



HAL
open science

ANALYZING SMALL SAMPLES OF REPEATED MEASURES DATA WITH THE MIXED-MODEL ADJUSTED F TEST

Jaime Arnau, Roser Bono, Guillermo Vallejo

► **To cite this version:**

Jaime Arnau, Roser Bono, Guillermo Vallejo. ANALYZING SMALL SAMPLES OF REPEATED MEASURES DATA WITH THE MIXED-MODEL ADJUSTED F TEST. *Communications in Statistics - Simulation and Computation*, 2009, 38 (05), pp.1083-1103. 10.1080/03610910902785746 . hal-00514348

HAL Id: hal-00514348

<https://hal.science/hal-00514348>

Submitted on 2 Sep 2010

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.



ANALYZING SMALL SAMPLES OF REPEATED MEASURES DATA WITH THE MIXED-MODEL ADJUSTED F TEST

Journal:	<i>Communications in Statistics - Simulation and Computation</i>
Manuscript ID:	LSSP-2008-0163.R2
Manuscript Type:	Original Paper
Date Submitted by the Author:	26-Jan-2009
Complete List of Authors:	Arnau, Jaime; University of Barcelona, Metodología de las Ciencias del Comportamiento Bono, Roser; University of Barcelona, Metodología de las Ciencias del Comportamiento Vallejo, Guillermo; University of Oviedo
Keywords:	Kenward-Roger method, Linear mixed model, Repeated measures, Type I error rate, Simulation
Abstract:	This research examines the Type I error rates obtained when using the mixed model with the Kenward-Roger correction and compares them with the Between-Within and Satterthwaite approaches in split-plot designs. A simulated study was conducted to generate repeated measures data with small samples under normal distribution. The data were obtained via three covariance matrices (unstructured, heterogeneous first-order auto-regressive and random coefficients), the one with the best fit being selected according to the Akaike criterion. The results of the simulation study showed the Kenward-Roger test to be more robust, particularly when the population covariance matrices were unstructured or heterogeneous first-order auto-regressive.

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 ANALYZING SMALL SAMPLES OF REPEATED MEASURES DATA WITH THE
5
6
7 MIXED-MODEL ADJUSTED F TEST
8
9

10
11
12
13
14 Jaime Arnau

15
16 University of Barcelona
17

18
19
20
21 Roser Bono

22
23 University of Barcelona
24

25
26
27
28 Guillermo Vallejo

29
30 University of Oviedo
31
32
33
34
35
36

37 Running head: ANALYZING SMALL SAMPLES OF REPEATED MEASURES
38
39 DATA
40
41
42
43
44
45
46
47

48 Received July 16, 2008
49

50
51 Correspondence concerning this article should be addressed to Roser Bono,
52
53 Departamento de Metodología de las Ciencias del Comportamiento, Facultad de
54
55 Psicología, Universidad de Barcelona, Passeig de la Vall d'Hebron 171, 08035-
56
57 Barcelona, Spain; e-mail: rbono@ub.edu
58
59
60

Abstract

This research examines the Type I error rates obtained when using the mixed model with the Kenward-Roger correction and compares them with the Between-Within and Satterthwaite approaches in split-plot designs. A simulated study was conducted to generate repeated measures data with small samples under normal distribution conditions. The data were obtained via three covariance matrices (unstructured, heterogeneous first-order auto-regressive and random coefficients), the one with the best fit being selected according to the Akaike criterion. The results of the simulation study showed the Kenward-Roger test to be more robust, particularly when the population covariance matrices were unstructured or heterogeneous first-order auto-regressive. The main contribution of this study lies in showing that the Kenward-Roger method corrects the liberal Type I error rates obtained with the Between-Within and Satterthwaite approaches, especially with positive pairings between group sizes and covariance matrices.

Key words: Kenward-Roger method; Linear mixed model; Repeated measures; Type I error rate; Simulation.

1. Introduction

Over the last fifteen years researchers have shown a growing interest in the analysis of repeated measures data in both experimental and longitudinal contexts. In the former the main objective aim is to analyse treatment effects, while in ~~the latter the idea is to test the effect across time~~ longitudinal studies interest lies in testing the time effect. In both cases the difficulty arises when applying traditional analytic procedures such as analysis of variance (ANOVA) or multivariate analysis of variance (MANOVA).

One of the recent approaches to the analysis of repeated measures data is based on the mixed model (Littell et al., 1996). Laird and Ware (1982) established the bases of the linear mixed model, which takes into account the possible correlation of within-subject errors. Subsequently, Cnaan et al. (1997) and Verbeke and Molenberghs (2000) reported a more complete specification and applied the multilevel model to longitudinal repeated measures data. Unlike ~~variance-based analyses (ANOVA and MANOVA)~~ the mixed model enables the structure of the covariance matrix to be specified on the basis of the data, rather than presupposing it. Thus, a more efficient estimate of the fixed effects is achieved and, consequently, more powerful statistical tests are obtained. This analysis can be performed using the PROC MIXED program of the SAS system (SAS Institute, 2000, 2004), which incorporates all the advantages of mixed model methodology with repeated measures data (Littell et al., 1998; Verbeke and Molenberghs, 1997). By means of this methodology the researcher models the covariance structure and achieves greater robustness when estimating the effects of repeated measures and interaction.

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

Various studies have reported that the mixed model offers better control of Type I error, despite the fact, as Wright and Wolfinger (1996) point out, that Type I error may show a positive bias with incorrect covariance structures and small sample sizes. Keselman et al. (1999a) applied the mixed model with degrees of freedom based on the residual variance, which are the same as the degrees of freedom of the conventional F test. These authors found that when the covariance matrix is not spherical the degrees of freedom associated with the conventional F test are too large. One alternative would thus be to apply the procedure developed by Satterthwaite (1941) to adjust the degrees of freedom.

The problem with estimating the fixed effects of the repeated-measures factor and the interaction with the group factor arises from the misspecification of the covariance matrix, especially when sample sizes are small. The PROC MIXED program of SAS includes several options for model specification with which it is possible to change the degrees of freedom when estimating these effects. In general, the inferences obtained with the PROC MIXED program of SAS are based on the Wald test, which is valid with large samples. Keselman et al. (1999b) point out that the ~~Satterthwaite approach has only recently been applied to covariance structures and it is still not entirely clear how the mixed model works. The~~ degrees of freedom with the Satterthwaite correction are more conservative than the residual degrees of freedom and they can be expected to yield more precise F tests. ~~This is the conclusion reached in the study by (Keselman et al., (1999a), which showed the F test to be much more liberal when the degrees of freedom were based on residuals.~~

The present study applied the mixed model (by means of SAS PROC MIXED, version 9.1.3) to compare tests of the fixed effects of repeated measures and the

1
2
3
4 interaction in split-plot designs using three degrees-of-freedom solutions: Between-
5
6 Within (BW), Satterthwaite (SW) and Kenward-Roger (KR). Although ~~the small~~
7 ~~sample distribution of~~ Wald statistics ~~is approximated by an~~ ~~may yield good~~
8 ~~approximations of the value of F with small samples~~ distribution (Kenward and Roger,
9 ~~1997~~), the statistical results are generally rather poor. The properties of small samples
10
11 can be improved by using the Satterthwaite approach to denominator degrees of
12
13 freedom (Satterthwaite, 1941), or through the Kenward-Roger adjusted degrees of
14
15 freedom solution (Kenward and Roger, 1977).
16
17
18
19
20
21
22

23 Keselman et al. (1999b) found that F tests combined with the Satterthwaite
24
25 approach, using PROC MIXED, were as robust as default tests when the true structure
26
27 of the covariance matrix was specified. Keselman et al. (1999a) based the degrees of
28
29 freedom of the mixed model on the variance of the residuals, which coincide with the
30
31 degrees of freedom of the error of the conventional F test. Through its *MODEL* option
32
33 the PROC MIXED program enables researchers to use the F approximation based on
34
35 the solution given by Satterthwaite (1946). This approximation corrects the degrees of
36
37 freedom for the test of fixed effects of repeated measures, and was investigated by
38
39 Keselman et al. (1999b) to determine its efficacy in controlling Type I error. However,
40
41 tests of the fixed effects of repeated measures using PROC MIXED were more robust
42
43 when the true covariance structure was known. With versions 8 and above of SAS it is
44
45 possible to use the Kenward-Roger solution (Kenward and Roger, 1997) as an
46
47 alternative way of calculating the degrees of freedom. This procedure appears to yield
48
49 more accurate ~~results of inferences~~ for the fixed effects with small samples (Kowalchuk
50
51 et al., 2004; Schaalje et al., 2001).
52
53
54
55
56
57
58
59
60

Assuming that statistical tests follow an approximate distribution of F with the corresponding calculation of the degrees of freedom, SAS PROC MIXED (version 9.1.3) now incorporates two methods for correcting this calculation. ~~The sample distribution of F approximately follows an F distribution with q and ν as degrees of freedom in the numerator and denominator of this ratio. When using the residual degrees of freedom, $n - \text{rank}(\mathbf{X})$, where n is the number of observations, the results obtained are too liberal. Hence it is better to use the two approaches for correcting degrees of freedom that are incorporated into SAS PROC MIXED:~~ DDFM=SATTERTH and DDFM=KENWARDROGER. Given the above, the present study aimed to compare the mixed-model procedure using the Satterthwaite and Kenward-Roger solutions for degrees of freedom with the system that separates the between and within degrees of freedom with small samples.

