson III Type 1 & 3 estimation. Simulation studies do not show superiority of any of the possible combinations of estimation methods and Satterthwaite approaches on three precision measures. However, the modified Satterthwaite approach with Henderson III Type 3 estimation is often preferred above the other combinations.

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60



For Peer Review Only

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

**A COMPARISON OF ESTIMATION METHODS ON THE COVERAGE  
PROBABILITY OF SATTERTHWAITE CONFIDENCE INTERVALS FOR ASSAY**

**PRECISION WITH UNBALANCED DATA**

Deleted: MISSING

Edwin R. van den Heuvel<sup>2</sup>

Deleted: <sup>3</sup>

Faculty of Mathematics & Natural Sciences,

Deleted: Statistical Department,

University of Groningen,

Formatted: Dutch (Netherlands)

Nijenborg 9, 9747 AG Groningen,

Deleted: Schering-Plough

The Netherlands

Deleted: Molenstraat 110

Deleted: 5340

Formatted: Dutch (Netherlands)

Deleted: BH Oss

Formatted: Dutch (Netherlands)

**Abstract**

Construction of closed-form confidence intervals on linear combinations of variance components were developed generically for balanced data and studied mainly for one-way and two-way random effects analysis of variance models. The Satterthwaite approach is easily generalized to unbalanced data and modified to increase its coverage probability. They are applied on measures of assay precision in combination with (restricted) maximum likelihood and Henderson III Type 1 & 3 estimation.

Deleted: classification

Deleted: using

Deleted: missing

Simulations of interlaboratory studies with unbalanced data and with small sample sizes do not show superiority of any of the possible combinations of estimation methods and Satterthwaite approaches on three measures of assay precision. However, the modified Satterthwaite approach with Henderson III Type 3 estimation is often preferred above the other combinations.

**KEY WORDS:**

Henderson III Moment Estimators, Likelihood Estimators, Measures of Precision, Satterthwaite Confidence Intervals, Unbalanced data, Variance Components.

Deleted: Missing Data,

**1. INTRODUCTION**

Statistical experimentation is frequently used to estimate measures of assay precision, which describe the variation induced by the measurement or assay processes. For example, method transfer studies

<sup>2</sup> This research has been sponsored by the Dutch Organization of Scientific Research (NWO) under Casimir project 018.002.029 in collaboration with the Probability and Statistics group of the Eindhoven University of Technology in the Netherlands.

(or interlaboratory studies) are performed regularly in pharmaceutical industry for biological and (bio)chemical assays to simulate the assay variation within and between laboratories for routine use of these assays. Studies like this focus on the measures of precision “repeatability”, “intermediate precision” and “reproducibility”, see for instance the international guideline ICH (1996). In statistical terms, measures of precision are defined by specific linear combinations of variance components which represent the random effects of (combinations of) assay sources of variation. The variance components are generally defined through the effects of sources of variation incorporated in the mixed or random effects analysis of variance models used to analyze the experimental data. The linear combinations are often the sum of individual variance components.

The choice of estimation method for the individual variance components in unbalanced designs has been debated in the statistical literature for a long time. Henderson (1953) provided three methods which are based on the moments of particular sets of sums of squares. The moment estimators are not unique because there are different ways of partitioning the sums of squares of sources of variation in unbalanced designs. Henderson’s third method is most generally applicable, since it can be applied to any random or mixed effects analysis of variance model. But also for this method there exist different sets of sums of squares. The order of including terms for the sources of variation in the analysis of variance model determines the different sets of sums of squares, see Searle, Casella and McCulloch (1992) and Littell, Milliken, Stroup, Wolfinger and Schabenberger (2006). All these choices reduce to the same unique set of sums of squares for balanced designs, which then possesses nice properties. Indeed, the sums of squares are mutually independently distributed and they are distributed as multiples of chi-square random variables. In unbalanced designs these properties no longer hold in general.

Deleted: , which may often be induced by missing data,

Deleted: of

Restricted maximum likelihood (REML) estimation is often preferred above other methods of estimation for unbalanced data, see McCulloch and Searle (2001). REML estimation provides an unique set of non-negative variance component estimates, but it does require the (strong) assumption of normality. The argument for REML estimation is for instance supported by Swallow and Monahan (1984). Their simulation study for the one-way (balanced) random effects analysis of variance model

Deleted: ors

Deleted: are

Deleted: for

Deleted: non-negative and

Deleted: solution

Deleted: s

concludes that the likelihood based estimators have in general similar or smaller biases and mean squared errors than the moment estimators.

Whatever estimation method is applied, it is good statistical practice to construct appropriate confidence intervals on assay precision. Burdick and Graybill (1992) give a nice overview of several generic approaches, but the closed-form approaches are essentially developed for balanced analysis of variance models because they require the nice properties of the sums of squares in balanced designs. Not much attention has yet been given to confidence intervals for measures of precision for unbalanced data, except though for the one-way random effects analysis of variance model, see Li and Li (2005) and Burdick, Quiroz and Iyer (2006), because then the mathematical complexity is limited. However, the Satterthwaite approach, described by for instance Satterthwaite (1946), Welch (1956) or Burdick and Graybill (1992), can easily be adapted to general unbalanced designs and it can also be applied to any method of variance component estimation. This original approach can then also be modified to increase its coverage probability and make it possibly conservative, because the original Satterthwaite approach is known to lead to too small coverage probabilities in some cases.

Deleted: y

Deleted: . For such models,

Deleted: only

Deleted: , but

Deleted: Furthermore,

Deleted: it will

Deleted: may

Deleted: in some cases

The goal of this article is to study *coverage probabilities* of the original and modified Satterthwaite approach to confidence intervals on measures of assay precision for unbalanced data using maximum likelihood, restricted maximum likelihood, Henderson III Type1 and Henderson III Type 3 variance component estimators. Simulation studies are conducted to investigate three measures of precision for a particular three-way classification random effects analysis of variance model. This model is often applied to data of interlaboratory studies, see Nijhuis and Van den Heuvel (2007). Emphasize is given to experiments with small sample sizes and to situations where individual variance components may be close to zero. A detailed description of the original and modified Satterthwaite approach and a detailed description of the four estimation methods are provided too.

Deleted: two

Deleted: es

Deleted: in the presence of missing

Deleted: (i.e. unbalanced data sets)

Deleted: four different estimation methods for individual variance components. The selected estimation methods are

Deleted: , see Littell *et al.* (2006).

Deleted: which is

## 2. GENERAL THEORY

Suppose a random or mixed effects analysis of variance model is used to describe the data from the (unbalanced) experimental (interlaboratory) study to estimate the assay precision. The random effects

are categorized into an “assay related” group and an “assay unrelated” group of variation sources.

The general statistical model is then presented by

$$(1) \quad y = X\beta + \sum_{i=1}^n U_i \gamma_i + \sum_{j=1}^m V_j \varepsilon_j,$$

with  $y$  the vector of  $N$  observations,  $X$  an  $(N \times p)$  matrix of known constants of rank  $p$ ,  $\beta$  a vector of  $p$  unknown fixed parameters,  $U_i$  an  $(N \times q_i)$  matrix of known constants,  $\gamma_i$  a random vector of  $q_i$  independent normally distributed “assay unrelated” variables with mean 0 and variance  $\sigma_{m+n-i+1}^2$ ,  $V_j$  an  $(N \times r_j)$  matrix of known constants,  $\varepsilon_j$  a random vector of  $r_j$  independent normally distributed “assay related” variables with mean 0 and variance  $\sigma_{m-j+1}^2$ , and all random vectors mutually independently distributed.

