The performance of a two-stage analysis of ABAB/BABA crossover trials

Paul Kabaila* and Matthew Vicendese
Department of Mathematics and Statistics, La Trobe University, Victoria 3086, Australia

Received 26 September 2011; revised 30 November 2011; accepted 23 January 2012

Freeman has considered the following two-stage procedure for finding a confidence interval for the treatment difference \( \theta \), using data from an AB/BA crossover trial. In the first stage, a preliminary test of the null hypothesis that the differential carryover is zero is carried out. If this hypothesis is accepted then the confidence interval for \( \theta \) is constructed assuming that the differential carryover is zero. If, on the other hand, this hypothesis is rejected then this confidence interval is constructed using only data from the first period. Freeman has shown that this confidence interval has minimum coverage probability far below nominal. He therefore concludes that this confidence interval should not be used. In the present paper, we analyze the performance of a similar two-stage procedure for an ABAB/BABA crossover trial. This trial differs in very significant ways from an AB/BA crossover trial, including the fact that for an ABAB/BABA crossover trial there is an unbiased estimator of the differential carryover that is unaffected by between-subject variation. Despite these great differences, we arrive at the same conclusion as Freeman. Namely, that the confidence interval resulting from the two-stage procedure should not be used.

Keywords: Crossover trials; Differential carryover; Preliminary hypothesis test; Two-stage procedure.

1 Introduction

Consider a two-treatment two-period crossover trial, with continuous responses. The purpose of this trial is to find a \( 1 - \alpha \) confidence interval for the difference \( \theta \) in the effects of two treatments, labeled A and B. Subjects are randomly allocated to either group 1 or group 2. Subjects in group 1 receive treatment A in the first period and then receive treatment B in the second period. Subjects in group 2 receive treatment B in the first period and then receive treatment A in the second period. This trial is called an AB/BA trial. To deal with the possibility of non-zero differential carryover, it was suggested (starting with Grizzle, 1965, 1974 and endorsed by Hills and Armitage, 1979; Armitage and Hills, 1982) that the following two-stage procedure be used. In the first stage, a preliminary test of the null hypothesis that the differential carryover is zero (against the alternative that it is non-zero) is carried out. If this null hypothesis is accepted then the confidence interval for \( \theta \) is constructed to have nominal coverage \( 1 - \alpha \), assuming that there is no differential carryover. If, on the other hand, this null hypothesis is rejected then this confidence interval is constructed using only data from the first period (since this is unaffected by carryover). As pointed out by Freeman (1989), accepting this null hypothesis is not equivalent to concluding that the differential carryover is exactly zero. Freeman (1989) shows that the confidence interval resulting from this two-stage procedure has minimum coverage probability far below \( 1 - \alpha \), demonstrating that this confidence interval should not be used. Senn (2006) states

*Corresponding author: e-mail: P.Kabaila@latrobe.edu.au, Phone: +61-03-9479-2594, Fax: +61-03-9479-2466

© 2012 WILEY-VCH Verlag GmbH & Co. KGaA, Weinheim
In my opinion the most important paper on cross-over trials in the 25 years of *Statistics in Medicine* is Peter Freeman’s paper."

What is the performance of this type of two-stage procedure for other crossover designs? Jones and Kenward (2003, pp. 123–125) analyze the performance of this type of procedure for Balaam’s design. This analysis makes the following two assumptions. The first assumption is that any carryover from a treatment in a given period is only into the next period, and not beyond (“first-order carryover” model). The second assumption is that the carryover from one period into the next period is determined only by the treatment applied in the first period and not the treatment applied in the second period. Thus, for example, according to this second assumption, the carryover from treatment A into the next period is the same, irrespective of whether the treatment in the next period is A or B. This assumption has rightly been criticized as being unrealistic by Fleiss (1986, 1989), Senn and Lambrou (1998), and Senn (2001, 2002, 2005). This severely limits the applicability of the analysis of Jones and Kenward (2003) of this type of procedure for Balaam’s design.