1.1. *Satterthwaite and Kenward-Roger approximations for correcting degrees of freedom*

The Satterthwaite approximation is a generalization of techniques proposed by Giesbrecht and Burns (1985), McLean and Sanders (1988) and Fai and Cornelius (1996). Let θ be a vector of unknown parameters in \mathbf{V} (variance/covariance matrix) and $\mathbf{C} = (\mathbf{C}'\mathbf{V}^{-1}\mathbf{X})^-$, where $^-$ denotes a generalized inverse. If $\hat{\mathbf{C}}$ and $\hat{\theta}$ are the corresponding estimates, the denominator degrees of freedom can be calculated by performing the spectral decomposition $\mathbf{L}\hat{\mathbf{C}}\mathbf{L}' = \mathbf{P}'\mathbf{D}\mathbf{P}$, where \mathbf{P} is an orthogonal matrix of eigenvectors and \mathbf{D} is a diagonal matrix of eigenvalues, both of dimension $q \times q$. Define l_m as the m th row of $\mathbf{P}\mathbf{L}$ and let

$$v_m = \frac{2(D_m)^2}{g'_m \mathbf{A} g_m} \quad (1)$$

where D_m is the m th diagonal element of \mathbf{D} and g_m is the gradient of $l_m \mathbf{C} l'_m$ with respect to $\boldsymbol{\theta}$, evaluated at $\hat{\boldsymbol{\theta}}$. \mathbf{A} is the asymptotic variance-covariance matrix of $\hat{\boldsymbol{\theta}}$ obtained from the second derivative matrix of the likelihood equations. Thus

$$E = \sum_{m=1}^q \frac{v_m}{v_m - 2} \mathbf{I}(v_m > 2) \quad (2)$$

where the indicator function eliminates terms for which $v_m \leq 2$. The degrees of freedom for F are then given by

$$v = \frac{2E}{E - q} \quad (3)$$

provided $E > q$; otherwise $v = 0$.

The Kenward-Roger approximation calculates the degrees of freedom in a similar way and is used for tests with mixed linear models ~~based on any covariance structure~~. If \mathbf{C} is a contrast matrix of range q , the Wald F for the hypothesis $H_0: \mathbf{C}\boldsymbol{\beta} = 0$ is given by $F = W/q$, where

$$W = (\mathbf{C}\hat{\boldsymbol{\beta}})'(\mathbf{C}(\mathbf{X}'\hat{\mathbf{V}}^{-1}\mathbf{X})^{-1}\mathbf{C}')^{-1}(\mathbf{C}\hat{\boldsymbol{\beta}}) \quad (4)$$

The next step is to calculate a scale factor δ and an approximate value of the degrees of freedom v . Thus, the Kenward-Roger F statistic is given by

$$F^* = \delta F_{\text{KR}} \quad (5)$$

where $F_{\text{FR}} = \frac{1}{q}(\mathbf{C}\hat{\boldsymbol{\beta}})'(\mathbf{C}(\mathbf{X}'\hat{\mathbf{V}}^{-1}\mathbf{X})^{-1}\mathbf{C}')^{-1}(\mathbf{C}\hat{\boldsymbol{\beta}})$. The moments of F^* are then generated

and equated to the moments of the F distribution to solve for δ and v . Under a null hypothesis it is assumed that F^* is approximately distributed as F with q degrees of freedom in the numerator and v degrees of freedom in the denominator. Thus, two

values have to be calculated from the data: the denominator degrees of freedom ν and a scale factor δ . Therefore,

$$\nu = 4 + \frac{q+2}{qy-1} \quad (6)$$

where

$$y = \frac{V[F_{KR}]}{2E[F_{KR}]^2}, \quad (7)$$

and

$$\delta = \frac{\nu}{E[F_{KR}](\nu-2)} \quad (8)$$

The degrees of freedom are calculated in a similar way to that used in the previous procedure and the Wald F statistic is adjusted to take into account the bias associated with small samples and the variability of the **estimated** variance matrix.

The inferences derived from simulation studies using these methods usually function well, even with complex covariance structures (Keselman et al., 1998; Schaalje et al., 2002). In the present study we compared the functioning of these models with normally-distributed small samples and covariance structures that violate the assumption of sphericity, a situation that occurs frequently in the context of longitudinal repeated measures. It has been demonstrated that in the case of normally-distributed data and heterogeneous within-group covariance structures the KR procedure meets the criterion of robustness (Livacic-Rojas et al., 2006). The aim of the present study was thus to **determine the functioning of heterogeneous covariance structures, both across groups and within measurement occasions, when estimating** estimate the fixed effects associated with time and their interaction with relatively small samples in split-plot designs, **using mixed models. This approach enables users to model the covariance**

structure for subjects for a repeated measures variable, and also allows them to fit different covariance structures to each group of subjects for each level of a between-subjects grouping variable.

2. Simulation method

The split-plot design includes a between-subjects factor, where the subjects ($i = 1, \dots, n_j$) are chosen at random for each group ($j = 1, \dots, J$), and a within-subjects factor of repeated measures ($k = 1, \dots, K$). The main objective of this type of design is to study the main repeated measures effect and the group x repeated measures interaction. Here it is assumed that the (y_{ijk}) data are normally distributed. The test statistics for the effects are based on the covariance structure selected by the Akaike Information Criterion (AIC).

2.1. Study variables

The statistical tests of the main repeated measures effect, as well as of the interaction with the group factor, have been conducted with both balanced and unbalanced designs, with a between-subjects factor and a within-subjects factor. Three values levels were taken used for the between-subjects factor, while the values levels four, six and eight levels were taken used for the within-subjects factor. In the simulation study the combinations of five variables were selected for each level value of K : a) covariance structure of the population; b) homogeneous and heterogeneous between-group covariance structures; c) total sample size; d) equal and unequal group sizes; and e) pairings of the covariance matrices and group sizes.

1
2
3
4
5 Three covariance structures with within-subject heterogeneity were used to
6 generate the simulated data: unstructured (UN), heterogeneous first-order auto-
7 regressive (ARH), and random coefficients (RC). At the same time we investigated
8 whether the assumption of sphericity was violated, taking the indices $\epsilon = 0.57$ and 0.75 .
9
10 When $\epsilon = 1$, the assumption of sphericity is met with designs $J \times K$, whereas with $\epsilon =$
11 $1/(K-1)$ the index takes the extreme value. Most studies use a value of 0.75 as a good
12 approximation of sphericity, and 0.57 as indicative of non-sphericity (Algina and
13 Keselman, 1998; Keselman et al., 1999a; Lix et al., 2003). Table 1 shows the values of
14 the covariance matrices for the corresponding sphericity indices. We have omitted the
15 values of the generating covariance matrices corresponding to structures of six and eight
16 repeated measures.
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32

33 [Table 1. Insert here approximately]
34
35
36
37

38 Of the eighteen covariance structures generated we analysed both the equal and
39 unequal between-group covariance matrices. With heterogeneous matrices the
40 inequality of the groups ~~fitted~~ followed the ratio 1:3:5, as in the studies by Keselman et
41 al. (1993), Vallejo et al. (2001), Lix et al. (2003), Livacic-Rojas et al. (2006), and
42 Vallejo and Ato (2006). Thus, $\Sigma_1 = \frac{1}{3}\Sigma_2$ and $\Sigma_3 = \frac{5}{3}\Sigma_2$, where Σ_j is the covariance
43 matrix for group j . In addition, we investigated the conditions of equality and inequality
44 for group size, considering total sample sizes of $N = 30, 36$ and 42 . Equal and unequal
45 group sizes were taken for each value of N . In the latter case we chose group sizes that
46 represented values of the variance coefficient of the group size, $\Delta n_j = 0.41$, and when
47 the group sizes were equal $\Delta n_j = 0$. With $\Delta n_j = 0.41$, the unequal group sizes were: a) 5,
48
49
50
51
52
53
54
55
56
57
58
59
60

1
2
3
4 10, 15 ($N = 30$); b) 6, 12, 18 ($N = 36$); and c) 7, 14, 21 ($N = 42$). With $\Delta n_j = 0$, the equal
5
6 group sizes were: a) 10, 10, 10 ($N = 30$); b) 12, 12, 12 ($N = 36$); and c) 14, 14, 14 ($N =$
7
8 42). Finally, we defined the type of pairing between group sizes and covariance matrix.
9
10 Positive pairing associates the sample of the larger group with the covariance matrix
11
12 whose values are larger. In contrast, negative pairing **relates combines** the largest group
13
14 size with the covariance matrix comprised of smaller elements. In the case of balanced
15
16 designs, pairing is null.
17
18
19

20
21 Table 2 summarises the various combinations of variables examined in the
22
23 present study. For each combination, 1000 replications were performed at a significance
24
25 level of 0.05: $K \times$ covariance structures $\times \varepsilon \times$ combinations $N(n_1 n_2 n_3)/\Delta n_j/\text{between-}$
26
27 group covariances/pairing \times replications ($3 \times 3 \times 2 \times 12 = 216$ simulation conditions).
28
29 For the simulations we studied the main effect of the repeated measures variable and the
30
31 interaction with the group variable when the data were normally distributed.
32
33
34
35
36
37

38 [Table 2. Insert here approximately]
39
40
41

42 The simulation data were generated by means of a macro from SAS 9.1.3 (SAS
43
44 Institute, 1997) and using the programming language IML (Interactive Matrix
45
46 Language), also from SAS (SAS Institute, 1999a). The first step involved generating the
47
48 covariance matrices from variances and correlations for values of $\varepsilon = 0.57$ and $\varepsilon = 0.75$.
49
50 Next, the RANNOR generator of SAS (SAS Institute, 1999b) was used to derive
51
52 normally-distributed pseudorandom observations by means of the Cholesky factor of
53
54 the covariance matrix Σ_j . Finally, each set of data was analysed with PROC MIXED
55
56 (SAS Institute, 2000, 2004), using the covariance structure selected according to the
57
58
59
60