The variance components for the “assay related” group are now defined by  $\sigma_1^2, \sigma_2^2, \dots, \sigma_m^2$ , with  $\sigma_1^2$  the “residual” variance component, representing the last term in model (1). The  $n$  variance components for the “assay unrelated” group are indexed by  $\sigma_{m+1}^2, \sigma_{m+2}^2, \dots, \sigma_{m+n}^2$ . The order in the indices for the variance components is directly related to the inverse order in the random effects in model (1). For all or most assay precision studies, the contribution in the variation of the observations  $y$  is firstly determined by “assay unrelated” sources of variation and secondly by “assay related” sources of variation. Indeed, differences between for instance batches will enter the response before it is measured and then it will be (possibly) increased when it is observed with the assay. Within the “assay related” group, a similar reasoning and ordering can often be made, but within the “assay unrelated” group this is less relevant since we are interested in the estimation of the assay precision.

Deleted: by

Deleted: . For

Deleted: however

The “assay related” group may contain random effects that represent interaction effects between “assay related” sources of variation and “assay unrelated” sources of variation, since they are considered to represent assay variation. For instance, the interaction effect between batches and laboratories is considered part of the assay variation, because it will be assumed that batches are homogeneous and their mean levels are constant during the time period of the study.

Deleted: also

Deleted: ing

A measure of assay precision is now defined by the linear combination of variance components for the first  $m_0$  random effects in the “assay related” group, i.e.

$$(2) \quad \gamma_{m_0}^2 = \sum_{q=1}^{m_0} c_q \sigma_q^2, \quad c_q \geq 0.$$

The coefficients in (2) are considered non-negative since we are interested in (part of) the total contribution of assay variation. Moreover, the coefficients are often equal to one, but other choices may be applicable in particular when a “reportable value” is some kind of average from replicated results. For repeatability,  $m_0$  is usually one, while for intermediate precision  $m_0$  is selected in such a way that all sources of variation contributing to the within laboratory variation (intermediate precision) are included in (2). For reproducibility  $m_0$  is equal to  $m$ , including all assay related sources of variation. Thus the different measures of assay precision are strictly ordered in size.

### 2.1 HENDERSON III MOMENT ESTIMATORS

For the terms contained in the analysis of variance model a set of “mean squares” may be defined, see Henderson (1953) or Searle *et al.* (1992). Let  $M_k$  be the mean square for the source of variation related to variance component  $\sigma_k^2$ . A characteristic of the set of mean squares for Henderson's third method is that the expectation (or expected mean square) is of the form

$$(3) \quad E(M_k) = \theta_k = \sum_{q=1}^k a_{k,q} \sigma_q^2.$$

Here E is the symbol for expectation and  $a_{k,1} = 1, a_{k,2} \geq 0, \dots, a_{k,k-1} \geq 0, a_{k,k} > 0$  are known constants related to the matrices for the random effects in model (1). Due to the order or hierarchy of the terms in the model, the variance components  $\sigma_{k+1}^2, \sigma_{k+2}^2, \dots, \sigma_{m+n}^2$  are not involved in the expectation of  $M_k$  and are therefore not present in (3), see page 104 of Searle *et al.* (1992). For balanced designs and normally distributed observations the statistic  $X_k \equiv n_k M_k / \theta_k$  follows a Chi-square distribution with  $n_k$  degrees of freedom. The number of degrees of freedom  $n_k$  is related to the number of levels  $q_k$  or  $r_k$  for the source of variation in model (1) for the “assay unrelated” or “assay related” group, respectively.

Deleted: variation

Deleted:

Deleted: III

Deleted: , depending on whether  $M_k$  belongs to

Deleted: sources of variation

Based on the expected mean squares in (3), the variance components can be expressed as linear combinations of the expected mean squares, i.e.

$$(4) \quad \sigma_k^2 = \sum_{q=1}^k \tilde{a}_{k,q} E(M_q),$$

with  $\tilde{a}_{k,1}, \tilde{a}_{k,2}, \dots, \tilde{a}_{k,k}$  known functions of the constants  $a_{1,1}, a_{2,1}, a_{2,2}, a_{3,1}, \dots, a_{3,3}, \dots, a_{m_0,1}, \dots, a_{m_0,m_0}$ . Then the *standard moment estimator* for the variance component  $\sigma_k^2$  denoted as  $\hat{\sigma}_k^2$  is obtained by substituting  $M_q$  for the expected mean square  $E(M_q)$  in (4).

Substituting the standard moment estimators for the variance components in (2) and using relation (4) we obtain the moment estimator  $\hat{\gamma}_{m_0}^2$  for the measure of precision in (2), i.e.

$$(5) \quad \hat{\gamma}_{m_0}^2 \equiv \sum_{q=1}^{m_0} c_q \hat{\sigma}_q^2 = \sum_{q=1}^{m_0} \tilde{c}_q M_q,$$

where  $\tilde{c}_1, \tilde{c}_2, \dots, \tilde{c}_{m_0}$  are known functions of the constants  $c_1, c_2, \dots, c_{m_0}$  and  $\tilde{a}_{1,1}, \tilde{a}_{2,1}, \tilde{a}_{2,2}, \tilde{a}_{3,1}, \dots, \tilde{a}_{3,3}, \dots, \tilde{a}_{m_0,1}, \dots, \tilde{a}_{m_0,m_0}$ . For balanced designs the constants  $\tilde{c}_1, \tilde{c}_2, \dots, \tilde{c}_{m_0}$  in (4) are non-negative when  $c_1 = c_2 = \dots = c_{m_0} = 1$ , but this is not necessarily true for unbalanced situations.

Deleted: a

The first moment of the estimator in (5) is of course the true measure of precision in (2), since the estimator is based on the first moments of the mean squares. The variance of the estimator in (5) is of the form

$$(6) \quad \sigma_{\hat{\gamma}_{m_0}^2}^2 = E(\hat{\gamma}_{m_0}^2 - \gamma_{m_0}^2)^2 = \sum_{q=1}^{m_0} c_q^2 \text{var}(\hat{\sigma}_q^2) + 2 \sum_{q=1}^{m_0-1} \sum_{r=q+1}^{m_0} c_q c_r \text{cov}(\hat{\sigma}_q^2, \hat{\sigma}_r^2).$$

which in case the experimental design is balanced reduces to

$$\sigma_{\hat{\gamma}_{m_0}^2}^2 = 2 \cdot \sum_{q=1}^{m_0} \tilde{c}_q^2 \theta_q^2 / n_q.$$

This follows from the Chi-square distribution related to the mean squares  $M_k$ .

It is well known that, except for the residual variance, the standard moment estimates for individual variance components can become negative for both balanced and unbalanced designs, see for instance Searle *et al.* (1992) and McCulloch and Searle (2001). Although the standard moment estimators are unbiased, they have the disadvantage that different and specifically ordered measures of precision may become incorrectly ordered when the standard moment estimators are used, due to the negative variance component estimates for individual variance components. For example, the estimator in (5) may be less than the standard moment estimator  $M_1$  for the residual variance  $\sigma_1^2$ . Nijhuis and Van den Heuvel (2007) proposed to use the *adjusted* or *truncated* moment estimators to maintain a proper order in the estimates for the measures of precision.

The adjusted moment estimators are defined by the standard moment estimators truncated to non-negative variance components estimators. This method of estimation is similar to the (but not exactly the same as) restricted maximum likelihood estimation method, see Subsection 2.2. The estimator for the measure of precision when the adjusted moment estimators are applied becomes

$$(7) \quad \tilde{\gamma}_{m_0}^2 \equiv \sum_{q=1}^{m_0} c_q \hat{\sigma}_q^2 \cdot 1_{[0, \infty)}(\hat{\sigma}_q^2),$$

with  $\hat{\sigma}_k^2$  the standard moment estimator for the variance component  $\sigma_k^2$  and  $1_A(x)$  the indicator function being 1 if  $x$  is an element of set  $A$  and zero elsewhere. The adjusted moment estimators will be used for the estimation of measures of precision here as well.