In the present paper, we consider an ABAB/BABA crossover trial. Subjects are randomly allocated to either group 1 or group 2. Subjects in group 1 receive treatments A, B, A, and B in the first, second, third, and fourth periods, respectively. Subjects in group 2 receive treatments B, A, B, and A in the first, second, third, and fourth periods, respectively. We assume that any carryover from a treatment in a given period is only into the next period, and not beyond. However, our analysis of this trial does not require us to assume that the carryover from one period into the next period is determined only by the treatment applied in the first period and not the treatment applied in the second period. This is because we never need to consider the carryover of a treatment from one period into the next period for which the same treatment is applied. Two major differences between the AB/BA and ABAB/BABA trials are the following. For an ABAB/BABA trial:

1. There is an unbiased estimator (which is unaffected by differential carryover) of $\theta$ that has the following properties. It is unaffected by the between-subject variation. Also, it is obtained without ignoring all of the data from periods 2, 3, and 4. This is the estimator $\hat{\Theta}_1$ described in Section 2.
2. There is an unbiased estimator of the differential carryover that is unaffected by the between-subject variation. This is the estimator $\hat{\Psi}_1$ described in Section 2.

There are two arguments against the adoption of $\hat{\Theta}_1$ as the standard estimator of $\theta$. First, as shown in Appendix A, this estimator is inefficient by comparison with the usual estimator of $\theta$ based on data from a completely randomized design, using the same number of measurements of the response, unless a restrictive condition holds. Second, there is an estimator of $\theta$, which we denote by $W$ and describe in Section 2, that is much more efficient than $\hat{\Theta}_1$, when the differential carryover is zero. We view $\hat{\Theta}_1$ as the analogue for an ABAB/BABA design of the estimator of $\theta$ constructed using only data from the first period of an AB/BA design.

To deal with the possibility of nonzero differential carryover, it is tempting to consider the use of the following two-stage procedure. In the first stage, a preliminary test of the null hypothesis that the differential carryover is zero (against the alternative that it is nonzero) is carried out. If this null hypothesis is accepted then the confidence interval for $\theta$ is constructed using the estimator $W$ (described in Section 2) and having nominal coverage $1 - \alpha$, assuming that there is no differential carryover. If, on the other hand, this null hypothesis is rejected then this confidence interval is constructed to have nominal coverage $1 - \alpha$, using the estimator $\hat{\Theta}_1$ (described in Section 2) that is based on data from all four periods. This two-stage procedure is described in detail in Section 2.

A computationally convenient formula for the coverage probability of the confidence interval that results from this procedure is presented in Section 2. In Section 3, we numerically evaluate the coverage properties of this confidence interval. We show that this confidence interval has minimum coverage probability far below $1 - \alpha$, demonstrating that this confidence interval should not be used. The coverage probability of this confidence interval depends only on the scaled differential carryover. This
is in sharp contrast to the coverage probability of the confidence interval resulting from the two-stage procedure applied to data from an AB/BA trial, found by Freeman (1989), which depends on both the scaled differential carryover and the ratio (error variance)/(subject variance).

Beginning with the work of Freeman (1989), the literature on the effect of preliminary model selection (using, e.g. hypothesis tests or minimizing a criterion such as AIC or Mallows's $C_p$) on confidence intervals has grown steadily. This literature is reviewed by Kabaila (2009). It is commonly the case that preliminary model selection has a highly detrimental effect on the coverage probability of these confidence intervals. However, each case needs to be considered individually on its merits.

2 The two-stage analysis of ABAB/BABA trials under consideration

We assume the following model for the ABAB/BABA trial. This model is similar to the model for an AB/BA crossover trial put forward by Grizzle (1965), as described by Grieve (1987). Let $n_1$ and $n_2$ denote the number of subjects in group 1 and group 2, respectively. Also let $Y_{ijk}$ be the response of the $j$-th subject in the $i$-th group and the $k$-th period ($i = 1, 2; j = 1, \ldots, n_i; k = 1, 2, 3, 4$). The model is

$$Y_{ijk} = \mu + \xi_{ij} + \pi_k + \phi_{ik} + \lambda_{jk} + \epsilon_{ijk},$$

where $\mu$ is the overall population mean,

$\xi_{ij}$ is the effect of the $j$-th subject in the $i$-th group,

$\pi_k$ is the effect of the $k$-th period,

$\phi_{ik}$ is the effect of the $i$-th treatment,

$\lambda_{jk}$ is the residual effect of the $q$-th treatment,

$\epsilon_{ijk}$ is the random error.

Note that both $\ell$ and $q$ are determined by the group $i$ and the period $k$. This model is described in less abbreviated form in Appendix B. We assume that the $\xi_{ij}$ and $\epsilon_{ijk}$ are independent and that the $\xi_{ij}$ are identically $N(0, \sigma^2)$ distributed and the $\epsilon_{ijk}$ are identically $N(0, \sigma^2)$ distributed, where $\sigma^2 > 0$ and $\sigma^2 > 0$. Let $m = (1/n_1) + (1/n_2)$.

