2004-05-17

Performance of the Kenward-Project when the Covariance Structure is Selected Using AIC and BIC

Elisa Valderas Gomez
Brigham Young University - Provo

PERFORMANCE OF THE KENWARD-ROGER METHOD WHEN THE COVARARIANCE STRUCTURE IS SELECTED USING AIC AND BIC

by

Elisa Valderas Gomez

A project submitted to the faculty of

Brigham Young University

in partial fulfillment of the requirements for the degree of

Master of Science

Department of Statistics

Brigham Young University

May 2004

BRIGHAM YOUNG UNIVERSITY

GRADUATE COMMITTEE APPROVAL

of a project submitted by

Elisa Valderas Gomez

This project has been read by each member of the following graduate committee and by majority vote has been found to be satisfactory.

_____	_____
Date	G. Bruce Schaalje, Chair
_____	_____
Date	Gilbert W. Fellingham
_____	_____
Date	William F. Christensen

BRIGHAM YOUNG UNIVERSITY

As chair of the candidate's graduate committee, I have read the project of Elisa Gomez Valderas in its final form and have found that (1) its format, citations, and bibliographical style are consistent and acceptable and fulfill university and department style requirements; (2) its illustrative materials including figures, tables and charts are in place; and (3) the final manuscript is satisfactory to the graduate committee and is ready for submission to the university library.

_____ _____

Date G. Bruce Schaalje
Chair, Graduate Committee

Accepted for the Department

_____

G. Bruce Schaalje
Graduate Coordinator

Accepted for the College

_____

G. Rex Bryce
Associate Dean, College of Physical and Mathematical Science

ABSTRACT

PERFORMANCE OF THE KENWARD-ROGER METHOD WHEN THE COVARARIANCE STRUCTURE IS SELECTED USING AIC AND BIC

Elisa Valderas Gomez

Department of Statistics

Masters of Science

Linear mixed models are frequently used to analyze data with random effects and/or repeated measures. A common approach to such analyses requires choosing a covariance structure. Information criteria, such as AIC and BIC, are often used by statisticians to help with this task. However, these criteria do not always point to the true covariance structure and therefore the wrong covariance structure is sometimes chosen. Once this step is complete, Wald statistics are used to test fixed effects. Degrees of freedom for these statistics are not known. However, there are approximation methods, such as Kenward and Roger (KR) and Satterthwaite (SW) that have been shown to work well in some situations. Schaalje et al. (2002) concluded that the KR method would perform at least as well as or better than the SW method in

many cases assuming that the covariance structure was known. On the other hand, Keselman et al. (1999) concluded that the performance of the SW method when the covariance structure was selected using AIC was poor for negative pairings of treatment sizes and covariance matrices and small sample sizes. Our study used simulations to investigate Type I error rates in test of fixed effects using Wald statistics with the KR adjustment method, incorporating the selection of the covariance structure using AIC and BIC. Performance of the AIC and BIC criteria in selecting the true covariance structure was also studied. The MIXED procedure (SAS v. 9) was used to analyze each simulated data set. Type I error rates from the best AIC and BIC models were always higher than target values. However, Type I error rates obtained by using the BIC criterion were better than those obtained by using the AIC criterion. Type I error rates for the correct models were often adequate depending on the sample size and complexity of covariance structure. Performance of AIC and BIC was poor. This could be a consequence of small sample sizes and the high number of covariance structures these criteria had to choose from.

ACKNOWLEDGMENTS

I would like to thank Dr. Bruce Schaalje for making possible the completion of this project by sharing his time, comments and feedback. Also, I would like to express thanks to Dr. Gilbert Fellingham and Dr. William Christensen for their help and suggestions with computer issues. I feel honored to have had the opportunity to work and learn in the Department of Statistics of Brigham Young University.

Special thanks go to my husband Ben for his unconditional help, love and support, to my mom for her faith and to my family in Spain and in the United States for their encouragement and trust. Finally, I want to express my gratitude to God for His Love and for giving me the strength to go on every day.

TABLES OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

1. INTRODUCTION

Linear mixed models are frequently used to analyze data with random effects and/or repeated measures. However, different techniques are used to implement linear mixed models. One of the steps in a common approach (e.g. SAS Proc Mixed) is to choose a covariance structure. The task of choosing a covariance structure is not simple, and the literature includes many examples of how important it is to choose the right structure (Grady and Helms, 1995; Singer, 1998; Littell et al., 2000), and the consequences of not doing so on the type I error rates for testing the fixed effects (Keselman et al., 1999; Ferron et al., 2002).

Information criteria (AIC, BIC) are often used by statisticians to choose the covariance structure (Singer, 1998; Keselman et al., 1999; Littell et al., 2000). Unfortunately, these criteria do not always point to the true covariance structure (Keselman et al., 1999; Ferron et al., 2002). Therefore, by only relying on these criteria, the wrong choice of covariance structure will be sometimes made.

Once the covariance structure has been chosen, the next step often involves tests and estimates of fixed effects using Wald statistics (Schaalje et al., 2001; Schaalje et al., 2002). Valid inferences about fixed effects in linear mixed models depend on the calculation of appropriate denominator degrees of freedom and the adjustment of the estimated covariance matrix and the Wald statistic. Such methods have been suggested, including the Satterthwaite (SW) and the Kenward-Roger (KR) procedures (Fai and Cornelius, 1996; Kenward and Roger, 1997). It has been shown, using simulation studies, that the SW and KR methods behave well in complicated

situations (Keselman et al., 1998; Schaalje et al, 2002). However, how well they behave depends on the true underlying covariance structure.

An alternative method of testing fixed effects in linear mixed models eliminates the necessity of choosing a covariance structure. This method is a non-pooled adjusted degrees of freedom multivariate test called the Welch-James-type test (WJ) (Johansen, 1980; Keselman et al., 1993). Simulations have shown the robustness of this method in several complex situations (Keselman et al., 1998). Some important restrictions apply to this method since it is appropriate only for repeated measures designs without covariates or missing values.

Schaalje et al. (2002) studied and compared Type I error rates in tests of fixed effects in linear mixed models. Wald statistics with the SW and KR methods were used. In their study, the true covariance structure was assumed to be known. On the other hand, Keselman et al. (1999) studied Type I error rates in tests of fixed effects of repeated measures. They used Wald statistics with the SW denominator degrees of freedom method and incorporated selection of the covariance structure into the error rates. In addition, they compared these results to the Type I error rates obtained using the WJ test.

This study investigates Type I error rates in tests of fixed effects in linear mixed models using Wald statistics with the KR method, incorporating the selection of covariance structure using AIC and BIC, as well as using the true covariance structure. Performance of the AIC and BIC criteria in selecting the true covariance structure will also be studied.

2. LITERATURE REVIEW

2.1. Mixed Model

The mixed model can be written as:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u} + \boldsymbol{\varepsilon},$$

where **X** and **Z** are the corresponding design matrices for the fixed and the random effects, $\boldsymbol{\beta}$ and **u** are the vectors of coefficients of the fixed and random effects, respectively, and $\boldsymbol{\varepsilon}$ denotes a vector of errors. The vector of coefficients of the random effects, **u,** is assumed to follow the normal distribution with mean **0** and covariance matrix **G**. $\boldsymbol{\varepsilon}$ is assumed to follow a normal distribution with mean **0** and covariance matrix **R**. Neither **G** nor **R** needs to be a diagonal matrix. The vectors **u** and $\boldsymbol{\varepsilon}$ are assumed independent; consequently, cov (**u**, $\boldsymbol{\varepsilon}$) = **0**, and Var (**y**) = **ZGZ'** + **R** = **V**.

If **V** is known, the generalized least squares estimator is the best linear unbiased estimator and can be written as:

$$\hat{\boldsymbol{\beta}}_{gls} = (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1}\mathbf{y}$$

If **V** is unknown, the estimated generalized least squares estimate of $\boldsymbol{\beta}$ is:

$$\hat{\boldsymbol{\beta}}_{egls} = (\mathbf{X}'\hat{\mathbf{V}}^{-1}\mathbf{X})^{-1}\mathbf{X}'\hat{\mathbf{V}}^{-1}\mathbf{y}$$

where $\hat{\mathbf{V}}$ is an estimate of the covariance matrix. This estimate is often obtained using maximum likelihood (ML) or restricted maximum likelihood (REML) methodology. The approximate covariance matrix of $\hat{\boldsymbol{\beta}}_{egls}$ is $(\mathbf{X}'\hat{\mathbf{V}}^{-1}\mathbf{X})^{-1}$. A commonly used test statistic for $H_0 : \mathbf{C}\boldsymbol{\beta} = \mathbf{0}$ is known as the Wald statistic. It can be written as:

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

where **C** is a matrix of contrasts of rank q. Asymptotically, W follows a chi-square distribution with q degrees of freedom if there is no variation in the term $\mathbf{(C(X'\hat{V}^{-1}X)^{-1}C')}$. Another option is to use the Wald F statistic, $F_{q,ddf} = W/q$, where ddf is the denominator degrees of freedom. This statistic takes into account that $\mathbf{(C(X'\hat{V}^{-1}X)^{-1}C')}$ is estimated and not known. A common method to calculate the ddf is the Satterthwaite (SW) method which computes the ddf using spectral decomposition of the hypothesis matrix together with repeated application of a method for single-degree-of-freedom-tests (Fai and Cornelius, 1996; Schaalje et al., 2002). Kenward and Roger (1997) have suggested a method which calculates the ddf, modifies the estimate of $\mathbf{V}$ and further adjusts the Wald F statistic to account for small sample bias and variability in $\hat{\mathbf{V}}$ (Schaalje et al., 2001).

2.2. Selecting Covariance Structure

The process of analyzing data using usually begins with the choice of **G** and **R**, often referred as the covariance structure specification. Commenting on the importance of this decision, Littell et al. (2000) noted that by using incorrect covariance structures we risk obtaining invalid estimators and inferences. Since the generalized least squares estimator is the best linear unbiased estimator, an incorrect covariance structure will affect the quality of the estimator. Littell et al. (2000) suggested that the first thing to do when choosing a covariance structure in repeated measures studies is to compute the unstructured sample covariance matrix and compare it to the covariance matrix estimates obtained using other structures including

compound symmetry (CS), autoregressive order 1 (AR(1)), autoregressive with random effects, and toeplitz. They observed which covariance matrices preserved the main patterns observed in the unstructured covariance matrix. They also used the correlogram (Cressie, 1993), a graphical tool which plots the correlation function, as a technique for visualizing patterns in the covariance structure.

Other tools in covariance structure selection were Akaike's information (Akaike, 1973) and Schwarz's Bayesian criteria (Schwarz, 1978), AIC and BIC respectively. These criteria are used as goodness-of-fit statistics to compare models with the same fixed effects but different covariance structures. Both criteria penalize for the number of parameters in the covariance structure in question, the penalty imposed by BIC being greater.

Littell et al.(2000) chose to rely more on BIC since their objective was parsimonious modeling. An advantage of parsimonious modeling is more powerful tests and more efficient estimates of fixed effects (Keselman et al., 1998). After considering all of the mentioned aids to covariance structure selection, Littell et. al. (2000) decided the autoregressive plus random coefficient structure was the best one for their study. This structure had the best BIC value, but more than that, it preserved the properties they were expecting to see in their study, that is, correlated observations in the same patient and correlation decreasing as the time lag increased. Later, they fitted other covariance structures and found that the estimates for the fixed effects were similar for some of these structures. However, the standard deviations changed, which resulted in unusually large or small F-values for tests of fixed effects.

Grady and Helms (1995) argued that a better understanding of the data is obtained as a consequence of the covariance-modeling process because the covariance structure followed by the data may contain much information of interest to the researcher. In their study, Grady and Helms dealt with missing data. One advantage of the covariance structure modeling approach is that it can deal with missing data without the necessity of case-wise deletion of data. They mentioned that a main aspect of choosing a covariance structure other than the unstructured one is the reduction in the number of parameters. This is an advantage when the chosen covariance structure is adequate, yielding good estimators and test statistics. However, if the structure is not adequate, ill-fitting estimators of the true variance will be obtained (Littell et al., 2000). Grady and Helms also recommended the use of correlograms to recognize correlation patterns to help in identifying an adequate structure. After considering several structures, likelihood ratio tests were computed and found to be helpful in recognizing a sensible structure. Grady and Helms pointed out that the high number of subjects in each group and repeated measures for each subject in their study could result in 'too much' power for the likelihood ratio tests.

2.3. Performance of AIC and BIC

Littell et al. (2000) recommended the use of the AIC and BIC criteria to help statisticians choose an adequate covariance structure. Keselman et al. (1998) studied the effectiveness of these two criteria in non-well-behaved, but common types of data in social sciences. These situations included nonspherical covariance structures, unequal sample sizes, unequal covariance structure by group, and normal/non-normal

data. They simulated data from three covariance structures, including Unstructured (UN), AR(1), and Random Coefficients (RC). These structures were used with and without heterogeneity within subjects and between groups.

Their results show that AIC chose the correct covariance structure 47% of the time while BIC was right 35% of the time. They also showed that the number of times AIC chose the right structure depended on the true covariance structure. For normal AR(1) and RC data with heterogeneity within-subjects and between groups, or log-normal UN data with heterogeneity between groups, AIC performed better. BIC picked the wrong covariance structure more often than the correct one for most types of data. The negative results obtained for BIC could be due to the severe penalty imposed for the number of parameters (Keselman et al., 1998). A possible explanation for the low success of both criteria could be that other simpler structures were good approximations to the true covariance structure.

Ferron et al. (2002) also looked at the performance of the AIC, BIC and likelihood ratio test (LRT) in helping the statistician choose an adequate covariance structure. Data were generated following an AR(1) structure with different sample sizes, numbers of repeated measures, and levels of autocorrelation. AIC, BIC and LRT were computed for the true covariance structure, AR(1), and for a single alternative, $\sigma^2\mathbf{I}$. The results show that AIC was more successful than the other two criteria for every combination of sample size, length of repeated measures and level of autocorrelation. AIC choose the right covariance structure 79% of the time versus 66% and 71% for BIC and LRT, respectively. AIC, BIC and LRT performed better when the sample size, the length of the repeated measures and the level of

autocorrelation were higher, the length of the repeated measures being the most influential characteristic. The effect of sample size was greater when the length of the repeated measures was short. The rates of success of these criteria also depended on the parameter values chosen to generate the data. It is important to notice that AIC, BIC and LRT may perform worse when there are more than two covariance structures to choose from, as in this case.

Since AIC and BIC may lead the researcher to fit an incorrect covariance structure, Ferron et al. (2002) studied the effects of a special case of misspecification of the **R** matrix on the estimates and tests of fixed effects. Data with linear and non-linear growth curves from two covariance structures, AR(1) and moving average (MA), were generated. Sample size, length of repeated measures and level of autocorrelation were varied. Data included missing and unequally spaced observations. An RC model was fitted with $\mathbf{R} = \sigma^2\mathbf{I}$ and an unstructured **G**. The Wald F statistic with the between/within method of SAS (Proc Mixed) for the ddf was used for tests of fixed effects (Ferron et al., 2002). As a consequence of the false assumption of $\mathbf{R} = \sigma^2\mathbf{I}$, estimates of the variance parameters were biased, the bias of the estimates being larger for shorter lengths of repeated measures. However, estimates of the fixed effects were not biased unless observations were unequally spaced and followed nonlinear growth curves. Similarly, inflated Type I error rates for the fixed effects were obtained when unequally spaced observations were present in nonlinear growth curves. Since misspecification of **R** occasioned problems on the estimates of the covariance, Ferron et al. (2002) advised caution in making

conclusions about covariance parameters when there is uncertainty about the true covariance structure.

2.4. Type I error rates

Schaalje et al. (2002) studied the Type I error rates for Wald F tests in mixed linear models with varying sample size, covariance structure and degree of imbalance. Their simulations included split-plot and repeated measures designs. The KR and SW methods were used for the tests of fixed effects. The SW method was proposed by Fai and Cornelius (1996), who tested it using simulations based on split-plot models varying the degree of imbalance and the value of intra-class correlation. Kenward and Roger (1997) proposed the KR method and studied its behavior under several covariance structures. Schaalje et al. (2002) expanded the study of the SW and KR methods to investigate their performance with small sample sizes and some complicated covariance structures, which were assumed to be known. These covariance structures included: compound symmetry (CS), Toeplitz (TOEP), heterogeneous compound symmetry (CSH), first order heterogeneous autoregressive (ARH(1)), and first-order-ante-dependence (ANTE(1)).

Type I error rates for the KR method were adequate for all CS simulations, for most of the TOEP simulations, and for CSH and ARH(1) simulations with larger sample sizes. CSH and ARH(1) simulations with small sample sizes and ANTE(1) simulations with large sample sizes produced reasonable Type I error rates. The KR method performed better with balanced designs for every covariance structure except CS. The KR method did as well or better than the FC method in every situation.

However, its performance varied depending on the sample size, the complexity of the covariance structure and the degree of imbalance of the data.

Robertson (1996) simulated data from three covariance structures including compound symmetry (CS), first order autoregressive plus common covariance (AR(1)+CC) and unstructured (UN). She studied the performance of the Wald F statistic using the between/within method of SAS (Proc Mixed) for cases with known covariance structure and for cases where the AIC, BIC and other covariance structure selection criteria were used. Her results were optimistic for CS data. However, Type I error rates obtained from AR(1)+CC and UN data were inflated, even when the true covariance structure was used. This could be a consequence of the fact that the between/within ddf method was used in all simulations.

Keselman et al. (1999) studied the performance of the SW method for tests of fixed effects when the covariance structure is known and when the data are used to choose the covariance structure. Their simulations included repeated measures designs with complete measurements on each subject. Three types of covariance structures were used to generate balanced and unbalanced designs containing one between-subject and one within-subject factor. Six conditions were varied, including type of population covariance structure, equal/unequal covariance structures among groups, total sample size, equal/unequal group sizes, "positive"/"negative" pairings of covariance matrices and group sizes (see Keselman et al., 1999) and normal/non-normal data.

Keselman et al. (1999) claimed that Type I error rates were adequate using the SW method with the true covariance structure unless the true covariance was UN.

When AIC was used to select a covariance structure, the performance of the SW method varied. Type I error rates tended to be adequate when the pairing was "positive" or the total sample size was large. Surprisingly, fitting the AR(1) or RC covariance structures with between-and within-group heterogeneity resulted in well behaved Type I error rates in all cases, even when the true covariance structure was UN. Results from non-normal data in general reflected those obtained from normal data. When the total sample size was increased, Type I error rates got closer to target values.

Keselman et al. (1999) also studied the Type I error rates produced by testing fixed effects using the WJ method. This is a non-pooled multivariate test used for testing the null hypothesis: $H_0 : \mathbf{C}\boldsymbol{\beta} = \mathbf{0}$, where $\boldsymbol{\beta}$ is a vector of means. The WJ test is:

$$T_{WJ} = (\mathbf{C}\overline{\mathbf{Y}})' (\mathbf{CSC}')^{-1} (\mathbf{C}\overline{\mathbf{Y}})$$

where $\overline{\mathbf{Y}}$ is the vector of sample means and $\mathbf{S} = \text{diag}\ (\frac{S_1}{n_1},..., \frac{\mathbf{S}_J}{n_J})$, where $\mathbf{S}_j$ is the sample covariance matrix of the jth-group and J is the number of groups. T_{WJ} / c follows an approximate F distribution with degrees of freedom equal to q and q(q+2)/(3A), where $c = q + 2A - 6A/(q+2)$ and A is equation 3.1.3 of Keselman et al. (1999).

The Type I error rates obtained from the WJ method were well behaved most of the time for testing the main effect. However, when testing the interaction, the WJ method produced inflated Type I error rates for negative pairings especially when the

total sample size was smaller. Overall, Type I error rates obtained with the SW method were comparable to the ones obtained by the WJ method.

2.5. Summary

The analysis of a mixed model requires the choice of a covariance structure. The quality of the estimator and inferences derived from the analysis depend upon this choice. Some of the tools that can help to make an adequate choice are AIC, BIC, and correlograms. For parsimonious modeling BIC is advised over AIC (Little et al., 2000). For a nested model the likelihood ratio test is also an appropriate tool (Grady and Helms, 1995). Unfortunately, neither AIC nor BIC always point to the true covariance structure, AIC having the higher performance in simulation studies (Keselman et al. 1998; Ferron et al. 2002). Misspecification influences estimates of the covariance structure and standard deviations of fixed effects.

The Wald F statistic is often used to test fixed effects in mixed models. SW is a common method to approximate the ddf, and KR further adjusts the estimates of the covariance matrix and thus adjusts the test statistic. The KR method always works as well or better than the SW method if the correct covariance structure is used. Type I error rates are adequate if covariance structures are known and not too complex (Schaalje et al. 2002). The performance of the SW method when the data are used to choose the covariance structure depends on the sample size and the "pairings". The WJ method produced adequate type I error rates in general, especially for large sample sizes.

3. METHODOLOGY

3.1. Design and conditions

Simulations were used to investigate Type I error rates of Kenward-Roger tests of fixed effects in repeated measures designs. The simulated data followed a repeated measure design with one between-subject factor with three treatment levels and one within-subject continuous covariate. The true parameter values for both factors were zero. The simulation study was divided in two studies depending on equality or inequality of treatment sample sizes.

3.2. Equal treatment sizes

The conditions considered in the equal treatment size study included: 3 or 5 subjects per treatment, 3 or 5 repeated measures per subject and fifteen covariance structures. The covariance structures are denoted CS, HCS, CSH, HCSH, ARRE, HARRE, ARHRE, HARHRE, RC, HRC, TOEP, HTOEP, TOEPH, HTOEPH, and UN, where the letter "H" at the beginning of the denotations indicates heterogeneity in covariance parameters values for the between-subject factor. Details on these covariance structures when the number of repeated measures was 5 are in Table 1. The upper left 3 x 3 submatrices were used in the simulations when the number of repeated measures was 3. Since the complete covariance matrix (**V**) is a block-diagonal matrix, only one block is presented for each covariance structure. Structures with heterogeneity for the between-subject factor are not shown in Table 1; for these cases the block covariance matrices for each treatment level were obtained by

multiplying the covariance parameters of the corresponding non-heterogeneous structures by 1/3, 1 and 5/3. Table 2 gives details on the heterogeneous between-treatment structures. The combinations of number of subjects, number of repeated measures, and covariance structures yielded 60 (2x2x5) situations.