1
2
3
4 ~~Akaike Information Criterion~~ (AIC). The DDFM options in the MODEL statement were
5
6
7 BETWITHIN, SATTERTH and KENWARDROGER. The TYPE statement was used to
8
9 specify the covariance structure which showed the best fit according to the AIC.
10
11

12 2.2. Covariance structures adjusted to the data

13
14
15
16
17

18 Given the possibility of heterogeneity, both within and between subjects, eleven
19 covariance structures were fit with PROC MIXED applying the AIC criteria, the aim
20 being to select the one with the lowest value. The following covariance matrices were
21 fit: a) compound symmetry (CS); b) unstructured (UN); c) first-order autoregressive
22 (AR); d) Huynh-Feldt spherical (HF); e) within-subjects heterogeneous compound
23 symmetry (CSH); f) within-subjects heterogeneous first-order autoregressive (ARH); g)
24 random coefficients (RC); h) between-subjects heterogeneous unstructured (UN_j); i)
25 between-subjects heterogeneous Huynh-Feldt spherical (HF_j); j) within- and between-
26 subjects heterogeneous first-order autoregressive (ARH_j); and k) between-subjects
27 heterogeneous random coefficients (RC_j), where the subscript j indicates that the
28 covariance matrices are not equal between the groups.
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50

51 3. Results

52
53
54
55
56
57
58
59
60

In an initial simulation study we ~~analysed~~ examined the covariance structure selected by the AIC from among the eleven covariance 11 structures (CS, UN, AR, HF, CSH, ARH, RC, UN_j, HF_j, ARH_j and RC_j). A second study was then conducted to calculate the p values for tests of fixed effects, taking into account the covariance structure selected by

1
2
3
4 AIC, and the different methods used to adjust the degrees of freedom (BW, SW and
5
6
7 KR).

11 3.1. *Selecting the covariance structure*

16 The AIC was used as the fit criterion due to its advantages over the Schwarz's Bayesian
17 Information Criterion (BIC). Keselman et al. (1998) demonstrated that the AIC chooses
18 the population covariance structure on 47% of occasions, while the BIC achieved a rate
19 of 35%. Similarly, Ferron et al. (2002) showed that the AIC selects the true covariance
20 structure on 79% of occasions, compared to a rate of 66% for the BIC. More recently,
21 Vallejo and Livacic-Rojas (2005) reported that the results of tests based on the AIC are
22 better at controlling Type I error rates than are those based on the BIC. These authors
23 found that with the BIC, PROC MIXED offers poor control over the estimated
24 probabilities of Type I error. Gomez et al. (2005) concluded that the AIC has a better
25 success rate with complex covariance structures, for example, UN. More recently,
26 Vallejo et al. (2008) found that with different group sizes the AIC is better at estimating
27 standard errors. Given these findings we conducted a simulation study using the AIC.
28 However, it should be remembered that the Akaike criterion does not always select the
29 **only** true structure, **since** and other structures may also provide adequate
30 approximations.
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50

51 Tables 3, 4 and 5 show **which of the eleven candidate covariance structures are**
52 **selected most often by the AIC** ~~the fit percentages of the eleven matrices to the~~
53 ~~covariance structures generated~~, according to the different combinations of variables
54 studied. It should be noted that with homogeneous between-groups covariances we used
55
56
57
58
59
60

1
2
3
4 the population matrices UN, ARH and RC, while UN_j , ARH_j and RC_j were used with
5
6 heterogeneous covariances.
7
8
9

10 11 3.1.1. Homogeneity of covariances 12 13

14
15
16 In general it can be seen that with homogeneous covariances the structures showing the
17
18 best fit are the same as the true UN (Table 3) and ARH (Table 4), with some exceptions
19
20 when $\varepsilon = 0.75$. With UN matrices the percentages oscillate between 63.3% and 84.3%
21
22 (Table 3), and with ARH matrices between 46.7% and 88.6% (Table 4). In contrast,
23
24 with RC covariance matrices the fit percentages of the true structure are practically null
25
26 (Table 5), the most frequently selected structure being CSH (40-76.3%).
27
28
29
30
31
32

33 3.1.2. Heterogeneity of covariances 34 35 36

37 ~~It can be seen in~~ Table 3 ~~shows~~ that with heterogeneous covariance matrices ~~and when~~
38
39 ~~UN is the true structure~~, the structures that ~~show the best fit to UN~~ are ~~selected most~~
40
41 ~~often are~~ UN_j (~~53.70.0-77.5%~~) and ARH_j (~~46.73.0-62.3%~~). When the covariance matrix
42
43 is ARH (Table 4) the ~~most frequently~~ frequently selected ~~ARH_j~~ covariance structure is
44
45 ~~selected more often~~ ARH_j (56.5-97.3%). However, with $K = 8$ and $\varepsilon = 0.75$ another
46
47 pattern emerges: with homogeneity of groups the ~~best-fitting most selected~~ matrix is
48
49 UN_j (90%), while with positive and negative pairings the ~~best-fitting most selected~~
50
51 matrix is UN (88.3% and 79.3%, respectively).
52
53
54
55

56 With RC covariance matrices, several structures show a ~~correct fit~~ higher
57
58 ~~percentage fit~~ (Table 5). Thus, when the covariance matrices are heterogeneous the fit
59
60

of the UN_j matrix oscillates between 27.3% and 87.9%, while the fit of the ARH_j matrix is in the range 34.6-52%. The RC_j structure only shows the best fit when $K = 6$, $\varepsilon = 0.57$ and there is positive pairing (35.5%).

[Table 3. Insert here approximately]

[Table 4. Insert here approximately]

[Table 5. Insert here approximately]

3.2. Type I error rates

The present study followed the criterion of Bradley (1978), by which a test is robust if the empirical error rate is within the range 0.025-0.075 for $\alpha = 0.05$. A test is considered to be liberal when the empirical Type I error rate exceeds the upper limit. In contrast, when the error rate is below the lower limit the test is conservative. According to Kowalchuk et al. (2004) it is important that applied researchers use procedures that control the Type I error rate within the limits established by Bradley (1978), particularly when these procedures must assume a series of violation conditions.

Tables 6-11 show the empirical Type I error rates according to the combination of the different variables indicated in Table 2, with respect to the population covariance structure. Bold entries correspond to values above the established upper limit while those in italics indicate values below the lower limit.

3.2.1. 0.57 sphericity

1
2
3
4 The values shown in Tables 6 and 7 indicate that the KR test is more robust than the
5
6 BW and SW tests. However, if the simulated covariance structure is UN (Table 6) the
7
8 KR method is liberal with negative pairing (0.077-0.272). It can be seen that when the
9
10 number of repeated measures increases to $K = 8$ the KR procedure remains robust even
11
12 with negative pairing. Table 7 shows that when the simulated covariance structure is
13
14 ARH the KR method ~~may still be~~ **is sometimes** robust with negative pairing, even with
15
16 $K = 4$ and $K = 6$. Finally, Table 8 depicts the results for RC simulated covariance
17
18 matrices. Note that in comparison to the other covariance structures the tests are
19
20 conservative when K increases and pairing is positive. In this case the KR method **often**
21
22 does not correct the Type I error rate.
23
24
25
26

27
28 [Table 6. Insert here approximately]
29

30 [Table 7. Insert here approximately]
31

32 [Table 8. Insert here approximately]
33
34
35
36
37

38 3.2.2. 0.75 sphericity 39 40 41

42 **For the more spherical covariance structures**, when the population covariance structure
43
44 is UN (Table 9) or ARH (Table 10) the KR test is more robust, although it is difficult
45
46 for Type I error rates to approach nominal values with negative pairing. This is
47
48 particularly so with the ~~effect of the~~ **interaction effect**. In these cases the KR test is
49
50 liberal. With an ARH structure, positive pairing and $K = 8$ the tests tend to be more
51
52 conservative, especially as regards the time effect (Table 10). With respect to the RC
53
54 structure (Table 11) the Type I error rates are much higher than the nominal value with
55
56 negative pairings. Furthermore, when the BW and SW tests are more conservative with
57
58
59
60

1
2
3
4 positive pairings, the same Type I error rate is obtained with the KR method. This is
5
6 observed to a greater extent with the time effect.
7
8
9

10
11 [Table 9. Insert here approximately]

12 [Table 10. Insert here approximately]

13
14 [Table 11. Insert here approximately]

20 21 4. Discussion

22
23
24
25 This research aimed to assess the robustness of the BW, SW and KR procedures in
26 split-plot designs with small **numbers of subjects per groups**. The first simulation study
27 enabled us to select the covariance matrices with the best fit according to the AIC, thus
28 taking into account any possible bias due to a wrong choice, which would affect the
29 robustness of the statistical test used (Vallejo et al., 2008). With homogeneous
30 covariances we found that the true UN and ARH structures showed a high fit
31 percentage. However, this was not the case for RC covariance matrices, where the fit
32 percentages of the true structure were close to zero. When the covariance matrices were
33 heterogeneous, several structures, in addition to the true one, showed a high fit
34 percentage.
35
36
37
38
39
40
41
42
43
44
45
46
47
48

49 As reported by Keselman and Keselman (1990) we found that heterogeneous
50 covariance matrices produce a greater discrepancy between empirical and nominal Type
51 I error rates. Some research has suggested that the mixed model with BW or SW
52 degrees of freedom increases Type I error rates when the between-groups covariance
53 matrices are heterogeneous and sample sizes are small, even when the groups are equal
54
55
56
57
58
59
60

1
2
3
4 in size (Keselman et al., 1999a, 1999b; Keselman et al., 2000; Wright and Wolfinger,
5
6 1996). This effect is heightened in the case of negative pairings. For example, Keselman
7
8 et al. (1999b) found that the empirical Type I error rate when using the SW adjustment
9
10 reached 8% when testing the main repeated-measures effect and 9.9% when testing the
11
12 interaction effect.
13