The parameter of interest is $\theta = \phi_1 - \phi_2$. It is convenient to define the parameter $\psi = 3(\lambda_1 - \lambda_2)/4$, which is proportional to the differential carryover. Let $\bar{Y}_{ik} = (1/n_i) \sum_{j=1}^{n_i} Y_{ijk}(i = 1, 2; k = 1, 2, 3, 4)$.

We reduce that data to $D_1, D_2, D_3, D_4$, where $D_1 = \bar{Y}_{11} - \bar{Y}_{21}, D_2 = \bar{Y}_{12} - \bar{Y}_{22}, D_3 = \bar{Y}_{13} - \bar{Y}_{23}$, and $D_4 = \bar{Y}_{14} - \bar{Y}_{24}$. The motivation for this data reduction is presented in Appendix B. Let

$$W = \frac{1}{4} (D_1 - D_2 + D_3 - D_4).$$

This is the usual estimator of $\theta$, when it is assumed that $\psi = 0$ (see, e.g., Table I of Senn and Lambrou, 1998). Let

$$\hat{\Theta} = D_1 - \frac{1}{4} D_2 - \frac{1}{2} D_3 - \frac{1}{4} D_4.$$

This is an unbiased estimator (which is unaffected by differential carryover) of $\theta$ that is unaffected by between-subject variation (cf. Table I of Senn and Lambrou, 1998). We will use $\hat{\Theta}$ as the estimator of $\theta$ when it cannot be assumed that necessarily $\psi = 0$. We view $\hat{\Theta}$ as the analogue for an ABAB/BABA design of the estimator of $\theta$ constructed using only data from the first period of an AB/BA design. As shown in Appendix A, $\hat{\Theta}$ is inefficient by comparison with the usual estimator of $\theta$ based on data from a completely randomized trial, using the same number of measurements of response, unless $\sigma^2 \geq 4.5 \sigma^2$. We will also make use of the following unbiased estimator of $\psi$:

$$\hat{\Psi} = \frac{3}{4} (D_1 - D_3).$$
As shown in Appendix B, these statistics have the following distributions: $W \sim N(\theta - \psi, m\sigma^2/4)$, $\hat{\Theta} \sim N(\theta, 11m\sigma^2/8)$, and $\hat{\Psi} \sim N(\psi, 9m\sigma^2/8)$. Note that when $\psi = 0$, $W$ is a much more efficient estimator of $\theta$ than $\hat{\Theta}$.

To deal with the possibility of nonzero differential carryover, it is tempting to consider the use of the following two-stage procedure. In the first stage, a preliminary test of the null hypothesis that $\psi = 0$ (against the alternative that $\psi \neq 0$) is carried out, using a test statistic based on $\hat{\Psi}$. If this null hypothesis is accepted then the confidence interval for $\theta$ is constructed using the estimator $W$ and having nominal coverage $1 - \alpha$, assuming that $\psi = 0$. If, on the other hand, this null hypothesis is rejected then this confidence interval is constructed to have nominal coverage $1 - \alpha$, using the estimator $\hat{\Theta}$.

To analyze the properties of this two-stage procedure, we make the simplification that $\sigma^2$ is known. Freeman (1989) makes the same simplification. So, in the first stage, we test the null hypothesis $H_0 : \psi = 0$ against the alternative hypothesis $H_1 : \psi \neq 0$ using the test statistic $\sqrt{8/9m} \hat{\Psi}/\sigma_\psi$. This test statistic has an $N(0,1)$ distribution under $H_0$. Define the quantile $c_\psi$ by the requirement that $P(-c_\psi \leq Z \leq c_\psi) = 1 - \alpha$ for $Z \sim N(0,1)$. The following is a test of $H_0$ against $H_1$, with level of significance $\alpha_\psi$. Accept $H_0$ if $|\sqrt{8/9m} \hat{\Psi}/\sigma_\psi| < c_\psi$; otherwise reject $H_0$. In the second stage, we proceed as follows. If $H_0$ is accepted then we construct a confidence interval for $\theta$, with nominal coverage $1 - \alpha$, assuming that $\psi = 0$. This confidence interval is

$$W - c_\psi \sqrt{m/4} \sigma_\epsilon, \ W + c_\psi \sqrt{m/4} \sigma_\epsilon. \hspace{1cm} (2)$$

