On Nearly Balanced Designs for Sensory Trials

Joachim Kunert, Oliver Sailer [1]

Fachbereich Statistik, Universität Dortmund, D-44221 Dortmund, Germany

August 2004

Abstract

In sensory experiments, often designs are used that are balanced for carryover effects. It is hoped that this controls for possible carryover effects, like, e.g., a lingering taste of the products.

Proper randomization is essential to guarantee the usual model assumption of independent identically distributed (i.i.d.) errors. We consider a randomization procedure that permutes treatment labels and assessors. This restricted randomization leaves the neighbour structure unchanged and validates the assumption of i.i.d. errors if the design used is a Generalized Youden Design (GYD).

However, the use of a neighbour balanced GYD may require too many assessors. The question arises, whether nearly balanced designs may be used without grossly violating the validity of the analysis. We therefore do a simulation study to assess the properties (under this restricted randomization) of nearly balanced designs like, e.g., the ones proposed by Périnel and Pagès (2004, Food Quality and Preference 15, 439–446).

We observe that, if there are no carryover effects, the variance estimates for treatment contrasts are not significantly biased whenever we use designs that are nearly GYD. Additionally, designs that are nearly carryover balanced still produce conservative variance estimates, even in the presence of large carryover effects.

In all, "nearly neighbour balanced nearly GYD" as proposed by Périnel and Pagès (2004) appear to be useful in experimental situations where the use of GYD is too restrictive. It should be stressed, however, that these results are true only if randomization is used as a protection against effects unaccounted for in the statistical model.

Key words: carryover balance, nearly balanced designs, randomization, validity

1   Introduction

There is a large number of papers in the sensometrics literature dealing with experimental designs if there is a danger of carryover effects. While the traditional recommendation had been to simply randomize the order in which the products are presented to the assessors, the publication of the seminal paper by Mac Fie, Greenhoff, Bratchell and Vallis (1989) has started a discussion about the use of neighbour balanced designs. While Mac Fie et al. (1989), Schlich (1993) and Wakeling and MacFie (1995) discussed construction methods for

---

[1] Corresponding author, e-mail: sailer@statistik.uni-dortmund.de, phone: +49-231-7555409

neighbour balanced designs, Ball (1997) and Périnel and Pagès (2004) deal with designs that are "near" such designs.

As already expressed in Kunert (1998), we are concerned about the consequences that this development might have for the practice of sensometrics. We are worried that experimenters might decide to use highly structured systematic designs for their experiments and neglect the benefits to be gained from randomization, see e.g. Kunert, Meyners and Erdbrügge (2002).

Another problem is that experimenters might think that the problem of carryover can be overcome with these designs. Hence they might spend less efforts in non-statistical methods to avoid carryover (like, e.g. washout periods). Neighbour designs are no panacea for problems concerned with carryover. They only work in a quite optimistic model, where the carryover effects are additive. Even in that model they do not avoid all bias due to carryover. It was, however, shown by several authors (see e.g. Azaïs and Druilhet, 1997) that these designs provide some robustness, they minimize some measures of the bias of the estimated treatment differences.

In Kunert (1998) it was demonstrated with the help of permutation tests, that the randomization by Bailey (1985) validates the model with assessor and order effects but leaves neighbour balance intact, provided the starting design is a special Generalized Youden Design.

Hence, we would support the use of neighbour balanced Generalized Youden Designs with this randomization. It will do no harm, if there are no carryover effects, but might provide some robustness if there are.

The construction methods in Ball (1997) and Périnel and Pagès (2004) try to achieve designs that are as similar as possible to such neighbour balanced Generalized Youden Desings. They do not claim to minimize some theoretical concept like bias due to carryover effects, or variance of the estimates. It was not clear to us, how these designs perform in practice.

The present paper has two aims. Firstly, we want to give a formal proof that the randomization of assessors and treatment labels introduced by Bailey (1985) for Generalized Latin Squares, ensures the validity of the model with assessor and order effects whenever the starting design is a special Generalized Youden Design and whenever there are no carryover effects. With that, we hope to remind experimenters of the need to randomize.

It is not possible to extend our proof to the "near" designs by Périnel and Pagès (2004).

So the second aim of our paper is to examine the performance of such designs under the randomization of assessors and treatment labels. Fortunately, our simulations appear to show that the non-validity is not very dramatic, at least for reasonable designs. For the designs that are near the neighbour balanced Generalized Youden Designs the bias of the variance estimates due to non-validity of the model appears to remain reasonably small.

2 Randomization Analysis when the Carryover Effects are Neglected

The treatments in a sensory experiment are assigned to period i and assessor j via a design δ. Formally, the design δ is a mapping from $\{1, \ldots, k\} \times \{1, \ldots, b\}$ on $\{1, \ldots, t\}$, where k is the number of periods (i.e. the number of assessments made by each assessor), b is the number of assessors and t is the number of treatments. Assessor j in period i receives treatment $\delta(i, j)$. As an example consider the case that we have two assessors, each of whom makes two assessments, and that there are three treatments. One possible mapping might be

$$d = \begin{bmatrix} 1 & 1 \\ 2 & 3 \end{bmatrix},$$

where rows indicate periods and columns assessors, while the numbers give the treatments. Then, for instance, assessor 1 in period 2 receives treatment 2, $d(2, 1) = 2$.

Assume η_{ij} is the hypothetical rating of assessor j at period i if he/she rated a standard product. The usual model for randomization analysis assumes that there is an additive treatment effect. More precisely, if the assessor rates product r, the randomization model assumes that his/her rating changes to

$$y_{ij} = \eta_{ij} + \tau_r, \tag{1}$$

where τ_r is the effect of product r. In (1) it is assumed that the treatment effects are additive and that there are no carryover effects.

In the row-column model each η_{ij} can be written as

$$\eta_{ij} = \alpha_i + \beta_j + e_{ij}, \tag{2}$$

where $\alpha_1, \ldots, \alpha_k$ are the fixed effects of the periods, $\beta_1, \ldots, \beta_b$ are the fixed effects of the assessors and the e_{ij} are independent random variables, each with expectation 0 and variance σ^2.

It is convenient to write the model in vector notation. We have

$$
\begin{aligned}
Y &= T_\delta \tau + \eta \\
&= T_\delta \tau + (1_b \otimes I_k)\alpha + (I_b \otimes 1_k)\beta + e,
\end{aligned}
$$

where $Y = [y_{11}, \ldots, y_{k1}, y_{12}, \ldots, y_{kb}]'$, T_δ is the treatment design matrix, τ is the vector of treatment effects and $\eta = [\eta_{11}, \ldots, \eta_{k1}, \eta_{12}, \ldots, \eta_{kb}]'$. Also, 1_k denotes the k-dimensional vector of ones, α and β are the vectors of period and assessor effects and e is the vector of errors.