Table 1: Parameter Values for Covariance Structures Used in the Simulations

Covariance Structures	Covariance parameters values
Compound Symmetry (CS)	
$\sigma^2\begin{pmatrix} 1 & & & & \\ \rho & 1 & & & \\ \rho & \rho & 1 & & \\ \rho & \rho & \rho & 1 & \\ \rho & \rho & \rho & \rho & 1 \end{pmatrix}$	$\begin{pmatrix} 1 & & & & \\ .5 & 1 & & & \\ .5 & .5 & 1 & & \\ .5 & .5 & .5 & 1 & \\ .5 & .5 & .5 & .5 & 1 \end{pmatrix}$
Heterogeneous Compound Symmetry (CSH)	
$\begin{pmatrix} \sigma_1^2 & & & & \\ \sigma_1\sigma_2\rho & \sigma_2^2 & & & \\ \sigma_1\sigma_3\rho & \sigma_2\sigma_3\rho & \sigma_3^2 & & \\ \sigma_1\sigma_4\rho & \sigma_2\sigma_4\rho & \sigma_3\sigma_4\rho & \sigma_4^2 & \\ \sigma_1\sigma_5\rho & \sigma_2\sigma_5\rho & \sigma_3\sigma_5\rho & \sigma_4\sigma_5\rho & \sigma_5^2 \end{pmatrix}$	$\begin{pmatrix} 1 & & & & \\ .84 & 2.81 & & & \\ 1.10 & 1.84 & 4.80 & & \\ 1.26 & 2.11 & 2.76 & 6.35 & \\ 1.30 & 2.18 & 2.85 & 3.28 & 6.79 \end{pmatrix}$
Autoregressive Order 1 plus Random Effect (ARRE)	
$\sigma^2\begin{pmatrix} 1 & & & & \\ \rho & 1 & & & \\ \rho^2 & \rho & 1 & & \\ \rho^3 & \rho^2 & \rho & 1 & \\ \rho^4 & \rho^3 & \rho^2 & \rho & 1 \end{pmatrix} + \begin{pmatrix} \sigma_R^2 & & & & \\ \sigma_R^2 & \sigma_R^2 & & & \\ \sigma_R^2 & \sigma_R^2 & \sigma_R^2 & & \\ \sigma_R^2 & \sigma_R^2 & \sigma_R^2 & \sigma_R^2 & \\ \sigma_R^2 & \sigma_R^2 & \sigma_R^2 & \sigma_R^2 & \sigma_R^2 \end{pmatrix}$	$\begin{pmatrix} 1 & & & & \\ .78 & 1 & & & \\ .62 & .78 & 1 & & \\ .51 & .62 & .78 & 1 & \\ .43 & .51 & .62 & .78 & 1 \end{pmatrix}$

Heterogeneous Autoregressive Order 1 plus Random Effect (ARHRE)

$$\begin{pmatrix} \sigma_1^2 & & & & \\ \sigma_1\sigma_2\rho & \sigma_2^2 & & & \\ \sigma_1\sigma_3\rho^2 & \sigma_2\sigma_3\rho & \sigma_3^2 & & \\ \sigma_1\sigma_4\rho^3 & \sigma_2\sigma_4\rho^2 & \sigma_3\sigma_4\rho & \sigma_4^2 & \\ \sigma_1\sigma_5\rho^4 & \sigma_2\sigma_5\rho^3 & \sigma_3\sigma_5\rho^2 & \sigma_4\sigma_5\rho & \sigma_5^2 \end{pmatrix} + \begin{pmatrix} \sigma_R^2 & & & & \\ \sigma_R^2 & \sigma_R^2 & & & \\ \sigma_R^2 & \sigma_R^2 & \sigma_R^2 & & \\ \sigma_R^2 & \sigma_R^2 & \sigma_R^2 & \sigma_R^2 & \\ \sigma_R^2 & \sigma_R^2 & \sigma_R^2 & \sigma_R^2 & \sigma_R^2 \end{pmatrix}$$

$$\begin{pmatrix} 1 & & & & \\ 1.22 & 2.81 & & & \\ 1.16 & 2.64 & 4.80 & & \\ .98 & 2.19 & 3.94 & 6.35 & \\ .78 & 1.65 & 2.92 & 4.67 & 6.79 \end{pmatrix}$$

Random Coefficients (RC)

$$\begin{pmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 2 \\ 1 & 3 \\ 1 & 4 \end{pmatrix} \begin{pmatrix} \sigma_I^2 & \sigma_{IS} \\ \sigma_{IS} & \sigma_s^2 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 2 \\ 1 & 3 \\ 1 & 4 \end{pmatrix}^T + \sigma^2 \begin{pmatrix} 1 & & & & \\ 0 & 1 & & & \\ 0 & 0 & 1 & & \\ 0 & 0 & 0 & 1 & \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}$$

$$\begin{pmatrix} 2.30 & & & & \\ .27 & 2.44 & & & \\ .24 & .61 & 2.98 & & \\ .21 & .78 & 1.35 & 3.92 & \\ .18 & .95 & 1.72 & 2.49 & 5.26 \end{pmatrix}$$

Toeplitz (TOEP)

$$\sigma^2 \begin{pmatrix} 1 & & & & \\ \rho_1 & 1 & & & \\ \rho_2 & \rho_1 & 1 & & \\ \rho_3 & \rho_2 & \rho_1 & 1 & \\ \rho_4 & \rho_3 & \rho_2 & \rho_1 & 1 \end{pmatrix}$$

$$\begin{pmatrix} 1 & & & & \\ .5 & 1 & & & \\ .3 & .5 & 1 & & \\ .2 & .3 & .5 & 1 & \\ .1 & .2 & .3 & .5 & 1 \end{pmatrix}$$

Heterogeneous Toeplitz (TOEPH)

$$\begin{pmatrix} \sigma_1^2 & & & & \\ \sigma_1\sigma_2\rho_1 & \sigma_2^2 & & & \\ \sigma_1\sigma_3\rho_2 & \sigma_2\sigma_3\rho_1 & \sigma_3^2 & & \\ \sigma_1\sigma_4\rho_3 & \sigma_2\sigma_4\rho_2 & \sigma_3\sigma_4\rho_1 & \sigma_4^2 & \\ \sigma_1\sigma_5\rho_4 & \sigma_2\sigma_5\rho_3 & \sigma_3\sigma_5\rho_2 & \sigma_4\sigma_5\rho_1 & \sigma_5^2 \end{pmatrix} \qquad \begin{pmatrix} 1 & & & & \\ .84 & 2.81 & & & \\ .66 & 1.84 & 4.80 & & \\ .50 & 1.27 & 2.76 & 6.35 & \\ .26 & .87 & 1.71 & 3.28 & 6.79 \end{pmatrix}$$

Unstructured (UN)

$$\begin{pmatrix} \sigma_1^2 & & & & \\ \sigma_{12} & \sigma_2^2 & & & \\ \sigma_{13} & \sigma_{23} & \sigma_3^2 & & \\ \sigma_{14} & \sigma_{24} & \sigma_{34} & \sigma_4^2 & \\ \sigma_{15} & \sigma_{25} & \sigma_{35} & \sigma_{45} & \sigma_5^2 \end{pmatrix} \qquad \begin{pmatrix} 1 & & & & \\ .92 & 2.81 & & & \\ .95 & 1.85 & 4.80 & & \\ .65 & .95 & 2.95 & 6.35 & \\ .25 & 1.50 & 2.50 & .95 & 6.79 \end{pmatrix}$$

Table 2: Values of covariance parameters used in the simulation of data with heterogeneity between treatment levels

Covariance Structures	Parameter Values for treatment 1	Parameter Values for treatment 2	Parameter Values for treatment 3
HCS	σ^2=.33, ρ=.5	σ^2=1, ρ=.5	σ^2=1.67, ρ=.5
HCSH	σ_1^2=.33, σ_2^2=.94, σ_3^2=1.6, σ_4^2=2.12, σ_5^2=2.26, ρ=.5	σ_1^2=1, σ_2^2=2.81, σ_3^2=4.80, σ_4^2=6.35, σ_5^2=6.79, ρ=.5	σ_1^2=1.67, σ_2^2=4.68, σ_3^2=8, σ_4^2=10.58, σ_5^2=11.32, ρ=.5
HARRE	σ^2=.25, ρ=.7, σ_R^2=.25	σ^2=.75, ρ=.7, σ_R^2=.25	σ^2=1.25, ρ=.7, σ_R^2=.25
HARHRE	σ_1^2=.25, σ_2^2=.85, σ_3^2=1.52, σ_4^2=2.03, σ_5^2=2.18, ρ=.7, σ_R^2=.25	σ_1^2=.75, σ_2^2=2.56, σ_3^2=4.55, σ_4^2=6.1, σ_5^2=6.54, ρ=.7, σ_R^2=.25	σ_1^2=1.25, σ_2^2=4.27, σ_3^2=7.58, σ_4^2=10.17, σ_5^2=10.9, ρ=.7, σ_R^2=.25
HRC	σ_I^2=.1, σ_S^2=.067, σ_{IS}=-.01, σ^2=2	σ_I^2=.3, σ_S^2=.2, σ_{IS}=-03, σ^2=2	σ_I^2=.5, σ_S^2=.33, σ_{IS}=-.05, σ^2=2
HTOEP	σ^2=.33, ρ_1=.5, ρ_2=.3, ρ_3=.2, ρ_4=.1	σ^2=1, ρ_1=.5, ρ_2=.3, ρ_3=.2, ρ_4=.1	σ^2=1.67, ρ_1=.5, ρ_2=.3, ρ_3=.2, ρ_4=.1
HTOEPH	σ_1^2=.33, σ_2^2=.94, σ_3^2=1.6, σ_4^2=2.12, σ_5^2=2.26, ρ_1=.5, ρ_2=.3, ρ_3=.2, ρ_4=.1	σ_1^2=1, σ_2^2=2.81, σ_3^2=4.80, σ_4^2=6.35, σ_5^2=6.79, ρ_1=.5, ρ_2=.3, ρ_3=.2, ρ_4=.1	σ_1^2=1.67, σ_2^2=4.68, σ_3^2=8, σ_4^2=10.58, σ_5^2=11.32, ρ_1=.5, ρ_2=.3, ρ_3=.2, ρ_4=.1

3.3. Unequal treatment size

Numbers of subjects per treatment were 3, 5 and 7 for the three treatment levels. As before, there were 3 or 5 repeated measures per subject and the same covariance structures as in the equal treatment size study were used. For those seven covariance structures with heterogeneity for the between-subject factor (Table 2), positive and negative pairings were considered. Positive pairing occurred when the treatment level with more subjects followed the covariance structure with larger variance parameters. Negative pairing occurred in the opposite situation. The number

of situations resulting from the varying the number of repeated measures, the covariance structures and the pairing was 44 (2x8 from the homogenous case plus 2x7x2 from the heterogeneous case).

3.4. Simulation Study

Ten thousand simulated data sets were generated for every situation. Data for the fifteen described covariance structures were generated following the two-step method of Ripley (1987). In the first step a random multivariate normal vector was generated with E(**y**)=**0** and Var(**y**)=**I,** where **I** is the identity matrix**.** In the second step this vector was multiplied by the Cholesky descomposition of the covariance matrix corresponding to the covariance structure in question. The resulting random vector had a mean vector of zero, and covariance matrix **V** whose blocks were the covariance matrices in Tables 1 and 2.

Data were simulated using PROC IML of SAS, and every simulated data set was analyzed with PROC MIXED. Some of the data sets were simulated using a Penguin Computing Dual Opteron Altus 1000E Linux machine using SAS v. 9. Other data sets were simulated using a Dell Power Edge 350 with a Pentium processor, using SAS v. 8.2. All the analyses were done on a Penguin Computing Dual Opteron Altus 1000E Linux machine using SAS v. 9.

The model used in the analysis of each data set was additive; that is, there was not an interaction term. The MODEL statement was y=treatment time, where the treatment effect was a categorical variable and the time effect was continuous. Fifteen covariance structures were fitted to each data set by appropriate specification in the

RANDOM and REPEATED statements. Table 3 specifies the appropriate commands for the covariance structures with homogeneity and heterogeneity between treatments.

Table 3: Commands used to model the data with the different covariance structures

Covariance Structures	**Commands for between-treatment homogeneity**	**Commands for between-treatment heterogeneity**
CS/ HCS	Repeated / type=CS subject=subject	Repeated / type=CS group=treatment subject=subject
CSH/ HCSH	Repeated / type=CSH subject=subject	Repeated / type=CSH group=treatment
ARRE/ HARRE	Random subject Repeated /type=AR(1) subject=subject	Random subject Repeated /type=AR(1) group=treatment subject=subject
ARHRE/ HARHRE	Random subject Repeated / type=ARH(1) subject=subject	Random subject Repeated / type=ARH(1) group=treatment subject=subject
RC/ HRC	Random intercept time/ type=UN subject=subject	Random intercept time/ type=UN group=treatment subject=subject
TOEP/ HTOEP	Repeated/ type=TOEP subject=subject	Repeated/ type=TOEP group=treatment subject=subject
TOEPH/ HTOEPH	Repeated/type=TOEPH subject=subject	Repeated/type=TOEPH group=treatment subject=subject
UN	Repeated/type=UN subject=subject	

The DDFM option in the MODEL statement was specified as KENWARDROGER. This option calculates the p-values for tests of fixed and random effects using the Kenward-Roger adjustments to the denominator degrees of freedom, the estimates of the covariance matrix, and F statistic.

The AIC and BIC criteria were used to choose a covariance structure for the data. The number of times these criteria selected the true covariance structure was recorded in each situation. Rates of success of these criteria were compared. P-values corresponding to the tests of the true null hypotheses regarding the effects of treatment and time were recorded separately for the best AIC, best BIC, and correct models. The proportion of times the p-values were less than or equal to $\alpha = 0.05$ and $\alpha = 0.01$ was recorded. The standard deviation for the estimated Type I error rates for n observations is equal to $\sqrt{\frac{p(1-p)}{n}}$. For 10,000 observations, the standard deviations would be approximately 0.002 and 0.001 for 0.05 and 0.01 respectively (Schaalje et al. 2002). Therefore, if the performance of the Kenward-Roger method were perfect, with 95% confidence, the expected proportions would be in the interval [0.046-0.054] for $\alpha = 0.05$ and [0.008-0.012] for $\alpha = 0.01$. A chi-square goodness-of-fit test was performed in each situation to verify if the p-values followed the Uniform (0,1).

4. RESULTS

4.1. Distributions of p-values from AIC, BIC and correct models

4.1.1. Equal treatment sizes

4.1.1.a. AIC and BIC best models

Distributions of p-values from fitting the best AIC and BIC models did not follow the uniform distribution for any covariance structure or sample size (Tables 4 to 7, Figures 1 to 8 in Appendix 1). The highest p-value for a goodness-of-fit test to the uniform (0,1) was .0131. This came from the 5x5 CS case for the within-subject effect (time) and the BIC selection criterion. Generally, distributions of p-values followed right skewed distributions, which will result in anti-conservative test statistics of the effects and thus increased Type I error rates.

Proportions of p-values less than or equal to α=0.05 or α =0.01 were always higher than the target values (Tables 4 to 7) and were never included in the expected 95% confidence intervals. However, there were some cases in which the empirical error rates were robust according to Bradley's criterion of robustness (Bradley, 1978); that is between .5α and 1.5α. For α =0.05 and 0.01, these intervals were [.025-.075] and [.005-.015], respectively. With one exception, all the situations in which observed error rates were robust came from the best BIC procedure and the 5x3 or 5x5 sample sizes (Tables 4 to 7). For the CSH 5x3 case and the time effect, the empirical Type I error rate was .0749 with α=0.05.

Table 4: Simulated Type I error rates (α=0.05 and 0.01), goodness of fit tests comparing the Uniform (0,1) to p-value distributions, and convergence rates for the 3x3 sample size studies

Covariance Structures	Effect	Best AIC model			Best BIC model			Correct model			Convergence rates for correct model
		Prop ≤ α=0.05	Prop ≤ α=0.01	χ^2 p-value	Prop ≤ α=0.05	Prop ≤ α=0.01	χ^2 p-value	Prop ≤ α=0.05	Prop ≤ α=0.01	χ^2 p-value	
CS	Treatment	.1491	.0589	<.0001	.1409	.0574	<.0001	.0525*	.0118*	.2941	100
	Time	.1311	.0564	<.0001	.1179	.0511	<.0001	.0493*	.0104*	.1572	
CSH	Treatment	.1556	.0623	<.0001	.1523	.0603	<.0001	.0728^	.0187	<.0001	100
	Time	.1437	.0681	<.0001	.132	.0632	<.0001	.0477*	.0088*	.5118	
ARRE	Treatment	.156	.0592	<.0001	.1509	.0562	<.0001	.1165	.0613	<.0001	100
	Time	.147	.0686	<.0001	.1363	.064	<.0001	.0369^	.0069^	<.0001	
ARHRE	Treatment	.1634	.0599	<.0001	.1618	.0598	<.0001	.0632^	.0168	.00013	92.74
	Time	.1653	.0785	<.0001	.154	.0716	<.0001	.0961	.047	<.0001	
RC	Treatment	.1385	.054	<.0001	.1336	.0526	<.0001	.0538*	.0117*	.1795	100
	Time	.1354	.0605	<.0001	.1233	.055	<.0001	.0363^	.0067^	.00087	
TOEP	Treatment	.1468	.058	<.0001	.1426	.0564	<.0001	.0537*	.0119*	.1305	100
	Time	.1368	.0586	<.0001	.1272	.0541	<.0001	.0513*	.0116*	.8607	
TOEPH	Treatment	.1508	.0567	<.0001	.1472	.0555	<.0001	.0867	.0239	<.0001	100
	Time	.1437	.0643	<.0001	.1319	.0583	<.0001	.0491*	.0088*	.8495	
UN	Treatment	.1559	.0554	<.0001	.1541	.0553	<.0001	.1143	.0374	<.0001	100
	Time	.1441	.0669	<.0001	.1344	.061	<.0001	.0568^	.0119*	.0983	
HCS	Treatment	.165	.0693	<.0001	.1615	.0689	<.0001	.0249^	.0028	<.0001	99.99
	Time	.1336	.0531	<.0001	.1254	.0488	<.0001	.1305	.0171	<.0001	
HCSH	Treatment	.1842	.0738	<.0001	.1811	.0731	<.0001	.0909	.0228	<.0001	77.63
	Time	.1537	.066	<.0001	.1434	.0621	<.0001	.1305	.052	<.0001	

Note1: * Proportions included in expected 95% intervals.
Note2: ^ Robust proportions ([.025-0.075] for α=.05 or [.005-.015] for α=.01) not included in the expected 95% confidence intervals.

Table 4 (Cont.): Simulated Type I error rates (α=0.05 and 0.01), goodness of fit tests comparing the Uniform (0,1) to p-value distributions, and convergence rates for the 3x3 sample size studies

Covariance Structures	Effect	Best AIC model			Best BIC model			Correct model			Convergence rates for correct model
		Prop ≤ α=0.05	Prop ≤ α=0.01	χ^2 p-value	Prop ≤ α=0.05	Prop ≤ α=0.01	χ^2 p-value	Prop ≤ α=0.05	Prop ≤ α=0.01	χ^2 p-value	
HARRE	Treatment	.1651	.0623	<.0001	.1629	.0628	<.0001	.0388^	.0101*	<.0001	100
	Time	.1576	.0697	<.0001	.1499	.0657	<.0001	.0615^	.0128^	.0053	
HARHRE	Treatment	.1773	.0722	<.0001	.178	.0725	<.0001	.1198	.0479	<.0001	65.41
	Time	.1637	.0791	<.0001	.1561	.0766	<.0001	.2218	.1084	<.0001	
HRC	Treatment	.1394	.0523	<.0001	.1372	.0512	<.0001	.0337^	.0068^	<.0001	90.68
	Time	.1268	.0546	<.0001	.1177	.0503	<.0001	.0393^	.0095*	<.0001	
HTOEP	Treatment	.1653	.0684	<.0001	.1655	.06	<.0001	.0401^	.0993	<.0001	99.61
	Time	.1359	.0593	<.0001	.131	.0564	<.0001	.101	.025	<.0001	
HTOEPH	Treatment	.1726	.0718	<.0001	.1738	.0725	<.0001	.0949	.02788	<.0001	22.95
	Time	.1479	.0654	<.0001	.14	.0618	<.0001	.1895	.0636	<.0001	

Note1: * Proportions included in expected 95% intervals.
Note2: ^ Robust proportions ([.025-0.075] for α=.05 or [.005-.015] for α=.01) not included in the expected 95% confidence intervals.

Table 5: Simulated Type I error rates (α=0.05 and 0.01), goodness of fit tests comparing the Uniform (0,1) to p-value distributions, and convergence rates for the 5x3 sample size studies

Covariance Structures	Effect	Best AIC model			Best BIC model			Correct model			Convergence rates for correct model
		Prop ≤ α=0.05	Prop ≤ α=0.01	χ^2 p-value	Prop ≤ α=0.05	Prop ≤ α=0.01	χ^2 p-value	Prop ≤ α=0.05	Prop ≤ α=0.01	χ^2 p-value	
CS	Treatment	.0961	.0275	<.0001	.0772	.0211	<.0001	.0479*	.0088*	.1854	100
	Time	.0763	.0227	<.0001	.0605^	.0177	<.0001	.0497*	.0104*	.8304	
CSH	Treatment	.0982	.0306	<.0001	.0863	.0249	<.0001	.0594^	.0128^	.0183	100
	Time	.0749^	.0248	<.0001	.0626^	.0174	<.0001	.0501*	.0107*	.6288	
ARRE	Treatment	.1048	.038	<.0001	.0855	.0293	<.0001	.0989	.0437	<.0001	100
	Time	.085	.0301	<.0001	.0772	.0242	<.0001	.0354^	.0056^	<.0001	
ARHRE	Treatment	.1035	.0305	<.0001	.0939	.0248	<.0001	.0589^	.0123*	.0054	97.49
	Time	.0935	.0354	<.0001	.0796	.0251	<.0001	.083	.0379	<.0001	
RC	Treatment	.0916	.0283	<.0001	.0745^	.0205	<.0001	.0495*	.0096*	.6332	100
	Time	.0801	.0251	<.0001	.0675^	.0183	<.0001	.0431^	.009*	.0631	
TOEP	Treatment	.0933	.0295	<.0001	.075^	.0225	<.0001	.046*	.009*	.5392	100
	Time	.0866	.0267	<.0001	.0764	.0202	<.0001	.0513*	.0092*	.27034	
TOEPH	Treatment	.1031	.0312	<.0001	.0891	.026	<.0001	.0629^	.0147^	.0009	100
	Time	.0823	.028	<.0001	.0691^	.0211	<.0001	.0482*	.0103*	.3397	
UN	Treatment	.0948	.0283	<.0001	.0826	.0229	<.0001	.0641^	.0162	<.0001	100
	Time	.0772	.0243	<.0001	.0655^	.0184	<.0001	.0499*	.0104*	.4552	
HCS	Treatment	.1087	.0359	<.0001	.0933	.0305	<.0001	.0428^	.0086*	.2580	100
	Time	.0873	.0278	<.0001	.0766	.0227	<.0001	.0572^	.0128^	.0525	
HCSH	Treatment	.1207	.0416	<.0001	.1151	.0364	<.0001	.0773	.0205	<.0001	100
	Time	.093	.0305	<.0001	.0806	.0236	<.0001	.0718^	.0181	<.0001	

Note1: * Proportions included in expected 95% intervals.
Note2: ^ Robust proportions ([.025-0.075] for α=.05 or [.005-.015] for α=.01) not included in the expected 95% confidence intervals.

Table 5 (Cont.): Simulated Type I error rates (α=0.05 and 0.01), goodness of fit tests comparing the Uniform (0,1) to p-value distributions, and convergence rates for the 5x3 sample size studies

Covariance Structures	Effect	Best AIC model			Best BIC model			Correct model			Convergence rates for correct model
		Prop ≤ α=0.05	Prop ≤ α=0.01	χ^2 p-value	Prop ≤ α=0.05	Prop ≤ α=0.01	χ^2 p-value	Prop ≤ α=0.05	Prop ≤ α=0.01	χ^2 p-value	
HARRE	Treatment	.1071	.0366	<.0001	.09	.0302	<.0001	.0612^	.0169	<.0001	100
	Time	.0945	.0294	<.0001	.0965	.0274	<.0001	.0466*	.0098*	<.0001	
HARHRE	Treatment	.1261	.0418	<.0001	.1204	.0396	<.0001	.0822	.0287	<.0001	76.25
	Time	.1031	.0375	<.0001	.0912	.0329	<.0001	.1141	.0497	<.0001	
HRC	Treatment	.0966	.0319	<.0001	.079	.0251	<.0001	.0376^	.0068^	<.0001	99.91
	Time	.0782	.0242	<.0001	.0664^	.0189	<.0001	.0371^	.0077^	<.0001	
HTOEP	Treatment	.1133	.0399	<.0001	.0982	.0321	<.0001	.0562^	.0112*	.2505	99.99
	Time	.0919	.0299	<.0001	.0872	.0277	<.0001	.071^	.0195	<.0001	
HTOEPH	Treatment	.1229	.0394	<.0001	.118	.0406	<.0001	.1014	.0265	<.0001	99.75
	Time	.0988	.0306	<.0001	.0894	.0278	<.0001	.0813	.0204	<.0001	

Note1: * Proportions included in expected 95% intervals.
Note2: ^ Robust proportions ([.025-0.075] for α=.05 or [.005-.015] for α=.01) not included in the expected 95% confidence intervals.

Table 6: Simulated Type I error rates (α=0.05 and 0.01), goodness of fit tests comparing the Uniform (0,1) to p-value distributions, and convergence rates for the 3x5 sample size studies

Covariance Structures	Effect	Best AIC model			Best BIC model			Correct model			Convergence rates for correct model
		Prop ≤ α=0.05	Prop ≤ α=0.01	χ^2 p-value	Prop ≤ α=0.05	Prop ≤ α=0.01	χ^2 p-value	Prop ≤ α=0.05	Prop ≤ α=0.01	χ^2 p-value	
CS	Treatment	.2024	.1153	<.0001	.176	.0991	<.0001	.0498*	.0092*	.5415	100
	Time	.1164	.053	<.0001	.103	.0452	<.0001	.0516*	.011*	.5129	
CSH	Treatment	.2526	.1569	<.0001	.2302	.1416	<.0001	.0747^	.0198	<.0001	99.99
	Time	.1313	.0602	<.0001	.1163	.0501	<.0001	.0636^	.016	.0072	
ARRE	Treatment	.2213	.1331	<.0001	.1964	.1156	<.0001	.0653^	.0197	<.0001	100
	Time	.1342	.066	<.0001	.1194	.0575	<.0001	.0447^	.0103*	.0012	
ARHRE	Treatment	.2695	.1577	<.0001	.2458	.1428	<.0001	.0644^	.0146^	.0084	91.19
	Time	.1521	.073	<.0001	.1396	.0652	<.0001	.0928	.0342	<.0001	
RC	Treatment	.2012	.124	<.0001	.175	.1051	<.0001	.0499*	.0116*	.0142	100
	Time	.1343	.0611	<.0001	.1222	.055	<.0001	.0463*	.012*	.041	
TOEP	Treatment	.2032	.1235	<.0001	.1775	.1055	<.0001	.0522*	.012*	.3575	100
	Time	.1246	.0549	<.0001	.1131	.0488	<.0001	.0466*	.0089*	.0304	
TOEPH	Treatment	.2485	.1435	<.0001	.2261	.1287	<.0001	.1153	.0370	<.0001	99.93
	Time	.141	.0662	<.0001	.1275	.0578	<.0001	.0661^	.0174	<.0001	
UN	Treatment	.29	.1736	<.0001	.2722	.1617	<.0001	.3383	.1997	<.0001	99.71
	Time	.141	.0639	<.0001	.1289	.0579	<.0001	.1136	.0422	<.0001	
HCS	Treatment	.2261	.1415	<.0001	.2004	.125	<.0001	.0251^	.0024	<.0001	100
	Time	.1284	.0598	<.0001	.1155	.0519	<.0001	.0561^	.0121*	.7905	
HCSH	Treatment	.2952	.1798	<.0001	.2748	.1669	<.0001	.1107	.0317	<.0001	74.44
	Time	.1528	.0652	<.0001	.1419	.0583	<.0001	.1871	.0674	<.0001	

Note1: * Proportions included in expected 95% intervals.
Note2: ^ Robust proportions ([.025-0.075] for α=.05 or [.005-.015] for α=.01) not included in the expected 95% confidence intervals.

Table 6 (Cont.): Simulated Type I error rates (α=0.05 and 0.01), goodness of fit tests comparing the Uniform (0,1) to p-value distributions, and convergence rates for the 3x5 sample size studies

Covariance Structures	Effect	Best AIC model			Best BIC model			Correct model			Convergence rates for correct model
		Prop ≤ α=0.05	Prop ≤ α=0.01	χ^2 p-value	Prop ≤ α=0.05	Prop ≤ α=0.01	χ^2 p-value	Prop ≤ α=0.05	Prop ≤ α=0.01	χ^2 p-value	
HARRE	Treatment	.2273	.1409	<.0001	.2038	.1251	<.0001	.039^	.0056^	.0027	100
	Time	.1567	.0754	<.0001	.1447	.0702	<.0001	.0656^	.0197	<.0001	
HARHRE	Treatment	.29	.1723	<.0001	.2735	.1607	<.0001	.1356	.0481	<.0001	57.43
	Time	.1693	.081	<.0001	.1586	.0753	<.0001	.200	.0898	<.0001	
HRC	Treatment	.2012	.1198	<.0001	.1796	.104	<.0001	.0305^	.0088*	<.0001	91.22
	Time	.1232	.0602	<.0001	.1132	.0515	<.0001	.0499*	.0145^	<.0001	
HTOEP	Treatment	.2184	.1337	<.0001	.1974	.1177	<.0001	.0061	.00047	<.0001	42.78
	Time	.1483	.068	<.0001	.138	.0632	<.0001	.2199	.1026	<.0001	
HTOEPH	Treatment	.2911	.1751	<.0001	.268	.1593	<.0001	0	0	.	.01
	Time	.1559	.0712	<.0001	.1448	.0646	<.0001	0	0	.	