14
15
16 In accordance with the studies of Kowalchuk et al. (2004) and Vallejo and Ato
17
18 (2006) we conclude that in many cases when the covariance structure selected by AIC is
19
20 applied the KR test of PROC MIXED enables, in many cases, Type I error rates to be
21
22 controlled. However, our results indicate that with negative pairings the KR method
23
24 tends to be liberal, while with positive pairings it proves to be conservative in some
25
26 cases. Similar results were reported by Vallejo and Livacic-Rojas (2005) and Vallejo
27
28 and Ato (2006) when estimating Type I error rates with normally-distributed data.
29
30 Furthermore, in our study, with the non-spherical covariance matrices UN, ARH and
31
32 RC we found that the KR method may be acceptable even with negative pairing. Gomez
33
34 et al. (2005) found no evidence that negative pairings had an adverse effect on Type I
35
36 error rates for the within-subject effect. In fact, they observed that for tests of time
37
38 effect, negative pairing produced slightly better Type I error rates than did positive
39
40 pairing. Our study could not confirm this conclusion because we did not generate the
41
42 same covariance structures.
43
44
45
46
47
48

49 In general, this study enables the following conclusions to be drawn: a) Type I
50
51 error rates are inflated liberal even when using the true covariance structure selected by
52
53 the AIC in the analysis; b) on most occasions the Type I error rates observed for the
54
55 interaction effect are higher than those for the main effect; c) the SW method is more
56
57 robust than the BW procedure, and the KR approximation is more robust than the SW
58
59
60

1
2
3
4 method; d) the BW and SW tests are more liberal; e) the tests are scarcely robust with
5
6 negative pairing; f) with positive pairing and ARH or RC covariance structures the tests
7
8 may be conservative; and g) the best outcomes with the KR test are achieved with UN
9
10 or ARH structures and $\varepsilon = 0.57$.
11
12

13
14 In sum, the main contribution of this study is to confirm that PROC MIXED
15
16 with the KR correction offers the best control of Type I error rates in most of the
17
18 conditions studied, particularly when the BW and SW tests are liberal with positive
19
20 pairings.
21
22

23 24 25 Acknowledgements 26

27
28
29
30 This research was supported by Grants MEC-SEJ2005-01923 and MEC-SEJ2005-
31
32 01883 from Spain's Ministry of Education and Science.
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

References

- Algina, J., Keselman, H. J. (1998). A power comparison of the Welch-James and Improved General Approximation tests in the split-plot design. *Journal of Educational and Behavioral Statistics* 23:152-169.
- Bradley, J. V. (1978). Robustness? *British Journal of Mathematical and Statistical Psychology* 31:144-152.
- Cnaan, A., Laird, N. M., Slasor, P. (1997). Using the general linear mixed model to analyze unbalanced repeated measures and longitudinal data. *Statistics in Medicine* 16:2349-2380.
- Fai, A. H. T., Cornelius, P. L. (1996). Approximate F-tests of multiple degree of freedom hypotheses in generalized least squares analyses of unbalanced split-plot experiments. *Journal of Statistical Computing and Simulation* 54:363-378.
- Ferron, J., Dailey, R., Yi, Q. (2002). Effects of misspecifying the first-level error structure in two-level models of change. *Multivariate Behavior Research* 37:379-403.
- Giesbrecht, F. G., Burns, J. C. (1985). Two-stage analysis based on a mixed model: large-sample asymptotic theory and small-sample simulation results. *Biometrics* 41:853-862.
- Gomez, E. V., Schaalje, G. B., Fellingham, G. W. (2005). Performance of the Kenward-Roger method when the covariance structure is selected using AIC and BIC. *Communication in Statistics-Theory and Methods* 34:377-392.
- Kenward, M. G., Roger, J. H. (1997). Small sample inference for fixed effects from restricted maximum likelihood. *Biometrics* 53:983-997.

1
2
3
4 Keselman, H. J., Algina, J., Kowalchuk R. K., Wolfinger, R. D. (1998). A comparison
5
6 of two approaches for selecting covariance structures in the analysis of repeated
7
8 measurements. *Communications in Statistics Simulation and Computation* 27:591-
9
10 604.
11
12

13 Keselman, H. J., Algina, J., Kowalchuk R. K., Wolfinger, R. D. (1999a). A comparison
14
15 of recent approaches to the analysis of repeated measurements. *British Journal of*
16
17 *Mathematical and Statistical Psychology* 52:63-78.
18
19

20 Keselman, H. J., Algina, J., Kowalchuk, R. K., Wolfinger, R. D. (1999b). The analysis
21
22 of repeated measurements: A comparison of mixed-model Satterthwaite F tests and
23
24 a nonpooled adjusted degrees of freedom multivariate test. *Communication in*
25
26 *Statistics-Theory and Methods* 28: 2967-2999.
27
28

29 Keselman, H. J., Carriere, K. C., Lix, L. M. (1993). Testing repeated measures
30
31 hypotheses when covariance matrices are heterogeneous. *Journal of Educational*
32
33 *Statistics* 18:305-319.
34
35

36 Keselman, J. C., Keselman, H. J. (1990). Analysing unbalanced repeated measures
37
38 designs. *British Journal of Mathematical and Statistical Psychology* 43:265-282.
39
40

41 Keselman, H. J., Kowalchuk, R. K., Algina, J., Lix, L. M., Wilcox, R. R. (2000).
42
43 Testing treatment effects in repeated measures designs: Trimmed means and
44
45 bootstrapping. *British Journal of Mathematical and Statistical Psychology* 53: 175-
46
47 191.
48
49

50 Kowalchuk, R. K., Keselman, H. J., Algina, J., Wolfinger, R. D. (2004). The analysis of
51
52 repeated measurements with mixed-model adjusted F tests. *Educational and*
53
54 *Psychological Measurement* 64:224-242.
55
56
57
58
59
60

- 1
2
3
4 Laird, N. M., Ware, J. H. (1982). Random effects models for longitudinal data.
5
6 *Biometrics* 38:963-974.
7
8
9 Littell, R. C., Henry, P. R., Ammerman, C. B. (1998). Statistical analysis of repeated
10
11 measures data using SAS procedures. *Journal of Animal Science* 76:1216-1231.
12
13 Littell, R. C., Milliken, G. A., Stroup, W. W., Wolfinger, R. D. (1996). *SAS System for*
14
15 *Mixed Models*. Cary, NC: SAS Institute.
16
17
18 Livacic-Rojas, P., Vallejo, G., Fernández, P. (2006). Procedimientos estadísticos
19
20 alternativos para evaluar la robustez mediante diseños de medidas repetidas. *Revista*
21
22 *Latinoamericana de Psicología* 38:579-599.
23
24
25 Lix, L. M., Algina, J., Keselman, H. J. (2003). Analyzing multivariate repeated
26
27 measures designs: A comparison of two approximate degrees of freedom
28
29 procedures. *Multivariate Behavioral Research* 38:403-431.
30
31
32 McLean, R. A., Sanders, W. L. (1988). Approximating degrees of freedom for standard
33
34 errors in mixed linear models. *Proceedings of the Statistical Computing Section,*
35
36 *American Statistical Association*. New Orleans, LA.
37
38
39 SAS Institute. (1997). *SAS Macro language: Reference*. Cary, NC: SAS Institute.
40
41
42 SAS Institute. (1999a). *SAS/IML Software, usage and reference, Version 8*. Cary, NC:
43
44 SAS Institute.
45
46
47 SAS Institute. (1999b). *SAS Language reference: Dictionary, Version 8*. Cary, NC: SAS
48
49 Institute.
50
51
52 SAS Institute. (2000). *SAS/STAT User's guide, Version 8*. Cary, NC: SAS Institute.
53
54
55 SAS Institute. (2004). *SAS Online Doc 9.1.3*. Cary, NC: SAS Institute.
56
57
58 Satterthwaite, F. E. (1941). Synthesis of variance. *Psychometrika* 6:309-316.
59
60

- 1
2
3
4 Satterthwaite, F. E. (1946). An approximate distribution of estimates of variance
5 components. *Biometrics Bulletin* 2:110-114.
6
7
8
- 9 Schaalje, G. B., McBride, J. B., Fellingham, G. W. (2001). Approximations to
10 distributions of test statistics in complex mixed linear models using SAS Proc
11 MIXED, *SUGI 26 Proceedings, Statistics, Data Analysis and Data Mining Section*,
12 Paper 262-26. Consulted on 15 November 2007 at:
13 <http://www2.sas.com/proceedings/sugi26/proceed.pdf>.
14
15
16
17
18
19
- 20 Schaalje, G. B., McBride, J. B., Fellingham, G. W. (2002). Adequacy of approximations
21 to distributions of test statistics in complex mixed linear models, *Journal of*
22 *Agricultural Biological and Environmental Statistics* 7:512-524.
23
24
25
26
27
- 28 Vallejo, G., Ato, M. (2006). Modified Brown-Forsythe procedure for testing
29 interactions effects in split-plot designs. *Multivariate Behavioral Research* 41:549-
30 578.
31
32
33
34
- 35 Vallejo, G., Ato, M., Valdés, T. (2008). Consequences of misspecifying the error
36 covariance structure in linear mixed models for longitudinal data. *Methodology*.
37 *European Journal of Research for the Behavioral and Social Sciences* 4:10-11.
38
39
40
41
- 42 Vallejo, G., Fidalgo, A., Fernández, P. (2001). Effects of covariance heterogeneity on
43 three procedures for analyzing multivariate repeated measures designs. *Multivariate*
44 *Behavioral Research* 36:1-27.
45
46
47
48
- 49 Vallejo, G., Livaci-Rojas, P. (2005). Comparison of two procedures for analyzing small
50 sets of repeated measures data. *Multivariate Behavioral Research* 40:179-205.
51
52
53
- 54 Verbeke, G., Molenberghs, G. (1997). *Linear Mixed Models in Practice: A SAS-*
55 *oriented Approach*. New York: Springer-Verlag.
56
57
58
59
60

1
2
3
4 Verbeke, G., Molenberghs, G. (2000). *Linear Mixed Models for Longitudinal Data*.