With a randomized design, δ is not fixed but randomly selected among a set of such mappings. The distribution of δ is chosen by the experimenter in such a way that it induces a distribution on the y_{ij} which is easy to treat. The idea is to consider the conditional distribution of the y_{ij} for given η_{ij} such that this conditional distribution is independent of all peculiarities which the distribution of the η_{ij} may have. The experimenter might decide to randomize assessors and treatment labels of a starting design d, i. e. he randomly selects a permutation π of $\{1, \ldots, b\}$ and a permutation γ of $\{1, \ldots, t\}$ and thus gets a design δ, such that

$$
\delta(i, j) = \gamma\Big(d\big(i, \pi(j)\big)\Big).
$$

Assume Π is the $b \times b$ permutation matrix corresponding to the permutation π of assessors and Γ is the $t \times t$ permutation matrix corresponding to the permutation γ of treatment labels. Then the design matrix after randomization can be written as

$$
T_\delta = (\Pi \otimes I_k)H_d\Gamma,
$$

where $\otimes$ indicates the Kronecker product of matrices, I_k is the identity matrix of size k, and H_d is the design matrix of the starting design d.

We are interested in estimating contrasts of the treatment effects, i. e. functions of the form $\ell'\tau$ where ℓ is a t-vector such that $\ell'1_t = 0$. Assuming model (2) we get a least squares estimate $\ell'\hat{\tau}$ for $\ell'\tau$ in that model, we get a variance $\mathrm{var}(\ell'\hat{\tau})$ of that estimate and we get an estimate $\hat{\mathrm{var}}(\ell'\hat{\tau})$ of that variance. A model is called (strongly) valid, see e.g. Bailey and Rowley (1987), for a given randomization if, for every fixed η, we have in model (1) that all

$$
\mathrm{E}_\delta\left(\ell'\hat{\tau}\right) = \ell'\tau \tag{3}
$$

and all

$$
\mathrm{var}_\delta\left(\ell'\hat{\tau}\right) = \mathrm{E}_\delta\left\{\hat{\mathrm{var}}(\ell'\hat{\tau})\right\}, \tag{4}
$$

4

where the index δ indicates that the expectation and the variance are calculated with respect to the distribution induced by the random selection of δ (while η is kept fixed !).

The usual randomization procedures to justify model (2) destroy any neighbour structure we might have given to the starting design d. They especially destroy balance for carryover effects. It is clear that we do not have to randomize if we *assume* that model (2) holds. In that case we can estimate treatment differences and the variance of the estimates even for fixed δ. However, we think this would be dangerous. If our assumptions on the η turn out to be wrong, then this might lead to grossly wrong estimates. Fortunately, for certain designs, there is a randomization which preserves balance for carryover effects and justifies model (2). This randomization procedure will be discussed in the rest of this section.

We use a starting design d with some desirable properties and apply some randomization which does not destroy these properties. Then we find out how the estimates from model (2) perform under the distribution induced by this randomization.

If we have a starting design which is balanced for carryover effects and do not want to destroy this balance, then it is only possible to randomize the assessors and the treatment labels. It is known, see Bailey (1985), that this randomization can justify model (2) if the starting design is a Latin Square. We show now that it can validate model (2) if the starting design belongs to a larger class. Consider the following properties:

(GYD1) The number of periods is not greater than the number of products and each assessor receives each product at most once.

(GYD2) For any pair of treatments the number of assessors receiving both treatments is a constant, namely $bk(k-1)/(t(t-1))$.

(GYD3) Each product appears in each period either $\left[\frac{b}{t}\right]$ or $\left[\frac{b}{t}\right]+1$ times.

(GYD4) The number of periods where any two distinct treatments occur $\left[\frac{b}{t}\right]+1$ times is a constant.

A design which fulfils conditions (GYD1) through (GYD4) is called a Generalized Youden Design. If for a Generalized Youden Design the number of assessors b is a multiple of the number of treatments t, then it is called uniform on the periods. A design which fulfils conditions (GYD1) and (GYD2) only is called a balanced incomplete block (BIB) design if

5

$k < t$ and a complete block design if $k = t$. Assume we start with a design d for which the design matrix is H_d and randomize assessors and treatment labels. As was indicated in the example, there are permutation matrices Π and Γ such that $T_\delta = (\Pi \otimes I_k)H_\delta\Gamma$, where T_δ is the design matrix of the design δ resulting from the randomization. Kunert (1998) has an example for the following proposition.

Proposition 1:

Assuming model (1), if we start with a Generalized Youden Design d that is uniform on the periods and randomize treatment labels and assessors, then the row-column model (2) is valid.

(The proof is in the appendix.)

Note that this result cannot be extended to the case that some (or all) products are tasted more than once by any assessor.

If the number of assessors is not a multiple of t, then Proposition 1 cannot be used. Yet the row-column model (2) is also valid for the special case where we consider a Generalized Youden Design d where $k = t$.

Proposition 2:

If we then assume model (1), start with a Generalized Youden Design d where $k = t$ and randomize treatment labels and assessors, then the row-column model (2) is valid.

(The proof is in the appendix.)

The results above hold for any of the Generalized Youden Designs considered above. They especially hold if the starting design d (and therefore every resulting δ) is balanced for carryover effects.

3 Influence of Carryover Effects on the Estimates

If there are carryover effects then the randomization model 1 transforms to

$$y_{ij} = \eta_{ij} + \tau_{\delta(i,j)} + \rho_{\delta(i-1,j)}, \tag{5}$$

where $\rho_{\delta(i-1,j)}$ is the carryover effect of the product tasted in the preceding period. In period 1 there is no carryover effect. If, for the η_{ij} in (5), the row-column model is *assumed*, then it

6

can be shown that designs which are balanced for carryover effects have excellent optimality properties, see e.g. Kunert (1984).

Unfortunately, assuming model (5), the row-column model cannot be validated by randomization, see Kunert and Utzig (1993).

Often the designs which are balanced for carryover effects are used together with the uncorrected estimate $\hat{\tau}$ from section 2. Then the balance is used as a precaution. It has to be pointed out, however, that these estimates in the case of balanced designs are not unbiased from carryover. We only get from balanced designs that the bias is in some sense minimal. Some considerations of the size of this bias in the case of fixed balanced designs can be found e.g. in Dyke (1983) or in Kunert (1985). We consider the bias of the estimate $\hat{\tau}$ under the viewpoint of randomization in the notation of section 2.

Model (5) in vector notation can be written as

$$Y = T_\delta \tau + F_\delta \rho + \eta,$$

where $\rho = [\rho_1, \ldots, \rho_t]'$. Here, the carryover design matrix F_δ is achieved from T_δ by premultiplying with the matrix $I_b \otimes W$, where

$$W = \begin{bmatrix} 0 & & \cdots & & 0 \\ 1 & & & & \\ 0 & \ddots & \ddots & & \vdots \\ \vdots & \ddots & \ddots & & \\ 0 & \cdots & 0 & 1 & 0 \end{bmatrix}.$$

The following proposition is proved in the appendix.