Note1: * Proportions included in expected 95% intervals.
Note2: ^ Robust proportions ([.025-0.075] for α=.05 or [.005-.015] for α=.01) not included in the expected 95% confidence intervals.

Table 7: Simulated Type I error rates (α=0.05 and 0.01), goodness of fit tests comparing the Uniform (0,1) to p-value distributions, and convergence rates for the 5x5 sample size studies

Covariance Structures	Effect	AIC model			BIC model			Correct model			Convergence rates for correct model
		Prop ≤ α=0.05	Prop ≤ α=0.01	χ^2 p-value	Prop ≤ α=0.05	Prop ≤ α=0.01	χ^2 p-value	Prop ≤ α=0.05	Prop ≤ α=0.01	χ^2 p-value	
CS	Treatment	.1027	.0346	<.0001	.0726^	.0199	<.0001	.0504*	.0089*	.1707	100
	Time	.0828	.028	<.0001	.0598^	.0168	.0131	.0526*	.0116*	.6907	
CSH	Treatment	.1202	.0437	<.0001	.0877	.0269	<.0001	.0564^	.0121*	<.0001	100
	Time	.0916	.0351	<.0001	.0614^	.0185	<.0001	.0507*	.0166	.6075	
ARRE	Treatment	.0961	.0353	<.0001	.0727^	.0226	<.0001	.0564^	.0133^	.0276	100
	Time	.0785	.0275	<.0001	.0548^	.0175	.0001	.0408^	.0084*	.0045	
ARHRE	Treatment	.1098	.0381	<.0001	.0861	.027	<.0001	.0513*	.0118*	.3743	95.58
	Time	.1028	.0385	<.0001	.081	.0277	<.0001	.0722^	.0228	<.0001	
RC	Treatment	.0942	.0309	<.0001	.0698^	.0188	<.0001	.0499*	.0103*	.708	100
	Time	.089	.0308	<.0001	.0751^	.0238	<.0001	.0461*	.0092*	.3493	
TOEP	Treatment	.0953	.0313	<.0001	.0678^	.0167	<.0001	.0475*	.0088*	.6891	100
	Time	.0939	.0358	<.0001	.0808	.0271	<.0001	.0519*	.0118*	.3504	
TOEPH	Treatment	.119	.0388	<.0001	.0859	.0258	<.0001	.0646^	.0139^	<.0001	100
	Time	.1016	.0389	<.0001	.0841	.0261	<.0001	.0565^	.0115*	.2255	
UN	Treatment	.1365	.0516	<.0001	.1109	.0383	<.0001	.1178	.0412	<.0001	100
	Time	.0968	.0318	<.0001	.0773	.0223	<.0001	.0665^	.0153^	<.0001	
HCS	Treatment	.0928	.0284	<.0001	.0733^	.0202	<.0001	.0437^	.0074^	.1815	100
	Time	.0779	.0241	<.0001	.0627^	.0168	<.0001	.0486*	.0096*	.004	
HCSH	Treatment	.1433	.052	<.0001	.1265	.0438	<.0001	.0979	.0263	<.0001	99.16
	Time	.1083	.0372	<.0001	.0828	.0268	<.0001	.1059	.0357	<.0001	

Note1: * Proportions included in expected 95% intervals.
Note2: ^ Robust proportions ([.025-0.075] for α=.05 or [.005-.015] for α=.01) not included in the expected 95% confidence intervals.

Table 7 (Cont.): Simulated Type I error rates (α=0.05 and 0.01), goodness of fit tests comparing the Uniform (0,1) to p-value distributions, and convergence rates for the 5x5 sample size studies

Covariance Structures	Effect	AIC model			BIC model			Correct model			Convergence rates for correct model
		Prop ≤ α=0.05	Prop ≤ α=0.01	χ^2 p-value	Prop ≤ α=0.05	Prop ≤ α=0.01	χ^2 p-value	Prop ≤ α=0.05	Prop ≤ α=0.01	χ^2 p-value	
HARRE	Treatment	.0934	.0319	<.0001	.0738^	.0218	<.0001	.0545^	.0105*	.37679	100
	Time	.0825	.0298	<.0001	.0701^	.0252	<.0001	.0494*	.0143^	.0002	
HARHRE	Treatment	.1314	.0465	<.0001	.1116	.0354	<.0001	.0916	.0337	<.0001	71.60
	Time	.1164	.0463	<.0001	.0896	.0312	<.0001	.1316	.0513	<.0001	
HRC	Treatment	.0908	.0291	<.0001	.0743^	.021	<.0001	.0420^	.0083*	<.0001	99.94
	Time	.0864	.0301	<.0001	.076	.0241	<.0001	.0523*	.0136^	<.0001	
HTOEP	Treatment	.0884	.0311	<.0001	.0713^	.0231	<.0001	.0576^	.0154^	.003	99.98
	Time	.0959	.0322	<.0001	.0914	.0276	<.0001	.0910	.0294	<.0001	
HTOEPH	Treatment	.1399	.0545	<.0001	.1149	.0423	<.0001	0	0	.	0
	Time	.1199	.0422	<.0001	.0978	.031	<.0001	0	0	.	

Note1: * Proportions included in expected 95% intervals.
Note2: ^ Robust proportions ([.025-0.075] for α=.05 or [.005-.015] for α=.01) not included in the expected 95% confidence intervals.

In general, sample sizes had an influence on the proportions obtained, the 3x5 (with proportions of up to .2952 for α=0.05) being the farthest proportions from expected values, followed by 3x3 (with proportions of up to.1842 for α=0.05), 5x5 (with proportions of up to .1433 for α=0.05) and 5x3 (with proportions of up to.1261 for α=0.05). When the within-subject effect was tested, proportions were usually closer to expected values than when the between-subject effect (treatment) was tested.

In addition, proportions from the best BIC models were usually closer to target values that the ones from the best AIC models. The covariance structure followed by the data also played an important role. Typically, the simpler the covariance structure, the closer the proportions were to expected values.

4.1.1.b. Correct Model

Convergence was achieved in most cases; however, the more complicated the covariance structure was, the lower the convergence rate (Tables 4 to 7). The most extreme cases of this fact happened for HTOEPH data and sizes 3x5 and 5x5 where the convergence rates were 0.01% and 0% respectively.

Distributions of p-values from fitting the correct model often followed the uniform distribution, depending on the sample size and covariance structure (Tables 4 to 7). P-values from data following the CS, TOEP and RC covariance structures had a uniform distribution for all the situations considered, except for the within-subject effect when the covariance structure was RC and the sample size was 3x3. However, this was a conservative case since the proportions less than α=0.05 and α =0.01 were .0363 and .0067 respectively. We would say that empirical Type I error rates are

conservative when they are less than the lower bound of the expected 95% confidence interval for α=0.05 or α =0.01. For 10,000 observations the lower bounds would be .046 and .008 for α=0.05 and 0.01 respectively.

Other instances of lack-of-fit to the uniform distribution were also of conservative nature (Table 8, Figures 1 to 8 in Appendix 1). This is important to mention because these distributions did not follow the uniform distribution, but they would cause less concern than the anti-conservative case because the Type I error obtained would be less than the expected. Many of the conservative p-values came from covariance structures with heterogeneity between treatment levels.

P-values obtained by fitting CSH, ARRE, ARHRE, TOEPH, UN, HCS, HARRE and HTOEP covariance structures followed the uniform distribution in some cases, depending on the sample size. The 5x3 and 5x5 cases were the most favorable. For the HCSH, HARHRE, HRC and HTOEPH covariance structures, the p-values did not follow the uniform distribution for any situation. These are complicated covariance structures with heterogeneity within- and between-subjects and did not produce conservative test statistics, except for the HRC covariance structure (Table 8).

There were some cases where the p-values did not follow the uniform distribution but the proportions less than or equal to α=0.05 or α =0.01 were included in the 95% confidence expected intervals or were considered robust according to Bradley's criterion (Bradley, 1978). Some of these cases include 5x3 ARHRE for treatment, 5x5 CSH for treatment, 5x5 HRC for time and 5x5 HTOEP for treatment (Table 4 to 7).

Proportions of p-values less than or equal to α=0.05 or α =0.01 for correct models were usually closer to target than those obtained by fitting the best AIC and BIC models. The exceptions to this pattern include the HARHRE and HTOEPH covariance structures for the within-subject effect (Tables 4 to 7).

Table 8: Situations producing conservative error rates (less than the lower bound of the expected 95% confidence interval for α=0.05 or α=0.01 for the correct model in the equal treatment size simulation studies

Size	Covariance structure	Effect	Prop ≤ 0.05	Prop ≤ 0.01
3x3	ARRE	Time	.0369	.0069
3x3	RC	Time	.0363	.0067
3x3	HCS	Treatment	.0249	.0028
3x3	HARRE	Treatment	.0388	.0101
3x3	HRC	Treatment	.0337	.0068
3x3	HRC	Time	.0393	.0095
3x3	HTOEP	Treatment	.0401	.0993
5x3	ARRE	Time	.0354	.0056
5x3	RC	Time	.0431	.009
5x3	HRC	Treatment	.0376	.0068
5x3	HRC	Time	.0371	.0077
3x5	ARRE	Time	.0447	.0103
3x5	HCS	Treatment	.0251	.0024
3x5	HARRE	Treatment	.039	.0056
3x5	HRC	Treatment	.0305	.0088
3x5	HTOEP	Treatment	.0061	.00047
5x5	ARRE	Time	.0408	.0084
5x5	HCS	Treatment	.0437	.0074
5x5	HRC	Treatment	.0420	.0083

4.1.2. Unequal treatment size

4.1.2.a. AIC and BIC best models

P-values from fitting the best AIC and BIC models in the unequal treatment size case were similar to those from the equal treatment size case. Distributions of these p-values did not follow the uniform distribution in any case (Tables 9 and 10, Figures 9 to 12 in Appendix 1). However, there were several cases in which the proportions of p-values less than or equal to α=0.05 or α =0.01 were considered robust according to Bradley's criterion (Bradley, 1978).

Most of the observed proportions less than or equal to α=0.05 or α =0.01 that were robust came from the (3, 5, 7)x5 sample size and best BIC models. Best BIC models usually presented closer proportions to expected values than those obtained from best AIC models.

Similarly, robust proportions occurred more often with 5 rather than 3 repeated measures. When the within-subject effect was tested, empirical Type I error rates were usually closer to expected values. However, for positive pairings, empirical Type I error rates for the between-subject effect were closer to expected values. Proportions obtained by fitting data with negative pairings were considerably more anti-conservative than those obtained from positive pairings for the between-subject effect. However, for the within-subject effect, the opposite situation obtained. Proportions for the positive pairings were slightly more anti-conservative than for the negative pairings.

Table 9: Simulated Type I error rates (α=0.05 and 0.01), goodness of fit tests comparing the Uniform (0,1) to p-value distributions, and convergence rates for the (3, 5, 7)x3 sample size studies

Covariance Structures	Effect	Best AIC model			Best BIC model			Correct model			Convergence rates for correct model
		Prop ≤ α=0.05	Prop ≤ α=0.01	χ^2 p-value	Prop ≤ α=0.05	Prop ≤ α=0.01	χ^2 p-value	Prop ≤ α=0.05	Prop ≤ α=0.01	χ^2 p-value	
CS	Treatment	.1193	.0477	<.0001	.0879	.0306	<.0001	.0514*	.0098*	.2446	100
	Time	.0861	.0298	<.0001	.067^	.0215	<.0001	.0497*	.0104*	.8304	
CSH	Treatment	.1216	.0444	<.0001	.0978	.0328	<.0001	.0591^	.0144^	.0257	99.99
	Time	.0858	.0308	<.0001	.069^	.0212	<.0001	.0498*	.0105*	.3506	
ARRE	Treatment	.1281	.0532	<.0001	.0937	.0354	<.0001	.0972	.0418	<.0001	100
	Time	.0946	.0345	<.0001	.0825	.0271	<.0001	.0352^	.0061^	<.0001	
ARHRE	Treatment	.1241	.0451	<.0001	.1062	.0348	<.0001	.0570^	.0127^	.4211	97.36
	Time	.1023	.0394	<.0001	.0836	.0284	<.0001	.0855	.0358	<.0001	
RC	Treatment	.1123	.0443	<.0001	.0848	.0299	<.0001	.0509*	.0098*	.5352	100
	Time	.0868	.0299	<.0001	.0721^	.021	<.0001	.0431^	.0083*	.17604	
TOEP	Treatment	.1168	.0455	<.0001	.0854	.0291	<.0001	.0502*	.0088*	.40832	100
	Time	.0974	.0355	<.0001	.0837	.0261	<.0001	.0511*	.0094*	.4788	
TOEPH	Treatment	.1184	.0472	<.0001	.098	.0354	<.0001	.062^	.016	.0017	100
	Time	.0932	.038	<.0001	.0755	.0281	<.0001	.0486*	.0099*	.4715	
UN	Treatment	.1171	.0422	<.0001	.0946	.0306	<.0001	.0657^	.0151^	<.0001	100
	Time	.0959	.0333	<.0001	.0722^	.0229	<.0001	.0499*	.0104*	.4418	
HCS positive	Treatment	.1004	.0356	<.0001	.0766	.0269	<.0001	.0366^	.0045^	.0003	99.99
	Time	.1111	.042	<.0001	.0935	.0325	<.0001	.062^	.0148^	.0004	
HCSH positive	Treatment	.0976	.0355	<.0001	.0781	.0273	<.0001	.0733^	.0181	<.0001	92.45
	Time	.1201	.047	<.0001	.099	.0379	<.0001	.0876	.0288	<.0001	

Note1: * Proportions included in expected 95% intervals.
Note2: ^ Robust proportions ([.025-0.075] for α=.05 or [.005-.015] for α=.01) not included in the expected 95% confidence intervals.

Table 9 (Cont.): Simulated Type I error rates (α=0.05 and 0.01), goodness of fit tests comparing the Uniform (0,1) to p-value distributions, and convergence rates for the (3, 5, 7)x3 sample size studies

Covariance Structures	Effect	Best AIC model			Best BIC model			Correct model			Convergence rates for correct model
		Prop ≤ α=0.05	Prop ≤ α=0.01	χ^2 p-value	Prop ≤ α=0.05	Prop ≤ α=0.01	χ^2 p-value	Prop ≤ α=0.05	Prop ≤ α=0.01	χ^2 p-value	
HARRE positive	Treatment	.1062	.0417	<.0001	.0861	.0329	<.0001	.0599^	.0166	<.0001	100
	Time	.1261	.0474	<.0001	.1098	.0387	<.0001	.0535*	.009*	.1156	
HARHRE positive	Treatment	.1118	.0425	<.0001	.0968	.0362	<.0001	.1020	.0491	<.0001	72.17
	Time	.1327	.0562	<.0001	.1087	.0448	<.0001	.1538	.074	<.0001	
HRC positive	Treatment	.1039	.0386	<.0001	.0767	.0243	<.0001	.0542*	.0131^	.0004	96.96
	Time	.0913	.0326	<.0001	.0752^	.0242	<.0001	.0369^	.0085*	<.0001	
HTOEP positive	Treatment	.096	.0367	<.0001	.0735^	.0263	<.0001	.0485*	.0126^	.3191	99.74
	Time	.1208	.047	<.0001	.1085	.0397	<.0001	.0839	.0227	<.0001	
HTOEPH positive	Treatment	.1012	.0381	<.0001	.081	.0285	<.0001	.0864	.0229	<.0001	60.33
	Time	.1276	.0473	<.0001	.1043	.0362	<.0001	.1094	.0328	<.0001	
HCS negative	Treatment	.1969	.0989	<.0001	.1649	.0785	<.0001	.0534*	.0152^	.00079	100
	Time	.0791	.0263	<.0001	.0751^	.0222	<.0001	.057^	.0137^	.2739	
HCSH negative	Treatment	.2217	.103	<.0001	.2049	.0911	<.0001	.1310	.0528	<.0001	92.64
	Time	.0831	.0277	<.0001	.0785	.0254	<.0001	.066^	.0175	<.0001	
HARRE negative	Treatment	.1742	.0817	<.0001	.1488	.0651	<.0001	.0647^	.0202	<.0001	100
	Time	.0846	.0313	<.0001	.0876	.0308	<.0001	.0429^	.0087*	<.0001	
HARHRE negative	Treatment	.2006	.0913	<.0001	.1879	.0828	<.0001	.1181	.0499	<.0001	74.75
	Time	.0965	.0357	<.0001	.0868	.0294	<.0001	.1019	.0429	<.0001	

Note1: * Proportions included in expected 95% intervals.
Note2: ^ Robust proportions ([.025-0.075] for α=.05 or [.005-.015] for α=.01) not included in the expected 95% confidence intervals.

Table 9 (Cont.): Simulated Type I error rates (α=0.05 and 0.01), goodness of fit tests comparing the Uniform (0,1) to p-value distributions, and convergence rates for the (3, 5, 7)x3 sample size studies

Covariance Structures	Effect	Best AIC model			Best BIC model			Correct model			Convergence rates for correct model
		Prop ≤ α=0.05	Prop ≤ α=0.01	χ^2 p-value	Prop ≤ α=0.05	Prop ≤ α=0.01	χ^2 p-value	Prop ≤ α=0.05	Prop ≤ α=0.01	χ^2 p-value	
HRC negative	Treatment	.1415	.0545	<.0001	.1132	.0379	<.0001	.0721^	.0219	<.0001	97.53
	Time	.088	.0302	<.0001	.0745^	.0236	<.0001	.0375^	.0084*	<.0001	
HTOEP negative	Treatment	.1925	.0929	<.0001	.1653	.0784	<.0001	.0927	.0402	<.0001	99.87
	Time	.0875	.0294	<.0001	.0837	.0276	<.0001	.0634	.0172	<.0001	
HTOEPH negative	Treatment	.219	.1068	<.0001	.204	.0974	<.0001	.1634	.0677	<.0001	62.66
	Time	.0855	.0278	<.0001	.0806	.0251	<.0001	.0723^	.0233	<.0001	

Note1: * Proportions included in expected 95% intervals.
Note2: ^ Robust proportions ([.025-0.075] for α=.05 or [.005-.015] for α=.01) not included in the expected 95% confidence intervals.

Table 10: Simulated Type I error rates (α=0.05 and 0.01), goodness of fit tests comparing the Uniform (0,1) to p-value distributions, and convergence rates for the (3, 5, 7)x5 sample size studies

Covariance Structures	Effect	Best AIC model			Best BIC model			Correct model			Convergence rates for correct model
		Prop ≤ α=0.05	Prop ≤ α=0.01	χ^2 p-value	Prop ≤ α=0.05	Prop ≤ α=0.01	χ^2 p-value	Prop ≤ α=0.05	Prop ≤ α=0.01	χ^2 p-value	
CS	Treatment	.1055	.0355	<.0001	.0797	.0225	<.0001	.0502*	.0096*	.8938	100
	Time	.0788	.0274	<.0001	.0601^	.0177	.0007	.0526*	.0116*	.6955	
CSH	Treatment	.1121	.037	<.0001	.0894	.026	<.0001	.0588^	.0129^	.0265	100
	Time	.0771	.0255	<.0001	.0605^	.0152^	.0002	.0514*	.0108*	.7298	
ARRE	Treatment	.1009	.033	<.0001	.0774	.0206	<.0001	.0561^	.0099*	.0159	100
	Time	.0797	.0309	<.0001	.0575^	.0178	.0003	.0406*	.0079*	.0113	
ARHRE	Treatment	.0993	.0357	<.0001	.0846	.0271	<.0001	.0494*	.0108*	.733	95.48
	Time	.0969	.0365	<.0001	.0810	.0291	<.0001	.0718^	.0230	<.0001	
RC	Treatment	.0984	.0316	<.0001	.0754^	.0182	<.0001	.0442^	.0073^	.2439	100
	Time	.0851	.0303	<.0001	.075^	.0232	<.0001	.0454^	.0096*	.7388	
TOEP	Treatment	.0962	.0361	<.0001	.0685^	.0213	<.0001	.048*	.0107*	.1771	100
	Time	.095	.0378	<.0001	.0792	.0277	<.0001	.0522*	.012*	.0046	
TOEPH	Treatment	.1098	.0382	<.0001	.0929	.0278	<.0001	.0718^	.0158	<.0001	100
	Time	.0922	.034	<.0001	.0802	.026	<.0001	.057^	.0112*	.1781	
UN	Treatment	.1316	.0452	<.0001	.1091	.0331	<.0001	.1188	.0398	<.0001	100
	Time	.0892	.0295	<.0001	.0766	.0215	<.0001	.0665^	.0153	<.0001	
HCS positive	Treatment	.0802	.0268	<.0001	.0601^	.0165	.0093	.0389^	.0054^	.0012	100
	Time	.0999	.0376	<.0001	.0744^	.0237	<.0001	.0536*	.0111*	.1107	
HCSH positive	Treatment	.1078	.0371	<.0001	.078	.0247	<.0001	.0915	.0286	<.0001	90.42
	Time	.125	.047	<.0001	.0855	.0277	<.0001	.1376	.0476	<.0001	

Note1: * Proportions included in expected 95% intervals.
Note2: ^ Robust proportions ([.025-0.075] for α=.05 or [.005-.015] for α=.01) not included in the expected 95% confidence intervals.

Table 10 (Cont.): Simulated Type I error rates (α=0.05 and 0.01), goodness of fit tests comparing the Uniform (0,1) to p-value distributions, and convergence rates for the (3, 5, 7)x5 sample size studies

Covariance Structures	Effect	Best AIC model			Best BIC model			Correct model			Convergence rates for correct model
		Prop ≤ α=0.05	Prop ≤ α=0.01	χ^2 p-value	Prop ≤ α=0.05	Prop ≤ α=0.01	χ^2 p-value	Prop ≤ α=0.05	Prop ≤ α=0.01	χ^2 p-value	
HARRE positive	Treatment	.0839	.0275	<.0001	.0666^	.0183	<.0001	.0542*	.0097*	.0499	100
	Time	.1062	.0448	<.0001	.0821	.0302	<.0001	.0542*	.0158	.0001	
HARHRE positive	Treatment	.1164	.0395	<.0001	.0864	.0261	<.0001	.1091	.0425	<.0001	64.18
	Time	.1353	.0549	<.0001	.1033	.0393	<.0001	.16	.0668	<.0001	
HRC positive	Treatment	.0742^	.0252	<.0001	.0592^	.0158	<.0001	.0454^	.0117*	<.0001	96.38
	Time	.0912	.0347	<.0001	.078	.0265	<.0001	.0556^	.0132^	<.0001	
HTOEP positive	Treatment	.0753^	.0235	<.0001	.0591^	.014^	.0033	.033^	.0063^	<.0001	74.71
	Time	.1211	.0512	<.0001	.102	.0382	<.0001	.1362	.0553	<.0001	
HTOEPH positive	Treatment	.1061	.0386	<.0001	.0798	.0261	<.0001	0	0	.	0
	Time	.1361	.05	<.0001	.1041	.0352	<.0001	0	0	.	
HCS negative	Treatment	.1242	.056	<.0001	.1012	.0404	<.0001	.053*	.0158	.0021	100
	Time	.0726^	.0248	<.0001	.0611^	.0178	.0023	.047*	.0104*	.7801	
HCSH negative	Treatment	.202	.0968	<.0001	.182	.0837	<.0001	.1336	.0539	<.0001	90.96
	Time	.0923	.0299	<.0001	.0754^	.0215	<.0001	.09	.0261	<.0001	
HARRE negative	Treatment	.1081	.0416	<.0001	.0892	.031	<.0001	.0542*	.012*	.0014	100
	Time	.0804	.0298	<.0001	.0677^	.0235	<.0001	.0427^	.0108*	<.0001	
HARHRE negative	Treatment	.1698	.0742	<.0001	.1532	.0653	<.0001	.1204	.0466	<.0001	69.34
	Time	.1143	.0444	<.0001	.0883	.0323	<.0001	.119	.0483	<.0001	

Note1: * Proportions included in expected 95% intervals.
Note2: ^ Robust proportions ([.025-0.075] for α=.05 or [.005-.015] for α=.01) not included in the expected 95% confidence intervals.

Table 10 (Cont.): Simulated Type I error rates (α=0.05 and 0.01), goodness of fit tests comparing the Uniform (0,1) to p-value distributions, and convergence rates for the (3, 5, 7)x5 sample size studies

Covariance Structures	Effect	Best AIC model			Best BIC model			Correct model			Convergence rates for correct model
		Prop ≤ α=0.05	Prop ≤ α=0.01	χ^2 p-value	Prop ≤ α=0.05	Prop ≤ α=0.01	χ^2 p-value	Prop ≤ α=0.05	Prop ≤ α=0.01	χ^2 p-value	
HRC negative	Treatment	.131	.0499	<.0001	.107	.0377	<.0001	.077	.0284	<.0001	97.29
	Time	.0896	.035	<.0001	.0794	.0273	<.0001	.05*	.0121*	<.0001	
HTOEP negative	Treatment	.1191	.0503	<.0001	.0985	.0358	<.0001	.0491*	.0113*	<.0001	75.19
	Time	.0916	.0312	<.0001	.0851	.0283	<.0001	.0928	.0305	<.0001	
HTOEPH negative	Treatment	.1921	.0888	<.0001	.1586	.0717	<.0001	0	0	.	0
	Time	.1052	.0382	<.0001	.0899	.0294	<.0001	0	0	.	

Note1: * Proportions included in expected 95% intervals.
Note2: ^ Robust proportions ([.025-0.075] for α=.05 or [.005-.015] for α=.01) not included in the expected 95% confidence intervals.

4.1.2.b. Correct Model

Distributions of p-values from fitting the correct model in the unequal treatment size case followed the uniform distribution for both effects and both numbers of repeated measures when the covariance structures were CS, CSH, or RC. Depending on the number of repeated measures and the effect tested, uniformly distributed p-values were also produced for the ARRE, ARHRE, TOEP, TOEPH, UN, HCS positive, HARRE positive, HTOEP positive, and HCS negative covariance structures. The rest of covariance structures never produced uniformly distributed p-values.