Henderson III moment estimators are not unique. Different ways of positioning the sums of squares are possible. The positioning has much to do with the order in which the statistical terms enter the statistical model (1), see for instance Searle *et al.* (1992). Although general rules may not be given, for assay precision studies the order of terms in the model is not at all complicated. Recall that the contribution of variation into the observations is induced by the “assay unrelated” sources of variation first and then followed by the “assay related” sources of variation. More specific, the order of terms in the model is usually related to the order in time in which variation may have entered the response. This Henderson III estimation method is referred to the Type 1 estimation method, see Searle *et al.* (1992) and Littell *et al.* (2006). Another way is the Type 3 method, which calculates the additional sum of squares of each term when it would be added to the model as the last term (but remaining the

Deleted: first

Deleted: been contributed.

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

identifiability of the terms of the full model). For this type of estimation the order of terms in model (1) is not relevant at all, see Littell *et al.* (2006). It should be clear that the residual variance component estimator  $\hat{\sigma}_1^2$  is the same for both Type 1 and Type 3 estimation method. It is also the only variance component estimator which always follows a Chi-square distribution (when appropriately normalized).

It is expected that the Type 3 estimation method is less sensitive to unbalanced data than the Type 1 method, since it is less susceptible to extreme results, in particular for small sample sizes.

Deleted: missing

Deleted: in small studies

## 2.2 MAXIMUM LIKELIHOOD BASED ESTIMATORS

Maximum likelihood estimation for random and mixed effects analysis of variance models is an estimation method that determines the fixed and random model parameters by maximizing the likelihood function for the selected linear model. A disadvantage of maximum likelihood estimation is that the variance component estimators are biased, see McCulloch and Searle (2001). The restricted maximum likelihood (REML) estimation method for estimating variance components in random and mixed effects analysis of variance models is the maximum likelihood estimation method applied to a linear transformation of the observations which do not contain any of the fixed effects anymore (expectation zero), see page 176 of McCulloch and Searle (2001).

Deleted: linear

Deleted: selects

For balanced data sets and non-negative standard moment estimates for variance components, the REML estimates are equal to the standard moment estimates, see page 177 of McCulloch and Searle (2001). REML estimates and maximum likelihood estimators though are always non-negative even if (some of) the standard moment estimates are negative. The non-negative estimates are essentially obtained by removing or dropping the appropriate terms from the random or mixed effects analysis of variance model, see Searle *et al.* (1992), page 235. This implies that the restricted maximum likelihood estimators are indeed similar to the adjusted moment estimators, with the small distinction that the adjusted moment estimation method does not drop any terms from the model.

Formatted: Font: Italic

Deleted: standard

Dropping terms from the model is in our opinion a (philosophical) disadvantage of the (restricted) maximum likelihood estimation method, in particular if small sample sizes are used. Indeed, for small sample sizes relevant contributions of variation sources may not necessarily be positively estimated since they could be less in size with respect to the contribution of other sources of variation.

Deleted: substantial

Deleted: are relatively

Deleted: substantial

Deleted: sources.

Eliminating this relevant source of variation on the basis of an observed negative variance component estimate is then actually incorrectly allocating the effect of this source of variation to other sources of variation. Thus a possible consequence of the (restricted) maximum likelihood estimation method is that not only the “truncated” variance component estimators are biased (due to the truncation itself), but also several other variance component estimators which has incorporated the variation of the eliminated sources of variation. This would be the case for both balanced as unbalanced designs. On the other hand, this disadvantage does not imply that the (restricted) maximum likelihood estimation method does not work very well in practice.

**Deleted:** Then altering the linear model due to  
**Deleted:** e variation to  
**Deleted:** s  
**Deleted:** with relative more substantial

**Deleted:** from other not related

Using the likelihood equations (cf. Searle *et al.* (1992)), the likelihood based estimation method would provide estimators  $s_1^2, s_2^2, \dots, s_{m_0}^2$ , which could represent ML or REML estimators, for the variance components  $\sigma_1^2, \sigma_2^2, \dots, \sigma_{m_0}^2$ . Let  $\hat{\gamma}_{m_0}^2$  be the (restricted) maximum likelihood estimator of the measure of precision given by

$$(8) \quad \hat{\gamma}_{m_0}^2 = \sum_{q=1}^{m_0} c_q s_q^2.$$

The first moment for the (restricted) maximum likelihood estimator in (8) is not necessarily an unbiased estimator for the true measure of precision. However, asymptotically the (restricted) maximum likelihood estimator in (8) will be unbiased when sample sizes converge to infinity in an appropriate way. The variance of the estimator in (8) is of the same form as (6), i.e.

$$(9) \quad \sigma_{\hat{\gamma}_{m_0}^2}^2 = E(\hat{\gamma}_{m_0}^2 - \gamma_{m_0}^2)^2 = \sum_{q=1}^{m_0} c_q^2 \text{var}(s_q^2) + 2 \sum_{q=1}^{m_0-1} \sum_{r=q+1}^{m_0} c_q c_r \text{cov}(s_q^2, s_r^2)$$

The variances and covariances in (9) are obtained by asymptotics using Fisher's Information matrix or the inverse Hessian matrix, which is defined by the second derivatives of the likelihood function with respect to the model parameters, see Searle *et al.* (1992) and Littell *et al.* (2006).

**Deleted:** based on the  
**Deleted:** variance  
**Deleted:** the  
**Deleted:**  
**Deleted:** of the

### 2.3 SATTERTHWAITE APPROACH

The approach of Satterthwaite (1946) is quite simple and is based on an approximation of the estimator as a Chi-square variable. Let  $T_{m_0}$  be an arbitrary estimator for the measure of precision in

(2) and let  $\tau_{T_{m_0}}$  be its corresponding standard error. The Satterthwaite approach approximates the

distribution of the estimator  $T_{m_0}$  with the distribution of the random variable  $E(T_{m_0}) \cdot X / df_S$ , where

$E(T_{m_0})$  is the expected value of the estimator  $T_{m_0}$ , with  $X$  a Chi-square distributed variable, and

with  $df_S$  the number of degrees of freedom defined by

$$df_S = 2 \cdot (E(T_{m_0}) / \tau_{T_{m_0}})^2.$$

Observe that this number of degrees of freedom makes the first two moments of  $T_{m_0}$  and

$E(T_{m_0}) \cdot X / df_S$  exactly equal and that the expected value  $E(T_{m_0})$  does not necessarily have to be

unbiased for the measure of precision  $\gamma_{m_0}^2$  in (2). Since the number of degrees of freedom is

unknown it should be estimated before it can be used in practice. The usual estimator is

$$(10) \quad \hat{df}_S = 2 \cdot (T_{m_0} / \hat{\tau}_{T_{m_0}})^2,$$

where  $\hat{\tau}_{T_{m_0}}$  is an estimator for the standard error  $\tau_{T_{m_0}}$ . This approximation can be applied to any

estimator and not just to estimators for the measures of precision, but the approximation seems to

make sense when the estimator is somewhat related to a Chi-square distributed variable as for

measures of precision.

Deleted: and

Deleted: The approximation works best when the tails of the distributions of  $T_{m_0}$  and  $E(T_{m_0}) \cdot X / df_S$  resemble each other.

Confidence intervals are now constructed using the standard theory for Chi-square distributed random

variables. A  $(1 - \alpha)100\%$  confidence interval for the measure of precision is defined by

$$(11) \quad \left[ \hat{df}_S \cdot T_{m_0} / \chi_{\hat{df}_S}^{-2}(1 - \alpha/2) ; \hat{df}_S \cdot T_{m_0} / \chi_{\hat{df}_S}^{-2}(\alpha/2) \right]$$

with  $\chi_d^{-2}(p)$  is the  $p^{\text{th}}$  percentile of the Chi-square distribution with  $d$  degrees of freedom. The

approximate confidence interval in (11) works best when the tails of the distributions of  $T_{m_0}$  and

$E(T_{m_0}) \cdot X / df_S$  resemble each other.