5
6
7 New York: Springer-Verlag.

8
9 Wright, S. P., Wolfinger, R. D. (1996, October). *Repeated measures analysis using*

10
11 *mixed models: Some simulation results*. Paper presented at the Conference on

12
13
14 Modelling Longitudinal and Spatially Correlated Data: Methods, Applications, and

15
16
17 Future Directions, Nantucket, MA.
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. Population covariance matrices.

UN				ARH				RC			
$\begin{bmatrix} 1 & 1.60000 & 1.80000 & 2.32379 \\ & 4 & 4.80000 & 4.64758 \\ & & 9 & 9.29516 \\ & & & 15 \end{bmatrix}$				$\begin{bmatrix} 1 & 1.5800000 & 1.8723000 & 1.9095318 \\ & 4 & 4.7400000 & 4.8342578 \\ & & 9 & 9.1789705 \\ & & & 15 \end{bmatrix}$				$\begin{bmatrix} 9 & 13 & 18 & 23 \\ & 27.7 & 34.4 & 45.1 \\ & & 59.8 & 67.2 \\ & & & 104.3 \end{bmatrix}$			
$\varepsilon = 0.57$				$\varepsilon = 0.57$				$\varepsilon = 0.57$			
$\begin{bmatrix} 1 & 1.6391 & 1.7913 & 3.7038 \\ & 4 & 2.9362 & 6.0712 \\ & & 9 & 6.6346 \\ & & & 15 \end{bmatrix}$				$\begin{bmatrix} 2 & .2449490 & 0.0424264 & 0.0054772 \\ & 3 & 0.5196152 & 0.0670820 \\ & & 9 & 1.1618950 \\ & & & 15 \end{bmatrix}$				$\begin{bmatrix} 1.30 & 0.90 & 1.50 & 2.10 \\ & 6.45 & 4 & 5.55 \\ & & 15.50 & 9 \\ & & & 27.45 \end{bmatrix}$			
$\varepsilon = 0.75$				$\varepsilon = 0.75$				$\varepsilon = 0.75$			

Note. UN = unstructured model; ARH = heterogeneous first-order autoregressive model; RC = random coefficients model; ε = sphericity index

Table 2. Group sizes of balanced and unbalanced designs with $J = 3$, $K = 4$, 6 and 8, UN, ARH and RC population covariance matrices, and $\varepsilon = 0.57$ and $\varepsilon = 0.75$

N	n_1	n_2	n_3	Δn_j	Covariances between groups	Pairing
30	10	10	10	0.00	=	Null
36	12	12	12	0.00	=	Null
42	14	14	14	0.00	=	Null
30	10	10	10	0.00	\neq	Null
36	12	12	12	0.00	\neq	Null
42	14	14	14	0.00	\neq	Null
30	5	10	15	0.41	\neq	+
36	6	12	18	0.41	\neq	+
42	7	14	21	0.41	\neq	+
30	15	10	5	0.41	\neq	-
36	18	12	6	0.41	\neq	-
42	21	14	7	0.41	\neq	-

Notes. J : groups, K : number of repeated measurements, ε : sphericity, N : total sample size, n_1 , n_2 and n_3 : group sizes, Δn_j : variance coefficient of the group size, $=/\neq$: homogeneity/heterogeneity of covariance matrices between groups, null/+/-: null/positive/negative pairing of group sizes and covariance matrices.

Table 3. Percentages of fit of the 11 candidate covariance structures selected by AIC for the UN population matrix selected by AIC.

Matrix of fit	K = 4							
	$\varepsilon = 0.57$				$\varepsilon = 0.75$			
	Covariances between groups				Covariances between groups			
	Pairing				Pairing			
	= Null	≠ Null	≠ +	≠ -	= Null	≠ Null	≠ +	≠ -
CS	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
UN	65.7	7.0	12.7	8.0	84.3	14.1	21.0	8.7
AR	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
HF	0.0	0.0	0.3	0.0	1.7	0.0	0.0	0.0
CSH	0.0	0.0	0.3	0.0	4.3	0.3	2.0	0.3
ARH	24.7	1.0	1.7	1.7	0.0	0.0	0.0	0.0
RC	0.3	0.0	0.0	0.0	0.0	0.0	0.0	0.0
UN _j	5.7	38	57.5	53.7	8.7	70.2	62.7	70.9
HF _j	0.0	0.7	4.0	0.7	1.0	15.4	14.0	20.1
ARH _j	3.7	53.0	23.4	36	0.0	0.0	0.3	0.0
RC _j	0.0	0.3	0.0	0.0	0.0	0.0	0.0	0.0
Matrix of fit	K = 6							
	$\varepsilon = 0.57$				$\varepsilon = 0.75$			
	Covariances between groups				Covariances between groups			
	Pairing				Pairing			
	= Null	≠ Null	≠ +	≠ -	= Null	≠ Null	≠ +	≠ -
CS	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
UN	54.0	19.5	5.1	58.4	22.7	2.0	5.3	10.7
AR	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
HF	0.0	0.0	1.3	0.0	0.0	0.3	1.0	0.0
CSH	0.0	0.0	0.0	0.0	47.3	5.3	12.0	5.7
ARH	0.0	0.0	32.6	0.0	16.0	4.7	1.0	0.3
RC	40.3	0.0	0.0	0.0	0.0	0.3	0.3	0.0
UN _j	5.8	77.5	1.3	30.8	8.3	12.3	12.3	20.7
HF _j	0.0	0.0	7.8	0.0	2.0	22.0	5.0	0.0
ARH _j	0.0	3.0	51.2	11.0	3.3	51.0	62.3	62.3
RC _j	0.0	0.0	0.7	0.0	0.3	2.0	0.7	0.3
Matrix of fit	K = 8							
	$\varepsilon = 0.57$				$\varepsilon = 0.75$			
	Covariances between groups				Covariances between groups			
	Pairing				Pairing			
	= Null	≠ Null	≠ +	≠ -	= Null	≠ Null	≠ +	≠ -
CS	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
UN	63.3	5.0	40.1	39.5	9.7	1.7	7.3	14.0
AR	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
HF	0.0	0.0	0.0	0.0	0.3	0.0	0.7	0.0
CSH	0.0	0.0	0.0	0.0	80.6	5.0	39.7	24.3
ARH	4.7	0.0	0.7	1.3	2.7	0.0	0.0	0.3
RC	0.0	0.0	0.0	0.0	0.7	0.0	0.3	0.0
UN _j	31.0	73.7	0.0	0.0	2.3	60.3	0.0	0.0
HF _j	0.7	0.0	0.0	0.0	3.0	13.0	0.0	0.0
ARH _j	0.3	21.3	59.2	59.2	0.3	20.0	46.7	61.0
RC _j	0.0	0.0	0.0	0.0	0.3	0.0	2.3	0.7

Note. In bold = selected matrix. The structure with the highest selection percentage is in bold

Table 4. Percentages of fit of the 11 candidate covariance structures selected by AIC for the ARH population matrix selected by AIC.

Matrix of fit	K = 4							
	$\varepsilon = 0.57$				$\varepsilon = 0.75$			
	Covariances between groups				Covariances between groups			
	Pairing				Pairing			
	= Null	≠ Null	≠ +	≠ -	= Null	≠ Null	≠ +	≠ -
CS	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
UN	11.7	1.3	1.3	2.4	10.7	3.0	1.7	1.3
AR	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
HF	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
CSH	0.7	0.3	0.0	0.0	34.0	2.7	3.0	2.0
ARH	78.0	4.0	4.7	3.7	46.7	1.0	3.7	1.0
RC	0.0	0.0	0.0	0.0	0.0	0.3	0.3	0.0
UN _j	4.3	10.7	26.7	28.7	3.7	12.0	24.0	24.3
HF _j	0.0	0.3	1.0	1.7	0.3	4.0	4.0	8.3
ARH _j	5.3	83.3	66.3	63.5	3.3	76.3	62.7	62.3
RC _j	0.0	0.0	0.0	0.0	1.3	0.7	0.7	0.7
Matrix of fit	K = 6							
	$\varepsilon = 0.57$				$\varepsilon = 0.75$			
	Covariances between groups				Covariances between groups			
	Pairing				Pairing			
	= Null	≠ Null	≠ +	≠ -	= Null	≠ Null	≠ +	≠ -
CS	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
UN	3.7	0.0	0.0	0.0	3.3	0.0	0.0	2.7
AR	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
HF	0.0	0.0	0.3	0.0	0.0	0.0	1.7	0.0
CSH	0.0	0.0	1.5	0.0	8.7	0.3	9.0	0.0
ARH	88.6	0.0	21.9	1.3	79.7	1.3	20.7	1.3
RC	1.7	0.0	0.3	0.0	0.0	0.0	3.0	0.0
UN _j	3.3	10.3	0.7	9.0	4.7	12.0	0.7	12.7
HF _j	0.0	0.3	1.3	0.3	0.0	1.7	2.3	1.3
ARH _j	2.7	89.3	66.4	89.0	3.3	84.7	56.5	81.7
RC _j	0.0	0.0	1.3	0.0	0.3	0.0	6.0	0.3
Matrix of fit	K = 8							
	$\varepsilon = 0.57$				$\varepsilon = 0.75$			
	Covariances between groups				Covariances between groups			
	Pairing				Pairing			
	= Null	≠ Null	≠ +	≠ -	= Null	≠ Null	≠ +	≠ -
CS	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
UN	1.3	0.0	0.0	0.3	40.7	7.0	88.3	79.3
AR	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
HF	0.0	0.0	0.0	0.0	2.3	0.3	1.3	3.3
CSH	0.0	0.0	0.0	0.0	18.7	2.0	6.3	9.7
ARH	79.7	0.3	2.7	4.3	0.0	0.0	0.0	0.0
RC	0.0	0.0	0.0	0.0	4.3	0.3	0.7	1.0
UN _j	16.7	22.0	0.0	10.7	27.7	90.0	0.0	0.0
HF _j	1.3	0.3	0.0	0.0	6.0	0.0	0.0	0.0
ARH _j	1.0	77.3	97.3	84.7	0.0	0.0	1.0	0.3
RC _j	0.0	0.0	0.0	0.0	0.3	0.3	2.3	6.3

Note. In bold = selected matrix. The structure with the highest selection percentage is in bold

Table 5. Percentages of fit of the 11 candidate covariance structures selected by AIC for the RC population matrix selected by AIC.