For the unequal treatment size simulations, there were also situations producing conservative error rates (Table 11, Figures 9 to 12 in Appendix 1). Once more almost all of these situations involve heterogeneity between treatments.

Table 11: Situations producing conservative error rates (less than the lower bound of the expected 95% confidence interval for α=0.05 or α=0.01) for the correct model in the unequal sample size simulation studies

Size	Covariance structure	Effect	Prop ≤ 0.05	Prop ≤ 0.01
(3, 5, 7)x3	ARRE	Time	.0352	.0061
(3, 5, 7)x3	RC	Time	.0431	.0083
(3, 5, 7)x3	HCS positive	Treatment	.0366	.0045
(3, 5, 7)x3	HRC positive	Time	.0369	.0085
(3, 5, 7)x3	HARRE negative	Time	.0429	.0087
(3, 5, 7)x3	HRC negative	Time	.0375	.0084
(3, 5, 7)x5	RC	Treatment	.0442	.0073
(3, 5, 7)x5	RC	Time	.0454	.0096
(3, 5, 7)x5	HCS positive	Treatment	.0389	.0054
(3, 5, 7)x5	HRC positive	Treatment	.0454	.0117
(3, 5, 7)x5	HTOEP positive	Treatment	.033	.0063
(3, 5, 7)x5	HARRE negative	Time	.0427	.0108

There were some situations in which the proportions of p-values less than or equal to α=0.05 or α =0.01 were robust according to Bradley's criterion (Bradley, 1978) or were included in the expected 95% confidence intervals (Table 9 to 10).

4.2. Performance of AIC and BIC criteria

4.2.1. Equal treatment sizes

The success rate, the percent of times that AIC or BIC chose the correct covariance structure, depended greatly on the sample size and covariance structure (Tables 12 to 15). The success rates were generally low. For instance, for the 3x3 sample case, the success rate ranged from 2.51 to 29.98%. The highest success rate was 73.91%, for the largest sample size and a simple covariance structure (ARRE). Success rates were higher for larger sample sizes and simpler covariance structures.

AIC had a higher success rate than BIC for complicated structures, especially for those with heterogeneity between-treatments. However, BIC had higher success rate than AIC for simpler structures.

A covariance structure was considered to be consistent if the correct covariance structure was chosen more often than any other specific structure. Consistency was more common for larger sample size cases. CS and CSH were always consistent independently of the situation. Other structures as ARRE, ARHRE, RC, UN, HCS, HCSH, HARRE, HARHRE, HRC and HTOEPH were sometimes consistent depending on the sample size and the type of criteria used. The TOEP, TOEPH and HTOEP covariance structures were never consistent.

Usually, when the number of repeated measures was larger, the "wrong" choices made by AIC and BIC made more sense. For instance, for the 5 repeated measures cases, AIC and BIC most often chose the ARRE and ARHRE covariance structures when data followed the TOEP and TOEPH covariance structures, respectively. This made sense since ARRE and ARHRE were specified with patterns similar to those of TOEP and TOEPH. However, for the 5x3 sample size, AIC and BIC most often chose the CS and CSH covariance structures when data followed the TOEP and TOEPH covariance structures respectively. TOEP and CS did not follow the same pattern, nor did TOEPH and CSH. Still, it is important to notice that TOEP was specified in such way that it decreased rapidly, while ARRE decreased much slower (Table 1).

Table 12: Success rates and details on covariance structures selected by AIC and BIC criteria for the 3x3 sample size simulation studies

Covariance Structures	Criterion	Success rate	Most often chosen model	Rate for most often chosen model
CS	AIC	23.46	CS	23.46
	BIC	29.98	CS	29.98
CSH	AIC	15.96	CSH	15.96
	BIC	17.79	CSH	17.79
ARRE	AIC	16.00	HRC	19.68
	BIC	19.99	CS	20.5
ARHRE	AIC	11.43	HRC	18.7
	BIC	12.01	HRC	17.1
RC	AIC	8.29	CS	17.84
	BIC	9.36	CS	22.5
TOEP	AIC	3.43	HRC	19.78
	BIC	3.61	CS	20.92
TOEPH	AIC	5.67	HRC	17.11
	BIC	5.69	HRC	15.57
UN	AIC	6.65	CSH	16.67
	BIC	5.89	CSH	18.19
HCS	AIC	9.32	HRC	15.62
	BIC	9.30	CS	15.62
HCSH	AIC	9.76	HRC	14.71
	BIC	7.56	HRC	13.68
HARRE	AIC	6.85	HRC	17.32
	BIC	6.59	HRC	15.74
HARHRE	AIC	12.59	HRC	17.42
	BIC	10.26	HRC	16.21
HRC	AIC	17.90	CS	18.12
	BIC	15.49	CS	22.51
HTOEP	AIC	16.45	HRC	17.87
	BIC	14.30	HRC	16.03
HTOEPH	AIC	3.42	HRC	16.64
	BIC	2.51	HRC	15.56

Table 13: Success rates and details on covariance structures selected by AIC and BIC criteria for the 5x3 sample size simulation studies

Covariance Structures	Criterion	Success rate	Most often chosen model	Rate for most often chosen model
CS	AIC	41.22	CS	41.22
	BIC	62.11	CS	62.11
CSH	AIC	35.42	CSH	35.42
	BIC	41.75	CSH	41.75
ARRE	AIC	27.15	ARRE	27.15
	BIC	37.08	ARRE	37.08
ARHRE	AIC	27.80	ARHRE	27.80
	BIC	29.83	ARHRE	29.83
RC	AIC	12.47	CS	26.67
	BIC	13.79	CS	40.03
TOEP	AIC	4.52	CS	27.84
	BIC	3.89	CS	41.30
TOEPH	AIC	6.81	CSH	24.96
	BIC	5.51	CSH	29.34
UN	AIC	5.76	CSH	33.41
	BIC	3.45	CSH	40.08
HCS	AIC	28.40	HCS	28.40
	BIC	26.49	CS	27.06
HCSH	AIC	16.53	HCSH	16.53
	BIC	6.57	CSH	27.51
HARRE	AIC	23.88	HARRE	23.88
	BIC	18.89	CS	21.22
HARHRE	AIC	12.60	HTOEPH	16.89
	BIC	4.47	ARHRE	22.73
HRC	AIC	8.18	CS	26.53
	BIC	3.28	CS	40.19
HTOEP	AIC	9.49	HARRE	21.49
	BIC	5.05	CS	19.28
HTOEPH	AIC	16.39	HTOEPH	16.39
	BIC	5.03	CSH	19.05

Table 14: Success rates and details on covariance structures selected by AIC and BIC criteria for the 3x5 sample size simulation studies

Covariance Structures	Criterion	Success rate	Most often chosen model	Rate for most often chosen model
CS	AIC	26.47	CS	26.47
	BIC	34.73	CS	34.73
CSH	AIC	23.02	CSH	23.02
	BIC	25.42	CSH	25.42
ARRE	AIC	30.45	ARRE	30.45
	BIC	38.14	ARRE	38.14
ARHRE	AIC	24.71	ARHRE	24.71
	BIC	26.91	ARHRE	26.91
RC	AIC	20.20	HRC	23.71
	BIC	23.29	RC	23.29
TOEP	AIC	6.23	ARRE	22.15
	BIC	6.32	ARRE	27.72
TOEPH	AIC	10.21	ARHRE	19.80
	BIC	9.46	ARHRE	21.24
UN	AIC	32.58	UN	32.58
	BIC	26.57	UN	26.57
HCS	AIC	25.98	HCS	25.98
	BIC	28.23	HCS	28.23
HCSH	AIC	10.57	UN	25.20
	BIC	7.59	UN	20.71
HARRE	AIC	28.64	HARRE	28.64
	BIC	30.07	HARRE	30.07
HARHRE	AIC	12.45	UN	23.94
	BIC	8.15	UN	19.67
HRC	AIC	25.17	HRC	25.17
	BIC	23.04	HRC	23.04
HTOEP	AIC	5.44	HARRE	23.76
	BIC	4.00	HARRE	24.37
HTOEPH	AIC	0	UN	24.46
	BIC	0	UN	19.91

Table 15: Success rates and details on covariance structures selected by AIC and BIC criteria for the 5x5 sample size simulation studies

Covariance Structures	Criterion	Success Rate	Most often chosen model	Rate for most often chosen model
CS	AIC	41.69	CS	41.69
	BIC	69.03	CS	69.03
CSH	AIC	50.38	CSH	50.30
	BIC	60.39	CSH	60.39
ARRE	AIC	51.26	ARRE	51.26
	BIC	73.91	ARRE	73.91
ARHRE	AIC	56.41	ARHRE	56.41
	BIC	68.46	ARHRE	68.46
RC	AIC	40.52	RC	40.52
	BIC	49.32	RC	49.32
TOEP	AIC	6.62	ARRE	38.28
	BIC	4.30	ARRE	54.08
TOEPH	AIC	10.49	ARHRE	45.87
	BIC	5.86	ARHRE	50.85
UN	AIC	25.62	UN	25.62
	BIC	9.01	CSH	38.08
HCS	AIC	51.39	HCS	51.39
	BIC	60.00	HCS	60.00
HCSH	AIC	34.07	HCSH	34.07
	BIC	8.60	CSH	36.10
HARRE	AIC	61.56	HARRE	61.56
	BIC	64.20	HARRE	64.20
HARHRE	AIC	25.96	HARHRE	25.96
	BIC	6.54	ARHRE	40.37
HRC	AIC	21.38	RC	35.29
	BIC	11.05	RC	45.68
HTOEP	AIC	15.33	HARRE	52.59
	BIC	4.15	HARRE	54.97
HTOEPH	AIC	0	HARHRE	20.43
	BIC	0	HARRE	27.83

4.2.2. Unequal treatment sizes

The success rates of AIC and BIC in the unequal treatment size situations were comparable to those of the equal treatment size cases with 5 repeated measures per subject. This is reasonable because the total number of subjects was the same. Success rates depended on the number of repeated measures and covariance structure (Tables 16 to 17). The success increased by increasing the number of repeated measures and using simple covariance structures.

Once again, AIC frequently had higher success rates than BIC for complicated structures, especially for those with heterogeneity between-treatments. However, BIC had higher success rates than AIC for simpler structures. Often AIC and BIC were able to recognize the correct structure other than the between-treatment heterogeneity.

CS, CSH, ARRE, and ARHRE covariance structures were consistent independently of the number of repeated measures and the type of criteria used. Other structures as RC, UN, HCS positive, HARRE positive, HCS negative, HCSH negative, HARRE negative and HARHRE negative were consistent depending on the number of repeated measures and the criteria used. The rest of the covariance structures were never consistent.

As in the equal treatment size case, AIC and BIC most often chose ARRE, ARHRE, HARRE, and HARHRE instead of TOEP, TOEPH, HTOEP, and HTOEPH respectively in the 5 repeated measure cases.

Pairing did not seem to have a great influence on the performance (success rate and consistency) of AIC and BIC. However, there was a slight improvement in the success rates of these criteria when the pairing was negative.

Table 16: Success rates and details on covariance structures selected by AIC and BIC criteria for the (3, 5, 7)x3 sample size simulation studies

Covariance Structures	Criterion	Success rate	Most often chosen model	Rate for most often chosen model
CS	AIC	38.35	CS	38.35
	BIC	60.13	CS	60.13
CSH	AIC	33.61	CSH	33.61
	BIC	40.57	CSH	40.57
ARRE	AIC	25.01	ARRE	25.01
	BIC	36.14	ARRE	36.14
ARHRE	AIC	25.95	ARHRE	25.95
	BIC	28.87	ARHRE	28.87
RC	AIC	12.20	CS	25.49
	BIC	13.03	CS	39.75
TOEP	AIC	4.02	CS	25.14
	BIC	3.53	CS	39.21
TOEPH	AIC	6.92	ARHRE	19.67
	BIC	5.47	CSH	28.28
UN	AIC	5.53	CSH	31.06
	BIC	3.31	CSH	39.17
HCS positive	AIC	20.33	HCS	20.33
	BIC	17.93	CS	35.23
HCSH positive	AIC	14.40	CSH	19.54
	BIC	5.49	CSH	31.45
HARRE positive	AIC	17.87	HARRE	17.89
	BIC	12.96	CS	25.12
HARHRE positive	AIC	15.78	ARHRE	18.06
	BIC	6.13	ARHRE	24.35
HRC positive	AIC	11.32	CS	24.03
	BIC	5.02	CS	37.77
HTOEP positive	AIC	11.80	HARRE	11.80
	BIC	6.01	CS	25.04
HTOEPH positive	AIC	8.50	ARHRE	12.60
	BIC	2.74	CSH	22.18
HCS negative	AIC	25.15	CS	25.15
	BIC	26.53	CS	/26.53
HCSH negative	AIC	17.39	CSH	17.39
	BIC	7.52	CSH	24.07
HARRE negative	AIC	21.49	HARRE	21.49
	BIC	18.27	CS	19.92

Table 16 (Cont.): Success rates and details on covariance structures selected by AIC and BIC criteria for the (3, 5, 7)x3 sample size simulation studies

Covariance Structures	Size (3, 5, 7)x3	Success rate	Most often chosen model	Rate for most often chosen model
HARHRE negative	AIC	15.20	HARHRE	15.20
	BIC	5.96	ARHRE	19.41
HRC negative	AIC	12.09	CS	24.95
	BIC	5.60	CS	39.38
HTOEP negative	AIC	13.36	HARRE	18.76
	BIC	7.54	HCS	19.15
HTOEPH negative	AIC	9.79	HARHRE	12.99
	BIC	3.09	CSH	17.71

Table 17: Success rates and details on covariance structures selected by AIC and BIC criteria for the (3, 5, 7)x5 sample size simulation studies

Covariance Structures	Size (3, 5, 7)x5	Success rate	Most often chosen model	Rate for most often chosen model
CS	AIC	41.44	CS	41.44
	BIC	68.17	CS	68.17
CSH	AIC	52.79	CSH	52.79
	BIC	59.68	CSH	59.68
ARRE	AIC	50.42	ARRE	50.42
	BIC	73.22	ARRE	73.22
ARHRE	AIC	59.22	ARHRE	59.22
	BIC	67.87	ARHRE	67.87
RC	AIC	40.00	RC	40.00
	BIC	48.62	RC	48.26
TOEP	AIC	6.86	ARRE	38.55
	BIC	4.42	ARR	53.62
TOEPH	AIC	11.52	ARHRE	48.20
	BIC	6.23	ARHRE	50.73
UN	AIC	26.62	UN	26.62
	BIC	9.12	CSH	37.81
HCS positive	AIC	45.90	HCS	HCS/45.90
	BIC	47.85	HCS	HCS/47.85
HCSH positive	AIC	25.29	CSH	CSH/27.38
	BIC	6.03	CSH	CSH/44.71
HARRE positive	AIC	55.08	HARRE	HARRE/55.08
	BIC	51.46	HARRE	51.46
HARHRE positive	AIC	22.38	ARHRE	33.61
	BIC	5.82	ARHRE	50.88

Table 17 (Cont.): Success rates and details on covariance structures selected by AIC and BIC criteria for the (3, 5, 7)x5 sample size simulation studies

Covariance Structures	Size (3, 5, 7)x5	Success rate	Most often chosen model	Rate for most often chosen model
HRC positive	AIC	23.77	RC	39.33
	BIC	12.68	RC	49.91
HTOEP positive	AIC	9.74	HARRE	48.22
	BIC	2.43	HARRE	44.89
HTOEPH positive	AIC	0	ARHRE	24.62
	BIC	0	ARHRE	34.89
HCS negative	AIC	51.37	HCS	51.37
	BIC	60.33	HCS	60.33
HCSH negative	AIC	33.60	HCSH	33.60
	BIC	10.04	CSH	31.83
HARRE negative	AIC	63.40	HARRE	63.40
	BIC	64.92	HARRE	64.92
HARHRE negative	AIC	26.66	HARHRE	26.66
	BIC	7.97	ARHRE	35.18
HRC negative	AIC	23.97	RC	29.33
	BIC	13.50	RC	38.06
HTOEP negative	AIC	9.84	HARRE	53.57
	BIC	2.46	HARRE	54.61
HTOEPH negative	AIC	0	HARHRE	23.29
	BIC	0	HARRE	27.28

5. CONCLUSION

5.1. Distributions of p-values from AIC, BIC and correct models.

Distributions of Kenward-Roger method p-values from fitting the best AIC and BIC models did not follow the uniform distribution for any sample size or covariance structure. These distributions were always right skewed which implies that Type I error rates were always higher than the target values. Type of selection criterion, number of subjects per treatment, number of repeated measures, effect tested, covariance structure and pairing affected the Type I error rates. Equality or inequality of number of subjects per treatment did not seem to affect the Type I error rates.

The best BIC models usually produced closer Type I error rates to target values than the best AIC models. Type I error rates were usually the farthest from target values for the 3x5 sample size situation and were closest for the 5x5 sample size situation. Tests for the within-subject effect generally produced closer Type I error rates to target values. Covariance structures with heterogeneity between- and within-treatment levels produced higher Type I error rates than those with only one type of heterogeneity or complete homogeneity. For tests of the between-subject effect, negative pairing produced dramatically higher Type I error rates than positive pairing. However, for tests of the within-subject effect, negative pairing produced slightly better Type I error rates than positive pairing. This disagrees with the results obtained by Keselman et al. (1999). They found that negative pairings had an adverse effect on Type I error rates for the within-subject effect. The disagreement could be due to the Kenward-Roger adjustment or the difference in sample sizes.

Convergence rates when the correct model was fitted depended on the complexity of the covariance structure. When the covariance structure was HTOEPH and the number of repeated measures was 5, the model converged only once in 20000 simulations. Therefore, it would be wise to not consider this covariance structure as a possibility when the available sample size is of the same order of magnitude as in this study.

Distributions of p-values from fitting the correct model often followed the uniform distribution. However, this depended on the sample size, equality of numbers of subjects per treatment, effect tested, pairing and covariance structure. P-values from data following the CS and RC covariance structures always followed the uniform distribution. P-values for TOEP data in the equal treatment size situation and CSH data in the unequal treatment size situation also were uniformly distributed. P-values for data with heterogeneity between- and within-treatment levels were seldom uniformly distributed, especially for negative pairings.

Distributions of p-values based on the correct model were sometimes left skewed, producing conservative type I error rates. This result did not occur with best AIC and best BIC models.

Unless sample sizes are large, if AIC and BIC are used, users should be aware that Type I error rates are higher than target values. Therefore, it is important to be cautious about declaring significance when AIC and BIC are the only tools used to select models. It would be wise to not rely just on AIC and BIC to choose a model, but to use the design and other practical knowledge to guide the choice of the covariance structure.

In general, results obtained by Keselman et al. (1999) were more optimistic than the ones obtained in this study for AIC models. Keselman et al. (1999) obtained conservative Type I error rates by using the best AIC models. This was probably due to the use of lager sample sizes.

Keselman et al. (1999) concluded that Type I error rates from fitting the HRC and HARH (similar to our HARHRE but without the random effect) covariance structures to every data set were adequate. They did not compare the distributions of p-values to the uniform, and their Type I error rates were only reasonably close to target values. Our study could not confirm the conclusions of Keselman et al. because only best AIC, best BIC and correct models were examined. However, even when the correct HARHRE structure was used, Type I error rates were far greater than target values. The smaller sample sizes used in our study could account for this difference.

Even if the correct covariance structure is known, Type I error rates are higher for complex structures and small sample size. This agrees with the results obtained by Schaalje et al. (2002).

5.2. Performance of AIC and BIC criteria

Percentages of times that the correct model was chosen by the AIC and BIC criteria were lower than those obtained by Keselman et al. (1998) and Ferron et al. (2002). However, Ferron et al. (2002) allowed choice between only two structures, and had much larger sample sizes. Keselman et al. (1998) allowed choice between 11 structures, but had much larger sample sizes than the ones used in our study.

AIC generally produced better success rates for data following covariance structures with heterogeneity between- and within-treatment levels. This result agrees with Keselman et al. (1998). However, for simpler covariance structures BIC had better success rates than AIC. Overall, both AIC and BIC had better success rates for simpler covariance structures and larger sample sizes.

For larger number of repeated measures and data following covariance structures with high number of parameters, AIC and BIC tended to most often choose covariance structures with the same patterns and lower numbers of parameters.

AIC and BIC are useful tools to help the research to choose a covariance structure. However, since they unfortunately do not always point to the correct covariance structure, it would be wise to not depend on them exclusively when choosing a covariance structure. It is important to be especially careful for small sample sizes because success rates were very low. Other resources such as correlograms (Little et al. (2002)), knowledge about the design and science should also be brought into play.