In the literature, the intuitive interpretation of the number of degrees of freedom has never been

connected with Satterthwaite approach, because Satterthwaite approach, has only been described as

a way of approximating the variance estimator at hand, whatever the outcome of the number of

degrees of freedom. But, the number of degrees of freedom also indicates the available amount of

Deleted: on Satterthwaite's approximation

Deleted: of

Deleted: referenced

Deleted: it

Deleted: However,

Deleted:

independent information used to estimate a variance component from observations of a normal distribution. Thus for a data set with  $N$  observations, the maximum number of degrees of freedom would be  $N - 1$ . Obviously, the minimum number of degrees of freedom to estimate a variance component would be at least equal to one since it takes at least two observations to estimate a variance component. Thus a consequence of the intuitive interpretation is that the number of degrees of freedom in (10) should be bounded from below by (at least) 1 and from above with (at most)  $N - 1$ , because the data does not contain any less or any more information than 1 and  $N - 1$ , respectively.

Deleted: information present in the data or the amount of

Deleted: se

Deleted: s

Deleted: indicate

The literature on Satterthwaite approximation has described its performance and it does not always provide good coverage probabilities for the estimation of linear combinations of variance components, see Burdick and Graybill (1992). The performance becomes less when one or more of the coefficients  $c_1, c_2, \dots, c_{m_0}$  are negative. However, for positive coefficients, the performance is diminished as well when sample sizes and variance components vary substantially in size. Some researchers have tried to improve the Satterthwaite approximation. Ames and Webster (1991) tried to correct the bias in the estimator for the number of degrees of freedom by changing the estimator in (10). They indeed improved the estimator, but their solution was developed for the situation that the estimator in (7) contains only two mean squares ( $m_0 = 2$ ). A generalization of this work seems not straightforward. Zou and McDermott (1999) extended the Satterthwaite approach by including higher order moments of the Chi-square distribution to obtain the best number of degrees of freedom. Their simulation study shows that the coverage probability becomes conservative, which may be positive for some situations. However, the complexity to unbalanced experimental designs will become too high for practical considerations.

The main disadvantage of Satterthwaite's approximation is in our view the use of the "estimate" for the number of degrees of freedom. We believe that the method would perform better when the number of degrees of freedom does not have to be estimated. Practically, this is of course impossible, but it indicates that the variation from the estimator may have a negative effect on the performance of the coverage probability. Thus limiting the variation of the estimator, by using the bounds 1 and

Deleted: tighter constraints than

$N-1$  or even tighter constraints may certainly be helpful in increasing the performance of Satterthwaite's approximation.

### **Henderson III Moment Estimators**

For the construction of confidence intervals for the measure of precision in (2) using Henderson III moment estimators, the estimator  $T_{m_0}$  in (10) and (11) should first be replaced by the adjusted moment estimator  $\tilde{\gamma}_{m_0}^2$  from (7). The first and second moments of this estimator can be established for balanced designs but they become too complicated for unbalanced designs, due to the loss of nice properties of the mean squares. Therefore, we will approximate the standard error by

$$(12) \quad \sigma_{\tilde{\gamma}_{m_0}^2}^2(1) = \sum_{q=1}^{m_0} c_q^2 \text{var}(\hat{\sigma}_q^2)_{[0,\infty)}(\hat{\sigma}_q^2) + 2 \sum_{q=1}^{m_0-1} \sum_{r=q+1}^{m_0} c_q c_r \text{cov}(\hat{\sigma}_q^2, \hat{\sigma}_r^2)_{[0,\infty)}(\hat{\sigma}_q^2)_{[0,\infty)}(\hat{\sigma}_r^2),$$

with  $\hat{\sigma}_q^2$  the standard moment estimator for the variance component  $\sigma_q^2$ . The variances and covariances in (12) can be estimated, see Searle *et al.* (1992). The estimates will be denoted by  $\hat{v}_{q,r}$  with  $v_{q,q} = \text{var}(\hat{\sigma}_q^2)$  and  $v_{q,r} = \text{cov}(\hat{\sigma}_q^2, \hat{\sigma}_r^2)$ , when  $q \neq r$ . The estimator for the standard error of the moment estimator  $\tilde{\gamma}_{m_0}^2$  in (7) is now denoted by  $\hat{\sigma}_{\tilde{\gamma}_{m_0}^2}^2(1)$ . This standard error reduces to the usual form in Satterthwaite (1946) for balanced designs, when all variance component estimates are non-negative.

The adjusted moment estimator  $\tilde{\gamma}_{m_0}^2$  in (7) is essentially a function of the mean squares  $M_1, M_2, \dots, M_{m_0}$ . This means that the number of degrees of freedom for the estimator  $\tilde{\gamma}_{m_0}^2$  is at least  $\min\{df_1, df_2, \dots, df_{m_0}\}$  and at most  $df_1 + df_2 + \dots + df_{m_0}$ , with  $df_k$  the "number of degrees of freedom" corresponding to the mean square  $M_k$ . These bounds indeed indicate the minimum and maximum amount of information present in the mean squares.

Using these constraints and the Satterthwaite approximation, the estimated number of degrees of freedom  $\hat{df}_S$  in (11) should be replaced by,

$$(13) \quad \hat{df}_M(1) = \max \left\{ \min \{df_1, df_2, \dots, df_{m_0}\}, \min \{df_1 + df_2 + \dots + df_{m_0}, 2 \cdot \tilde{\gamma}_{m_0}^4 / \hat{\sigma}_{\tilde{\gamma}_{m_0}}^2(1)\} \right\}.$$

It should be noted that Nijhuis and Van den Heuvel (2007) did use a different estimator for the number of degrees of freedom in (13). They replaced  $\hat{\sigma}_{\tilde{\gamma}_{m_0}}^2(1)$  in (13) by the estimator  $\hat{\sigma}_{\tilde{\gamma}_{m_0}}^2$ , which can be obtained by substituting the estimators  $\hat{v}_{q,r}$  in (6). The constraints though were still the same because it is a property of the Satterthwaite approximation for balanced designs. It is expected that the estimated number of degrees of freedom in (13) would be lower than or equal to the choice given by Nijhuis and Van den Heuvel (2007). Hence, the approach here will result usually in larger or equal coverage probabilities. We consider this as beneficial, because the literature reports that one of the weaknesses of Satterthwaite approach is that the confidence intervals produces coverage probabilities which are too low in some cases, e.g. Burdick and Graybill (1992).

To accommodate this even further, one could argue that the standard error for the moment estimator  $\tilde{\gamma}_{m_0}^2$  in (7) is better approximated by only the first sum in (12), because this term usually dominates the standard error and because the covariances in (12) are expected to be negative as well. Thus an alternative or modified approximate standard error can be defined by

$$(14) \quad \hat{\sigma}_{\tilde{\gamma}_{m_0}}^2(2) = \sum_{q=1}^{m_0} c_q^2 \text{var}(\hat{\sigma}_q^2) \mathbb{I}_{[0,\infty)}(\hat{\sigma}_q^2).$$

Still using the same constraints as in (13), the modified Satterthwaite approach provides the following number of degrees of freedom.

Deleted: would then become

$$(15) \quad \hat{df}_M(2) = \max \left\{ \min \{df_1, df_2, \dots, df_{m_0}\}, \min \{df_1 + df_2 + \dots + df_{m_0}, 2 \cdot \tilde{\gamma}_{m_0}^4 / \hat{\sigma}_{\tilde{\gamma}_{m_0}}^2(2)\} \right\},$$

with  $\hat{\sigma}_{\tilde{\gamma}_{m_0}}^2(2)$  the estimator for  $\sigma_{\tilde{\gamma}_{m_0}}^2(2)$  in (14), by substituting  $\hat{v}_{q,q}$  for  $\text{var}(\hat{\sigma}_q^2)$  in (14).

It should be noted that for repeatability ( $m_0 = 1$ ), the lower and upper constraint on the number of degrees of freedom become equal to  $df_1$ , the number of degrees of freedom for the residual variance. Thus the selected Satterthwaite approximation with (13) and (15) both become exact in case the estimator is a true mean square, such as for the residual variance component.