If, on the other hand, $H_0$ is rejected then we do not assume that $\psi = 0$ and we construct a confidence interval for $\theta$, with nominal coverage $1 - \alpha$, based on $\hat{\Theta}$. This confidence interval is

$$\left[ \hat{\Theta} - c_\psi \sqrt{11m/8} \sigma_\epsilon, \ \hat{\Theta} + c_\psi \sqrt{11m/8} \sigma_\epsilon \right].$$

Let $J$ denote the confidence interval for $\theta$ that results from this two-stage procedure. Also let $\gamma = \sqrt{8/9m} \hat{\Psi}/\sigma_\theta$. As shown in Appendix B, the coverage probability of the confidence interval $J$ is

$$P(\theta \in J) = P(|H| < c_\psi)P(|X| \leq c_\psi) + P(|G| \leq c_\psi, |H| \geq c_\psi). \hspace{1cm} (4)$$

where

$$\left[ \begin{array}{c} G \\ H \end{array} \right] \sim N \left( \left[ \begin{array}{c} 0 \\ \gamma \end{array} \right], \left[ \begin{array}{cc} 1 & 3/\sqrt{11} \\ 3/\sqrt{11} & 1 \end{array} \right] \right) \hspace{1cm} (5)$$

and $X \sim N(-3\gamma/\sqrt{2}, 1)$. Note that, for given $\alpha_\psi$ and $\alpha$, the coverage probability (4) is a function of the scaled differential carryover $\gamma$. The right-hand side of (4) is easily computed (using, e.g., R or MATLAB programs), for each given $\gamma$. The last term on the right-hand side of (4) can be computed by evaluating the cumulative distribution function of the bivariate normal distribution (5). Alternatively, this term can be computed by numerically evaluating the integral (C2), derived in Appendix C.

### 3 Numerical evaluation of the coverage probability as a function of $\gamma$

Consider the two-stage procedure, for an AB/BA trial, based on a preliminary test with given level of significance and resulting in a confidence interval with a given nominal coverage. As shown by Freeman (1989), the actual coverage probability of this confidence interval depends on both the scaled differential carryover ($\lambda \sqrt{m}/\sigma$ in Freeman’s notation) and $\rho = \sigma^2/(\sigma_\theta^2 + \sigma_\epsilon^2)$. For each different value of $\rho$, there is a different graph of this coverage probability as a function of the scaled differential carryover $\gamma$. This term can be computed by numerically evaluating the integral (C2), derived in Appendix C.
Figure 1  Plot of the coverage probability of the confidence interval for $\theta$, resulting from the two-stage procedure, against $\gamma$. This confidence interval has nominal coverage $1 - \alpha = 0.95$. The preliminary hypothesis test has significance level $\alpha_1 = 0.1$. The horizontal dashed line has vertical axis intercept 0.95.

carryover. The larger the value of $\rho$, the smaller the minimum coverage probability of this confidence interval.

Now consider the two-stage procedure described in the previous section, for an ABAB/BABA trial, based on a preliminary test with given level of significance and resulting in a confidence interval with a given nominal coverage. In sharp contrast to the AB/BA trial, the actual coverage probability (given by (4)) of this confidence interval depends only on the scaled differential carryover $\gamma$. This coverage is uninfluenced by the between-subject variability (which is described by the parameter $\sigma^2$).

For level of significance $\alpha_1 = 0.1$ of the preliminary test and nominal coverage $1 - \alpha = 0.95$, this coverage probability as a function of $\gamma$ is shown in Fig. 1. The minimum coverage probability of this confidence interval is 0.4711, showing that this confidence interval is completely inadequate. The minimum coverage probability of this confidence interval was computed for a wide range of values of $\alpha_1$ and $1 - \alpha$. In every case, this confidence interval was found to have minimum coverage probability far below nominal, showing that it is completely inadequate. Note that for a given level of significance $\alpha_1$ of the preliminary test and given nominal coverage $1 - \alpha$, the minimum coverage probability of
this confidence interval does not depend on either of the sample sizes \(n_1\) and \(n_2\). The only effect of an increase in \(n_1\) and \(n_2\) is to change the scaling (via \(m = (1/n_1) + (1/n_2)\)) of the differential carryover \(\psi\). Consequently, the harmful effect of preliminary hypothesis testing does not disappear with an increase in sample sizes \(n_1\) and \(n_2\).