APPENDIX 1: Histograms

FIGURE 1: HISTOGRAMS OF P-VALUES FOR THE TREATMENT EFFECT -3X3

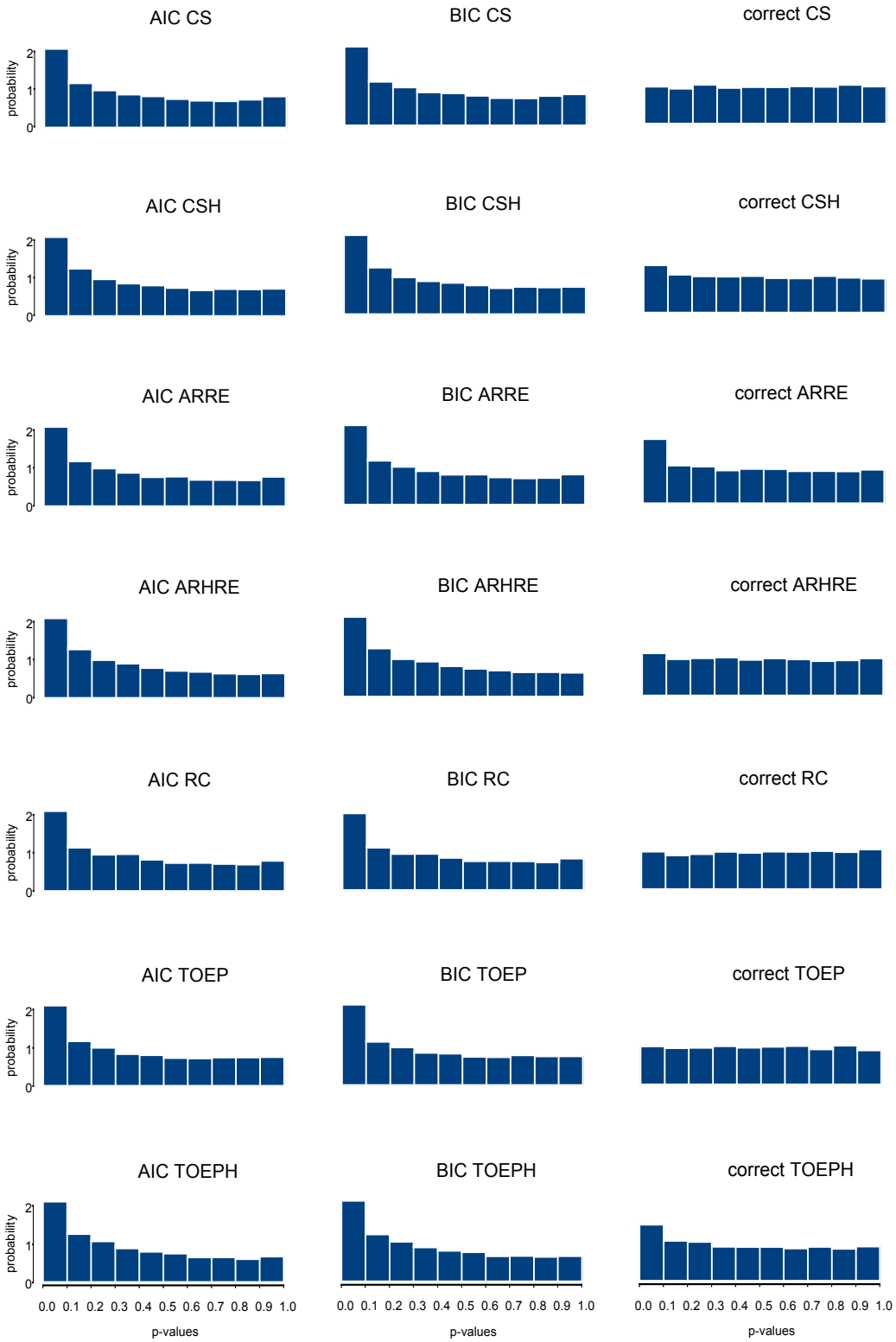


FIGURE 1 (Cont): HISTOGRAMS OF P-VALUES FOR THE TREATMENT EFFECT -3X3

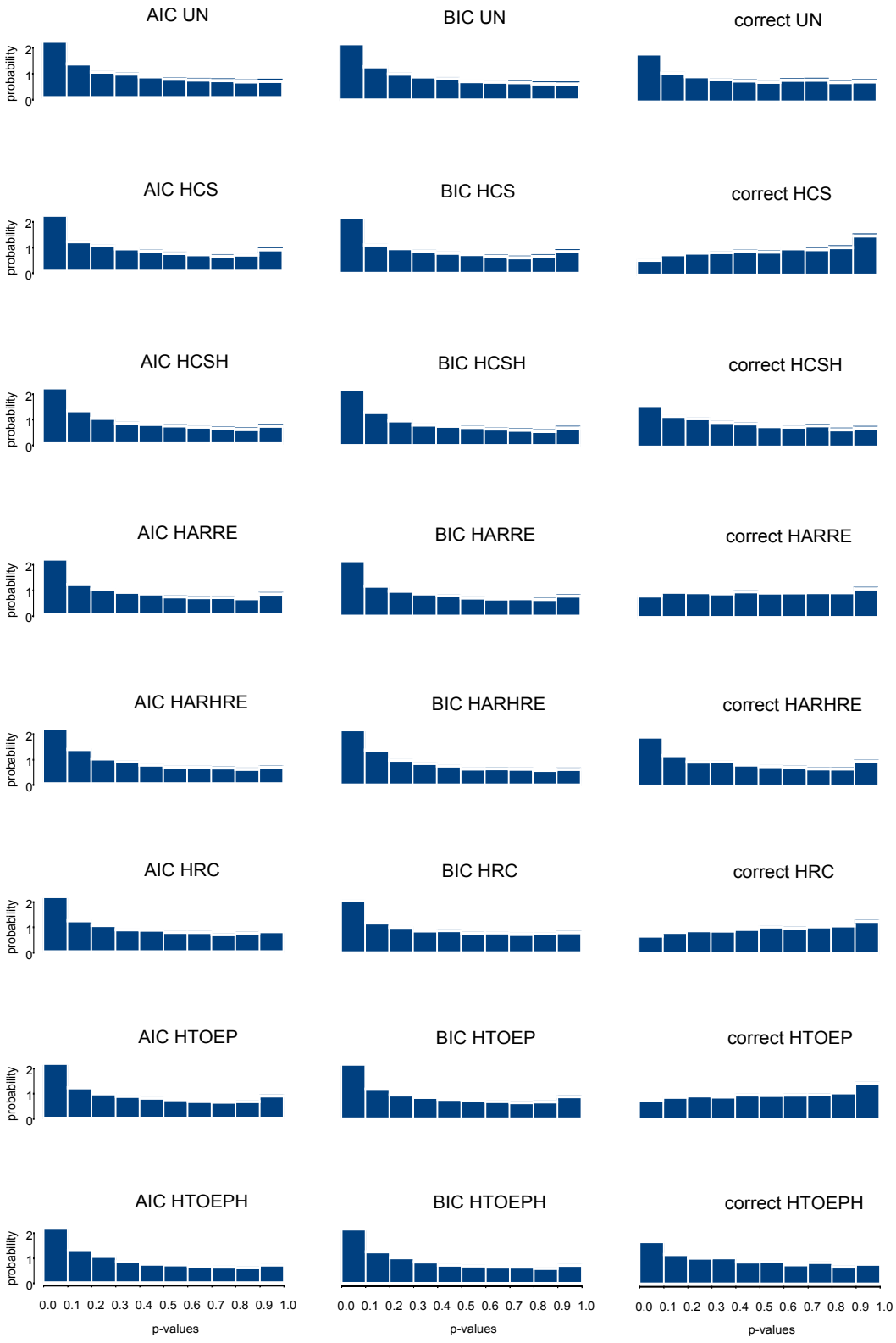


FIGURE 2: HISTOGRAMS OF P-VALUES FOR THE TIME EFFECT -3X3

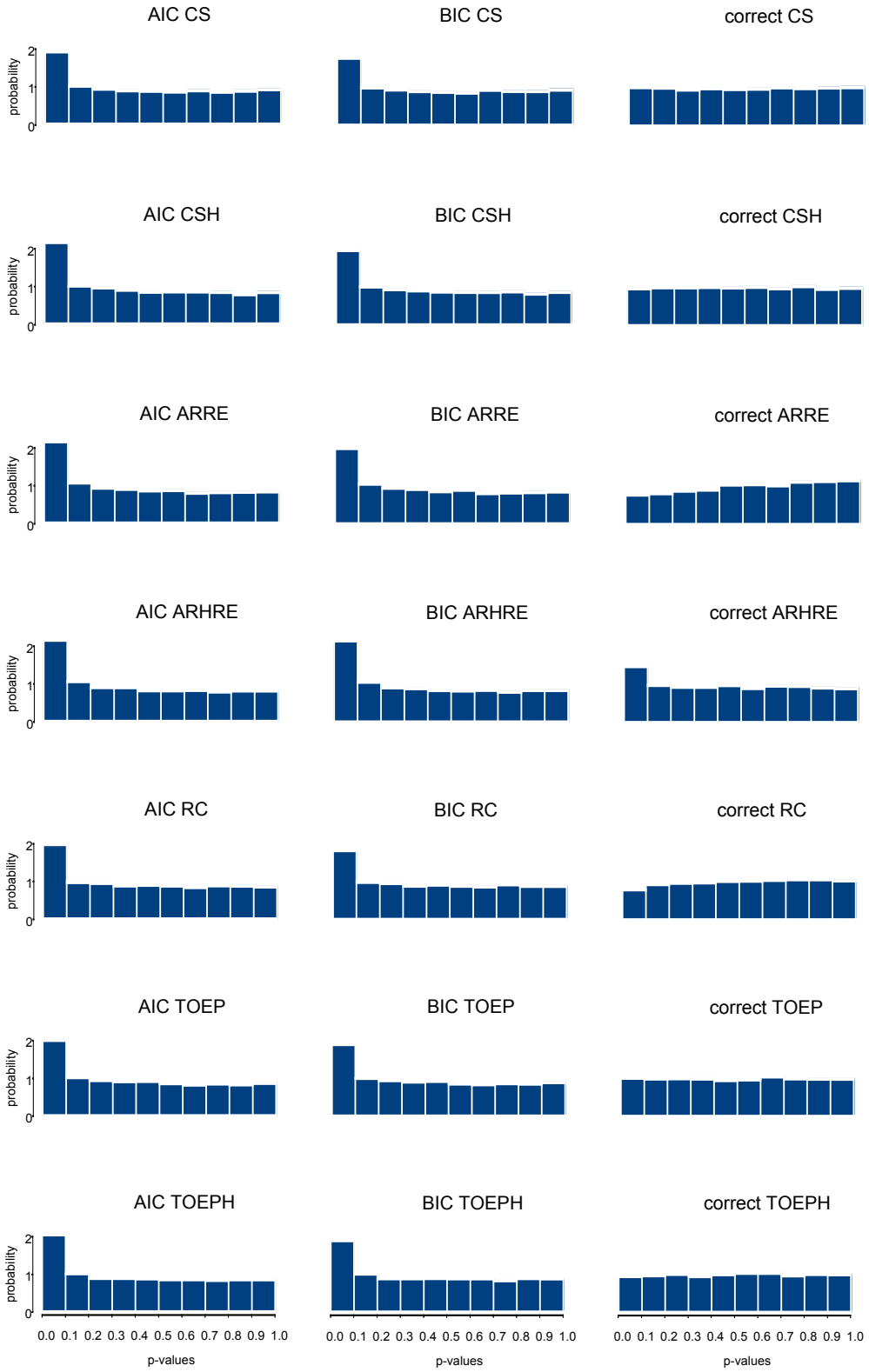


FIGURE 2 (Cont): HISTOGRAMS OF P-VALUES FOR THE TIME EFFECT -3X3

FIGURE 3: HISTOGRAMS OF P-VALUES FOR THE TREATMENT EFFECT - 5X3

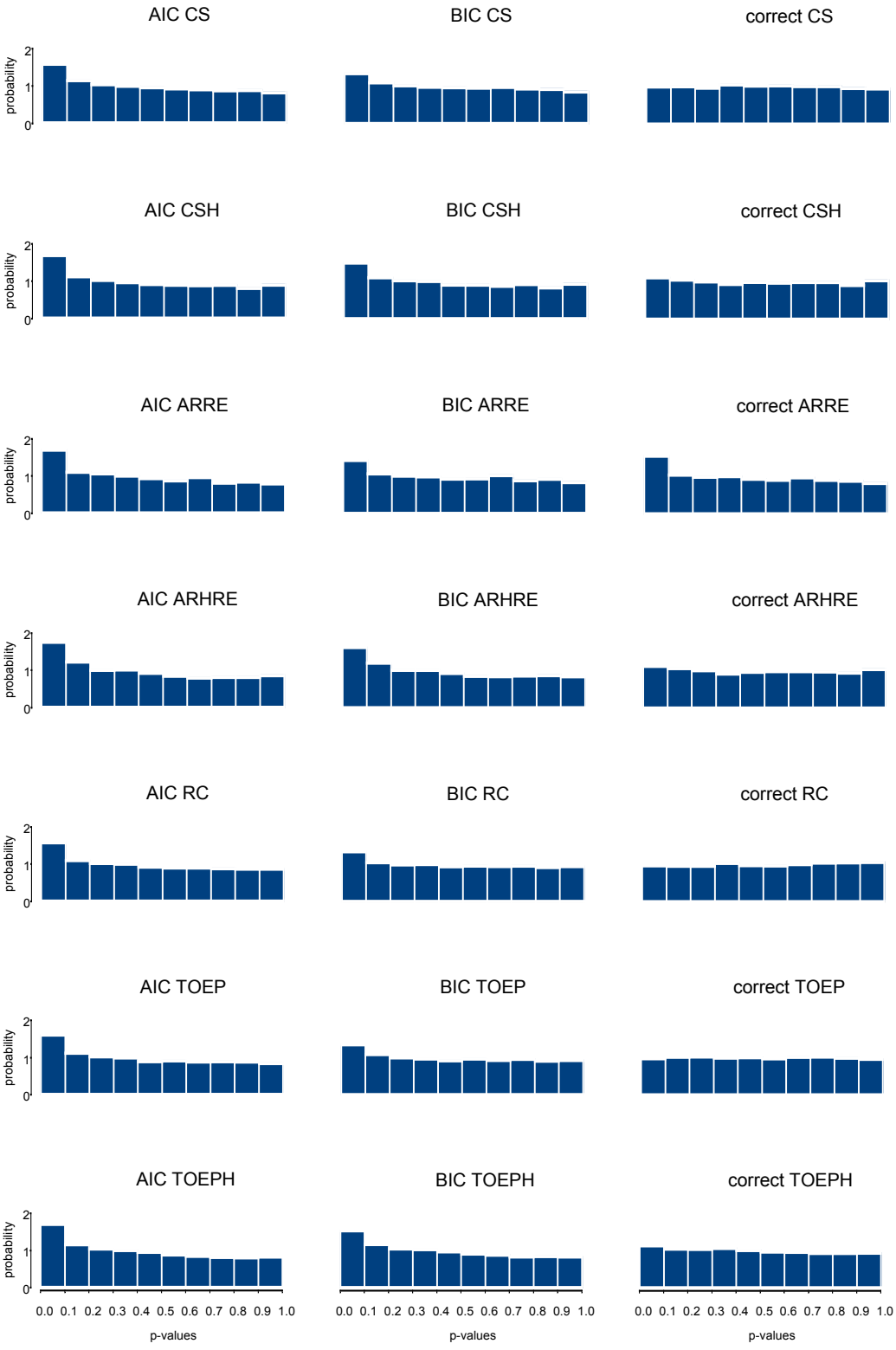


FIGURE 3 (Cont): HISTOGRAMS OF P-VALUES FOR THE TREATMENT EFFECT -5X3

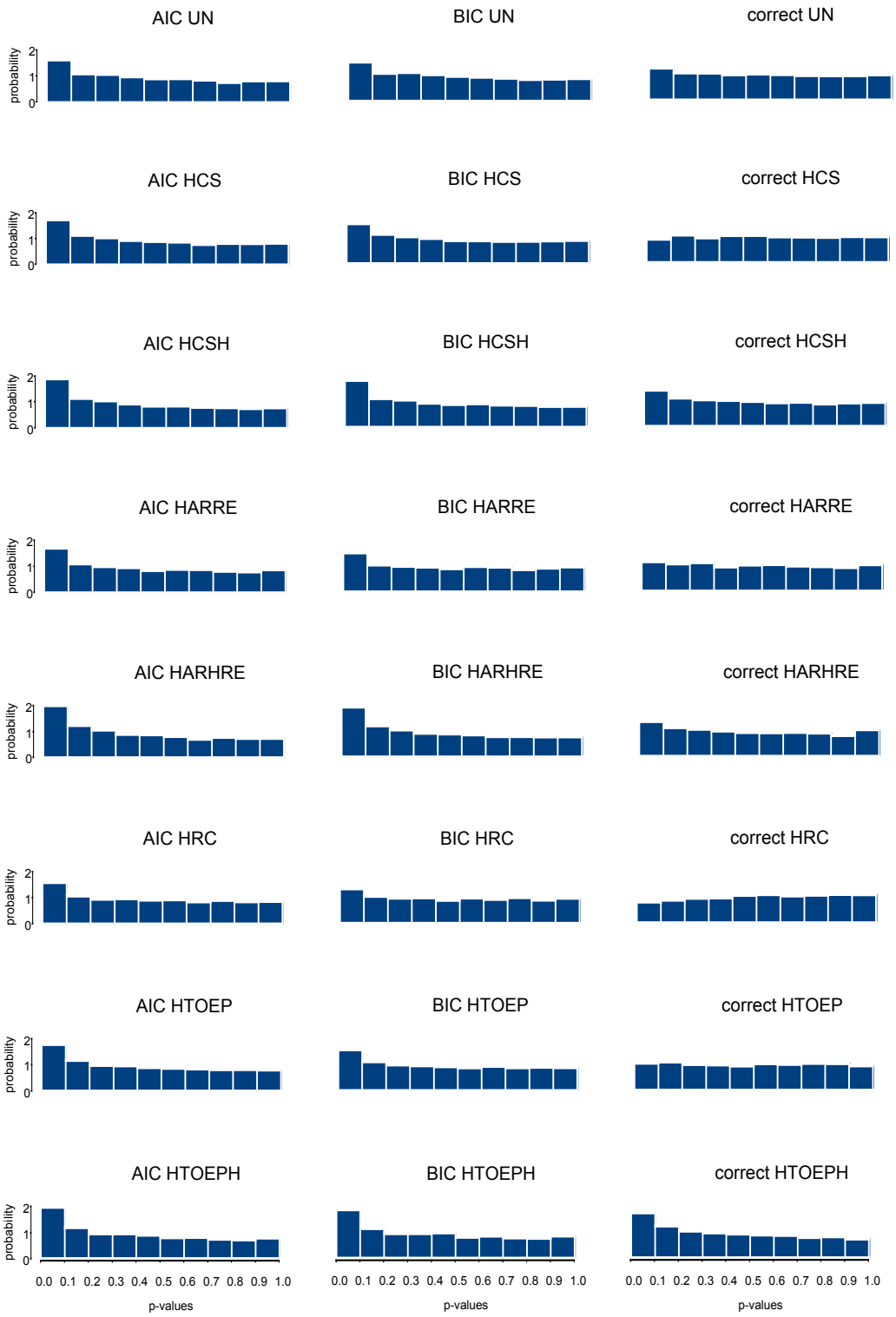


FIGURE 4: HISTOGRAMS OF P-VALUES FOR THE TIME EFFECT -5X3

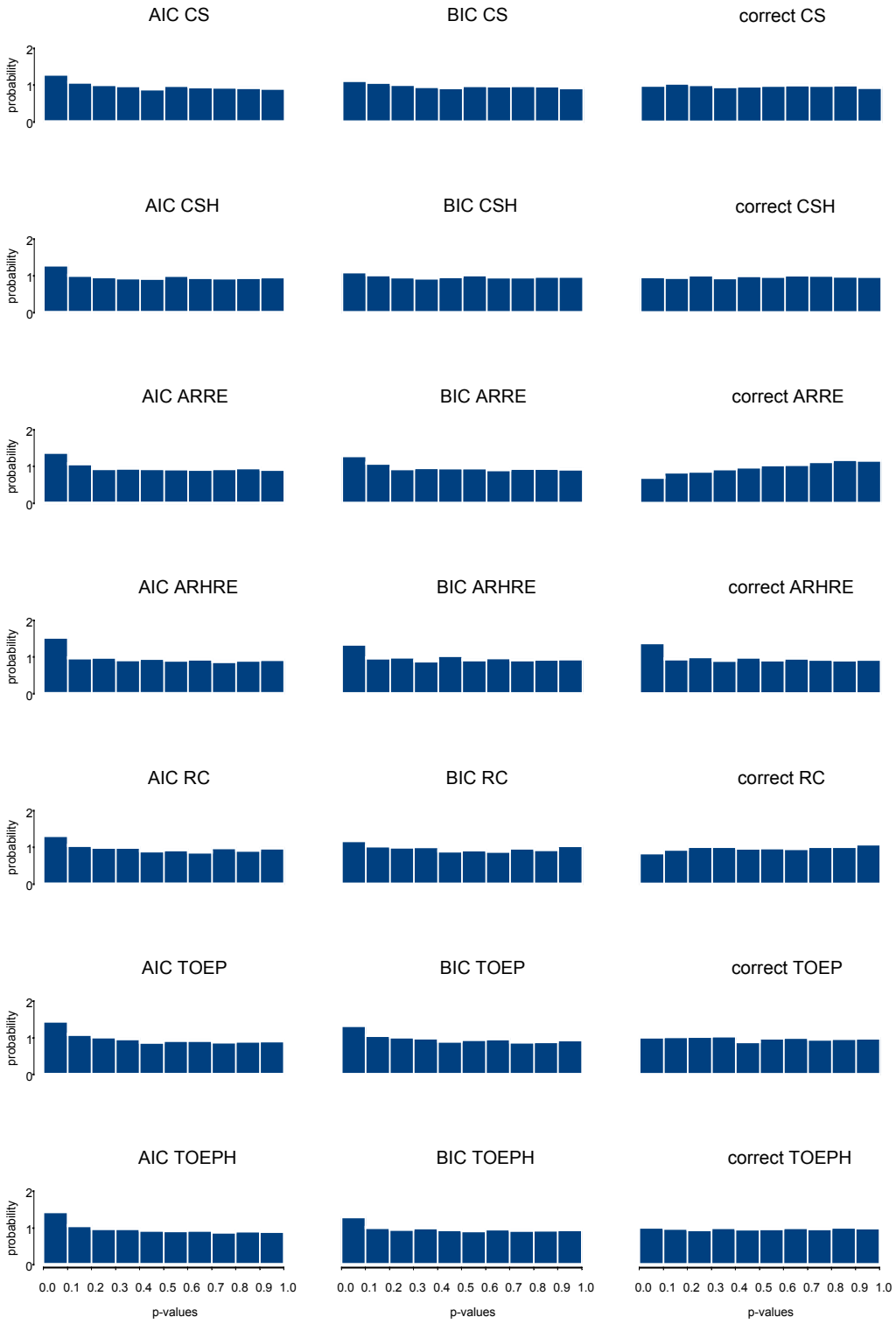
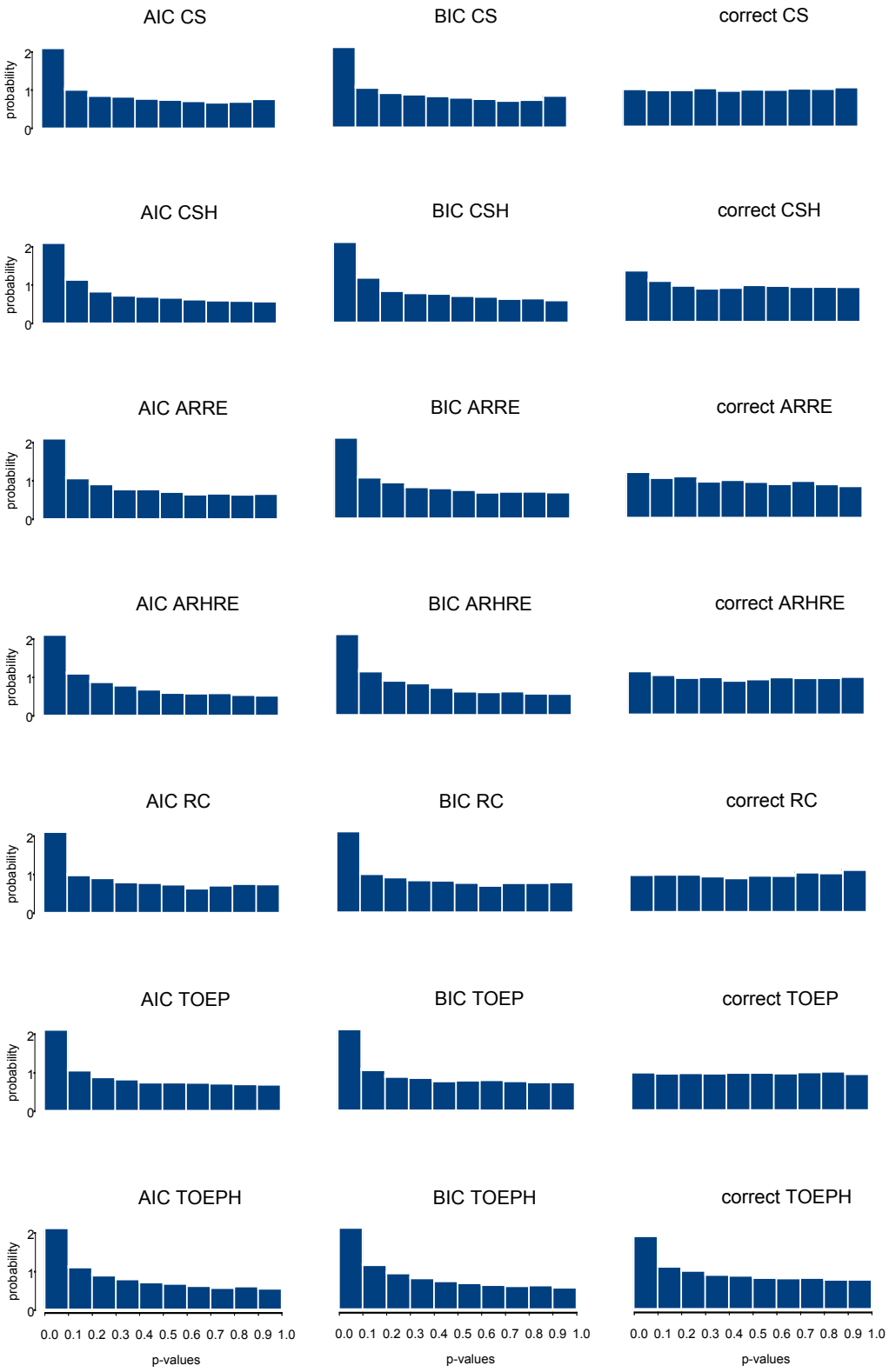