$K = 4$									
$\varepsilon = 0.57$					$\varepsilon = 0.75$				
Matrix of fit	Covariances between groups				Covariances between groups				
	Pairing				Pairing				
	= Null	≠ Null	≠ +	≠ -	= Null	≠ Null	≠ +	≠ -	
CS	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
UN	30.1	4.0	5.0	2.7	14.3	3.0	3.0	2.3	
AR	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	
HF	0.0	0.0	0.0	0.0	0.3	0.3	0.0	0.0	
CSH	40.0	8.0	8.7	8.7	66.0	8.7	12.3	7.0	
ARH	17.7	1.0	0.7	1.0	9.0	0.3	0.3	0.7	
RC	0.0	2.0	2.7	1.0	0.3	0.0	0.0	0.0	
UN _j	3.3	36.7	42.7	52.7	6.0	26.7	30.7	40.0	
HF _j	0.0	0.0	0.7	2.0	0.7	9.0	12.7	11.0	
ARH _j	3.0	45.0	36.0	29.3	3.0	52.0	40.7	39.0	
RC _j	5.7	3.3	3.7	2.7	0.3	0.0	0.3	0.0	
$K = 6$									
$\varepsilon = 0.57$					$\varepsilon = 0.75$				
Matrix of fit	Covariances between groups				Covariances between groups				
	Pairing				Pairing				
	= Null	≠ Null	≠ +	≠ -	= Null	≠ Null	≠ +	≠ -	
CS	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	
UN	17.0	4.3	0.0	13.7	11.7	1.7	6.3	11.3	
AR	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	
HF	0.0	0.0	4.0	0.0	0.3	0.0	1.0	0.0	
CSH	59.7	10.3	27.1	24.7	76.3	11.3	37.1	22.3	
ARH	0.0	1.3	1.7	0.0	0.3	0.0	0.7	0.0	
RC	0.0	0.6	14.7	3.7	1.7	0.7	4.1	0.7	
UN _j	6.7	55.7	0.3	27.3	4.3	39.0	0.0	21.7	
HF _j	0.0	2.3	6.6	0.0	1.7	22.7	9.2	5.0	
ARH _j	0.0	14.3	10.0	21.7	0.3	23.7	33.8	36.4	
RC _j	16.7	11.0	35.5	9.0	3.3	1.0	7.4	2.3	
$K = 8$									
$\varepsilon = 0.57$					$\varepsilon = 0.75$				
Matrix of fit	Covariances between groups				Covariances between groups				
	Pairing				Pairing				
	= Null	≠ Null	≠ +	≠ -	= Null	≠ Null	≠ +	≠ -	
CS	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	
UN	15.4	5.7	24.3	33.9	49.0	8.0	65.3	63.0	
AR	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	
HF	0.0	0.0	0.0	0.0	2.7	0.3	5.7	1.7	
CSH	52.7	6.0	43.0	27.8	17.7	2.7	14.0	21.3	
ARH	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	
RC	4.3	0.0	4.0	1.0	0.0	0.0	2.3	5.3	
UN _j	22.7	76.6	0.0	0.0	29.0	87.9	0.0	0.0	
HF _j	3.7	0.0	0.0	0.0	0.0	0.0	0.0	0.0	
ARH _j	0.0	11.4	22.7	34.6	0.0	0.0	0.3	0.3	
RC _j	1.0	0.3	6.0	2.7	1.7	1.0	12.3	11.7	

Note. In bold = selected matrix. The structure with the highest selection percentage is in bold

Table 6. Empirical rates of Type I error for the time and interaction effects (nominal value 0.05). Simulated covariance structure UN with $K = 4, 6$ and 8 repeated measures and $\varepsilon = 0.57$

								Time effect								
N	n_1	n_2	n_3	Δn_j	Covariances between groups	Pairing	$K = 4$			$K = 6$			$K = 8$			
							BW	SW	KR	BW	SW	KR	BW	SW	KR	
30	10	10	10	0.00	=	Null	0.087	0.092	0.075	0.091	0.093	0.069	0.102	0.086	0.068	
36	12	12	12	0.00	=	Null	0.081	0.086	0.067	0.073	0.077	0.063	0.072	0.074	0.060	
42	14	14	14	0.00	=	Null	0.070	0.071	0.061	0.082	0.082	0.072	0.079	0.079	0.068	
30	10	10	10	0.00	\neq	Null	0.097	0.084	0.070	0.126	0.088	0.064	0.123	0.097	0.071	
36	12	12	12	0.00	\neq	Null	0.085	0.071	0.061	0.110	0.088	0.061	0.107	0.079	0.070	
42	14	14	14	0.00	\neq	Null	0.088	0.072	0.060	0.091	0.067	0.058	0.095	0.075	0.064	
30	5	10	15	0.41	\neq	+	0.117	0.089	0.063	0.097	0.083	0.073	0.090	0.080	0.064	
36	6	12	18	0.41	\neq	+	0.113	0.089	0.073	0.090	0.075	0.059	0.063	0.057	0.049	
42	7	14	21	0.41	\neq	+	0.093	0.082	0.067	0.083	0.078	0.071	0.077	0.067	0.060	
30	15	10	5	0.41	\neq	-	0.231	0.150	0.109	0.268	0.272	0.243	0.158	0.107	0.077	
36	18	12	6	0.41	\neq	-	0.192	0.127	0.082	0.224	0.228	0.206	0.118	0.076	0.054	
42	21	14	7	0.41	\neq	-	0.189	0.133	0.098	0.182	0.132	0.105	0.129	0.101	0.072	
								Interaction effect								
N	n_1	n_2	n_3	Δn_j	Covariances between groups	Pairing	$K = 4$			$K = 6$			$K = 8$			
							BW	SW	KR	BW	SW	KR	BW	SW	KR	
30	10	10	10	0.00	=	Null	0.083	0.089	0.071	0.097	0.101	0.084	0.211	0.144	0.076	
36	12	12	12	0.00	=	Null	0.080	0.083	0.066	0.067	0.071	0.059	0.074	0.077	0.062	
42	14	14	14	0.00	=	Null	0.079	0.081	0.067	0.090	0.093	0.075	0.065	0.068	0.060	
30	10	10	10	0.00	\neq	Null	0.139	0.096	0.068	0.228	0.151	0.066	0.218	0.144	0.074	
36	12	12	12	0.00	\neq	Null	0.134	0.099	0.075	0.193	0.126	0.069	0.197	0.128	0.071	
42	14	14	14	0.00	\neq	Null	0.093	0.069	0.047	0.162	0.105	0.060	0.176	0.123	0.075	
30	5	10	15	0.41	\neq	+	0.236	0.162	0.074	0.135	0.098	0.062	0.144	0.101	0.066	
36	6	12	18	0.41	\neq	+	0.197	0.129	0.059	0.127	0.099	0.070	0.105	0.083	0.053	
42	7	14	21	0.41	\neq	+	0.153	0.113	0.070	0.130	0.101	0.071	0.122	0.087	0.070	
30	15	10	5	0.41	\neq	-	0.429	0.276	0.153	0.294	0.301	0.272	0.238	0.139	0.087	
36	18	12	6	0.41	\neq	-	0.348	0.229	0.131	0.265	0.267	0.233	0.186	0.131	0.073	
42	21	14	7	0.41	\neq	-	0.303	0.202	0.126	0.297	0.198	0.120	0.175	0.116	0.074	

Note. BW, SW and KR = methods for calculating the degrees of freedom (Between-Within, Satterthwaite and Kenward-Roger) In bold = liberal

Table 7. Empirical rates of Type I error for the time and interaction effects (nominal value 0.05). Simulated covariance structure ARH with $K = 4, 6$ and 8 repeated measures and $\epsilon = 0.57$