4 Conclusion

For an ABAB/BABA trial, we have shown that the minimum coverage probability of the confidence interval resulting from the two-stage procedure is far below the nominal coverage, showing that this confidence interval is completely inadequate. Increasing the sample sizes \(n_1\) and \(n_2\) does not improve the situation. Our conclusion is that this confidence interval should not be used. This is similar to the conclusion of Freeman (1989) for confidence intervals resulting from a two-stage procedure applied to an AB/BA trial. In other words, we provide further support for the rejection by Senn (2002, p. 12) of analyses of data from any two-treatment crossover trial based on a preliminary test of the null hypothesis that the differential carryover is zero.

Appendix A: The efficiency of \(\hat{\Theta}\) by comparison with an estimator from a completely randomized trial

In this appendix, we consider the efficiency of \(\hat{\Theta}\) by comparison with the usual estimator of \(\theta\) based on data from a completely randomized trial. For an ABAB/BABA crossover trial, the total number of measurements of the response is \(4(n_1 + n_2)\). We therefore compare \(\hat{\Theta}\) with the usual estimator of \(\theta\) based on data from a completely randomized trial, with \(2(n_1 + n_2)\) randomly chosen subjects in each group.

Let \(Y_1^A, \ldots, Y_{2(n_1 + n_2)}^A\) denote the responses of the \(2(n_1 + n_2)\) subjects given treatment A. Also let \(Y_1^B, \ldots, Y_{2(n_1 + n_2)}^B\) denote the responses of the \(2(n_1 + n_2)\) subjects given treatment B. Consistently with the model (1), we suppose that \(Y_1^A, \ldots, Y_{2(n_1 + n_2)}^A, Y_1^B, \ldots, Y_{2(n_1 + n_2)}^B\) are independent random variables, where \(Y_1^A, \ldots, Y_{2(n_1 + n_2)}^A\) are identically \(N(\mu + \phi, \sigma_1^2 + \sigma_2^2)\) distributed and \(Y_1^B, \ldots, Y_{2(n_1 + n_2)}^B\) are identically \(N(\mu + \phi_2, \sigma_1^2 + \sigma_2^2)\) distributed. The usual estimator of \(\theta\) is

\[
\hat{\Theta} = \frac{1}{2(n_1 + n_2)}(Y_1^A + \ldots Y_{2(n_1 + n_2)}^A) = \frac{1}{2(n_1 + n_2)}(Y_1^B + \ldots Y_{2(n_1 + n_2)}^B).
\]

This estimator has an \(N(\theta, (\sigma_1^2 + \sigma_2^2)/(n_1 + n_2))\) distribution. Suppose, for simplicity, that \(n_1 = n_2 = n\). Thus, \(\text{Var}(\hat{\Theta}) = (\sigma_1^2 + \sigma_2^2)/(2n)\) and \(\text{Var}(\hat{\Theta}) = 11\sigma_2^2/(4n)\). Thus, \(\text{Var}(\hat{\Theta}) \leq \text{Var}(\hat{\Theta})\) if and only if \(\sigma_2^2 \geq 4.5\sigma_1^2\).

Appendix B: Details for Section 2

This appendix consists of 3 sections. In the first section, we carry out data reduction. In the second section, we derive the distributions of the statistics \(W, \hat{\Theta}\) and \(\hat{\psi}\). In the third section, we derive the formula (4) for the coverage probability of the confidence interval \(J\) resulting from the two-stage procedure.
B.1 Data reduction

It follows from the model (1) that

\[ Y_{ij1} = \mu + \xi_{ij1} + \tau_{i1} + \phi_{1} + \epsilon_{1j1} \]
\[ Y_{ij2} = \mu + \xi_{ij1} + \tau_{i2} + \phi_{2} + \lambda_{i} + \epsilon_{1j2} \]
\[ Y_{ij3} = \mu + \xi_{ij1} + \tau_{i3} + \phi_{1} + \lambda_{i2} + \epsilon_{1j3} \]
\[ Y_{ij4} = \mu + \xi_{ij1} + \tau_{i4} + \phi_{1} + \lambda_{i3} + \epsilon_{1j4} \]
\[ Y_{ij1} = \mu + \xi_{ij2} + \tau_{i1} + \phi_{2} + \epsilon_{2j1} \]
\[ Y_{ij2} = \mu + \xi_{ij2} + \tau_{i2} + \phi_{1} + \lambda_{i2} + \epsilon_{2j2} \]
\[ Y_{ij3} = \mu + \xi_{ij2} + \tau_{i3} + \phi_{2} + \lambda_{i3} + \epsilon_{2j3} \]
\[ Y_{ij4} = \mu + \xi_{ij2} + \tau_{i4} + \phi_{1} + \lambda_{i2} + \epsilon_{2j4} \]