For the Henderson III moment estimators there are now four ways of constructing confidence intervals

for the measure of precision in (2). The two estimation methods Type 1 and Type 3 will be combined with both the original and modified Satterthwaite approach for constructing standard errors of the estimator of the measure of precision, see the number of degrees of freedom in (13) and (15).

Deleted: , through t

Deleted: combination of the

Deleted: and

Deleted: es

Deleted: to

Deleted: for

### Maximum Likelihood Based Estimators

The variances and covariances in (9) can be estimated based on the inverse of the Hessian of the likelihood function or the Fisher Information matrix, see Searle et al. (1992). If these estimators are (also) denoted by  $\hat{v}_{q,r}$ , with  $v_{q,q} = \text{var}(s_q^2)$  and  $v_{q,r} = \text{cov}(s_q^2, s_r^2)$  when  $q \neq r$ , then the estimator for the standard error in (9) for the (restricted) maximum likelihood estimator  $\hat{\gamma}_{m_0}^2$  in (8) is then  $s_{\hat{\gamma}_{m_0}^2}^2(1)$ , which is obtained by substituting the estimators  $\hat{v}_{q,r}$  into (9), i.e.

$$(16) \quad s_{\hat{\gamma}_{m_0}^2}^2(1) = \sum_{q=1}^{m_0} c_q^2 \hat{v}_{q,q} + 2 \sum_{q=1}^{m_0-1} \sum_{r=q+1}^{m_0} c_q c_r \hat{v}_{q,r}.$$

The confidence interval for the measure of precision with the (restricted) maximum likelihood estimator is now obtained by (11), with  $T_{m_0}$  replaced by  $\hat{\gamma}_{m_0}^2$  and the number of degrees of freedom  $df$  in (11) replaced by

$$(17) \quad df_L(1) = \max\left\{1, \min\left\{N-1, 2 \cdot \hat{\gamma}_{m_0}^4 / s_{\hat{\gamma}_{m_0}^2}^2(1)\right\}\right\}.$$

Similar to the reasoning for the Henderson III moment estimators, the dominating term in (16) would be the first sum alone. The confidence interval would then be constructed with the modified number of degrees of freedom

$$(18) \quad df_L(2) = \max\left\{1, \min\left\{N-1, 2 \cdot \hat{\gamma}_{m_0}^4 / s_{\hat{\gamma}_{m_0}^2}^2(2)\right\}\right\}$$

with  $s_{\hat{\gamma}_{m_0}^2}^2(2)$  given by the modified estimator  $s_{\hat{\gamma}_{m_0}^2}^2(2) = c_1^2 \hat{v}_{1,1} + c_2^2 \hat{v}_{2,2} + \dots + c_{m_0}^2 \hat{v}_{m_0,m_0}$ .

For the likelihood based estimators, the constraints 1 and  $N-1$  are used instead of "degrees of freedom" which are used for the moment estimators. The likelihood based estimators do not have the

1 same interpretation as the moment estimators in unbalanced designs, see Searle *et al.* (1992).  
 2  
 3 Indeed, due to the possibility of dropping terms from the model for variance components which are  
 4 estimated at zero, changes the “number of degrees of freedom” for the estimation of other variance  
 5 components. This would also hold for repeatability, i.e.  $m_0 = 1$ . Another reason is that even the  
 6  
 7 residual variance component estimator, under the assumption that the “full” analysis of variance  
 8 model would represent the data, does not follow a chi-square distribution in general and the  
 9 expectation of the estimator may include variance components which are unrelated to the residual  
 10 variance component. Although it may be possible to narrow the constraints in (17) and (18) with the  
 11 use of the “number of levels” for each term in the model, we have chosen to use only the boundaries  
 12 in (17) and (18), which would always be correct in all situations (as described in the beginning of this  
 13 subsection).  
 14  
 15  
 16  
 17  
 18  
 19  
 20  
 21

22 For the likelihood based estimation method there are also four ways to construct confidence intervals  
 23 for the measure of precision in (2). Both the maximum likelihood and the restricted maximum  
 24 likelihood method may be used with the original and modified Satterthwaite approach for the standard  
 25 error, of the estimator of the measure of precision, see the number of degrees of freedom in (17) and  
 26  
 27  
 28  
 29  
 30  
 31  
 32  
 33

Deleted: es

Deleted: s

Deleted: for

### 3. SIMULATION STUDY

34 A simulation study was conducted using SAS<sup>®</sup> software package to investigate the performance of the  
 35 estimation methods with the original and modified Satterthwaite approach to construct confidence  
 36 intervals on the measures of assay precision with unbalanced data. The performance measure for the  
 37 simulation study is the coverage probability for 95% confidence intervals. In line with Jennings (1987)  
 38 the left tail and right tail probabilities are investigated too, to determine the symmetry or bias of the  
 39 confidence intervals. Of particular interest is the performance of these methods in small but frequently  
 40 applied interlaboratory studies.  
 41  
 42  
 43  
 44  
 45  
 46  
 47  
 48  
 49  
 50  
 51  
 52  
 53  
 54  
 55  
 56  
 57  
 58  
 59  
 60

Deleted: when missing data occurs.

Deleted: often

The simulated statistical model is a three-way random effects analysis of variance model which was  
 already discussed by Nijhuis and Van den Heuvel (2007) for balanced designs. For unbalanced data  
the model is described by

Deleted: When missing data

Deleted: is presented, it

(19)  $y_{hijk} = \mu + \alpha_h + \beta_i + (\alpha\beta)_{hi} + \delta_{j(i)} + (\alpha\delta)_{hj(i)} + \varepsilon_{hijk}$ ,

with  $h = 1, 2, \dots, H$ ,  $i = 1, 2, \dots, I$ ,  $j = 1, 2, \dots, J$ , and  $k = 1, 2, \dots, K_{hij}$  and with

$\mu$ : the overall mean,

$\alpha_h$ : the effect of batch  $h$ ,  $\alpha_h \sim N(0, \sigma_6^2)$ ,

$\beta_i$ : the effect of laboratory  $i$ ,  $\beta_i \sim N(0, \sigma_5^2)$ ,

$(\alpha\beta)_{hi}$ : the interaction effect of batch  $h$  and laboratory  $i$ ,  $(\alpha\beta)_{hi} \sim N(0, \sigma_4^2)$ ,

$\delta_{j(i)}$ : the effect of analytical run  $j$  within laboratory  $i$ ,  $\delta_{j(i)} \sim N(0, \sigma_3^2)$ ,

$(\alpha\delta)_{hj(i)}$ : the interaction effect for batch  $h$  with analytical run  $j$  within laboratory  $i$ ,

$(\alpha\delta)_{hj(i)} \sim N(0, \sigma_2^2)$

$\varepsilon_{hijk}$ : the residual,  $\varepsilon_{hijk} \sim N(0, \sigma_1^2)$ .

and with all random terms mutually independently distributed. Note that the imbalance of the design is incorporated by the dependency of the repeated number of observations  $K_{hij}$  on batch, laboratory and run.

The five variance components  $\sigma_1^2, \sigma_2^2, \sigma_3^2, \sigma_4^2$ , and  $\sigma_5^2$  all belong to the "assay related" group,

because it is assumed that the batches are homogeneous. The variance component  $\sigma_6^2$  belongs to

the "assay unrelated" group, because it quantifies the variation between production batches. The

measures of precision repeatability, intermediate precision and reproducibility are now defined by,

Repeatability:  $\gamma_1^2 = \sigma_1^2$ ,

(20) Intermediate Precision:  $\gamma_3^2 = \sigma_1^2 + \sigma_2^2 + \sigma_3^2$ ,

Reproducibility:  $\gamma_5^2 = \sigma_1^2 + \sigma_2^2 + \sigma_3^2 + \sigma_4^2 + \sigma_5^2$ .