								Time effect								
N	n_1	n_2	n_3	Δn_j	Covariances between groups	Pairing	K = 4			K = 6			K = 8			
							BW	SW	KR	BW	SW	KR	BW	SW	KR	
30	10	10	10	0.00	=	Null	0.096	0.083	0.073	0.095	0.079	0.070	0.088	0.078	0.073	
36	12	12	12	0.00	=	Null	0.081	0.071	0.064	0.083	0.067	0.065	0.073	0.064	0.063	
42	14	14	14	0.00	=	Null	0.065	0.056	0.054	0.083	0.072	0.068	0.066	0.062	0.062	
30	10	10	10	0.00	≠	Null	0.094	0.074	0.064	0.096	0.074	0.061	0.112	0.084	0.065	
36	12	12	12	0.00	≠	Null	0.080	0.062	0.051	0.065	0.055	0.046	0.079	0.067	0.064	
42	14	14	14	0.00	≠	Null	0.093	0.079	0.071	0.070	0.063	0.053	0.076	0.067	0.064	
30	5	10	15	0.41	≠	+	0.095	0.082	0.065	0.094	0.078	0.066	0.092	0.074	0.065	
36	6	12	18	0.41	≠	+	0.068	0.054	0.047	0.093	0.078	0.071	0.079	0.068	0.064	
42	7	14	21	0.41	≠	+	0.069	0.058	0.054	0.068	0.057	0.052	0.077	0.065	0.065	
30	15	10	5	0.41	≠	-	0.174	0.112	0.092	0.149	0.091	0.065	0.147	0.096	0.082	
36	18	12	6	0.41	≠	-	0.158	0.108	0.083	0.133	0.087	0.074	0.140	0.088	0.075	
42	21	14	7	0.41	≠	-	0.116	0.069	0.052	0.108	0.075	0.060	0.124	0.086	0.071	
								Interaction effect								
N	n_1	n_2	n_3	Δn_j	Covariances between groups	Pairing	K = 4			K = 6			K = 8			
							BW	SW	KR	BW	SW	KR	BW	SW	KR	
30	10	10	10	0.00	=	Null	0.098	0.081	0.078	0.093	0.076	0.072	0.089	0.069	0.070	
36	12	12	12	0.00	=	Null	0.080	0.066	0.062	0.090	0.067	0.067	0.074	0.055	0.055	
42	14	14	14	0.00	=	Null	0.075	0.066	0.064	0.084	0.076	0.074	0.085	0.070	0.069	
30	10	10	10	0.00	≠	Null	0.139	0.089	0.065	0.136	0.087	0.057	0.228	0.155	0.079	
36	12	12	12	0.00	≠	Null	0.113	0.080	0.062	0.121	0.082	0.058	0.126	0.091	0.073	
42	14	14	14	0.00	≠	Null	0.111	0.085	0.074	0.112	0.077	0.055	0.112	0.078	0.067	
30	5	10	15	0.41	≠	+	0.146	0.101	0.054	0.135	0.100	0.065	0.148	0.096	0.077	
36	6	12	18	0.41	≠	+	0.126	0.084	0.056	0.135	0.088	0.053	0.110	0.075	0.058	
42	7	14	21	0.41	≠	+	0.107	0.078	0.061	0.095	0.071	0.052	0.102	0.073	0.061	
30	15	10	5	0.41	≠	-	0.261	0.141	0.099	0.225	0.122	0.072	0.238	0.129	0.091	
36	18	12	6	0.41	≠	-	0.208	0.129	0.094	0.189	0.109	0.072	0.187	0.124	0.095	
42	21	14	7	0.41	≠	-	0.158	0.085	0.056	0.163	0.097	0.071	0.157	0.089	0.071	

Note. BW, SW and KR = methods for calculating the degrees of freedom (Between-Within, Satterthwaite and Kenward-Roger) In bold = liberal

Table 8. Empirical rates of Type I error for the time and interaction effects (nominal value 0.05).
Simulated covariance structure RC with $K = 4, 6$ and 8 repeated measures and $\varepsilon = 0.57$

								Time effect								
N	n_1	n_2	n_3	Δn_j	Covariances between groups	Pairing	$K = 4$			$K = 6$			$K = 8$			
							BW	SW	KR	BW	SW	KR	BW	SW	KR	
30	10	10	10	0.00	=	Null	0.066	0.053	0.051	0.079	0.065	0.064	0.099	0.084	0.071	
36	12	12	12	0.00	=	Null	0.063	0.053	0.049	0.075	0.067	0.064	0.071	0.063	0.061	
42	14	14	14	0.00	=	Null	0.063	0.067	0.059	0.078	0.071	0.067	0.072	0.067	0.066	
30	10	10	10	0.00	\neq	Null	0.105	0.092	0.076	0.129	0.100	0.076	0.131	0.098	0.075	
36	12	12	12	0.00	\neq	Null	0.101	0.082	0.063	0.108	0.086	0.074	0.119	0.100	0.078	
42	14	14	14	0.00	\neq	Null	0.087	0.072	0.064	0.075	0.059	0.050	0.097	0.086	0.069	
30	5	10	15	0.41	\neq	+	0.105	0.085	0.071	0.108	0.083	0.095	<i>0.017</i>	<i>0.015</i>	<i>0.015</i>	
36	6	12	18	0.41	\neq	+	0.089	0.069	0.056	0.068	0.046	0.064	<i>0.013</i>	<i>0.013</i>	<i>0.013</i>	
42	7	14	21	0.41	\neq	+	0.089	0.082	0.068	<i>0.001</i>	<i>0.001</i>	<i>0.001</i>	<i>0.015</i>	<i>0.015</i>	<i>0.015</i>	
30	15	10	5	0.41	\neq	-	0.226	0.141	0.103	0.127	0.095	0.069	0.274	0.276	0.249	
36	18	12	6	0.41	\neq	-	0.208	0.136	0.090	0.223	0.205	0.202	0.111	0.077	0.065	
42	21	14	7	0.41	\neq	-	0.169	0.118	0.085	0.177	0.126	0.088	0.114	0.080	0.067	
								Interaction effect								
N	n_1	n_2	n_3	Δn_j	Covariances between groups	Pairing	$K = 4$			$K = 6$			$K = 8$			
							BW	SW	KR	BW	SW	KR	BW	SW	KR	
30	10	10	10	0.00	=	Null	0.066	0.056	0.055	0.097	0.079	0.075	0.194	0.133	0.056	
36	12	12	12	0.00	=	Null	0.055	0.043	0.041	0.081	0.064	0.064	0.091	0.077	0.078	
42	14	14	14	0.00	=	Null	0.072	0.072	0.061	0.075	0.060	0.059	0.080	0.065	0.066	
30	10	10	10	0.00	\neq	Null	0.155	0.106	0.084	0.227	0.162	0.091	0.204	0.143	0.076	
36	12	12	12	0.00	\neq	Null	0.211	0.131	0.073	0.187	0.125	0.078	0.190	0.129	0.069	
42	14	14	14	0.00	\neq	Null	0.127	0.092	0.079	0.149	0.103	0.059	0.156	0.104	0.068	
30	5	10	15	0.41	\neq	+	0.254	0.172	0.087	0.244	0.185	0.012	0.027	<i>0.020</i>	<i>0.021</i>	
36	6	12	18	0.41	\neq	+	0.193	0.143	0.072	0.251	0.069	0.098	<i>0.023</i>	<i>0.015</i>	<i>0.015</i>	
42	7	14	21	0.41	\neq	+	0.156	0.105	0.071	0.031	0.025	<i>0.024</i>	0.028	<i>0.023</i>	<i>0.023</i>	
30	15	10	5	0.41	\neq	-	0.425	0.287	0.178	0.222	0.118	0.082	0.291	0.295	0.273	
36	18	12	6	0.41	\neq	-	0.337	0.225	0.125	0.246	0.229	0.224	0.177	0.083	0.072	
42	21	14	7	0.41	\neq	-	0.271	0.173	0.112	0.306	0.200	0.121	0.150	0.082	0.071	

Note. BW, SW and KR = methods for calculating the degrees of freedom (Between-Within, Satterthwaite and Kenward-Roger) In bold = liberal; in italics = conservative

Table 9. Empirical rates of Type I error for the time and interaction effects (nominal value 0.05). Simulated covariance structure UN with $K = 4, 6$ and 8 repeated measures $\varepsilon = 0.75$

Time effect															
N	n_1	n_2	n_3	Δn_j	Covariances between groups	Pairing	K = 4			K = 6			K = 8		
							BW	SW	KR	BW	SW	KR	BW	SW	KR
30	10	10	10	0.00	=	Null	0.079	0.081	0.067	0.096	0.082	0.079	0.089	0.072	0.070
36	12	12	12	0.00	=	Null	0.082	0.083	0.073	0.076	0.066	0.063	0.077	0.070	0.067
42	14	14	14	0.00	=	Null	0.070	0.073	0.057	0.055	0.050	0.049	0.067	0.057	0.056
30	10	10	10	0.00	≠	Null	0.107	0.083	0.070	0.090	0.076	0.063	0.103	0.071	0.053
36	12	12	12	0.00	≠	Null	0.112	0.089	0.075	0.056	0.048	0.045	0.104	0.080	0.059
42	14	14	14	0.00	≠	Null	0.070	0.055	0.048	0.078	0.068	0.064	0.095	0.077	0.061
30	5	10	15	0.41	≠	+	0.109	0.079	0.067	0.076	0.064	0.055	0.077	0.064	0.059
36	6	12	18	0.41	≠	+	0.107	0.090	0.073	0.065	0.059	0.052	0.080	0.072	0.065
42	7	14	21	0.41	≠	+	0.098	0.084	0.064	0.061	0.054	0.046	0.084	0.075	0.071
30	15	10	5	0.41	≠	-	0.239	0.165	0.133	0.156	0.110	0.092	0.156	0.105	0.083
36	18	12	6	0.41	≠	-	0.237	0.150	0.107	0.140	0.090	0.068	0.139	0.096	0.080
42	21	14	7	0.41	≠	-	0.186	0.129	0.098	0.116	0.077	0.064	0.104	0.076	0.065
Interaction effect															
N	n_1	n_2	n_3	Δn_j	Covariances between groups	Pairing	K = 4			K = 6			K = 8		
							BW	SW	KR	BW	SW	KR	BW	SW	KR
30	10	10	10	0.00	=	Null	0.097	0.102	0.080	0.106	0.091	0.091	0.087	0.072	0.069
36	12	12	12	0.00	=	Null	0.079	0.081	0.067	0.080	0.068	0.067	0.083	0.070	0.072
42	14	14	14	0.00	=	Null	0.084	0.089	0.071	0.077	0.069	0.065	0.081	0.073	0.073
30	10	10	10	0.00	≠	Null	0.223	0.150	0.074	0.143	0.094	0.065	0.193	0.129	0.071
36	12	12	12	0.00	≠	Null	0.180	0.121	0.064	0.099	0.068	0.053	0.087	0.123	0.073
42	14	14	14	0.00	≠	Null	0.168	0.123	0.069	0.106	0.085	0.072	0.163	0.116	0.071
30	5	10	15	0.41	≠	+	0.230	0.160	0.090	0.126	0.088	0.064	0.124	0.086	0.064
36	6	12	18	0.41	≠	+	0.208	0.163	0.080	0.126	0.097	0.072	0.114	0.091	0.071
42	7	14	21	0.41	≠	+	0.155	0.115	0.071	0.094	0.070	0.057	0.095	0.072	0.054
30	15	10	5	0.41	≠	-	0.435	0.300	0.195	0.198	0.109	0.088	0.225	0.126	0.096
36	18	12	6	0.41	≠	-	0.379	0.253	0.147	0.198	0.126	0.088	0.199	0.103	0.081
42	21	14	7	0.41	≠	-	0.288	0.188	0.124	0.145	0.087	0.077	0.150	0.097	0.076