Proposition 3:

Assume model (5). We start with a Generalized Youden Design and randomize assessors and treatment labels. If we use the estimate $\ell'\hat{\tau}$ and if there are carryover effects $\rho \neq 0$ then generally this estimate will be biased. The bias equals $-\frac{\ell'\rho}{k}$ and is independent of the treatment order in the starting design d. The carryover effects generally increase the variance of the estimate.

Assume now that the starting design d and the design d^* which consists of the first $k - 1$ periods of d are both Generalized Youden Designs that are uniform on the periods and that d is balanced for carryover effects. Then the variance of the estimate $\ell'\hat{\tau}$ is the same with or without carryover effects.

4 Simulation Study

4.1 Introduction

The results of propositions 1, 2 and 3 are demonstrated using a simulation study. A number of designs were randomized and then applied to different data sets. This data represents the results of the application of the designs in sensory trials. One aim of the simulation study is to provide examples for the usefulness of randomization in sensory trials.

Some authors have suggested using nearly balanced designs that do not require restrictions on t, b and k, see e.g. Ball (1997) or Périnel and Pagès (2004). It is unclear whether using such designs leads to grossly biased results when carryover effects are present. Therefore, another goal of the simulation study is to assess the validity of designs which do not fulfil the requirements of the propositions above.

All computations were done in R, version 1.8.1. Programm code for doing the analysis can be obtained from the authors.

The data were generated by sampling 150 realizations, 5 for each of 30 virtual assessors, from the log-normal distribution. Artificial non-additive period effects were created by sorting the data for each assessor in ascending order. The data was then censored such that values larger than 10 were set to 10. This means that the assessors rate the products on a continuous scale from 0 to 10. The rounded data matrix is shown in appendix D.

This data set was used as a basis for all simulations reported in this paper. For design d_3 of section 4.2 for example, we were using the first 15 columns and the last 3 rows of this data.

We also simulated normal data with additive period effects. Since the row-column model (2) corrects for additive period effects, we get very good approximations to the theoretical results. Also, in the case of normal data, the randomization t-statistics are actually t distributed. The non-additive non-normal data are more informative as they also show how good our method of analysis works if the model assumptions are violated. Therefore we will only discuss the results we get when making more realistic assumptions about the nature of the data.

4.2 Analysis of Randomization Distributions

In the following we will arbitrarily focus on the elementary contrast $\ell'\tau = \tau_1 - \tau_t$, where t again is the number of treatments in the starting design d, i.e. $\ell = (1, 0, \ldots, 0, -1)'$. We will randomize treatment labels and assessors and thereby generate $n = 10000$ designs $d^{(i)}$, $i = 1, \ldots, 10000$. We then compute an estimate $\ell'\hat{\tau}^{(i)}$ of the treatment contrast in the row-column model (2) for every permutation $i = 1, \ldots, 10000$ of d. Note that the data remains unchanged in each permutation as long as there are no treatment effects at all. If we want to simulate direct treatment effects or first order carryover effects, the data has to be edited according to the permutation of the starting design.

To check the validity of the design we need to compute estimates for the mean and variance of the contrast estimates under the randomization distribution δ as well as an estimate of the mean of the variance estimate in model (2). We estimate $\mathrm{E}_\delta(\ell'\hat{\tau})$ by the arithmetic mean $\tilde{\mathrm{E}}_\delta(\ell'\hat{\tau})$ of the estimated contrasts. A good estimate for $\mathrm{var}_\delta(\ell'\hat{\tau})$ is the empirical variance

$$\mathrm{v\tilde{a}r}_\delta(\ell'\hat{\tau}) = \frac{1}{n-1} \sum_{i=1}^{n} \left(\ell'\hat{\tau}^{(i)} - \frac{1}{n} \sum_{i=1}^{n} \ell'\hat{\tau}^{(i)} \right)^2$$

of the estimated contrasts. Our estimate for $\mathrm{E}_\delta\left\{\mathrm{v\hat{a}r}(\ell'\hat{\tau})\right\}$ will be the arithmetic mean $\tilde{\mathrm{E}}_\delta\left\{\mathrm{v\hat{a}r}(\ell'\hat{\tau})\right\}$ of the estimated variances of the estimated contrasts $\hat{\tau}_1^{(i)} - \hat{\tau}_5^{(i)}$, $i = 1, \ldots, 10000$.

If (3) holds, then $\tilde{\mathrm{E}}_\delta(\ell'\hat{\tau})$ should be close to $\ell'\tau$. If (4) holds, $\mathrm{v\tilde{a}r}_\delta(\ell'\hat{\tau})$ should be close to $\tilde{\mathrm{E}}_\delta\left\{\mathrm{v\hat{a}r}(\ell'\hat{\tau})\right\}$.

It follows from the central limit theorem that

$$Z = \sqrt{n} \frac{\tilde{\mathrm{E}}_\delta(\ell'\hat{\tau}) - \ell'\tau}{\mathrm{v\tilde{a}r}_\delta(\ell'\hat{\tau})} \sim \mathrm{t}_{9999} \approx N(0, 1),$$

if indeed $\mathrm{E}_\delta(\ell'\hat{\tau}) = \ell'\tau$. We may reject the null hypothesis that (3) holds for model (2) if $|Z|$ is larger than $u_{0.975} = 1.96$.

Since we will only consider designs where all treatment contrasts are estimable, we don't have to do this test as long as there are no carryover effects.

In order to assess whether (4) holds for our model, we compute a confidence interval for

$$\mathrm{var}_\delta(\ell'\hat{\tau}) - \mathrm{E}_\delta\left\{\mathrm{v\hat{a}r}(\ell'\hat{\tau})\right\}. \tag{6}$$

9

For the construction of the confidence interval we divide the set of permutation estimates for the contrasts, $\ell'\hat\tau^{(i)}$, and the variance estimates of the contrasts, $\mathrm{v\hat{a}r}(\ell'\hat\tau)^{(i)}$, $i = 1, \ldots, 10000$, into 100 groups of 100 statistics each. For every group j we then compute the empirical variance $S_{\ell'\hat\tau}^{2\,[j]}$ of the contrast estimates and the arithmetic mean $\bar{X}_{v\hat{a}r(\ell'\hat\tau)}^{[j]}$ of the variance estimates. If the randomization is valid, the differences

$$\lambda^{[j]} = S_{\ell'\hat\tau}^{2\,[j]} - \bar{X}_{v\hat{a}r(\ell'\hat\tau)}^{[j]}, \; j = 1, \ldots, 100$$

are unbiased i.i.d. estimates for (6). Thus we can use the central limit theorem to show that

$$(\bar{X}_{\lambda^{[j]}} - u_{0.975}\sqrt{S_{\lambda^{[j]}}^2/100}, \bar{X}_{\lambda^{[j]}} + u_{0.975}\sqrt{S_{\lambda^{[j]}}^2/100}) \tag{7}$$

is an approximate 95% confidence interval for (6). Here, $\bar{X}_{\lambda^{[j]}}$ and $S_{\lambda^{[j]}}^2$ denote the arithmetic mean and empirical variance of the values of $\lambda^{[j]}$.