Five "practical cases" of parameter choices for the variance components of the statistical model are used in the simulation study and they are presented in Table 1. Table 1 also presents the true measures of precision. The variance component for batch was always kept equal to 2 and the

Deleted: a constant

Formatted: Font: (Default) Arial, 10 pt, English (U.S.), Lowered by 6 pt

Deleted:

Formatted: Font: (Default) Arial, 10 pt, English (U.S.), Lowered by 6 pt

Formatted: Font: (Default) Arial, 10 pt, English (U.S.), Lowered by 6 pt

Formatted: Font: (Default) Arial, 10 pt, English (U.S.), Lowered by 3 pt

Formatted: Font: (Default) Arial, 10 pt, English (U.S.), Lowered by 3 pt

Formatted: Font: (Default) Arial, 10 pt, English (U.S.), Lowered by 6 pt

Deleted: and

Deleted: the main random effects for batch  $h$ , laboratory  $i$ , and

Formatted: Font: (Default) Arial, 10 pt, English (U.S.), Lowered by 7 pt

Deleted: respectively,  $(\alpha\beta)_{hi}$  and

Deleted: random

Deleted: s for batch  $h$  and laboratory  $i$  and

Deleted: respectively and with

Formatted: Font: (Default) Arial, 10 pt, English (U.S.), Lowered by 7 pt

Formatted: Font: (Default) Arial, 10 pt, English (U.S.), Lowered by 7 pt

Deleted: .

Formatted: Font: (Default) Arial, 10 pt, English (U.S.), Lowered by 7 pt

Deleted: It is assumed that  $\alpha_h, \beta_i, \delta_{j(i)}, (\alpha\beta)_{hi}, (\alpha\delta)_{hj(i)}$  ... [1]

Deleted: first f

Formatted: ... [2]

Formatted: ... [3]

Formatted: ... [4]

Formatted: ... [5]

Formatted: ... [6]

Deleted: and

Deleted: t

Deleted: sixth

Formatted: ... [7]

Deleted: , respectively,

Deleted: (20) .

Deleted:

Formatted: ... [8]

Deleted: and

residual variance is omitted in the table because it is represented by repeatability. Each case represents a possible situation in practice. Case E represents a method with a dominating repeatability. Case D has a dominating lab-to-lab variation. Case B has a dominating run-to-run variation, which could be caused by analysts. Case C represents a dominating interaction between batch and run, indicating that the assay is sensitive to the batches. This case is sometimes observed for biological assays. Case A has all of these effects without being one dominant.

< Table 1 >

The selected sample sizes or experimental designs for the different variation sources used in the simulation study are presented in Table 2. These sample sizes were all combined with the practical cases from Table 1. The last column in Table 2 indicates the total number of test samples in the study to give an impression of the size of the intralaboratory experiment.

< Table 2 >

Unbalanced data in the simulation study is implemented by introducing missing data to the complete statistical model in (19) that has been simulated first. The selected mechanism is "completely missing at random" (cf. Little and Rubin (2003)) at individual results. The missing data is generated with a Bernoulli random variable indicating which result is omitted from the full study. The selected probabilities for introducing missing data were set at 0, 0.1 and 0.2. For each selected probability all combinations of sample sizes in Table 2 with all practical cases in Table 1 were considered. The number of simulation runs for a particular set of input parameter was selected at 10000, i.e. the number of times model (19) is generated. This would result in a standard error of approximately 0.0022 (or 0.22%) for the targeted coverage probability of the simulated confidence intervals.

Deleted: Missing

Deleted: after

Deleted: mechanism

Deleted: the

The coverage probabilities from the simulation study are summarized in Tables 3 to 9. Instead of presenting huge tables of coverage probabilities, only the ranges of coverage probabilities for the different sample sizes or experimental designs are given. Tables 3 and 4 present the results for intermediate precision and reproducibility for balanced data and Tables 5 and 6 present the same information for unbalanced data with 20% missing data. Each table presents the four estimation methods applied with the original (OS) and modified (MS) Satterthwaite approach and the five

different practical cases (A to E). Repeatability is not presented, because the coverage probabilities were very close to the target of 95%. Tables 7, 8, and 9 present the ranges on the left and right tail probabilities when 10% missing data is generated for repeatability, intermediate precision and reproducibility, respectively. Only the modified Satterthwaite approach is presented, but again the four estimation methods and the five practical cases are included.

Deleted: are

< Table 3 >

< Table 4 >

< Table 5 >

< Table 6 >

< Table 7 >

< Table 8 >

< Table 9 >

4. CONCLUSIONS

Deleted: 5

A few generic rules were observed from the simulation study. The original and modified Satterthwaite approaches are generally reasonable close to the target coverage probability of 95%, despite the small sample sizes and the complexity of the structure of the data. The modified Satterthwaite approach showed higher coverage probabilities than the original Satterthwaite approach and in some cases even substantially higher coverages are observed. The maximum likelihood estimation method seems to produce a little bit lower coverage probabilities than the restricted maximum likelihood method. The Henderson III Type 3 method has somewhat larger coverage probabilities than the Henderson III Type 1 method for intermediate precision and reproducibility (for unbalanced data of course). The coverage probabilities for the Henderson III methods reduce with the amount of missing data while the maximum likelihood methods are less affected. Henderson III Type 1 method is affected most by the imbalance, clearly more than any of the other estimation methods. The likelihood based estimation methods gave left tail probabilities which are mostly underestimated (asymmetry), indicating lower confidence limits that are too low, even for repeatability in some of the simulation cases. The Henderson III estimation methods gave less extreme asymmetrical coverage probabilities than the likelihood based estimation methods. When the coverage probabilities for the Henderson III

Deleted: in general

Deleted: in case of

Deleted: missing

Deleted: Type 1 method seem to strongly

Deleted:

Deleted: ,

Deleted: The modified approach showed higher coverage probabilities than the original Satterthwaite approach, which were in some case even substantially higher.

Deleted: for repeatability

Deleted: often

Deleted: e

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

methods were asymmetrical, the right tail would generally give lower probabilities than the left tail, making the upper confidence limit conservative, except for some particular simulation cases,

Deleted: trol  
Deleted: .

For balanced data, two sets of simulation cases clearly stand out, namely practical case B for intermediate precision and practical case D for reproducibility. The ML and REML estimation methods do not seem to perform very well, not even when the modified Satterthwaite approach is used, although the modified approach does increase the coverage probabilities with respect to the original Satterthwaite approach. The modified Satterthwaite approach for the Henderson III Type 1 & 3 estimation methods do seem to give more appropriate coverage probabilities for these practical cases, although the modified Satterthwaite approach shows too high coverage probabilities for the other simulation cases in general with all estimation methods.

Deleted: the  
Deleted: situation (no missing data)

Deleted: it  
Deleted: s

Deleted: es  
Deleted: reasonable  
Deleted: . However,  
Deleted: s  
Deleted: in general

For the intermediate precision in simulation case B and for reproducibility in simulation case D, the likelihood based estimation methods with the modified Satterthwaite approach improve in coverage probability when missing data occurs. Contrary to this, Henderson III Type 3 estimation method reduces in coverage probability for these simulation cases and the coverage probabilities become similar to the coverage probabilities for the REML estimation method when both methods use the modified Satterthwaite approach. For the original Satterthwaite approach though, Henderson III Type 3 estimation method remains better than REML estimation method for these two sets of simulation cases B and D for unbalanced data.

Deleted: surprisingly seem to

Deleted: almost to the

Deleted: This seems to hold true also for the other cases when missing data is present.  
Deleted: i

It is quite difficult to make a firm choice between the two Satterthwaite approaches or between any of the estimation methods. However, REML is preferred above ML estimation and Henderson III Type 3 is preferred above Henderson III Type 1 estimation. Thus either REML or Henderson III Type 3 should be selected. We believe that the simulation results give a small preference to Henderson III Type 3 estimation method using the modified Satterthwaite approach. For particular simulation cases it provides similar or better results as any other combination of estimation method with one of the two Satterthwaite approaches and it provides in general higher coverage probabilities than aimed for, which makes it a good but conservative method to construct confidence intervals. Furthermore, it provides somewhat more symmetrical coverage probabilities than the maximum likelihood based

Deleted: However, w

Deleted: s

1  
2 estimation methods, which produces asymmetrical confidence intervals even, for repeatability. The  
3  
4 REML estimation method with the modified Satterthwaite approach would be an appropriate  
5  
6 alternative in our opinin.