FIGURE 4 (Cont): HISTOGRAMS OF P-VALUES FOR THE TIME EFFECT -5X3

FIGURE 5: HISTOGRAMS OF P-VALUES FOR THE TREATMENT EFFECT -



3X5

FIGURE 5 (Cont): HISTOGRAMS OF P-VALUES FOR THE TREATMENT EFFECT -3X5

FIGURE 6: HISTOGRAMS OF P-VALUES FOR THE TIME EFFECT -3X5

FIGURE 6 (Cont): HISTOGRAMS OF P-VALUES FOR THE TIME EFFECT -3X5

FIGURE 7: HISTOGRAMS OF P-VALUES FOR THE TREATMENT EFFECT -5X5

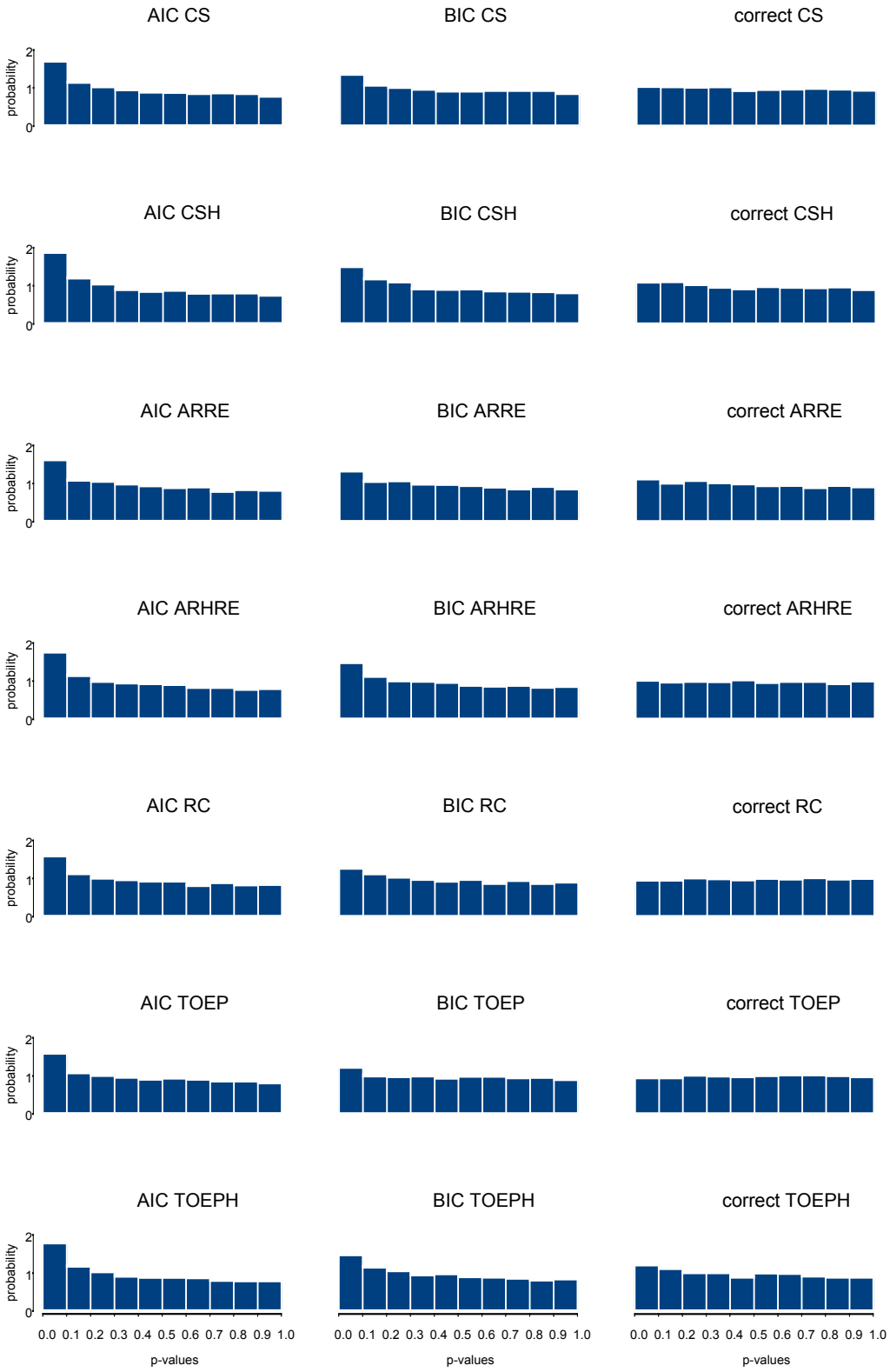


FIGURE 7 (Cont): HISTOGRAMS OF P-VALUES FOR THE TREATMENT EFFECT -5X5

FIGURE 8: HISTOGRAMS OF P-VALUES FOR THE TIME EFFECT -5X5

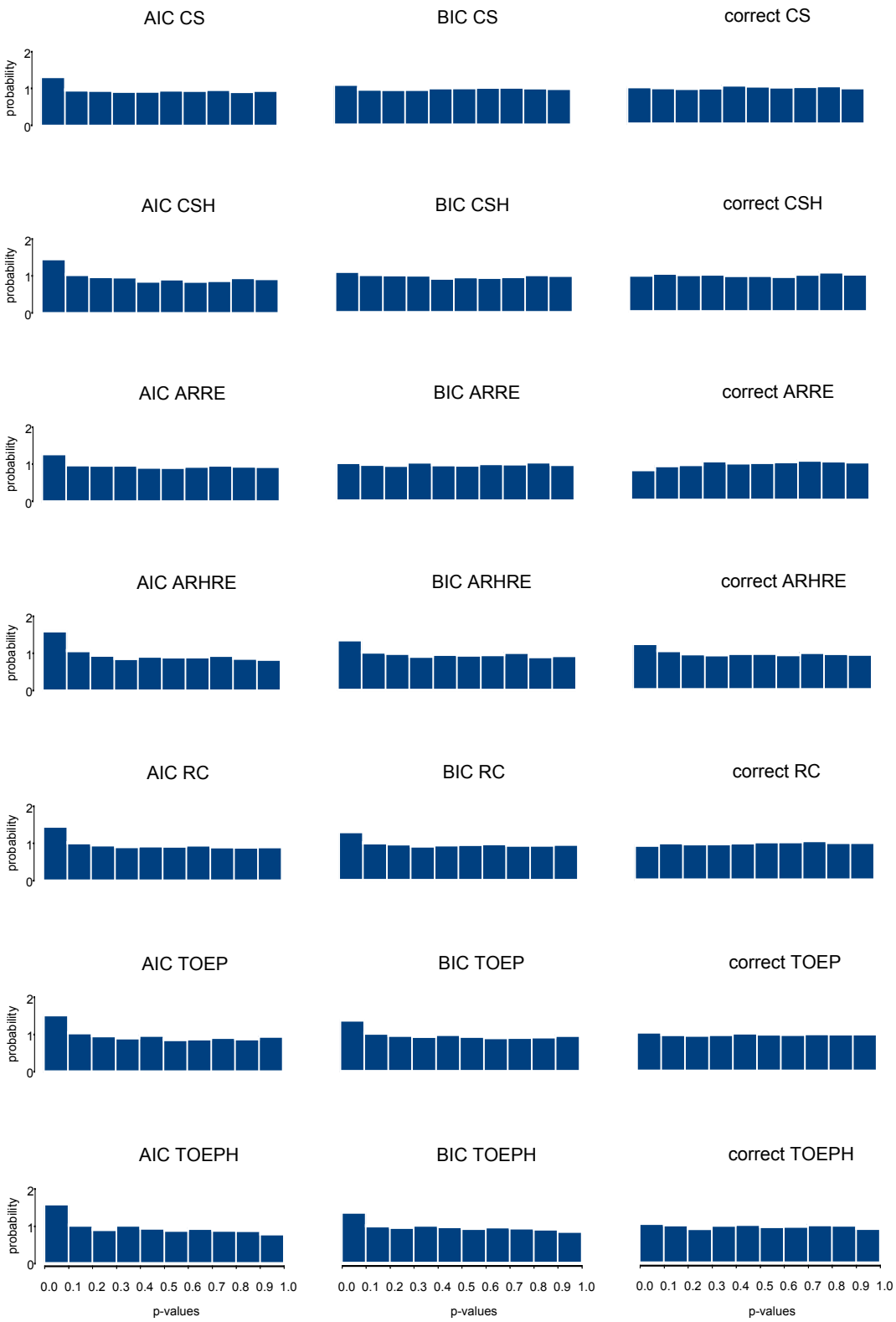


FIGURE 8 (Cont): HISTOGRAMS OF P-VALUES FOR THE TIME EFFECT -5X5

FIGURE 9: HISTOGRAMS OF P-VALUES FOR THE TREATMENT EFFECT -(3, 5, 7)x3

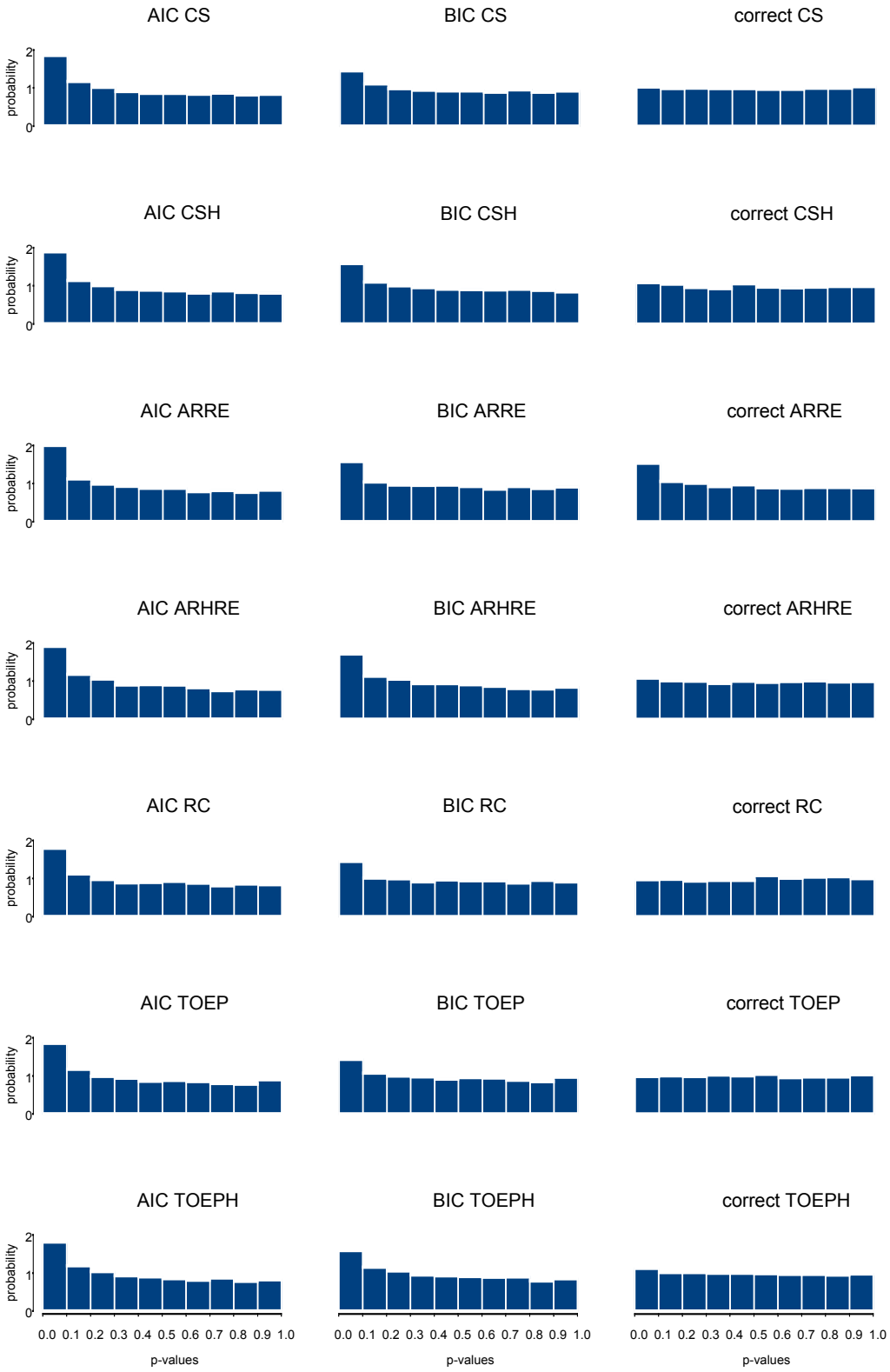


FIGURE 9 (Cont): HISTOGRAMS OF P-VALUES FOR THE TREATMENT EFFECT -(3, 5, 7)x3