Another way to demonstrate the validity and usefulness of a design by way of simulation is to compute the randomization t-statistics

$$\mathrm{t}^{(i)} = \frac{\ell'\hat\tau^{(i)}}{\sqrt{\mathrm{v\hat{a}r}(\ell'\hat\tau)^{(i)}}}, \; i = 1, \ldots, 10000. \tag{8}$$

If the design is valid and if there are no product effects at all, then these permutation t-statistics are asymptotically t distributed with ν degrees of freedom, where ν is the degrees of freedom associated with estimating treatment contrasts in the row-column model, i.e. $\nu = (b-1)(k-1) - t + 1$. This will be assessed using probability function plots. Kunert (1998) also uses the following fact to assess the validity of the assumed model. The number L of permutation t-statistics smaller than $t_{\nu,0.05}$, the 5% quantile of the t distribution with ν degrees of freedom, is binomial with expectation 10000π and variance $10000\pi(1-\pi)$. If the permutation t-statistics were in fact t distributed with ν degrees of freedom, then $\pi = 0.05$ and

$$\frac{L - 500}{\sqrt{475}} \overset{approx.}{\underset{\sim}{}} N(0,1), \tag{9}$$

It follows that values of L smaller than 458 or larger than 542 indicate at the 5% level that the permutation statistics do not follow the hypothesized t distribution. However, it should be noted that for some data generating processes the asymptotic distribution of the permutation t-statistics is not equal to the t distribution. Since the test above assumes that the permutation t-statistics are t distributed we may get misleading results if, for example, we use data with an extreme outlier. Only the results for the mean and variance of the contrasts hold independently of the data.

10

4.3   Example 1: A carryover balanced Generalized Latin Square

Kunert (1998) demonstrates that the Generalized Youden Design d_1, where

$$
d_1 \quad = \quad
\begin{pmatrix}
1 & 2 & 3 & 4 & 5 & 3 & 4 & 5 & 1 & 2 \\
5 & 1 & 2 & 3 & 4 & 4 & 5 & 1 & 2 & 3 \\
2 & 3 & 4 & 5 & 1 & 2 & 3 & 4 & 5 & 1 \\
4 & 5 & 1 & 2 & 3 & 5 & 1 & 2 & 3 & 4 \\
3 & 4 & 5 & 1 & 2 & 1 & 2 & 3 & 4 & 5
\end{pmatrix},
$$

validates the row-column model (2) if there aren't any product effects and treatment labels and assessors are randomized.

We redo the analysis with our data and generate 10000 designs $d_1^{(1)}, \ldots, d_1^{(10000)}$ by randomizing product labels 1 through 5 and by randomizing assessors 1 through 10. For $\tau = \rho = 0$ we have $\tilde{\mathrm{E}}_\delta(\ell'\hat{\tau}) = -0.0018$.

$\tilde{\mathrm{var}}_\delta(\ell'\hat{\tau}) = 0.2446$ is close enough to $\tilde{\mathrm{E}}\{\hat{\mathrm{var}}(\ell'\hat{\tau})\} = 0.2486$ to demonstrate that the second condition for the validity of the design is fulfilled. This is because the realization $(-0.0110, 0.0024)$ of (7) includes 0.

Another indication that the use of the design d_1 is valid is given in figure 1a. The empirical cumulative distribution function of the t-statistics (8) almost perfectly fits the cumulative distribution function of the t distribution with $\nu = 32$ degrees of freedom. For example, $L = 476 \in [458, 542]$ of the 10000 t-statistics are smaller than $t_{32, 0.05}$. According to the remark after (9) we do not reject the hypothesis that the true proportion is 0.05.

With exactly the same procedures we analyze the case where the first product has a direct effect of -1. $H_0 : \ell'\tau = 0$ is no longer true. $L = 6154$ of the permutation t-statistics are smaller than $t_{32, 0.05}$. This means that the empirical power of the test of the hypothesis $H_0 : \ell'\tau = 0$ vs. $H_1 : \ell'\tau < 0$ is 0.6154.

The validity however is destroyed if there are first order carryover effects. This was also shown in Kunert (1998). Let's consider the case where product 2 has a carryover effect of 10 and there are no other direct or carryover effects. Then our estimates $\ell'\hat{\tau}^{(i)}$ don't change and we still get $\tilde{\mathrm{E}}_\delta(\ell'\hat{\tau}) = -0.0018$. According to proposition 3 we get a bias of $-\ell'\rho/k = -(1, 0, 0, 0, -1)(0, 10, 0, 0, 0)'/5 = 0$, i.e. the estimate is unbiased. Also, since d_1 is carryover balanced and the first $k - 1$ periods of d_1 also form a Generalized Youden Design, the variance of the estimate does not increase. This is verified in the simulation, where we

11

get $\tilde{\text{var}}_\delta(\ell'\hat{\tau}) = 0.2446$. However, the estimate of the variance of the contrast estimate is increased. Therefore, the test of $H_0 : \ell'\tau = 0$ vs. $H_1 : \ell'\tau < 0$ becomes conservative. There is no $\text{t}^{(i)}$ which is larger than $\text{t}_{32,\,0.05}$. This is also shown in figure 1b.
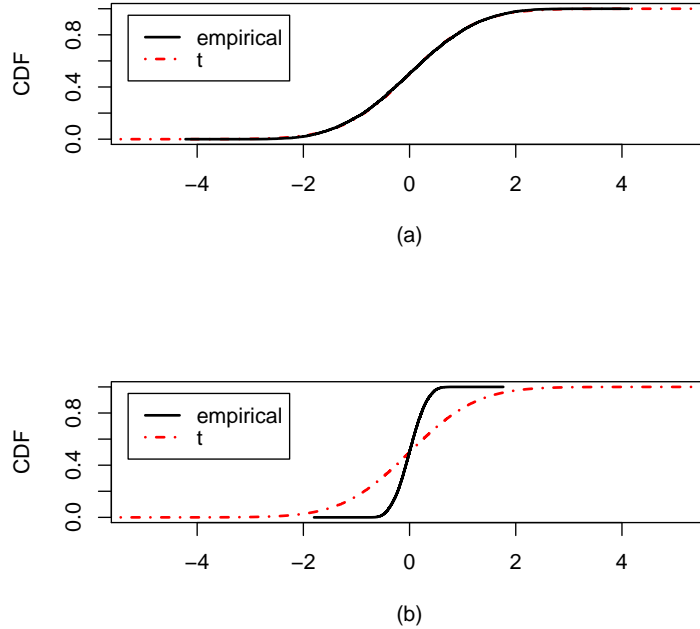

Figure 1: *Empirical and theoretical CDF of the permutation t-statistics for the starting design d_1 (a) if there are no product effects and (b) for $\rho_2 = 10$.*

If we add an additional direct treatment effect of $\tau = -1$, the power of the test of $H_0 : \ell'\tau = 0$ vs. $H_1 : \ell'\tau \neq 0$ via the permutation t-statistics is reduced compared to the situation without carryover effects. This is also because the variance estimate increases while the true variance of the contrast estimate does not increase. Our model is no longer valid. This result is supported by the fact that the confidence interval (7) takes the value $(-3.6110, -3.5910)$ and doesn't include 0.