Let \( \bar{Y}_{ik} = (1/\eta_{k}) \sum_{j=1}^{n} Y_{ijk} \) \( (i = 1, 2; k = 1, 2, 3, 4) \). We first reduce the data to \( \bar{Y}_{1.1}, \bar{Y}_{1.2}, \bar{Y}_{1.3}, \bar{Y}_{1.4}, \bar{Y}_{2.1}, \bar{Y}_{2.2}, \bar{Y}_{2.3}, \) and \( \bar{Y}_{2.4} \). Note that

\[ \bar{Y}_{1.1} = \mu + \bar{\xi}_{11} + \tau_{1} + \phi_{1} + \bar{\epsilon}_{11} \]
\[ \bar{Y}_{1.2} = \mu + \bar{\xi}_{11} + \tau_{2} + \phi_{2} + \lambda_{1} + \bar{\epsilon}_{12} \]
\[ \bar{Y}_{1.3} = \mu + \bar{\xi}_{11} + \tau_{3} + \phi_{1} + \lambda_{2} + \bar{\epsilon}_{13} \]
\[ \bar{Y}_{1.4} = \mu + \bar{\xi}_{11} + \tau_{4} + \phi_{1} + \lambda_{3} + \bar{\epsilon}_{14} \]
\[ \bar{Y}_{2.1} = \mu + \bar{\xi}_{21} + \tau_{1} + \phi_{2} + \bar{\epsilon}_{21} \]
\[ \bar{Y}_{2.2} = \mu + \bar{\xi}_{21} + \tau_{2} + \phi_{1} + \lambda_{2} + \bar{\epsilon}_{22} \]
\[ \bar{Y}_{2.3} = \mu + \bar{\xi}_{21} + \tau_{3} + \phi_{2} + \lambda_{3} + \bar{\epsilon}_{23} \]
\[ \bar{Y}_{2.4} = \mu + \bar{\xi}_{21} + \tau_{4} + \phi_{1} + \lambda_{2} + \bar{\epsilon}_{24} \]

where \( \bar{\xi}_{1} = (1/\eta_{1}) \sum_{j=1}^{n} \xi_{ij1} \) and \( \bar{\xi}_{21} = (1/\eta_{1}) \sum_{j=1}^{n} \epsilon_{ij1} \). Note that \( \bar{\xi}_{11}, \bar{\xi}_{21}, \bar{\epsilon}_{11}, \ldots, \bar{\epsilon}_{14}, \bar{\epsilon}_{21}, \ldots, \bar{\epsilon}_{24} \) are independent, \( \bar{\xi}_{11} \sim N(0, \sigma_{1}^{2}/\eta_{1}), \bar{\xi}_{21} \sim N(0, \sigma_{1}^{2}/\eta_{1}), \bar{\epsilon}_{11}, \ldots, \bar{\epsilon}_{14} \) are identically \( N(0, \sigma_{1}^{2}/\eta_{1}) \) distributed and \( \bar{\epsilon}_{21}, \ldots, \bar{\epsilon}_{24} \) are identically \( N(0, \sigma_{1}^{2}/\eta_{1}) \) distributed.

The only way to remove the influence of the parameters \( \tau_{1}, \ldots, \tau_{4} \) on the reduced data \( \bar{Y}_{1.1}, \bar{Y}_{1.2}, \ldots, \bar{Y}_{2.4} \) is to perform a further data reduction to \( D_{1}, \ldots, D_{4} \), where \( D_{1} = \bar{Y}_{1.1} - \bar{Y}_{2.1}, D_{2} = \bar{Y}_{1.2} - \bar{Y}_{2.2}, D_{3} = \bar{Y}_{1.3} - \bar{Y}_{2.3}, \) and \( D_{4} = \bar{Y}_{1.4} - \bar{Y}_{2.4} \). Note that