Note. BW, SW and KR = methods for calculating the degrees of freedom (Between-Within, Satterthwaite and Kenward-Roger) In bold = liberal

Table 10. Empirical rates of Type I error for the time and interaction effects (nominal value 0.05).
Simulated covariance structure ARH with $K = 4, 6$ and 8 repeated measures and $\varepsilon = 0.75$

								Time effect								
N	n_1	n_2	n_3	Δn_j	Covariances between groups	Pairing	$K = 4$			$K = 6$			$K = 8$			
							BW	SW	KR	BW	SW	KR	BW	SW	KR	
30	10	10	10	0.00	=	Null	0.072	0.065	0.063	0.081	0.075	0.075	0.106	0.087	0.069	
36	12	12	12	0.00	=	Null	0.075	0.065	0.062	0.104	0.093	0.092	0.081	0.084	0.069	
42	14	14	14	0.00	=	Null	0.084	0.075	0.072	0.064	0.055	0.056	0.076	0.076	0.066	
30	10	10	10	0.00	\neq	Null	0.074	0.055	0.053	0.095	0.080	0.075	0.120	0.087	0.067	
36	12	12	12	0.00	\neq	Null	0.073	0.056	0.051	0.089	0.073	0.069	0.108	0.087	0.073	
42	14	14	14	0.00	\neq	Null	0.077	0.066	0.065	0.072	0.062	0.058	0.095	0.076	0.063	
30	5	10	15	0.41	\neq	+	0.087	0.071	0.068	0.069	0.060	0.058	0.025	0.027	<i>0.021</i>	
36	6	12	18	0.41	\neq	+	0.068	0.053	0.054	0.091	0.076	0.075	<i>0.014</i>	<i>0.014</i>	<i>0.010</i>	
42	7	14	21	0.41	\neq	+	0.068	0.053	0.072	0.069	0.059	0.061	<i>0.015</i>	<i>0.015</i>	<i>0.014</i>	
30	15	10	5	0.41	\neq	-	0.197	0.122	0.098	0.169	0.107	0.094	0.284	0.285	0.254	
36	18	12	6	0.41	\neq	-	0.134	0.099	0.086	0.100	0.062	0.053	0.254	0.258	0.230	
42	21	14	7	0.41	\neq	-	0.126	0.088	0.082	0.123	0.092	0.074	0.243	0.248	0.225	
								Interaction effect								
N	n_1	n_2	n_3	Δn_j	Covariances between groups	Pairing	$K = 4$			$K = 6$			$K = 8$			
							BW	SW	KR	BW	SW	KR	BW	SW	KR	
30	10	10	10	0.00	=	Null	0.088	0.076	0.076	0.092	0.071	0.072	0.191	0.125	0.066	
36	12	12	12	0.00	=	Null	0.071	0.058	0.058	0.080	0.066	0.068	0.084	0.089	0.075	
42	14	14	14	0.00	=	Null	0.072	0.058	0.057	0.069	0.064	0.064	0.073	0.074	0.066	
30	10	10	10	0.00	\neq	Null	0.122	0.078	0.058	0.111	0.078	0.069	0.218	0.136	0.071	
36	12	12	12	0.00	\neq	Null	0.102	0.068	0.056	0.105	0.067	0.054	0.178	0.119	0.064	
42	14	14	14	0.00	\neq	Null	0.104	0.074	0.060	0.098	0.070	0.058	0.167	0.118	0.073	
30	5	10	15	0.41	\neq	+	0.119	0.089	0.072	0.126	0.084	0.058	0.040	0.044	0.036	
36	6	12	18	0.41	\neq	+	0.117	0.085	0.061	0.122	0.090	0.075	0.032	0.033	0.028	
42	7	14	21	0.41	\neq	+	0.117	0.085	0.056	0.109	0.079	0.071	<i>0.024</i>	<i>0.024</i>	<i>0.020</i>	
30	15	10	5	0.41	\neq	-	0.244	0.137	0.116	0.222	0.129	0.101	0.285	0.289	0.257	
36	18	12	6	0.41	\neq	-	0.177	0.099	0.079	0.169	0.094	0.078	0.292	0.299	0.272	
42	21	14	7	0.41	\neq	-	0.166	0.110	0.093	0.143	0.078	0.070	0.255	0.259	0.244	

Note. BW, SW and KR = methods for calculating the degrees of freedom (Between-Within, Satterthwaite and Kenward-Roger) In bold = liberal; in italics = conservative

Table 11. Empirical rates of Type I error for the time and interaction effects (nominal value 0.05). Simulated covariance structure RC with $K = 4, 6$ and 8 repeated measures and $\varepsilon = 0.75$

								Time effect								
N	n_1	n_2	n_3	Δn_j	Covariances between groups	Pairing	K = 4			K = 6			K = 8			
							BW	SW	KR	BW	SW	KR	BW	SW	KR	
30	10	10	10	0.00	=	Null	0.079	0.072	0.071	0.095	0.088	0.084	0.105	0.080	0.068	
36	12	12	12	0.00	=	Null	0.071	0.058	0.058	0.083	0.069	0.065	0.066	0.068	0.060	
42	14	14	14	0.00	=	Null	0.083	0.077	0.076	0.074	0.066	0.064	0.073	0.073	0.065	
30	10	10	10	0.00	≠	Null	0.092	0.070	0.069	0.115	0.087	0.066	0.119	0.081	0.066	
36	12	12	12	0.00	≠	Null	0.076	0.060	0.057	0.084	0.071	0.054	0.114	0.092	0.070	
42	14	14	14	0.00	≠	Null	0.055	0.046	0.046	0.093	0.074	0.061	0.098	0.081	0.069	
30	5	10	15	0.41	≠	+	0.105	0.084	0.062	0.096	0.097	0.097	<i>0.023</i>	<i>0.023</i>	<i>0.018</i>	
36	6	12	18	0.41	≠	+	0.070	0.063	0.061	<i>0.011</i>	<i>0.001</i>	<i>0.001</i>	<i>0.019</i>	<i>0.020</i>	<i>0.017</i>	
42	7	14	21	0.41	≠	+	0.061	0.051	0.050	<i>0.001</i>	<i>0.001</i>	<i>0.001</i>	<i>0.017</i>	<i>0.017</i>	<i>0.015</i>	
30	15	10	5	0.41	≠	-	0.241	0.152	0.094	0.140	0.990	0.083	0.255	0.240	0.236	
36	18	12	6	0.41	≠	-	0.141	0.098	0.088	0.144	0.092	0.079	0.270	0.273	0.245	
42	21	14	7	0.41	≠	-	0.131	0.089	0.078	0.158	0.114	0.083	0.230	0.235	0.211	
								Interaction effect								
N	n_1	n_2	n_3	Δn_j	Covariances between groups	Pairing	K = 4			K = 6			K = 8			
							BW	SW	KR	BW	SW	KR	BW	SW	KR	
30	10	10	10	0.00	=	Null	0.097	0.076	0.079	0.090	0.076	0.075	0.198	0.129	0.063	
36	12	12	12	0.00	=	Null	0.100	0.074	0.075	0.096	0.087	0.085	0.083	0.086	0.072	
42	14	14	14	0.00	=	Null	0.074	0.062	0.064	0.070	0.059	0.058	0.071	0.073	0.061	
30	10	10	10	0.00	≠	Null	0.124	0.085	0.075	0.216	0.138	0.070	0.214	0.146	0.066	
36	12	12	12	0.00	≠	Null	0.116	0.073	0.060	0.186	0.123	0.076	0.176	0.114	0.067	
42	14	14	14	0.00	≠	Null	0.103	0.065	0.058	0.159	0.109	0.070	0.156	0.112	0.065	
30	5	10	15	0.41	≠	+	0.223	0.155	0.095	0.095	0.090	0.089	0.042	0.043	0.036	
36	6	12	18	0.41	≠	+	0.109	0.077	0.067	0.026	<i>0.017</i>	<i>0.017</i>	0.035	0.035	0.030	
42	7	14	21	0.41	≠	+	0.097	0.067	0.059	0.028	0.025	0.025	0.034	0.038	0.028	
30	15	10	5	0.41	≠	-	0.429	0.283	0.171	0.196	0.118	0.100	0.281	0.248	0.243	
36	18	12	6	0.41	≠	-	0.207	0.118	0.104	0.177	0.097	0.082	0.275	0.283	0.260	
42	21	14	7	0.41	≠	-	0.166	0.092	0.082	0.292	0.182	0.112	0.254	0.256	0.235	

Note. BW, SW and KR = methods for calculating the degrees of freedom (Between-Within, Satterthwaite and Kenward-Roger) In bold = liberal; in italics = conservative