The validity is also destroyed by carryover effects of treatments occurring in the treatment contrasts. In addition to the situation above we also get a biased estimate of $\ell'\tau$. If there are no treatment effects at all and $\rho = (10, 0, 0, 0, 0)'$, we get $\tilde{\text{E}}_\delta(\ell'\hat{\tau}) = -2.0018$, $\tilde{\text{var}}_\delta(\ell'\hat{\tau}) = 0.2446$ and $\tilde{\text{E}}_\delta\{\hat{\text{var}}(\ell'\hat{\tau})\} = 3.8433$. This again perfectly fits into the theoretical results of proposition 3. Since the variance estimate is way too large, only 48 of the 10000 t-statistics are smaller than the 5% quantile of the t_{32}-distribution although the contrast estimates are biased.

Finally, if we add a direct effect of $\tau_1 = -1$ to the data we get an empirical power of detecting this departure of the null hypothesis of 0.3005. This is because the bias introduced by the carryover effect of product 1 draws the t-statistics farther from zero compared to the situation where there is no carryover effect.

The results for design d_1 are summarized in table 1.

	τ_1	ρ_1	ρ_2	Proportion of $t^{(i)} < t_{32,0.05}$	Mean of $\ell'\hat{\tau}_{(3)\,i}$	$\ell'\tau$	Emp. var. of $\ell'\hat{\tau}_{(3)\,i}$	Mean of $\hat{\text{var}}(\ell'\hat{\tau})_i$
Case 1	0	0	0	0.0476	-0.0018	0	0.2446	0.2486
Case 2	-1	0	0	0.6154	-1.0018	-1	0.2446	0.2486
Case 3	0	0	10	0.0000	-0.0018	0	0.2446	3.8453
Case 4	-1	0	10	0.0000	-1.0018	-1	0.2446	3.8453
Case 5	0	10	0	0.0048	-2.0018	0	0.2446	3.8433
Case 6	-1	10	0	0.3005	-3.0018	-1	0.2446	3.8433

Table 1: *Results for 10000 permutations of the starting design d_1.*

4.4   Example 2: A nearly carryover balanced GYD

Let us now take a look at a design that fulfils the conditions of proposition 2. We obtain d_2 from design d_1 by deleting the last assessor. This design is no longer a Generalized Youden Design and it is not carryover balanced any more.

Table 2 shows that the design is valid if there are no product effects at all. The realization of the confidence interval (6) is $(-0.0082, 0.0067)$ which includes 0. That supports our claim that the variance estimate is unbiased.

Neither the mean of the variance estimate nor the variability of the contrast estimate change if we add a direct effect of -1. If carryover effects are present, the variance of the estimated contrast changes. This is due to the fact that d_2 is not carryover balanced. For example, product 3 follows product 1 twice, whereas product 4 follows product 1 only once. If we have $\rho_2 = 10$, the variability of $\ell'\tau$ is about three times as large as in case 1. It is about twice as large as in case 1 if we have a carryover effect of 10 on the first product. Since neither the values for the contrast estimates nor the values for the variance estimates change dramatically compared to those of design d_1, we get similar results for the proportion of

13

t-statistics smaller than the respective critical values. The power in cases 2 and 6 is a little smaller than the power for design d_1. But that may be the case just because we have one assessor less to observe the difference in treatments.

These results suggest that d_2 is almost as good as d_1 for practical purposes even if there are carryover effects.

	τ_1	ρ_1	ρ_2	Proportion of $t^{(i)} < t_{28,0.05}$	Mean of $\ell'\hat{\tau}_{(3)\,i}$	$\ell'\tau$	Emp. var. of $\ell'\hat{\tau}_{(3)\,i}$	Mean of $\hat{\mathrm{var}}(\ell'\hat{\tau})_i$
Case 1	0	0	0	0.0471	-0.0075	0	0.3104	0.3103
Case 2	-1	0	0	0.5343	-1.0075	-1	0.3104	0.3103
Case 3	0	0	10	0.0000	-0.0159	0	0.9009	4.3654
Case 4	-1	0	10	0.0017	-1.0159	-1	0.9009	4.3654
Case 5	0	10	0	0.0199	-2.0100	0	0.5438	4.3558
Case 6	-1	10	0	0.2555	-3.0100	-1	0.5438	4.3558

Table 2: *Results for 10000 permutations of the starting design d_2.*

4.5 Example 3: A nearly balanced design by Périnel and Pagès (2004)

The design d_3, where

$$d_3 = \begin{pmatrix} 2 & 10 & 7 & 3 & 8 & 4 & 9 & 1 & 2 & 2 & 4 & 1 & 8 & 6 & 9 & 7 & 6 & 5 & 5 & 10 & 3 & 7 & 3 & 6 & 1 & 4 & 10 & 9 & 1 & 8 \\ 3 & 2 & 4 & 9 & 4 & 7 & 10 & 9 & 6 & 10 & 1 & 7 & 2 & 8 & 1 & 8 & 5 & 3 & 6 & 3 & 10 & 5 & 1 & 5 & 9 & 7 & 4 & 8 & 2 & 6 \\ 5 & 9 & 8 & 4 & 3 & 2 & 7 & 5 & 4 & 1 & 10 & 9 & 5 & 3 & 6 & 6 & 8 & 1 & 10 & 7 & 2 & 1 & 8 & 9 & 3 & 10 & 6 & 2 & 4 & 7 \\ 7 & 7 & 10 & 5 & 6 & 8 & 1 & 6 & 10 & 8 & 9 & 3 & 4 & 2 & 2 & 3 & 9 & 4 & 1 & 5 & 9 & 4 & 7 & 2 & 8 & 6 & 1 & 5 & 3 & 10 \\ 6 & 3 & 5 & 10 & 9 & 1 & 4 & 7 & 8 & 5 & 6 & 8 & 9 & 1 & 7 & 4 & 10 & 2 & 3 & 2 & 8 & 6 & 10 & 4 & 5 & 2 & 3 & 7 & 9 & 5 \end{pmatrix},$$

was suggested by Périnel and Pagès (2004). Their design does not fulfil the conditions of proposition 1, it is not even a balanced incomplete block design.