Deleted: also

Deleted: estimation

Deleted: in our opinin

## 8 ACKNOWLEDGEMENTS

10 I would like to thank E. Talens for verifying the SAS programs for the simulation study and O. Almalik  
11 and H. van Zwam for verifying the SAS output with the article. Furthermore, I would like to thank A. Di  
12  
13  
14 Bucchianico for critically reading the article and providing good suggestions.

Deleted: and the example

## 17 REFERENCES

- 18  
19 Ames, M.H. and Webster, J.T. (1991), On estimating approximate degrees of freedom, *The American*  
20  
21 *Statistician*, **45** (2), 45- 50.
- 22  
23 Burdick, R.K. and Graybill, F.A. (1992), *Confidence Intervals on Variance Components*, Marcel  
24  
25 Dekker, New York.
- 26  
27 Burdick, R.K., Quiroz, J. and Iyer, H.K. (2006), The present status of confidence interval estimation for  
28  
29 one-factor random models, *Journal of Statistical Planning and Inference*, **136**, 4307-4325.
- 30  
31 Henderson C.R. (1953), Estimation of variance and covariance components, *Biometrics*, **9**, 226-252.
- 32  
33 ICH Steering Committee (1996), *Q2B: Validation of Analytical Procedures: Methodology*, ICH  
34  
35 Harmonised Tripartite Guideline
- 36  
37 Jennings, D.E. (1987), How Do We Judge Confidence-Interval Adequacy, *The American Statistician*,  
38  
39 **41** (4), 335-337.
- 40  
41 Li, X. and Li, G. (2005), Confidence intervals on sum of variance components with unbalanced  
42  
43 designs, *Communications in Statistics: Theory and Methods*, **34**, 833-845.
- 44  
45 Littell, R.C., Milliken, G.A., Stroup, W.W., Wolfinger, R.D. and Schabenberger, O. (2006), *SAS® for*  
46  
47 *mixed models*, 2<sup>nd</sup> ed., SAS Institute Inc., Cary, NC, USA.
- 48  
49 Little, R.J.A. and Rubin, D.B. (2002), *Statistical Analysis with Missing Data*, 2<sup>nd</sup> ed., Wiley, New York,  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60 McCulloch, C.E. and Searle, S.R. (2001), *Generalized, Linear, and Mixed Models*, Wiley, New York.
- Nijhuis, M.B. and Van den Heuvel, E.R. (2007), Closed-form confidence intervals on measures of  
precision for an interlaboratory study, *Journal of Biopharmaceutical Statistics*, **17**, 123-142.

Deleted: , 2002.

- 1  
2 Satterthwaite, F.E. (1946), An approximate distribution of estimates of variance components,  
3 *Biometrics Bulletin*, **2**, 110-114.  
4  
5 Searle, S.R., Casella, G. and McCulloch, C.E. (1992), *Variance Components*, Wiley, New York.  
6  
7 Swallow, W.H. and Monahan, J.F. (1984), Monte Carlo Comparison of ANOVA, MIVQUE, REML, and  
8 ML Estimators of Variance Components, *Technometrics*, **26** (2), 47- 57.  
9  
10 Welch, B.L. (1956), On linear combinations of several variances, *Journal of the American Statistical*  
11 *Association*, **52** (273), 132-148.  
12  
13 Zou, K.H. and McDermott, M.P. (1999), Higher-moment approaches to approximate interval  
14 estimation for a certain intraclass correlation coefficient, *Statistics in Medicine*, **18**, 2051-2061.  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

Table 1: Choices of model parameter in the simulation study.

Variance Components	Case A	Case B	Case C	Case D	Case E
Lab	0.6	0.6	0.2	1.0	0.2
Batch*Lab	0.3	0.1	0.1	0.1	0.1
Run(Lab)	0.7	1.2	0.2	0.2	0.2
Batch*Run(Lab)	0.6	0.1	1.1	0.1	0.1
Repeatability	0.5	0.5	0.5	0.5	0.5
Intermediate Precision	1.8	1.8	1.8	0.8	0.8
Reproducibility	2.7	2.4	2.1	1.9	1.1

Table 2: Selected samples sizes for batches, laboratories, runs, and replicates used in the simulation study.

	Number of batches (H)	Number of laboratories (I)	Number of analytical runs (J)	Number of replicates (K)	Total number of test samples (HIJK)
1	2	2	2	3	24
2	2	2	3	2	24
3	3	2	2	2	24
4	2	2	3	3	36
5	3	2	2	3	36
6	3	2	3	2	36
7	2	2	4	3	48
8	3	2	4	2	48
9	3	2	3	3	54
10	2	3	2	3	36
11	2	3	3	2	36
12	3	3	2	2	36
13	2	3	3	3	54
14	3	3	2	3	54
15	3	3	3	2	54
16	2	3	4	3	72
17	3	3	4	2	72
18	3	3	3	3	81

Table 3: Ranges of coverage probabilities for intermediate precision for interlaboratory studies with balanced designs.

CI-Method		Case A	Case B	Case C	Case D	Case E
REML	OS	88.67 – 94.04	82.20 – 91.96	90.90 – 94.98	93.04 – 95.45	93.18 – 95.31
	MS	92.38 – 97.70	86.03 – 95.34	94.24 – 98.15	95.71 – 98.44	95.60 – 98.38
ML	OS	87.77 – 93.95	81.26 – 91.86	89.83 – 94.95	92.97 – 95.37	93.04 – 95.19
	MS	91.84 – 97.73	85.53 – 95.40	93.67 – 98.16	95.61 – 98.43	95.50 – 98.37
TYPE 1&3	OS	92.93 – 95.96	87.20 – 94.35	92.39 – 96.04	96.02 – 98.15	96.19 – 98.37
	MS	94.28 – 98.19	90.39 – 96.54	93.63 – 98.37	98.05 – 99.02	97.55 – 99.06

Deleted: in

Deleted: 24

Deleted: 43

Deleted: 05

Table 4: Ranges of coverage probabilities for reproducibility for interlaboratory studies with balanced designs.

CI-Method		Case A	Case B	Case C	Case D	Case E
REML	OS	92.91 – 94.92	90.35 – 94.85	95.39 – 96.69	78.56 – 88.28	92.75 – 95.56
	MS	95.96 – 98.39	93.20 – 97.32	97.52 – 99.14	84.56 – 94.22	96.45 – 98.58
ML	OS	89.67 – 93.41	86.98 – 92.97	92.83 – 96.06	72.68 – 86.04	90.69 – 94.85
	MS	94.12 – 98.04	91.41 – 96.61	96.40 – 98.98	81.52 – 93.29	95.65 – 98.33
TYPE 1&3	OS	94.72 – 98.00	94.51 – 97.26	91.35 – 97.92	85.18 – 95.87	96.62 – 98.80
	MS	95.76 – 99.07	96.38 – 98.58	91.23 – 99.02	89.98 – 97.46	97.55 – 99.48

Deleted: in

Deleted: 11

Deleted: 5

Deleted: 00

Deleted: 3

Deleted: 91

Deleted: 50

Deleted: 79

Deleted: 4

Deleted: 0

Deleted: 35

**Table 5:** Ranges of coverage probabilities for intermediate precision for interlaboratory studies with 20% missing data.