FIGURE 9 (Cont): HISTOGRAMS OF P-VALUES FOR THE TREATMENT EFFECT -(3, 5, 7)x3

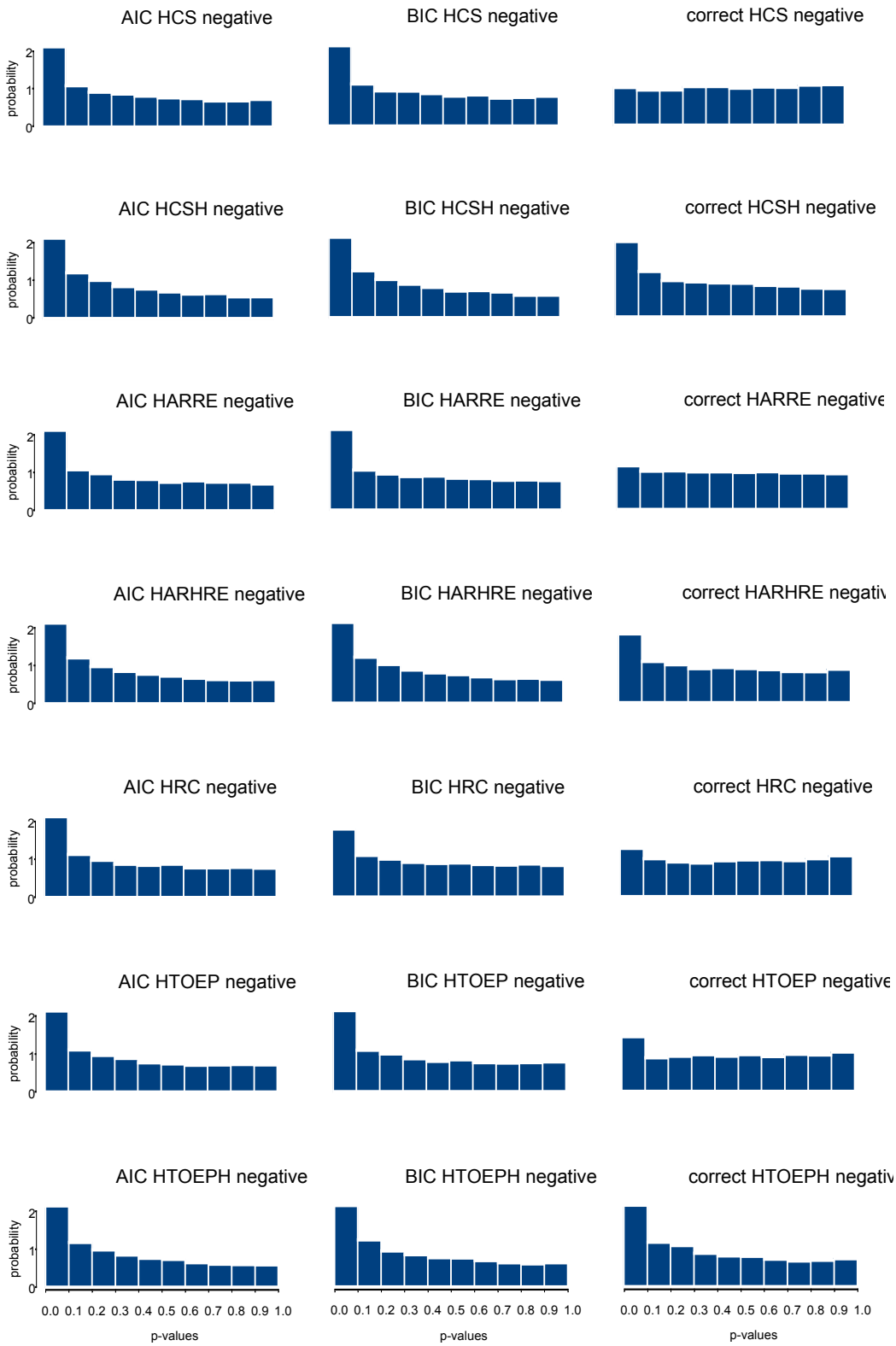


FIGURE 10: HISTOGRAMS OF P-VALUES FOR THE TIME EFFECT -(3, 5, 7)x3

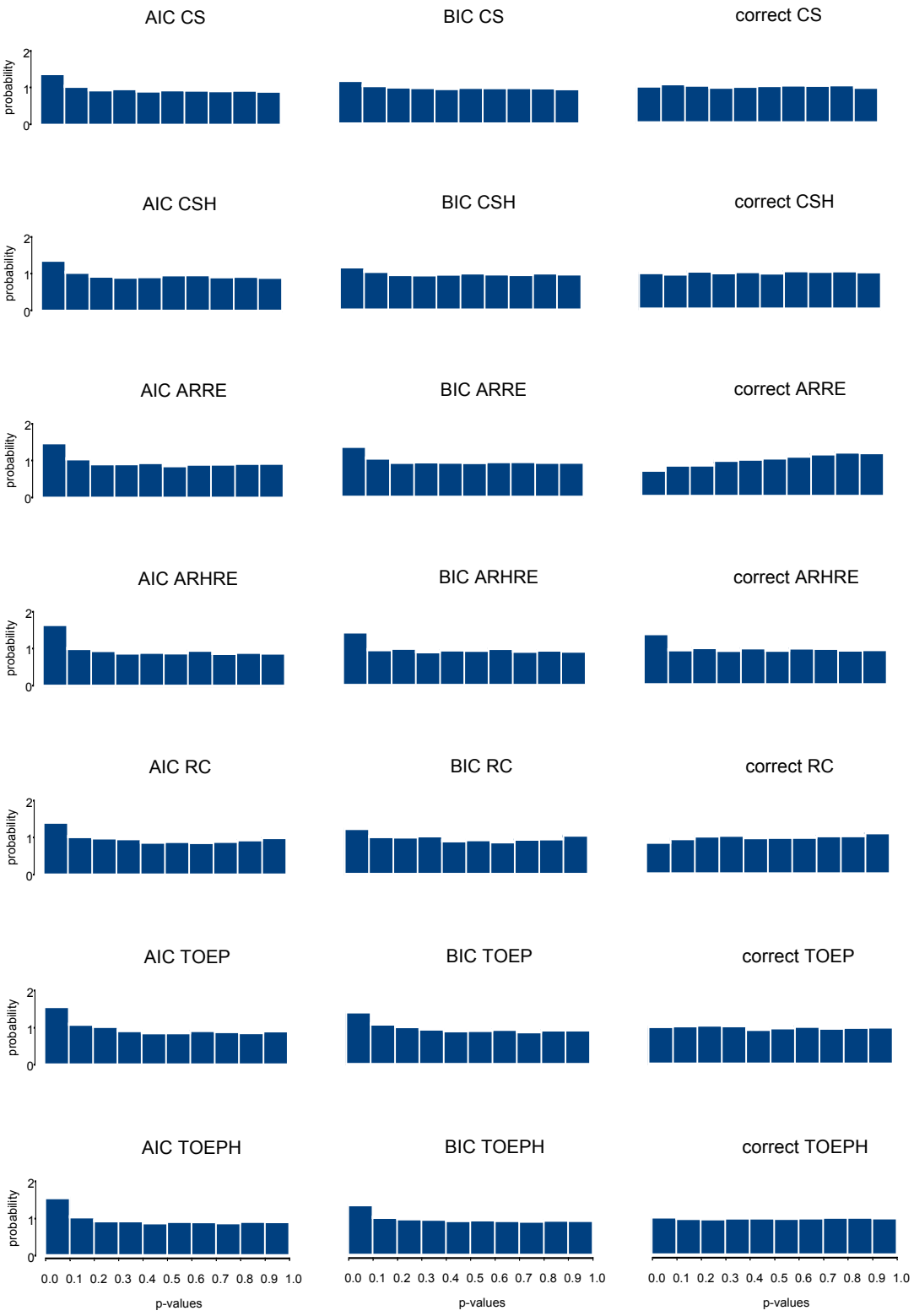


FIGURE 10 (Cont): HISTOGRAMS OF P-VALUES FOR THE TIME EFFECT -(3, 5, 7)x3

FIGURE 10 (Cont): HISTOGRAMS OF P-VALUES FOR THE TIME EFFECT -(3, 5, 7)x3

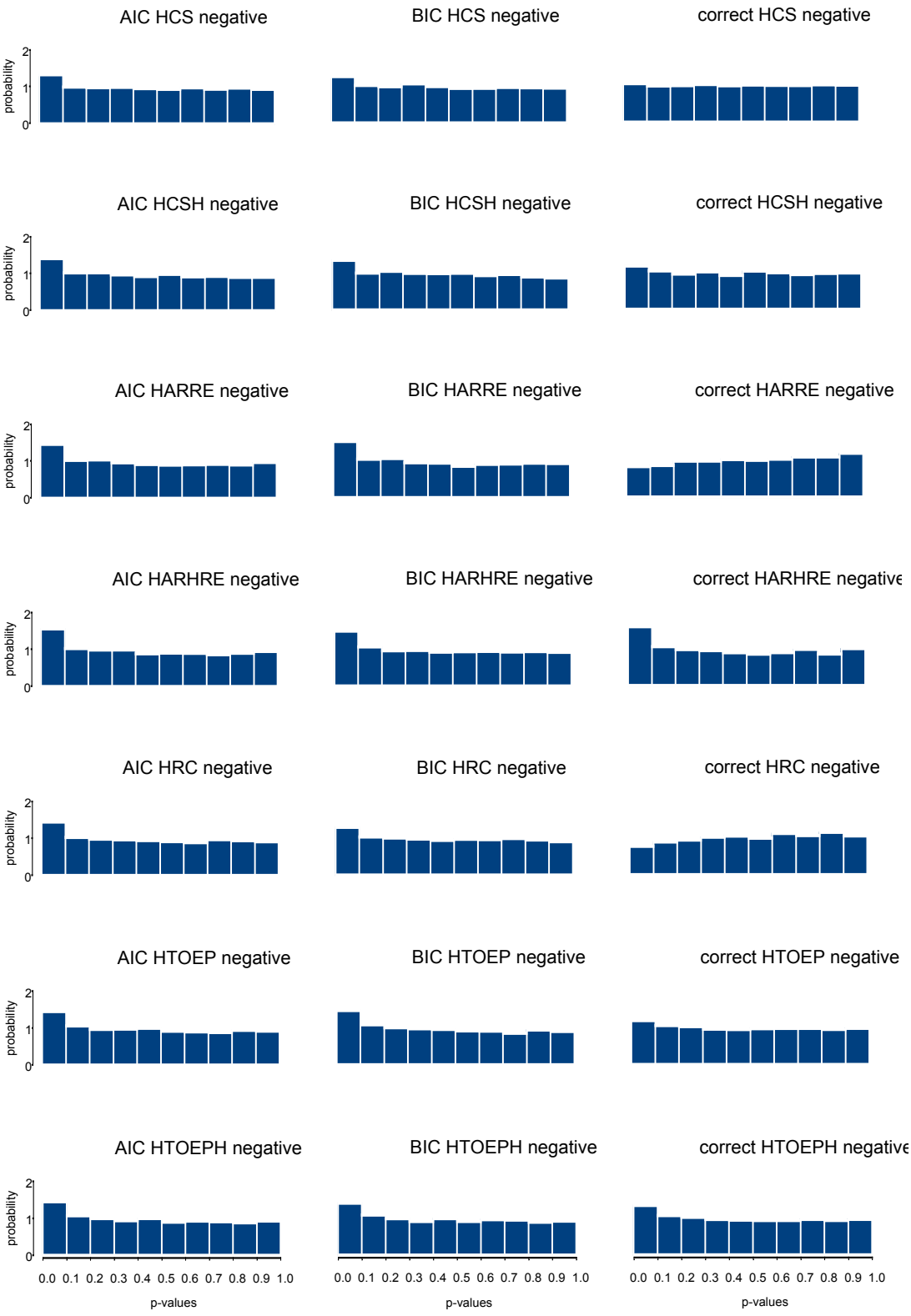


FIGURE 11: HISTOGRAMS OF P-VALUES FOR THE TREATMENT EFFECT -(3, 5, 7)x5

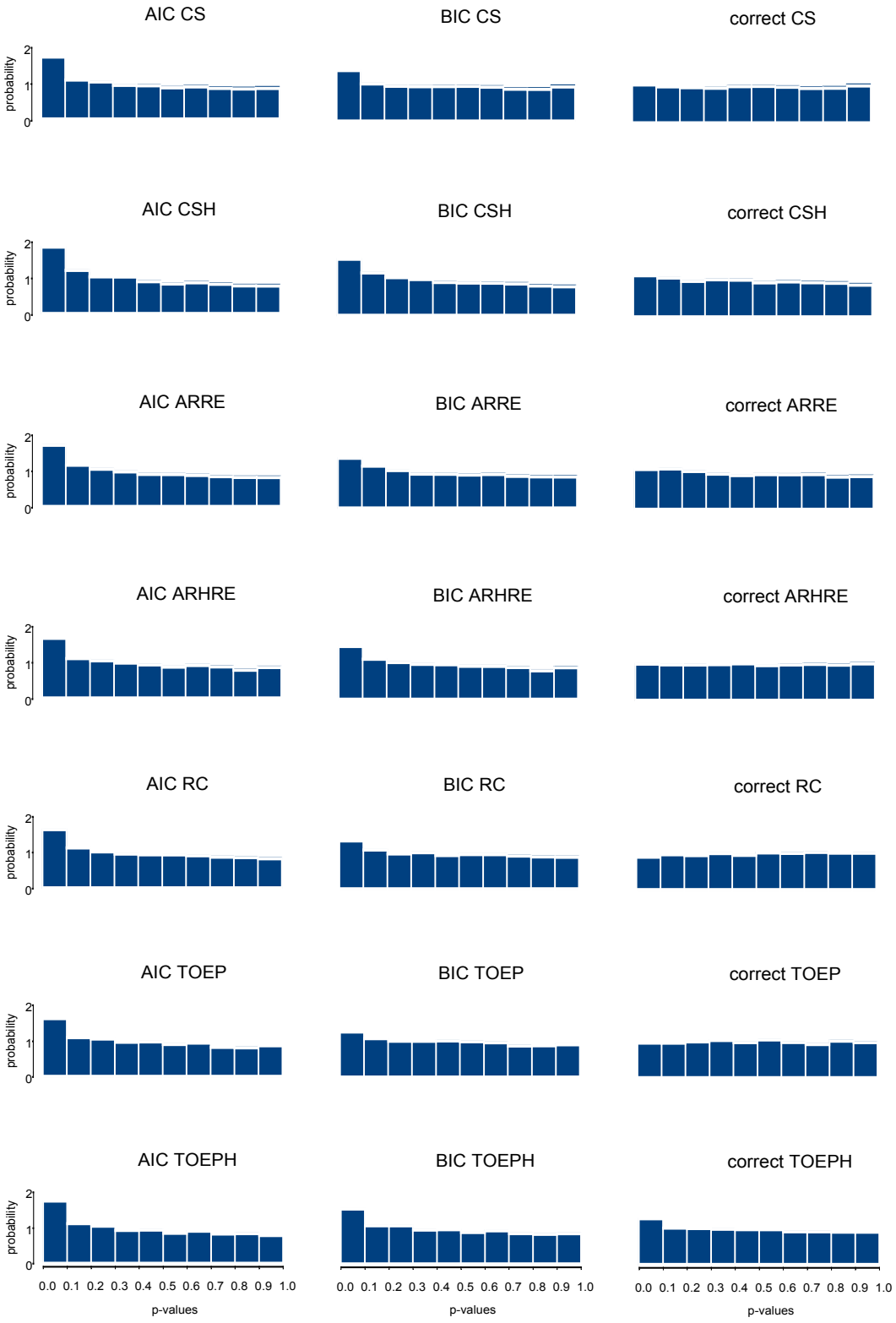


FIGURE 11 (Cont): HISTOGRAMS OF P-VALUES FOR THE TREATMENT EFFECT -(3, 5, 7)x5

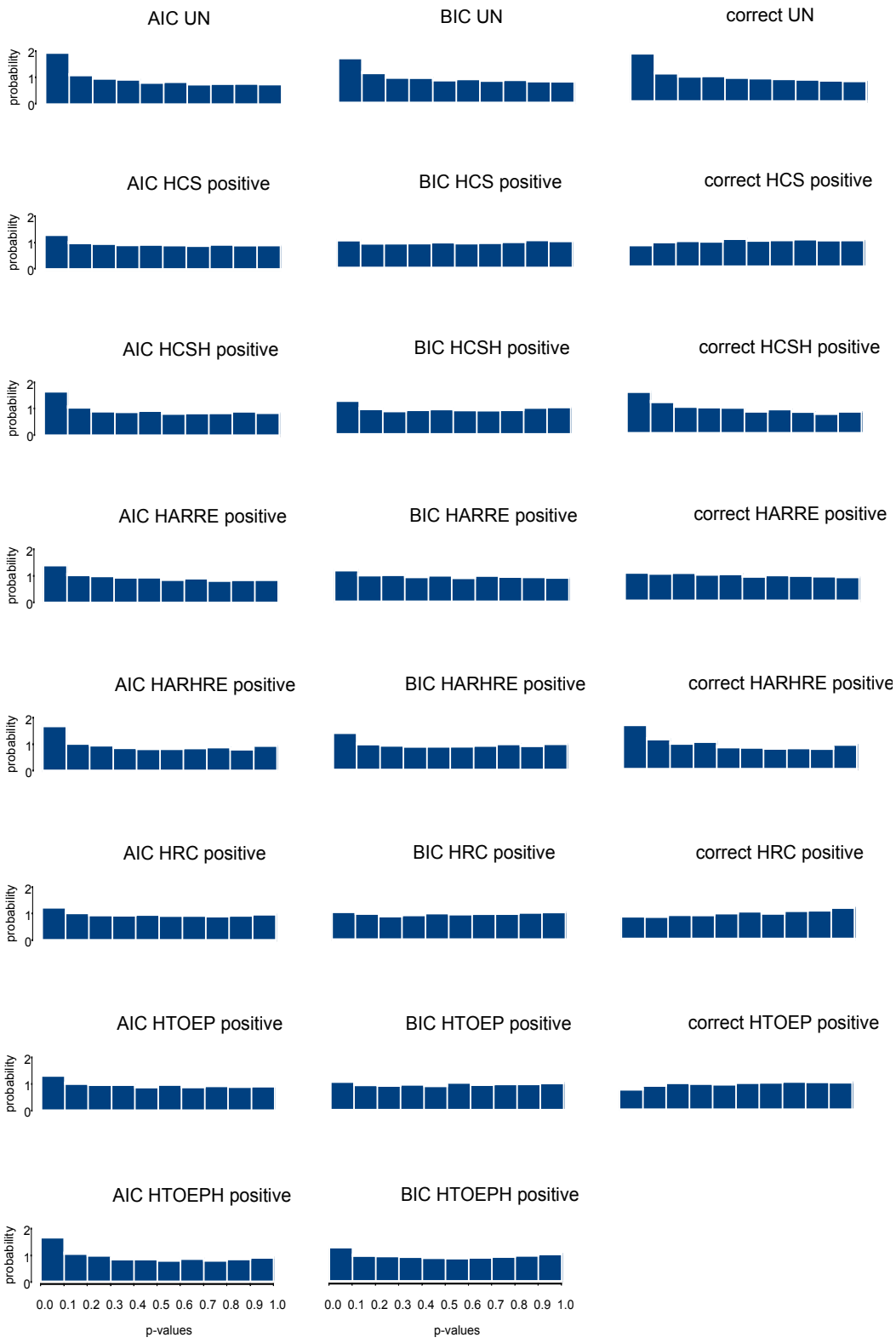


FIGURE 11 (Cont): HISTOGRAMS OF P-VALUES FOR THE TREATMENT EFFECT -(3, 5, 7)x5

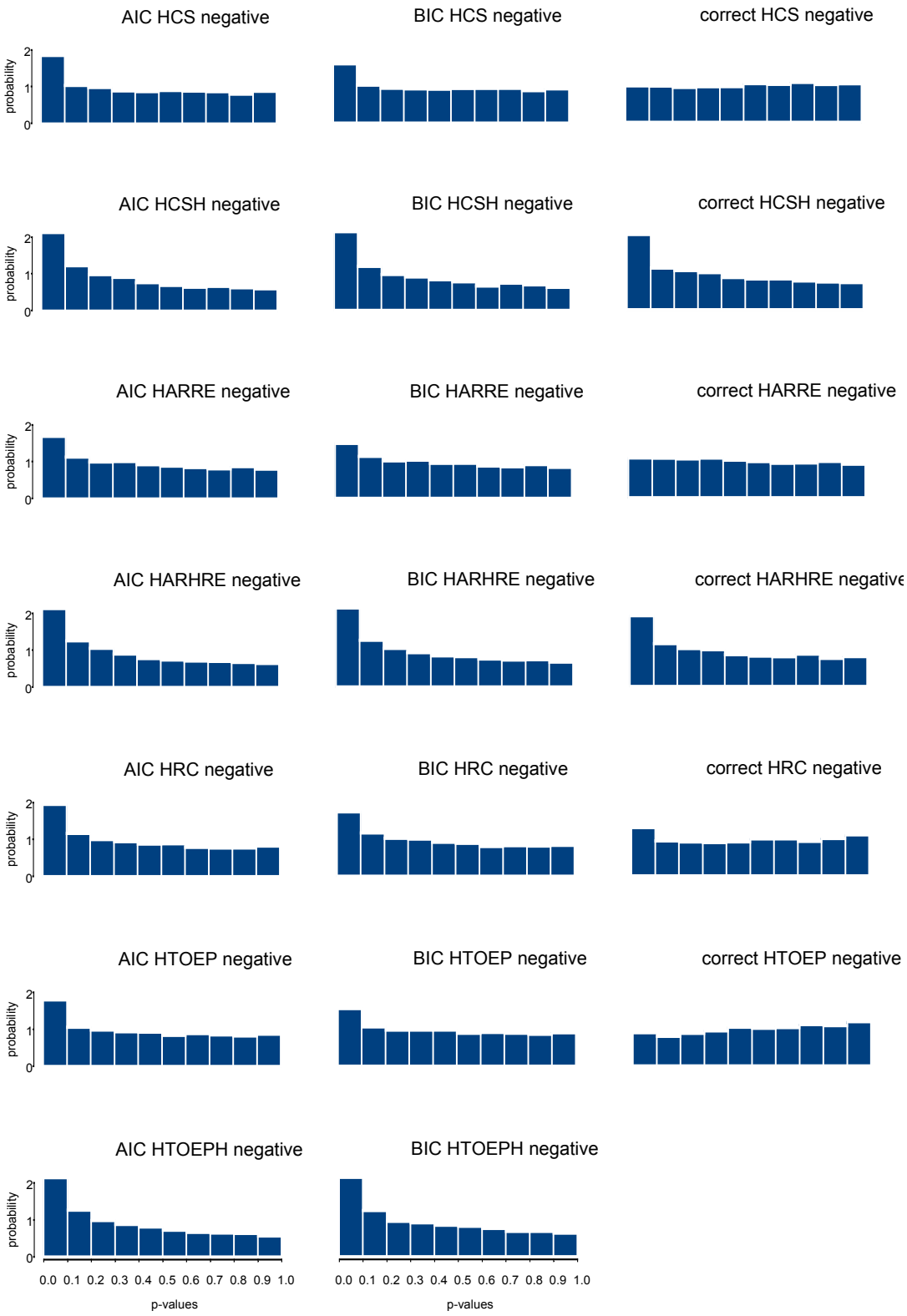


FIGURE 12: HISTOGRAMS OF P-VALUES FOR THE TIME EFFECT -(3, 5, 7)x5

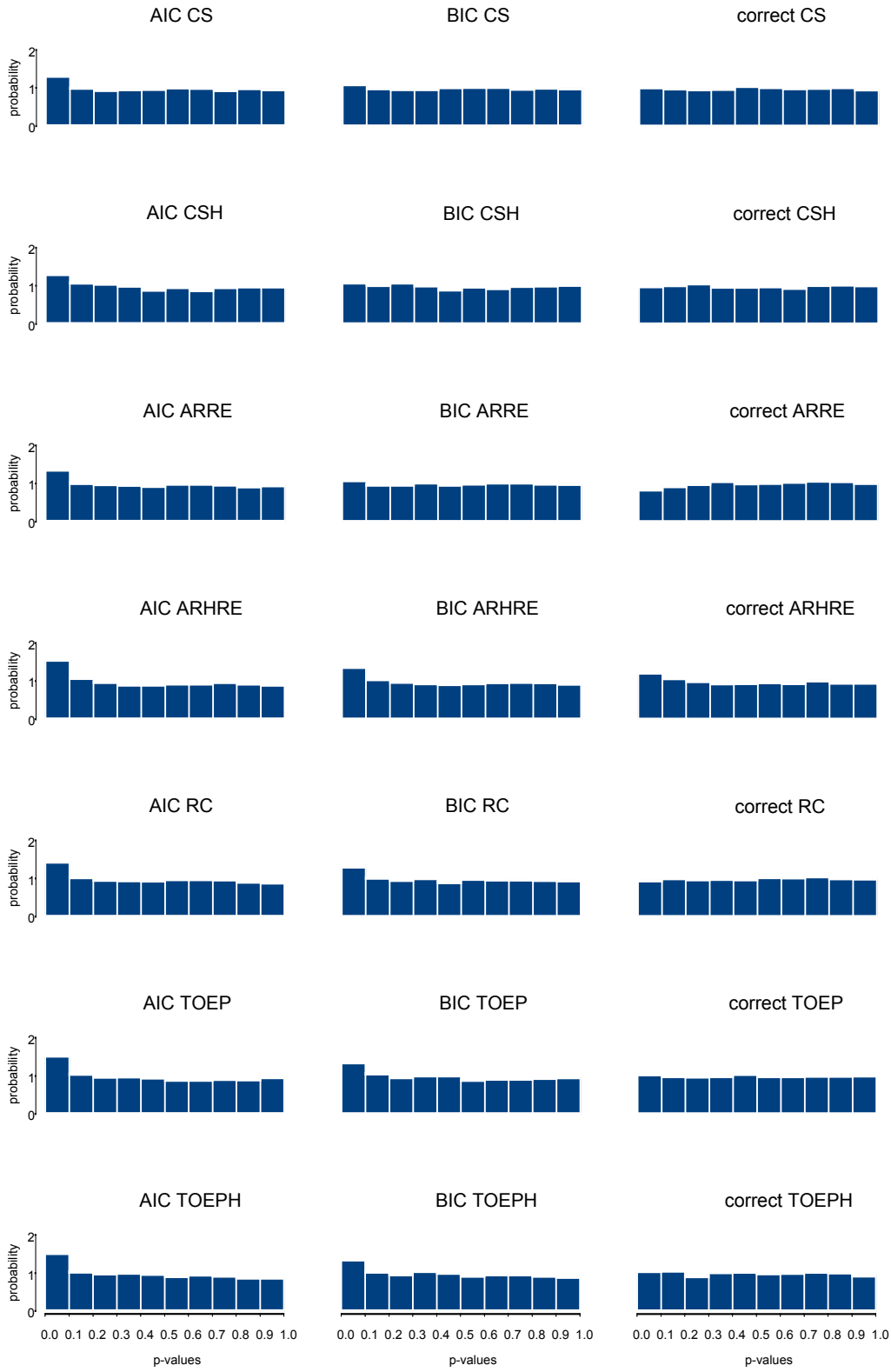
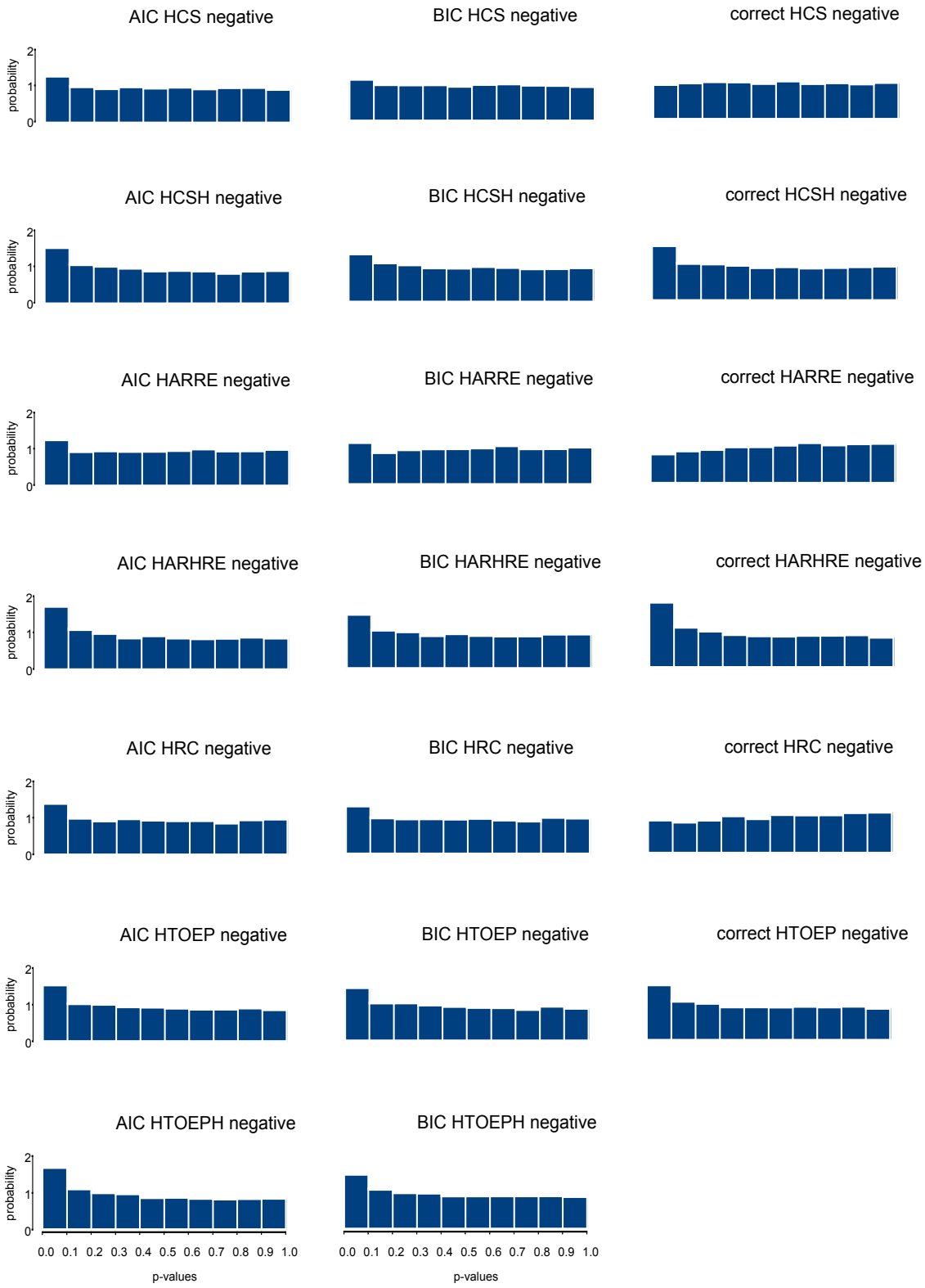