Does the simulation suggest that it is not valid in case 1? The realization of (6) is close to zero. The confidence interval for the difference between the contrast variance and the expectation of its estimate is $(-0.0072, 0.0018)$. So there is no indication that the design may be invalid. If it isn't valid, than this is of no practical effect if there are no carryover effects.

14

In case 3 the mean of the variance estimates for the contrast is 7 times as high as in case 1 while the variance of the contrast estimate increases only by as factor of 4. We overestimate the true variance of $\ell'\tau$ although the design is not fully carryover balanced. As in the analysis of design d_2, the t-test based on (9) is conservative. The CDF plots for design d_3, cases 1 and 3 in figures 2a and 2b below, are very similar to the respective figures for d_1.

In case 5, our estimate of the contrast is biased. The mean of $\ell'\tau = -1.9865$ is close to $-10/k = -2$. Although we once again overestimate the variance of the contrast, we falsely reject the hypothesis that $\ell'\tau = 0$ against the alternative that $\ell'\tau < 0$ too often. Because of the huge bias, 51% of the permutation t-statistics are less than $t_{107,\,0.05}$.

The results for d_3 are summarized in table 3.
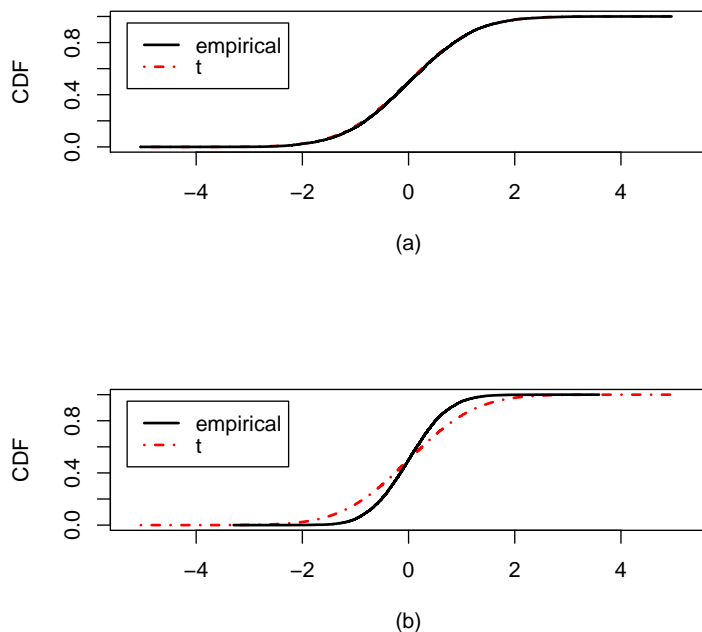

(a)

(b)

Figure 2: *Empirical and theoretical CDF of the permutation t-statistics for the starting design d_3 (a) if there are no product effects and (b) for $\rho_2 = 10$.*

4.6 Example 4: An unbalanced design

In order to examine what happens if we use designs that clearly aren't uniform on the periods, let's consider the design d_4 where

$$
d_4 \;=\; \begin{pmatrix}
1 & 1 & 1 & 1 & 1 & 1 & 1 & 3 & 2 & 2 & 2 & 2 & 4 & 5 & 4 \\
2 & 2 & 2 & 2 & 3 & 3 & 3 & 1 & 3 & 3 & 3 & 3 & 5 & 4 & 5 \\
4 & 4 & 5 & 5 & 4 & 4 & 5 & 5 & 4 & 4 & 5 & 5 & 1 & 2 & 3
\end{pmatrix}.
$$

15

	τ_1	ρ_1	ρ_2	Proportion of $t^{(i)} < t_{107,0.05}$	Mean of $\ell'\hat{\tau}_{(3)\,i}$	$\ell'\tau$	Emp. var. of $\ell'\hat{\tau}_{(3)\,i}$	Mean of $\hat{\text{var}}(\ell'\hat{\tau})_i$
Case 1	0	0	0	0.0466	0.0090	0	0.1658	0.1682
Case 2	-1	0	0	0.7728	-0.9910	-1	0.1658	0.1682
Case 3	0	0	10	0.0028	0.0062	0	0.5201	1.3896
Case 4	-1	0	10	0.0925	-0.9938	-1	0.5201	1.3896
Case 5	0	10	0	0.5096	-1.9865	0	0.3160	1.3895
Case 6	-1	10	0	0.9707	-2.9865	-1	0.3160	1.3895

Table 3: *Results for 10000 permutations of the starting design d_3.*

Note that d_4 is not even a balanced incomplete block design. The rows of the first 12 columns of d_4 are not even connected. The design is also not carryover balanced as treatment 2 follows treatment 1 four times whereas treatment 1 never follows treatment 2. However, we don't reject (3). The confidence interval for (6) is (0.0164, 0.0609), which does not include 0. Therefore we conclude that the design is not valid because the second condition, (4), is violated significantly at the 5% level. Still, the variance estimates don't differ much. We underestimate the variance by approximately 5%. This is why the empirical cumulative distribution function of the t-statistics almost coincides with the CDF of the t_{24}-distribution. This is shown in figure 3a.

In case 3 the variance of the contrast estimate gets inflated by the carryover effect of product two. This is because d_4 is not carryover balanced. The empirical variance of our samples of $\ell'\tau_{(2)}$ is about 8 times as high as in case 1. The average value of the least squares variance estimate is only about 60% of the empirical variance of the contrast estimates. This means that we underestimate the true variance and the test of $H_0 : \ell'\tau = 0$ vs. $H_1 : \ell'\tau < 0$ becomes anti-conservative. The empirical α-level of the test is $0.089 > 0.050$ as 890 of the 10000 permutation t-statistics are less than the critical value of $t_{24,0.05}$. Note that in figure 3b the empirical CDF lies above the CDF of the t-distribution for $x < 0$ whereas in figures 1b and 2b the empirical CDF lies below the theoretical CDF.

Finally in case 5, we also get an anti-conservative t-test. The reason for this is that the contrast estimate is biased by $-10/k = -3.3333$ as indicated by the mean of the contrast estimates which is -3.3571. As long as there are no carryover effects, the randomization may not validate the use of the design but this might still be of little practical interest. However, if there are carryover effects then the use of highly unbalanced designs like this one is discouraged.