\[ D_{1} = \bar{\xi}_{11} - \bar{\xi}_{21} + \theta + \eta_{1} \]
\[ D_{2} = \bar{\xi}_{21} - \bar{\xi}_{21} - \theta + \frac{1}{2} \psi + \eta_{2} \]
\[ D_{3} = \bar{\xi}_{31} - \bar{\xi}_{31} + \theta - \frac{1}{2} \psi + \eta_{3} \]
\[ D_{4} = \bar{\xi}_{41} - \bar{\xi}_{42} - \theta + \frac{1}{2} \psi + \eta_{4} \]

where \( \eta_{1} = \bar{\epsilon}_{11} - \bar{\epsilon}_{21}, \eta_{2} = \bar{\epsilon}_{21} - \bar{\epsilon}_{22}, \eta_{3} = \bar{\epsilon}_{13} - \bar{\epsilon}_{23}, \) and \( \eta_{4} = \bar{\epsilon}_{14} - \bar{\epsilon}_{24} \). Note that \( \eta_{1}, \ldots, \eta_{4} \) are independent and identically \( N(0, m\sigma_{1}^{2}) \) distributed.

B.2 Derivation of the distributions of the statistics \( W, \hat{\Theta}, \) and \( \hat{\Psi} \)

Note that

\[ W = \theta - \psi + \frac{1}{4} (\eta_{1} - \eta_{2} + \eta_{3} - \eta_{4}). \]  \hspace{1cm} (B1)

Thus, \( W \sim N(\theta - \psi, m\sigma_{1}^{2}/4) \). Note that

\[ \hat{\Theta} = \theta + \frac{1}{4} \eta_{1} - \frac{1}{2} \eta_{2} - \frac{1}{4} \eta_{3} - \frac{1}{4} \eta_{4} \]  \hspace{1cm} (B2)

© 2012 WILEY-VCH Verlag GmbH & Co. KGaA, Weinheim  www.biometrical-journal.com
and that

\[ \hat{\Psi} = \psi + \frac{3}{4}(\eta_1 - \eta_3). \]  

(B3)

It follows from (B1) and (B3) that \((W, \hat{\Psi})\) has a bivariate normal distribution and that Cov\((W, \hat{\Psi}) = 0\). Thus, \(W\) and \(\hat{\Psi}\) are independent random variables. It follows from (B2) and (B3) that

\[
\begin{bmatrix}
\hat{\Omega} \\
\hat{\Psi}
\end{bmatrix}
\sim \mathcal{N}
\begin{bmatrix}
\theta \\
\psi
\end{bmatrix}
\begin{bmatrix}
\sigma_w^2 \\
\sigma_{\omega}^2
\end{bmatrix}
\begin{bmatrix}
11 \\
9
\end{bmatrix}
\]

(B4)

B.3 Derivation of the formula (4) for the coverage probability

Define the event

\[ E = \left\{ \left| \sqrt{\frac{8}{9m}} \hat{\Psi} \right| < c_a \right\}. \]

If this event occurs then \(J\) is equal to (2) and if \(E^c\) occurs then \(J\) is equal to (3). By the law of total probability, the coverage probability \(P(\theta \in J)\) is equal to

\[
P(E \cap \{\theta \in J\}) + P(E^c \cap \{\theta \in J\})
= P\left(E \cap \left\{ \frac{W - \theta}{\sigma_e} \sqrt{\frac{4}{m}} \leq c_a \right\}\right) + P\left(E^c \cap \left\{ \frac{\hat{\phi} - \theta}{\sigma_e} \sqrt{\frac{8}{11m}} \leq c_a \right\}\right)
= P(E) P\left(\frac{W - \theta}{\sigma_e} \sqrt{\frac{4}{m}} \leq c_a \right) + P\left(\frac{\hat{\phi} - \theta}{\sigma_e} \sqrt{\frac{8}{11m}} \leq c_a \right)
\]

since \(W\) and \(\hat{\Psi}\) are independent random variables. Now define

\[ \gamma = \sqrt{\frac{8}{9m}} \psi, \quad H = \sqrt{\frac{8}{9m}} \hat{\Psi}, \quad G = \frac{(\hat{\phi} - \theta)}{\sigma_e} \sqrt{\frac{8}{11m}}, \quad \text{and} \quad X = \frac{(W - \theta)}{\sigma_e} \sqrt{\frac{4}{m}}. \]

Thus, the coverage probability \(P(\theta \in J)\) is given by (4). Note that it follows from (B4) that the distribution of \((G, H)\) is given by (5).