Deleted: for

CI-Method		Case A	Case B	Case C	Case D	Case E
REML	OS	90.09 – 94.04	83.18 – 92.60	90.74 – 94.91	94.38 – 95.75	93.49 – 95.63
	MS	93.61 – 98.05	87.83 – 96.07	94.38 – 98.65	96.29 – 98.30	95.90 – 98.75
ML	OS	89.25 – 93.93	82.50 – 92.40	89.85 – 94.83	94.23 – 95.76	93.39 – 95.55
	MS	93.29 – 98.10	87.55 – 96.01	93.90 – 98.64	96.14 – 98.32	95.87 – 98.73
TYPE 1	OS	88.02 – 95.41	82.47 – 91.66	86.11 – 95.37	92.90 – 96.85	92.47 – 96.88
	MS	88.47 – 97.68	84.04 – 94.15	86.53 – 97.74	92.83 – 98.45	92.31 – 98.38
TYPE 3	OS	90.89 – 95.66	85.11 – 92.91	89.43 – 95.61	94.00 – 96.93	93.07 – 96.97
	MS	91.60 – 97.98	87.17 – 95.45	89.92 – 97.94	94.04 – 98.31	93.24 – 98.55

Deleted: 95.75 –

**Table 6:** Ranges of coverage probabilities for reproducibility for interlaboratory studies with 20% missing data.

Deleted: for

CI-Method		Case A	Case B	Case C	Case D	Case E
REML	OS	93.30 – 95.54	91.54 – 94.56	95.42 – 97.02	80.74 – 89.95	93.39 – 95.94
	MS	96.47 – 98.77	94.35 – 97.64	97.69 – 99.28	87.29 – 95.14	96.87 – 98.68
ML	OS	90.41 – 94.28	88.32 – 92.68	93.11 – 96.27	75.16 – 87.95	91.55 – 95.41
	MS	95.25 – 98.69	92.90 – 97.13	96.55 – 99.20	85.21 – 95.00	96.26 – 98.53
TYPE 1	OS	87.58 – 96.73	85.51 – 94.01	82.41 – 96.77	82.23 – 94.10	90.14 – 97.67
	MS	87.71 – 98.13	85.74 – 95.61	82.45 – 97.67	84.88 – 95.96	90.13 – 98.91
TYPE 3	OS	91.36 – 97.33	89.62 – 95.41	86.96 – 97.17	82.03 – 94.15	91.17 – 97.61
	MS	91.59 – 98.58	90.07 – 97.26	86.87 – 98.16	84.16 – 95.40	91.12 – 98.80

Deleted: 5

Deleted: 3

Deleted: 78

**Table 7:** Ranges of left and right tail probabilities for repeatability for interlaboratory studies with 10% missing data based on the modified Satterthwaite method.

Deleted: based on the modified Satterthwaite method

CI-Method		Case A	Case B	Case C	Case D	Case E
REML	Left	2.31 – 2.80	1.66 – 2.31	2.27 – 3.04	1.57 – 2.57	1.74 – 2.39
	Right	2.31 – 2.87	2.35 – 3.54	2.25 – 2.85	2.56 – 3.77	2.53 – 3.62
ML	Left	2.32 – 2.80	1.69 – 2.31	2.26 – 3.01	1.62 – 2.57	1.78 – 2.41
	Right	2.29 – 2.88	2.35 – 3.53	2.25 – 2.85	2.55 – 3.76	2.53 – 3.67
TYPE 1&3	Left	2.31 – 2.76	2.24 – 2.65	2.26 – 2.93	2.16 – 2.88	2.16 – 2.76
	Right	2.23 – 2.82	2.22 – 2.67	2.27 – 2.85	2.24 – 2.66	2.34 – 2.66

Deleted: for

Deleted: 71

Deleted: 27

Deleted: 29

**Table 8:** Ranges of left and right tail probabilities for intermediate precision for interlaboratory studies with 10% missing data based on the modified Satterthwaite method.

Deleted: based on the modified Satterthwaite method

CI-Method		Case A	Case B	Case C	Case D	Case E
REML	Left	0.50 – 0.85	0.85 – 1.47	0.35 – 0.55	0.28 – 0.61	0.28 – 0.53
	Right	1.28 – 6.45	2.85 – 12.21	0.97 – 5.36	1.02 – 4.27	1.09 – 4.13
ML	Left	0.40 – 0.87	0.68 – 1.50	0.34 – 0.55	0.29 – 0.62	0.30 – 0.54
	Right	1.30 – 6.90	2.96 – 13.02	1.01 – 5.82	1.02 – 4.24	1.09 – 4.22
TYPE 1	Left	1.11 – 5.94	2.53 – 6.85	1.20 – 7.84	0.60 – 4.21	0.77 – 4.70
	Right	0.85 – 4.74	2.13 – 9.12	0.49 – 4.10	0.16 – 1.32	0.22 – 1.61
TYPE 3	Left	0.87 – 3.26	1.44 – 3.90	1.03 – 4.62	0.53 – 3.50	0.69 – 4.03
	Right	0.87 – 4.52	2.10 – 9.16	0.45 – 3.90	0.22 – 1.34	0.21 – 1.75

Deleted: for

Deleted: 8.10

**Table 9:** Ranges of left and right tail probabilities for reproducibility for interlaboratory studies with 10% missing data based on the modified Satterthwaite method.

Deleted: based on the modified Satterthwaite method

CI-Method		Case A	Case B	Case C	Case D	Case E
REML	Left	0.19 – 0.62	0.35 – 0.86	0.12 – 0.39	0.68 – 1.05	0.07 – 0.33
	Right	1.22 – 3.44	1.93 – 5.77	0.57 – 2.25	4.20 – 13.16	1.13 – 3.28
ML	Left	0.15 – 0.28	0.23 – 0.45	0.14 – 0.35	0.23 – 0.50	0.07 – 0.28
	Right	1.48 – 5.09	2.60 – 7.67	0.77 – 3.21	5.12 – 16.26	1.29 – 4.10
TYPE 1	Left	0.57 – 7.07	1.06 – 6.51	1.03 – 13.77	0.97 – 4.62	0.27 – 5.65
	Right	0.40 – 4.13	1.02 – 6.32	0.17 – 1.57	1.46 – 10.70	0.15 – 1.78
TYPE 3	Left	0.33 – 4.44	0.46 – 3.63	0.92 – 9.92	0.96 – 3.96	0.28 – 4.70
	Right	0.43 – 2.85	1.01 – 4.97	0.16 – 1.44	1.94 – 11.18	0.22 – 2.10

Deleted: for

Deleted: 73

Deleted: 69

Deleted: 31

Deleted: 64

**Page 16: [1] Deleted** **heuvele3** **1/5/2010 1:47:00 PM**

It is assumed that  $\alpha_h$ ,  $\beta_i$ ,  $\delta_{j(i)}$ ,  $(\alpha\beta)_{hi}$ ,  $(\alpha\delta)_{hj(i)}$ , and  $\varepsilon_{hijk}$  are all mutually normally and independently distributed with mean zero and variances  $\sigma_6^2$ ,  $\sigma_5^2$ ,  $\sigma_4^2$ ,  $\sigma_3^2$ ,  $\sigma_2^2$ , and  $\sigma_1^2$ , respectively.

**Page 16: [2] Formatted** **heuvele3** **1/15/2010 6:23:00 PM**

Font: (Default) Arial, 10 pt, English (U.S.), Lowered by 5 pt

**Page 16: [3] Formatted** **heuvele3** **1/15/2010 6:23:00 PM**

Font: (Default) Arial, 10 pt, English (U.S.), Lowered by 5 pt

**Page 16: [4] Formatted** **heuvele3** **1/15/2010 6:24:00 PM**

Font: (Default) Arial, 10 pt, English (U.S.), Lowered by 5 pt

**Page 16: [5] Formatted** **heuvele3** **1/15/2010 6:24:00 PM**

Font: (Default) Arial, 10 pt, English (U.S.), Lowered by 5 pt

**Page 16: [6] Formatted** **heuvele3** **1/15/2010 6:24:00 PM**

Font: (Default) Arial, 10 pt, English (U.S.), Lowered by 5 pt

**Page 16: [7] Formatted** **heuvele3** **1/18/2010 9:19:00 AM**

Font: (Default) Arial, 10 pt, English (U.S.), Lowered by 5 pt

**Page 16: [8] Formatted** **heuvele3** **1/5/2010 1:53:00 PM**

Indent: First line: 35.4 pt