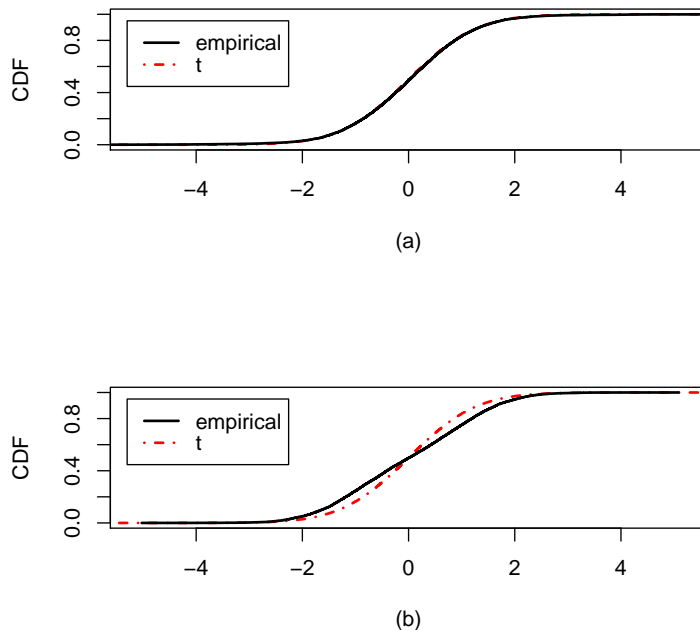The results for d_4 are summarized in table 4.

(a)

(b)

Figure 3: *Empirical and theoretical CDF of the permutation t-statistics for the starting design d_4 (a) if there are no product effects and (b) for $\rho_2 = 10$.*

	τ_1	ρ_1	ρ_2	Proportion of $t^{(i)} < t_{24,0.05}$	Mean of $\ell'\hat{\tau}_{(3)\,i}$	$\ell'\tau$	Emp. var. of $\ell'\hat{\tau}_{(3)\,i}$	Mean of $\hat{\text{var}}(\ell'\hat{\tau})_i$
Case 1	0	0	0	0.0492	0.0121	0	0.7084	0.6699
Case 2	-1	0	0	0.3104	-0.9879	-1	0.7084	0.6699
Case 3	0	0	10	0.0890	0.0072	0	5.6120	3.4831
Case 4	-1	0	10	0.2054	-0.9928	-1	5.6120	3.4831
Case 5	0	10	0	0.5078	-3.3571	0	3.8175	3.4567
Case 6	-1	10	0	0.7324	-4.3571	-1	3.8175	3.4567

Table 4: *Results for 10000 permutations of the starting design d_4.*

4.7 Example 5: A carryover balanced design for which proposition 3 does not hold

There is one additional design of interest in this study. The design d_5, where

$$
d_5 = \begin{pmatrix}
1 & 2 & 3 & 4 & 1 & 2 & 3 & 4 & 1 & 2 & 3 & 4 \\
2 & 3 & 4 & 1 & 3 & 4 & 1 & 2 & 2 & 1 & 4 & 3 \\
4 & 1 & 2 & 3 & 2 & 3 & 4 & 1 & 3 & 4 & 1 & 2
\end{pmatrix}
$$

is a Generalized Youden Design which is also carryover balanced yet does not fulfil the conditions of proposition 3. Three assessors get treatments 1 and 2 in the first two periods of d_5 but only one assessor gets treatments 1 and 4 in the first two periods.

As can be seen from table 5 the empirical variance is larger in cases 3–6 than in the case with no carryover effects. But since the design is carryover balanced the variance does not get inflated much as was the case with d_4 for example and the t-test is still conservative in case 3.

	τ_1	ρ_1	ρ_2	Proportion of $t^{(i)} < t_{19,0.05}$	Mean of $\ell'\hat{\tau}_{(3)\,i}$	$\ell'\tau$	Emp. var. of $\ell'\hat{\tau}_{(3)\,i}$	Mean of $\hat{\mathrm{var}}(\ell'\hat{\tau})_i$
Case 1	0	0	0	0.0500	0.0005	0	0.6028	0.6019
Case 2	-1	0	0	0.3331	-0.9995	-1	0.6028	0.6019
Case 3	0	0	10	0.0000	0.0071	0	0.7665	4.3203
Case 4	-1	0	10	0.0027	-0.9929	-1	0.7665	4.3203
Case 5	0	10	0	0.3896	-3.3307	0	0.6636	4.2956
Case 6	-1	10	0	0.7749	-4.3307	-1	0.6636	4.2956

Table 5: *Results for 10000 permutations of the starting design d_5.*

5 Conclusions

The simulation study demonstrates that the row-column model (2) is valid for Generalized Youden Designs if model (1) holds. It may be valid even if we don't use Generalized Youden Designs as long as the conditions aren't too badly violated. Even if our analysis strongly suggests that the row-column model (2) is invalid for a design, the t-test can still work in practice if the differences between the randomization distribution of the t-statistics and the model t distribution are small.

If there are carryover effects of treatments, we may get a conservative analysis even if we don't use Generalized Youden Designs that are carryover balanced. If there are carryover effects of products that do not cancel out, we are likely to get into trouble whether we use balanced designs or not.

In many practical situations the use of Generalized Youden Designs, especially of those that are also balanced for carryover, may be too restrictive. Designs which are constructed from such designs by omitting a few blocks (assessors) may be applied in such situations without getting extremely wrong results.

The extent of damage that is caused by carryover effects should always be reduced by both non-statistical methods also. Any carryover effects should be minimized by the way the experiment is conducted. Then one may want to use a balanced or nearly balanced design together with the randomization of row and treatment labels. If there are no carryover effects, then the analysis which ignores carryover effects is valid or nearly valid. If there are carryover effects then the additional variance due to carryover effects is reduced by a large degree.

References

Azaïs, J. M. and Druilhet, P. (1997). Optimality of neighbour balanced designs when neighbour effects are neglected, *Journal of Statistical Planning and Inference* **64**: 353–367.

Bailey, R. A. (1985). Restricted randomization for neighbour-balanced designs, *Statistics and Decisions Supplement* **2**: 237–248.

Bailey, R. A. and Rowley, C. A. (1987). Valid randomization, *Proceedings of the Royal Society London A* **410**: 105–124.

Ball, R. D. (1997). Incomplete block designs for the minimisation of order and carry-over effects in sensory analysis, *Food Quality and Preference* **8**(2): 111–118.

Dyke, G. V. (1983). Discussion of a paper by wilkinson et al, *Journal of the Royal Statistical Society B* **45**: 191–192.

Kunert, J. (1984). Optimality of balanced uniform repeated measurements designs, *Annals of Statistics* **12**: 1006–1017.

Kunert, J. (1985). Optimal experimental design when the errors are assumed to be correlated, *Statistics and Decisions Supplement* **2**: 287–298.

Kunert, J. (1998). Sensory experiments as crossover studies, *Food Quality and Preference* **9**: 243–253.