FIGURE 12 (Cont): HISTOGRAMS OF P-VALUES FOR THE TIME EFFECT -(3, 5, 7)x5

FIGURE 12 (Cont): HISTOGRAMS OF P-VALUES FOR THE TIME EFFECT - (3, 5, 7)x5



APPENDIX 2: Example code

```
/*****************************************************************
Generating data following the 8 covariance structures with
homogeneity between-treatments.
*****************************************************************/

proc iml;
seed=7;
nit=10000;
dat=j(1,12);
cs={1 .5 .5,
    .5 1 .5,
    .5 .5 1};
r_cs=I(9)@cs;

csh={1  .84  1.10,
   .84  2.81 1.84,
    1.10 1.84 4.80};
r_csh=I(9)@csh;

arre={1 .78 .62,
      .78  1 .78,
      .62 .78 1};
r_arre=I(9)@arre;

arhre={1  1.22  1.16,
    1.22  2.81 2.64,
    1.16 2.64 4.80};
r_arhre=I(9)@arhre;

rc={1 0, 1 1, 1 2}*{.3 -.03, -.03  .2}*{1 1 1 , 0 1 2}+ 2*i(3);
r_rc=I(9)@rc;

toe={1 .5 .3,
     .5 1 .5,
     .3 .5 1};
r_toe=I(9)@toe;

toeh={1 .84 .66,
   .84 2.81 1.84,
   .66 1.84 4.80};
r_toeh=I(9)@toeh;

un={1 .92 .95,
   .92 2.81 1.85,
   .95 1.85 4.80};
r_un=I(9)@un;

*generating data;
do iter=1 to nit;
 iter2=j(27,1,iter);
 s={1, 2, 3, 4, 5, 6, 7, 8, 9};
 subject=s@j(3,1,1);
 trt=j(9,1,1)//j(9,1,2)//j(9,1,3);
 t={0,1,2};
 time=j(9,1,1)@t;
```

```
 y1=normal(repeat(seed,27,1));
 y2=normal(repeat(seed,27,1));
 y3=normal(repeat(seed,27,1));
 y4=normal(repeat(seed,27,1));
 y5=normal(repeat(seed,27,1));
 y6=normal(repeat(seed,27,1));
 y7=normal(repeat(seed,27,1));
 y8=normal(repeat(seed,27,1));
 response_cs=t(root(r_cs))*y1;
 response_csh=t(root(r_csh))*y2;
 response_arre=t(root(r_arre))*y3;
 response_arhre=t(root(r_arhre))*y4;
 response_rc=t(root(r_rc))*y5;
 response_toe=t(root(r_toe))*y6;
 response_toeh=t(root(r_toeh))*y7;
 response_un=t(root(r_un))*y8;


dat=dat//(iter2||subject||trt||time||response_cs||response_csh||respo
nse_arre||response_arhre||response_rc||response_toe
           ||response_toeh||response_un);
end;
tot=nit*3*3*3+1;
dat=dat[2:tot,];
create one from dat;
append from dat;
quit iml;
run;


data one (rename=(col1=iter col2=subject col3=trt col4=time
col5=response_1 col6=response_2 col7=response_3
          col8=response_4 col9=response_5 col10=response_6
col11=response_7 col12=response_8));
set one;
run;

/*****************************************************************
Macro to fit every data set with all the considered models
*****************************************************************/


%macro names;

/*initializing the data set from where the results will be obtained*/

data cs1;aic=.; bic=.; model_type='        '; covstruct=' '; iter=.;
output; run;
data cs; iter=.; descr='       '; value=.; output; run;
data csh1;aic=.; bic=.; model_type='        '; covstruct=' '; iter=.;
output; run;
data csh; iter=.; descr='       '; value=.; output; run;
data arre1;aic=.; bic=.; model_type='        '; covstruct=' ';
iter=.; output; run;
data arre; iter=.; descr='       '; value=.; output; run;
```

```
data arhre1;aic=.; bic=.; model_type='        '; covstruct=' ';
iter=.; output; run;
data arhre; iter=.; descr='       '; value=.; output; run;
data rc1;aic=.; bic=.; model_type='        '; covstruct=' '; iter=.;
output; run;
data rc; iter=.; descr='       '; value=.; output; run;
data toe1;aic=.; bic=.; model_type='        '; covstruct=' '; iter=.;
output; run;
data toep; iter=.; descr='       '; value=.; output; run;
data toeh1;aic=.; bic=.; model_type='        '; covstruct=' ';
iter=.; output; run;
data toeph; iter=.; descr='       '; value=.; output; run;
data un1;aic=.; bic=.; model_type='        '; covstruct=' '; iter=.;
output; run;
data un; iter=.; descr='       '; value=.; output; run;
data hcs1;aic=.; bic=.; model_type='        '; covstruct=' '; iter=.;
output; run;
data hcs; iter=.; descr='       '; value=.; output; run;
data hcsh1;aic=.; bic=.; model_type='        '; covstruct=' ';
iter=.; output; run;
data hcsh; iter=.; descr='       '; value=.; output; run;
data harre1;aic=.; bic=.; model_type='        '; covstruct=' ';
iter=.; output; run;
data harre; iter=.; descr='       '; value=.; output; run;
data harhre1;aic=.; bic=.; model_type='        '; covstruct=' ';
iter=.; output; run;
data harhre; iter=.; descr='       '; value=.; output; run;
data hrc1;aic=.; bic=.; model_type='        '; covstruct=' '; iter=.;
output; run;
data hrc; iter=.; descr='       '; value=.; output; run;
data htoe1;aic=.; bic=.; model_type='        '; covstruct=' ';
iter=.; output; run;
data htoep; iter=.; descr='       '; value=.; output; run;
data htoeh1; aic=.; bic=.; model_type='        '; covstruct=' ';
iter=.; output; run;
data htoeph; iter=.; descr='       '; value=.; output; run;
data pval_cs1;effect='    '; NumDF=.; DenDF=.; Fvalue=.; probf=.;
model_type='        '; covstruct=' '; iter=.; output; run;
data pval_cs; iter=.; effect='      '; NumDF=.; DenDF=.; Fvalue=.;
probf=.; output;run;
data pval_csh1;effect='    '; NumDF=.; DenDF=.; Fvalue=.; probf=.;
model_type='        '; covstruct=' '; iter=.; output; run;
data pval_csh; iter=.; effect='      '; NumDF=.; DenDF=.; Fvalue=.;
probf=.; output;run;
data pval_arre1;effect='    '; NumDF=.; DenDF=.; Fvalue=.; probf=.;
model_type='        '; covstruct=' '; iter=.; output; run;
data pval_arre; iter=.; effect='      '; NumDF=.; DenDF=.; Fvalue=.;
probf=.; output;run;
data pval_arhre1;effect='    '; NumDF=.; DenDF=.; Fvalue=.; probf=.;
model_type='        '; covstruct=' '; iter=.; output; run;
data pval_arhre; iter=.; effect='      '; NumDF=.; DenDF=.; Fvalue=.;
probf=.; output;run;
data pval_rc1;effect='    '; NumDF=.; DenDF=.; Fvalue=.; probf=.;
model_type='        '; covstruct=' '; iter=.; output; run;
```

```
data pval_rc; iter=.; effect='      '; NumDF=.; DenDF=.; Fvalue=.;
probf=.; output;run;
data pval_toe1;effect='    '; NumDF=.; DenDF=.; Fvalue=.; probf=.;
model_type='        '; covstruct=' '; iter=.; output; run;
data pval_toep; iter=.; effect='      '; NumDF=.; DenDF=.; Fvalue=.;
probf=.; output;run;
data pval_toeh1;effect='    '; NumDF=.; DenDF=.; Fvalue=.; probf=.;
model_type='        '; covstruct=' '; iter=.; output; run;
data pval_toeph; iter=.; effect='      '; NumDF=.; DenDF=.; Fvalue=.;
probf=.; output;run;
data pval_un1;effect='    '; NumDF=.; DenDF=.; Fvalue=.; probf=.;
model_type='        '; covstruct=' '; iter=.; output; run;
data pval_un; iter=.; effect='      '; NumDF=.; DenDF=.; Fvalue=.;
probf=.; output;run;
data pval_hcs1;effect='    '; NumDF=.; DenDF=.; Fvalue=.; probf=.;
model_type='        '; covstruct=' '; iter=.; output; run;
data pval_hcs; iter=.; effect='      '; NumDF=.; DenDF=.; Fvalue=.;
probf=.; output;run;
data pval_hcsh1;effect='    '; NumDF=.; DenDF=.; Fvalue=.; probf=.;
model_type='        '; covstruct=' '; iter=.; output; run;
data pval_hcsh; iter=.; effect='      '; NumDF=.; DenDF=.; Fvalue=.;
probf=.; output;run;
data pval_harre1;effect='    '; NumDF=.; DenDF=.; Fvalue=.; probf=.;
model_type='        '; covstruct=' '; iter=.; output; run;
data pval_harre; iter=.; effect='      '; NumDF=.; DenDF=.; Fvalue=.;
probf=.; output;run;
data pval_harhre1;effect='    '; NumDF=.; DenDF=.; Fvalue=.; probf=.;
model_type='        '; covstruct=' '; iter=.; output; run;
data pval_harhre; iter=.; effect='      '; NumDF=.; DenDF=.;
Fvalue=.; probf=.; output;run;
data pval_hrc1;effect='    '; NumDF=.; DenDF=.; Fvalue=.; probf=.;
model_type='        '; covstruct=' '; iter=.; output; run;
data pval_hrc; iter=.; effect='      '; NumDF=.; DenDF=.; Fvalue=.;
probf=.; output;run;
data pval_htoe1;effect='    '; NumDF=.; DenDF=.; Fvalue=.; probf=.;
model_type='        '; covstruct=' '; iter=.; output; run;
data pval_htoep; iter=.; effect='      '; NumDF=.; DenDF=.; Fvalue=.;
probf=.; output;run;
data pval_htoeh1; effect='    '; NumDF=.; DenDF=.; Fvalue=.; probf=.;
model_type='        '; covstruct=' '; iter=.; output; run;
data pval_htoeph; iter=.; effect='      '; NumDF=.; DenDF=.;
Fvalue=.; probf=.; output;run;

/*analyzing every data sets with all the considered models and using
ods to obtain the p-values and AIC and BIC values for each model*/

%do j=1 %to 8;

ods listing close;

proc mixed data=one;
by iter;
class trt subject;
parms (.5) (.5);
model response_&j= trt time /ddfm=kenwardroger;
```

```
repeated /type=cs subject=subject;
ods output fitstatistics=cs;
ods output tests3=pval_cs;
run;

ods listing;

data cs2;set cs; retain aic bic; keep aic bic model_type covstruct
iter; model_type='CS    ';covstruct="&j";
      if descr='AIC (smaller is better)' then do; aic=value;
bic=.;end;
      if descr='BIC (smaller is better)' then do; bic=value; output;
end; run;

data cs1;
set cs1 cs2;run;

data pval_cs2; set pval_cs; model_type='CS    ';covstruct="&j"; keep
effect probf model_type
covstruct iter;

data pval_cs1;
set pval_cs1 pval_cs2;run;


ods listing close;

proc mixed data=one;
by iter;
class trt subject;
parms (1) (2.81) (4.80) (.5);
model response_&j =trt time /ddfm=kenwardroger;
repeated /type=csh subject=subject;
ods output fitstatistics=csh;
ods output tests3=pval_csh;
run;

ods listing;

data csh2;set csh;retain aic bic;keep aic bic model_type covstruct
iter;model_type='CSH   ';covstruct="&j";
     if descr='AIC (smaller is better)' then do; aic=value;
bic=.;end;
     if descr='BIC (smaller is better)' then
do;bic=value;output;end;run;
data csh1;
set csh1 csh2; run;

data pval_csh2; set pval_csh; model_type='CSH   ';covstruct="&j";keep
effect probf model_type
covstruct iter;
data pval_csh1;
set pval_csh1 pval_csh2; run;

ods listing close;
```

```
proc mixed data=one;
by iter;
class trt subject;
parms (.25) (.7) (.75);
model response_&j= trt time /ddfm=kenwardroger;
repeated /type=ar(1) subject=subject;
random subject;
ods output fitstatistics=arre;
ods output tests3=pval_arre;
run;

ods listing;

data arre2;set arre;retain aic bic;keep aic bic model_type covstruct
iter;model_type='ARRE  ';covstruct="&j";
       if descr='AIC (smaller is better)' then do; aic=value;
bic=.;end;
       if descr='BIC (smaller is better)' then
do;bic=value;output;end;run;
data arre1;
set arre1 arre2; run;

data pval_arre2; set pval_arre; model_type='ARRE
';covstruct="&j";keep effect probf model_type
covstruct iter;
data pval_arre1;
set pval_arre1 pval_arre2; run;

ods listing close;

proc mixed data=one;
by iter;
class trt subject;
parms (.25) (.75) (2.56) (4.55) (.7);
model response_&j= trt time /ddfm=kenwardroger;
repeated /type=arh(1) subject=subject;
random subject;
ods output fitstatistics=arhre;
ods output tests3=pval_arhre;
run;

ods listing;

data arhre2;set arhre;retain aic bic;keep aic bic model_type
covstruct iter;model_type='ARHRE ';covstruct="&j";
     if descr='AIC (smaller is better)' then do; aic=value;
bic=.;end;
     if descr='BIC (smaller is better)' then
do;bic=value;output;end;run;
data arhre1;
set arhre1 arhre2; run;

data pval_arhre2; set pval_arhre; model_type='ARHRE
';covstruct="&j";keep effect probf model_type
```

```
covstruct iter;
data pval_arhre1;
set pval_arhre1 pval_arhre2; run;

ods listing close;

proc mixed data=one;
by iter;
class trt subject;
parms (.3) (-.03) (.2) (2);
model response_&j= trt time /ddfm=kenwardroger;
random intercept time /type=un subject=subject;
ods output fitstatistics=rc;
ods output tests3=pval_rc;
run;

ods listing;

data rc2;set rc;retain aic bic;keep aic bic model_type covstruct
iter;model_type='RC    ';covstruct="&j";
     if descr='AIC (smaller is better)' then do; aic=value;
bic=.;end;
     if descr='BIC (smaller is better)' then
do;bic=value;output;end;run;
data rc1;
set rc1 rc2; run;

data pval_rc2; set pval_rc; model_type='RC    ';covstruct="&j";keep
effect probf model_type
covstruct iter;
data pval_rc1;
set pval_rc1 pval_rc2; run;

ods listing close;

proc mixed data=one;
by iter;
class trt subject;
parms (.5) (.3) (1);
model response_&j= trt time /ddfm=kenwardroger;
repeated /type=toep subject=subject;
ods output fitstatistics=toep;
ods output tests3=pval_toep;
run;

ods listing;

data toe2;set toep;retain aic bic;keep aic bic model_type covstruct
iter;model_type='TOEP  ';covstruct="&j";
      if descr='AIC (smaller is better)' then do; aic=value;
bic=.;end;
      if descr='BIC (smaller is better)' then
do;bic=value;output;end;run;
data toe1;
set toe1 toe2; run;
```

```
data pval_toe2; set pval_toep; model_type='TOEP
';covstruct="&j";keep effect probf model_type
covstruct iter;
data pval_toe1;
set pval_toe1 pval_toe2; run;

ods listing close;

proc mixed data=one;
by iter;
class trt subject;
parms (1) (2.81) (4.80) (.5) (.3);
model response_&j= trt time /ddfm=kenwardroger;
repeated /type=toeph subject=subject;
ods output fitstatistics=toeph;
ods output tests3=pval_toeph;
run;

ods listing;

data toeh2;set toeph;retain aic bic;keep aic bic model_type covstruct
iter;model_type='TOEPH ';covstruct="&j";
     if descr='AIC (smaller is better)' then do; aic=value;
bic=.;end;
     if descr='BIC (smaller is better)' then
do;bic=value;output;end;run;
data toeh1;
set toeh1 toeh2; run;

data pval_toeh2; set pval_toeph; model_type='TOEPH
';covstruct="&j";keep effect probf model_type
covstruct iter;
data pval_toeh1;
set pval_toeh1 pval_toeh2; run;

ods listing close;

proc mixed data=one;
by iter;
class trt subject;
parms (1) (.92) (2.81) (.95) (1.85) (4.80);
model response_&j= trt time /ddfm=kenwardroger;
repeated /type=un subject=subject;
ods output fitstatistics=un;
ods output tests3=pval_un;
run;

ods listing;

data un2;set un;retain aic bic;keep aic bic model_type covstruct
iter;model_type='UN    ';covstruct="&j";
     if descr='AIC (smaller is better)' then do; aic=value;
bic=.;end;
```

```
     if descr='BIC (smaller is better)' then
do;bic=value;output;end;run;
data un1;
set un1 un2; run;

data pval_un2; set pval_un; model_type='UN    ';covstruct="&j";keep
effect probf model_type
covstruct iter;
data pval_un1;
set pval_un1 pval_un2; run;

ods listing close;

proc mixed data=one;
by iter;
class trt subject;
parms (.5) (.5) (.5) (.5) (.5) (.5);
model response_&j= trt time /ddfm=kenwardroger;
repeated /type=cs subject=subject group=trt;
ods output fitstatistics=hcs;
ods output tests3=pval_hcs;
run;

ods listing;

data hcs2;set hcs; retain aic bic;keep aic bic model_type covstruct
iter;model_type='HCS   ';covstruct="&j";
      if descr='AIC (smaller is better)' then do; aic=value;
bic=.;end;
      if descr='BIC (smaller is better)' then do; bic=value; output;
end; run;
data hcs1;
set hcs1 hcs2; run;

data pval_hcs2; set pval_hcs; model_type='HCS   ';covstruct="&j";
keep effect probf model_type
covstruct iter;
data pval_hcs1;
set pval_hcs1 pval_hcs2; run;

ods listing close;

proc mixed data=one;
by iter;
class trt subject;
parms (1) (2.81) (4.80) (.5) (1) (2.81) (4.80) (.5) (1) (2.81) (4.80)
(.5);
model response_&j =trt time /ddfm=kenwardroger;
repeated /type=csh subject=subject group=trt;
ods output fitstatistics=hcsh;
ods output tests3=pval_hcsh;
run;

ods listing;
```

```
data hcsh2;set hcsh;retain aic bic;keep aic bic model_type covstruct
iter;model_type='HCSH  ';covstruct="&j";
     if descr='AIC (smaller is better)' then do; aic=value;
bic=.;end;
     if descr='BIC (smaller is better)' then
do;bic=value;output;end;run;
data hcsh1;
set hcsh1 hcsh2; run;

data pval_hcsh2; set pval_hcsh; model_type='HCSH
';covstruct="&j";keep effect probf model_type
covstruct iter;
data pval_hcsh1;
set pval_hcsh1 pval_hcsh2; run;


ods listing close;

proc mixed data=one;
by iter;
class trt subject;
parms (.25) (.75) (.7) (.75) (.7) (.75) (.7);
model response &j= trt time /ddfm=kenwardroger;
repeated /type=ar(1) subject=subject group=trt;
random subject;
ods output fitstatistics=harre;
ods output tests3=pval_harre;
run;

ods listing;

data harre2;set harre;retain aic bic;keep aic bic model_type
covstruct iter;model_type='HARRE ';covstruct="&j";
       if descr='AIC (smaller is better)' then do; aic=value;
bic=.;end;
       if descr='BIC (smaller is better)' then
do;bic=value;output;end;run;
data harre1;
set harre1 harre2; run;

data pval_harre2; set pval_harre; model_type='HARRE
';covstruct="&j";keep effect probf model_type
covstruct iter;
data pval_harre1;
set pval_harre1 pval_harre2; run;

ods listing close;

proc mixed data=one;
by iter;
class trt subject;
parms (.25) (.75) (2.56) (4.55) (.7)
            (.75) (2.56) (4.55) (.7)
            (.75) (2.56) (4.55) (.7);
model response_&j= trt time /ddfm=kenwardroger;
```

```
repeated /type=arh(1) subject=subject group=trt;
random subject;
ods output fitstatistics=harhre;
ods output tests3=pval_harhre;
run;

ods listing;

data harhre2;set harhre;retain aic bic;keep aic bic model_type
covstruct iter;model_type='HARHRE';covstruct="&j";
     if descr='AIC (smaller is better)' then do; aic=value;
bic=.;end;
     if descr='BIC (smaller is better)' then
do;bic=value;output;end;run;
data harhre1;
set harhre1 harhre2; run;

data pval_harhre2; set pval_harhre;
model_type='HARHRE';covstruct="&j";keep effect probf model_type
covstruct iter;
data pval_harhre1;
set pval_harhre1 pval_harhre2; run;

ods listing close;

proc mixed data=one;
by iter;
class trt subject;
parms (.3) (-.03) (.2) (.3) (-.03) (.2) (.3) (-.03) (.2) (2);
model response_&j= trt time /ddfm=kenwardroger;
random intercept time /type=un subject=subject group=trt;
ods output fitstatistics=hrc;
ods output tests3=pval_hrc;
run;

ods listing;

data hrc2;set hrc;retain aic bic;keep aic bic model_type covstruct
iter;model_type='HRC   ';covstruct="&j";
     if descr='AIC (smaller is better)' then do; aic=value;
bic=.;end;
     if descr='BIC (smaller is better)' then
do;bic=value;output;end;run;
data hrc1;
set hrc1 hrc2; run;

data pval_hrc2; set pval_hrc; model_type='HRC   ';covstruct="&j";keep
effect probf model_type
covstruct iter;
data pval_hrc1;
set pval_hrc1 pval_hrc2; run;

ods listing close;

proc mixed data=one;
```

```
by iter;
class trt subject;
parms (1) (.5) (.3) (1) (.5) (.3) (1) (.5) (.3);
model response_&j= trt time /ddfm=kenwardroger;
repeated /type=toep subject=subject group=trt;
ods output fitstatistics=htoep;
ods output tests3=pval_htoep;
run;

ods listing;

data htoe2;set htoep;retain aic bic;keep aic bic model_type covstruct
iter;model_type='HTOEP ';covstruct="&j";
      if descr='AIC (smaller is better)' then do; aic=value;
bic=.;end;
      if descr='BIC (smaller is better)' then
do;bic=value;output;end;run;
data htoe1;
set htoe1 htoe2; run;

data pval_htoe2; set pval_htoep; model_type='HTOEP
';covstruct="&j";keep effect probf model_type
covstruct iter;
data pval_htoe1;
set pval_htoe1 pval_htoe2; run;

ods listing close;

proc mixed data=one;
by iter;
class trt subject;
parms (1) (2.81) (4.80) (.5) (.3) (1) (2.81) (4.80) (.5) (.3) (1)
(2.81) (4.80) (.5) (.3);
model response_&j= trt time /ddfm=kenwardroger;
repeated /type=toeph subject=subject group=trt;
ods output fitstatistics=htoeph;
ods output tests3=pval_htoeph;
run;

ods listing;

data htoeh2;set htoeph;retain aic bic;keep aic bic model_type
covstruct iter;model_type='HTOEPH';covstruct="&j";
     if descr='AIC (smaller is better)' then do; aic=value;
bic=.;end;
     if descr='BIC (smaller is better)' then
do;bic=value;output;end;run;
data htoeh1;
set htoeh1 htoeh2; run;

data pval_htoeh2; set pval_htoeph;
model_type='HTOEPH';covstruct="&j";keep effect probf model_type
covstruct iter;
data pval_htoeh1;
set pval_htoeh1 pval_htoeh2; run;
```

```
ods listing close;
ods listing;

/**merging all the resulted data sets together and separating them
depending on the covariance structure used to generate the data**/

data all1 all2 all3 all4 all5 all6 all7 all8;
merge cs1(rename=(  aic=aic1 bic=bic1))
      csh1 (rename=( aic=aic2 bic=bic2))
      arre1 (rename=( aic=aic3 bic=bic3))
      arhre1 (rename=( aic=aic4 bic=bic4))
      rc1 (rename=( aic=aic5 bic=bic5))
      toe1 (rename=( aic=aic6 bic=bic6))
      toeh1 (rename=( aic=aic7 bic=bic7))
      un1 (rename=( aic=aic8 bic=bic8))
      hcs1 (rename=( aic=aic9 bic=bic9))
      hcsh1 (rename=( aic=aic10 bic=bic10))
      harre1 (rename=( aic=aic11 bic=bic11))
      harhre1 (rename=( aic=aic12 bic=bic12))
      hrc1 (rename=( aic=aic13 bic=bic13))
      htoe1 (rename=( aic=aic14 bic=bic14))
      htoeh1(rename=( aic=aic15 bic=bic15));
by covstruct iter;
drop model_type;
if iter ne .;
if covstruct=1 then output all1;
if covstruct=2 then output all2;
if covstruct=3 then output all3;
if covstruct=4 then output all4;
if covstruct=5 then output all5;
if covstruct=6 then output all6;
if covstruct=7 then output all7;
if covstruct=8 then output all8;
run;


data pval_all1 pval_all2 pval_all3 pval_all4 pval_all5 pval_all6
pval_all7 pval_all8 ;
merge pval_cs1 (rename=( probf=probf1))
      pval_csh1 (rename=( probf=probf2))
      pval_arre1 (rename=( probf=probf3))
      pval_arhre1 (rename=( probf=probf4))
      pval_rc1 (rename=( probf=probf5))
      pval_toe1 (rename=( probf=probf6))
      pval_toeh1 (rename=( probf=probf7))
      pval_un1(rename=( probf=probf8))
      pval_hcs1 (rename=( probf=probf9))
      pval_hcsh1 (rename=( probf=probf10))
      pval_harre1 (rename=( probf=probf11))
      pval_harhre1 (rename=( probf=probf12))
      pval_hrc1 (rename=( probf=probf13))
      pval_htoe1 (rename=( probf=probf14))
      pval_htoeh1 (rename=( probf=probf15));
by covstruct iter;
```

```
if covstruct=1 then output pval_all1;
if covstruct=2 then output pval_all2;
if covstruct=3 then output pval_all3;
if covstruct=4 then output pval_all4;
if covstruct=5 then output pval_all5;
if covstruct=6 then output pval_all6;
if covstruct=7 then output pval_all7;
if covstruct=8 then output pval_all8;
run;

/** final data sets for data following CS**/

data dum1;
set all1;
minaic=min(of aic1-aic15);
minbic=min(of bic1-bic15);
array a(15) aic1-aic15;
array b(15) bic1-bic15;
do i=1 to 15;
if a(i)=minaic then aicm=i;
if b(i)=minbic then bicm=i;
end;
keep iter aicm bicm;
run;

data finaltr1 finaltm1;
merge dum1 pval_all1;
by iter;
array p(15) probf1-probf15;

if effect='trt' then do;
var='trt_cs    ';
correctp=probf1;
do i=1 to 15;
 if aicm=i then bestaicp=p(i);
 if bicm=i then bestbicp=p(i);
end;
keep var iter aicm bicm bestaicp bestbicp correctp;
output finaltr1;
end;

if effect='time' then do;
correctp=probf1;
var='time_cs';
do i=1 to 15;
 if aicm=i then bestaicp=p(i);
 if bicm=i then bestbicp=p(i);
end;
keep var iter aicm bicm bestaicp bestbicp correctp;
output finaltm1;
end;
```

```
/** final data sets for data following CSH**/

data dum2;
set all2;
minaic=min(of aic1-aic15);
minbic=min(of bic1-bic15);
array a(15) aic1-aic15;
array b(15) bic1-bic15;
do i=1 to 15;
if a(i)=minaic then aicm=i;
if b(i)=minbic then bicm=i;
end;
keep iter aicm bicm;
run;

data finaltr2 finaltm2;
merge dum2 pval_all2;
by iter;
array p(15) probf1-probf15;

if effect='trt' then do;
correctp=probf2;
var='trt csh';
do i=1 to 15;
 if aicm=i then bestaicp=p(i);
 if bicm=i then bestbicp=p(i);
end;
keep var iter aicm bicm bestaicp bestbicp correctp;
output finaltr2;
end;

if effect='time' then do;
correctp=probf2;
var='tim_csh';
do i=1 to 15;
 if aicm=i then bestaicp=p(i);
 if bicm=i then bestbicp=p(i);
end;
keep var iter aicm bicm bestaicp bestbicp correctp;
output finaltm2;
end;

/** final data sets for data following ARRE**/

data dum3;
set all3;
minaic=min(of aic1-aic15);
minbic=min(of bic1-bic15);
array a(15) aic1-aic15;
array b(15) bic1-bic15;
do i=1 to 15;
if a(i)=minaic then aicm=i;
if b(i)=minbic then bicm=i;
end;
keep iter aicm bicm;
```

```
run;

data finaltr3 finaltm3;
merge dum3 pval_all3;
by iter;
array p(15) probf1-probf15;

if effect='trt' then do;
correctp=probf3;
var='trt_arre';
do i=1 to 15;
 if aicm=i then bestaicp=p(i);
 if bicm=i then bestbicp=p(i);
end;
keep var iter aicm bicm bestaicp bestbicp correctp;
output finaltr3;
end;

if effect='time' then do;
correctp=probf3;
var='tim_arre';
do i=1 to 15;
 if aicm=i then bestaicp=p(i);
 if bicm=i then bestbicp=p(i);
end;
keep var iter aicm bicm bestaicp bestbicp correctp;
output finaltm3;
end;

/** final data sets for data following ARHRE**/
data dum4;
set all4;
minaic=min(of aic1-aic15);
minbic=min(of bic1-bic15);
array a(15) aic1-aic15;
array b(15) bic1-bic15;
do i=1 to 15;
if a(i)=minaic then aicm=i;
if b(i)=minbic then bicm=i;
end;
keep iter aicm bicm;
run;

data finaltr4 finaltm4;
merge dum4 pval_all4;
by iter;
array p(15) probf1-probf15;

if effect='trt' then do;
correctp=probf4;
var='trt_arhre';
do i=1 to 15;
 if aicm=i then bestaicp=p(i);
 if bicm=i then bestbicp=p(i);
end;
```

```
keep var iter aicm bicm bestaicp bestbicp correctp;
output finaltr4;
end;

if effect='time' then do;
correctp=probf4;
var='tim_arhre';
do i=1 to 15;
 if aicm=i then bestaicp=p(i);
 if bicm=i then bestbicp=p(i);
end;
keep var iter aicm bicm bestaicp bestbicp correctp;
output finaltm4;
end;

/** final data sets for data following RC**/

data dum5;
set all5;
minaic=min(of aic1-aic15);
minbic=min(of bic1-bic15);
array a(15) aic1-aic15;
array b(15) bic1-bic15;
do i=1 to 15;
if a(i)=minaic then aicm=i;
if b(i)=minbic then bicm=i;
end;
keep iter aicm bicm;
run;

data finaltr5 finaltm5;
merge dum5 pval_all5;
by iter;
array p(15) probf1-probf15;

if effect='trt' then do;
correctp=probf5;
var='trt_rc';
do i=1 to 15;
 if aicm=i then bestaicp=p(i);
 if bicm=i then bestbicp=p(i);
end;
keep var iter aicm bicm bestaicp bestbicp correctp;
output finaltr5;
end;

if effect='time' then do;
correctp=probf5;
var='tim_rc';
do i=1 to 15;
 if aicm=i then bestaicp=p(i);
 if bicm=i then bestbicp=p(i);
end;
keep var iter aicm bicm bestaicp bestbicp correctp;
output finaltm5;
```

```
end;

/** final data sets for data following TOEP **/

data dum6;
set all6;
minaic=min(of aic1-aic15);
minbic=min(of bic1-bic15);
array a(15) aic1-aic15;
array b(15) bic1-bic15;
do i=1 to 15;
if a(i)=minaic then aicm=i;
if b(i)=minbic then bicm=i;
end;
keep iter aicm bicm;
run;

data finaltr6 finaltm6;
merge dum6 pval_all6;
by iter;
array p(15) probf1-probf15;

if effect='trt' then do;
correctp=probf6;
var='trt_toep';
do i=1 to 15;
 if aicm=i then bestaicp=p(i);
 if bicm=i then bestbicp=p(i);
end;
keep var iter aicm bicm bestaicp bestbicp correctp;
output finaltr6;
end;

if effect='time' then do;
correctp=probf6;
var='tim_toep';
do i=1 to 15;
 if aicm=i then bestaicp=p(i);
 if bicm=i then bestbicp=p(i);
end;
keep var iter aicm bicm bestaicp bestbicp correctp;
output finaltm6;
end;

/** final data sets for data following TOEPH **/

data dum7;
set all7;
minaic=min(of aic1-aic15);
minbic=min(of bic1-bic15);
array a(15) aic1-aic15;
array b(15) bic1-bic15;
do i=1 to 15;
if a(i)=minaic then aicm=i;
if b(i)=minbic then bicm=i;
```

```
end;
keep iter aicm bicm;
run;

data finaltr7 finaltm7;
merge dum7 pval_all7;
by iter;
array p(15) probf1-probf15;

if effect='trt' then do;
correctp=probf7;
var='trt_toeph';
do i=1 to 15;
 if aicm=i then bestaicp=p(i);
 if bicm=i then bestbicp=p(i);
end;
keep var iter aicm bicm bestaicp bestbicp correctp;
output finaltr7;
end;

if effect='time' then do;
correctp=probf7;
var='tim toeph';
do i=1 to 15;
 if aicm=i then bestaicp=p(i);
 if bicm=i then bestbicp=p(i);
end;
keep var iter aicm bicm bestaicp bestbicp correctp;
output finaltm7;
end;

/** final data sets for data following UN **/

data dum8;
set all8;
minaic=min(of aic1-aic15);
minbic=min(of bic1-bic15);
array a(15) aic1-aic15;
array b(15) bic1-bic15;
do i=1 to 15;
if a(i)=minaic then aicm=i;
if b(i)=minbic then bicm=i;
end;
keep iter aicm bicm;
run;

data finaltr8 finaltm8;
merge dum8 pval_all8;
by iter;
array p(15) probf1-probf15;

if effect='trt' then do;
correctp=probf8;
var='trt_un';
do i=1 to 15;
```

```
 if aicm=i then bestaicp=p(i);
 if bicm=i then bestbicp=p(i);
end;
keep  var iter aicm bicm bestaicp bestbicp correctp;
output finaltr8;
end;

if effect='time' then do;
correctp=probf8;
var='tim_un';
do i=1 to 15;
 if aicm=i then bestaicp=p(i);
 if bicm=i then bestbicp=p(i);
end;
keep  var iter aicm bicm bestaicp bestbicp correctp;
output finaltm8;
end;

run;
%end;

%mend names;

%names;


/*****************************************************************
creating external data set with the results
*****************************************************************/
data tr1hom33;
set finaltr1;
file 'tr1hom33';
put var iter aicm bicm bestaicp bestbicp correctp;
run;
data tr2hom33;
set finaltr2;
file 'tr2hom33';
put var iter aicm bicm bestaicp bestbicp correctp;

data tr3hom33;
set finaltr3;
file 'tr3hom33';
put var iter aicm bicm bestaicp bestbicp correctp;
run;
data tr4hom33;
set finaltr4;
file 'tr4hom33';
put var iter aicm bicm bestaicp bestbicp correctp;

data tr5hom33;
set finaltr5;
file 'tr5hom33';
put var iter aicm bicm bestaicp bestbicp correctp;
   ;
data tr6hom33;
```

```
set finaltr6;
file 'tr6hom33';
put var iter aicm bicm bestaicp bestbicp correctp;
run;
data tr7hom33;
set finaltr7;
file 'tr7hom33';
put var iter aicm bicm bestaicp bestbicp correctp;

data tr8hom33;
set finaltr8;
file 'tr8hom33';
put var iter aicm bicm bestaicp bestbicp correctp;
run;
data tm1hom33;
set finaltm1;
file 'tm1hom33';
put var iter aicm bicm bestaicp bestbicp correctp;
run;
data tm2hom33;
set finaltm2;
file 'tm2hom33';
put var iter aicm bicm bestaicp bestbicp correctp;

data tm3hom33;
set finaltm3;
file 'tm3hom33';
put var iter aicm bicm bestaicp bestbicp correctp;
run;
data tm4hom33;
set finaltm4;
file 'tm4hom33';
put var iter aicm bicm bestaicp bestbicp correctp;
run;
data tm5hom33;
set finaltm5;
file 'tm5hom33';
put var iter aicm bicm bestaicp bestbicp correctp;

data tm6hom33;
set finaltm6;
file 'tm6hom33';
put var iter aicm bicm bestaicp bestbicp correctp;
run;
data tm7hom33;
set finaltm7;
file 'tm7hom33';
put var iter aicm bicm bestaicp bestbicp correctp;

data tm8hom33;
set finaltm8;
file 'tm8hom33';
put var iter aicm bicm bestaicp bestbicp correctp;
```

6. REFERENCES

Akaike, H. (1973), "Information theory and an extension of the maximum likelihood principle," In *Second International Symposium on Information Theory*, B. N. Petrov and F. Csaki, eds. Akademiai Kiado, Budapest, 267-281.

Bradley, J. V. (1978), "Robustness?," *British Journal of Mathematical and Statistical Psychology*, 31, 144-152.

Brown, H., and Prescott, R. (1999), *Applied Mixed Models in Medicine*, Wiley, UK.

Cressie, N.A.C. (1993), *Statistics for Spatial Data*, Wiley, New York.

Dawson, K. S., Gennings, C., and Carter, W. H. (1997) " Two graphical Techniques useful in detecting correlation structure in repeated measures data," *The American Statistician*, 51, 275-283.

Ferron, J., Dailey, R., and Yi, Q. (2002), "Effects of misspecifying the first-level error structure in two-level models of change," *Multivariate Behavioral Research*, 37 (3) 379-403.

Grady, J. J., and Helms, R. W. (1995), "Model selection techniques for the covariance-matrix for incomplete longitudinal data," *Statistics in Medicine*, 14 (13) 1397-1416.

Keselman, H. J., Algina, J., Kowalchuk, R. K., and Wolfinger, R. D. (1998), "A comparison of two approaches for selecting covariance structures in the analysis of repeated measurements," *Communications in Statistics: Simulation and Computation*, 27 (3) 591-604.

Keselman, H. J., Algina, J., Kowalchuk, R. K., and Wolfinger, R. D. (1999), "The analysis of repeated measurements: A comparison of mixed-model satterthwaite F test and a nonpooled adjusted degrees of freedom multivariate test," *Communications in Statistics-Theory and Methods*, 28 (12) 2967-2999.

Littell R. C., Rendergast J., and Natarajan, R. (2000), "Modeling covariance structure in the analysis of repeated measures data," *Statistics in Medicine*, 19 (13) 1793-1819.

Schaalje, G. B., McBride, J. B., and Fellingham, G. W. (2001), "Approximations to Distributions of Test Statistics in Complex Mixed Linear Models Using SAS Proc MIXED," Paper 262-26 in Proceedings of the SAS Users Group International 26th Anunual Conference.

Schwarz,G. (1978), "Estimating the dimension of a model," Annals of Statistics, 6, 461-464.

Schaalje, G. B., McBride, J. B., and Fellingham, G. W. (2002), "Adequacy of Approximations to Distributions of Test Statistics in Complex Mixed Linear Models," *Journal of Agricultural, Biological, and Environmental Statistics*, 7, 512-524.

SAS Institute, Inc. (1999), SAS version 8, Online Help, Cary, NC: SAS Institute.

Singer, J. D. (1998), "Using SAS PROC MIXED to fit multilevel models, hierarchical models, and individual growth models," *Journal of Educational and Behavioral Statistics*, 24 (4) 323-325.