Appendix C: Alternative expression for \(P(\{|G| \leq c_a, \{|H| \geq c_a\})\)

In this appendix, we present an alternative expression for \(P(\{|G| \leq c_a, \{|H| \geq c_a\})\) that may be convenient for the computation of the coverage probability (4). By the law of total probability,

\[
P(\{|G| \leq c_a\} = P(\{|G| \leq c_a, \{|H| \geq c_a\}) + P(\{|G| \leq c_a, \{|H| < c_a\}),
\]

so that

\[
P(\{|G| \leq c_a, \{|H| \geq c_a\}) = P(\{|G| \leq c_a\}) - P(\{|G| \leq c_a, \{|H| < c_a\})
= 1 - \alpha - P(\{|G| \leq c_a, \{|H| < c_a\})
\]

since \(G \sim \mathcal{N}(0, 1)\).
Let \( f_{G,H}(g, h) \) denote the probability density function of \((G, H)\), evaluated at \((g, h)\). Also, let \( f_{H|G}(h|g) \) denote the probability density function of \(H\) conditional on \(G = g\), evaluated at \(h\). Let \( \phi \) denote the \(N(0, 1)\) probability density function. Observe that

\[
P(\{|G| \leq c_α, |H| < c_α|1\}) = \int_{-c_α}^{c_α} \int_{-c_α}^{c_1} f_{G,H}(g, h) \, dh \, dg = \int_{-c_α}^{c_α} \int_{-c_α}^{c_1} f_{H|G}(h|g) \, dh \, \phi(g) \, dg.
\]  

(C1)

It follows from (5) that the distribution of \(H\) conditional on \(G = g\) is \(N(\mu(g), \nu)\), where \(\mu(g) = \gamma + (3g/\sqrt{11})\) and \(\nu = 2/11\). Thus, (C1) is equal to

\[
\int_{-c_α}^{c_α} (\Phi(c_α, \mu(g), \nu) - \Phi(-c_α, \mu(g), \nu)) \, \phi(g) \, dg.
\]  

(C2)

where \(\Phi(x; \mu, \nu)\) denotes the \(N(\mu, \nu)\) cumulative distribution function, evaluated at \(x\). The integral (C2) is readily evaluated using the numerical quadrature functions available in either \texttt{R} or \texttt{MATLAB}.

References

Armitage, P. and Hills, M. (1982). The two-period crossover trial. \textit{Statistician} 31, 119–131.

Fleiss, J. L. (1986). On multiperiod crossover trials. Letter to the editor. \textit{Biometrics} 42, 449–450.

Fleiss, J. L. (1989). A critique of recent research on the two-treatment crossover design. \textit{Controlled Clinical Trials} 10, 237–243.

Freeman, P. (1989). The performance of the two-stage analysis of two-treatment, two-period crossover trials. \textit{Statistics in Medicine} 8, 1421–1432.

Grieve, A. P. (1987). A note on the analysis of the two-period crossover design when the period-treatment interaction is significant. \textit{Biometrical Journal} 7, 771–775.

Grizzle, A. P. (1965). The two-period change-over design and its use in clinical trials. \textit{Biometrics} 21, 467–480.

Grizzle, A. P. (1974). Corrigenda to Grizzle (1965). \textit{Biometrics} 42, 459.

Hills, M. and Armitage, P. (1979). The two-period crossover clinical trial. \textit{British Journal of Clinical Pharmacology} 8, 7–20.

Jones, B. and Kenward, M. G. (2003). \textit{Design and Analysis of Cross-Over Trials} (2nd edn.). Chapman & Hall, Boca Raton, FL.

Kabaila, P. (2009). The coverage properties of confidence regions after model selection. \textit{International Statistical Review} 77, 405–414.

Senn, S. (2001). Cross-over trials in drug development: theory and practice. \textit{Journal of Statistical Planning and Inference} 96, 29–40.

Senn, S. (2002). \textit{Cross-Over Trials in Clinical Research} (2nd edn.). Wiley, Chichester.

Senn, S. (2005). Misunderstandings regarding clinical cross-over trials. Letter to the editor. \textit{Statistics in Medicine} 24, 3675–3678.

Senn, S. (2006). Crossover trials in Statistics in Medicine: the first ‘25’ years. \textit{Statistics in Medicine} 25, 3430–3442.

Senn, S. and Lambrou, D. (1998). Robust and realistic approaches to carry-over. \textit{Statistics in Medicine} 17, 2849–2864.