Kunert, J. and Utzig, B. (1993). Estimation of variance in cross-over designs, *Journal of the Royal Statistical Society B* **55**: 919–927.

Kunert, J., Meyners, M. and Erdbrügge, M. (2002). On the applicability of anova for the analysis of sensory data, 7è Journées Européennes Agro-Industrie et Méthodes Statistiques, pp. 129–134.

Mac Fie, H. J. H., Greenhoff, K., Bratchell, N. and Vallis, L. (1989). Designs to balance the effect of order of presentation and first-order carry-over effects in hall tests, *Journal of Sensory Studies* **4**: 129–148.

Périnel, E. and Pagès, J. (2004). Optimal nested cross-over designs in sensory analysis, *Food Quality and Preference* **15**: 439–446.

Schlich, P. (1993). Use of change-over designs and repeated measurements in sensory and consumer studies, *Food Quality and Preference* **4**: 223–235.

Wakeling, I. N. and MacFie, H. J. H. (1995). Designing consumer trials for first and higher orders of carry-over effects when only a subset of k samples from p may be tested, *Food Quality and Preference* **6**: 299–308.

A   Proof of Proposition 1

Define $Q_x = I_x - \frac{1}{x}1_x1_x'$. Then correction for row- and column-effects corresponds to multiplication with $Q_b \otimes Q_k$. Hence, in the row-column model (2) a contrast $\ell'\tau$ is estimated by

$\ell'\hat{\tau}$, where

$$\hat{\tau} = C_\delta^+ T_\delta'(Q_b \otimes Q_k)Y,$$

C_δ^+ is the Moore-Penrose generalized inverse of $C_\delta = T_\delta'(Q_b \otimes Q_k)T_\delta$ and $\ell'1_t = 0$. In model (2) the variance of $\ell'\hat{\tau}$ equals $\mathrm{var}(\ell'\hat{\tau}) = \sigma^2 \ell' C_\delta^+ \ell$ and is estimated by $\mathrm{v\hat{a}r}(\ell'\hat{\tau}) = \hat{\sigma}^2 \ell' C_\delta^+ \ell$, with

$$\hat{\sigma}^2 = \frac{1}{(b-1)(k-1)-t+1} \left\{ Y'(Q_b \otimes Q_k)Y - \hat{\tau}'C_\delta\hat{\tau} \right\}.$$

If the starting design d is a Generalized Youden Design that is uniform on the periods, then with our randomization each design δ will be another Generalized Youden Design that is uniform on the periods. It then follows that

$$C_\delta = c_1 Q_t,$$

where $c_1 = \frac{b(k-1)}{t-1}$. Since $Q_t T_\delta'(Q_b \otimes Q_k) = T_\delta'(Q_b \otimes Q_k)$ we get

$$\hat{\tau} = \frac{1}{c_1} T_\delta'(Q_b \otimes Q_k)Y.$$

We explore the performance of the estimates $\ell'\hat{\tau}$ and $\mathrm{v\hat{a}r}(\ell'\hat{\tau})$ in model (1) under the randomization distribution of δ. There are permutation matrices Π and Γ such that $T_\delta = (\Pi \otimes I_k)H_d\Gamma$, see section 2. Thus,

$$
\begin{aligned}
\ell'\hat{\tau} &= \frac{1}{c_1} \ell' T_\delta'(Q_b \otimes Q_k)(T_\delta\tau + \eta) \\
&= \frac{1}{c_1} \ell' \left\{ C_\delta\tau + \Gamma'H_d'(\Pi' \otimes I_k)(Q_b \otimes Q_k)\eta \right\} \\
&= \ell'\tau + \frac{1}{c_1} \ell'\Gamma'H_d'(\Pi'Q_b \otimes Q_k)\eta \quad\quad (10)
\end{aligned}
$$

and

$$\mathrm{E}_\delta(\ell'\hat{\tau}) = \ell'\tau + \frac{1}{t!b!c_1} \ell' \sum_\Gamma \sum_\Pi \Gamma'H_d'(\Pi'Q_b \otimes Q_k)\eta.$$

For any t-dimensional vector x we have $\frac{1}{t!} \sum_\Gamma \Gamma'x = \frac{1}{t}1_t 1_t'x$. This implies that

$$
\begin{aligned}
\mathrm{E}_\delta(\ell'\hat{\tau}) &= \ell'\tau + \frac{1}{tb!c_1} \ell'1_t 1_t' \sum_\Pi H_d'(\Pi'Q_b \otimes Q_k)\eta \\
&= \ell'\tau. \quad\quad (11)
\end{aligned}
$$

To calculate $\mathrm{var}_\delta(\ell'\hat{\tau})$ we use (10) and get

$$\mathrm{var}_\delta(\ell'\hat{\tau}) = \frac{1}{t!b!c_1^2} \ell' \sum_\Gamma \sum_\Pi \Gamma'H_d'(\Pi'Q_b \otimes Q_k)\eta\eta'(Q_b\Pi \otimes Q_k)H_d\Gamma\ell.$$

21

In order to simplify this expression we define

$$S_\eta = \frac{1}{b-1} \left\{ \sum_{i=1}^{b} Q_k \eta_i \eta_i' Q_k - \frac{1}{b} \left(\sum_{i=1}^{b} Q_k \eta_i \right) \left(\sum_{i=1}^{b} Q_k \eta_i \right)' \right\}.$$

It can be shown that

$$\operatorname{tr} S_\eta = \eta'(Q_b \otimes Q_k)\eta/(b-1) \tag{12}$$

and

$$\frac{1}{b!} \sum_{\Pi} (\Pi' Q_b \otimes Q_k)\eta\eta'(Q_b\Pi \otimes Q_k) = Q_b \otimes S_\eta. \tag{13}$$

Here, (12) follows from

$$
\begin{aligned}
(b-1) \operatorname{tr} S_\eta &= \operatorname{tr} \left\{ \sum_{i=1}^{b} \eta_i' Q_k \eta_i - \frac{1}{b} \left(\sum_{i=1}^{b} \eta_i' \right) Q_k \left(\sum_{i=1}^{b} \eta_i \right) \right\} \\
&= \operatorname{tr} \left\{ \eta'(Q_b \otimes Q_k)\eta \right\}.
\end{aligned}
$$

Recall that Π is a permutation matrix. If we write $\Pi'\eta = [\eta_{\Pi^{-1},1}', \ldots, \eta_{\Pi^{-1},b}']'$ we get

$$\frac{1}{b!} \sum_{\Pi} (\Pi' Q_b \otimes Q_k)\eta\eta'(Q_b\Pi \otimes Q_k)$$

$$= \frac{1}{b!} \sum_{\Pi} \left(\left(\left(\Pi' - \frac{1}{b}1_b 1_b' \right) \otimes Q_k \right) \begin{bmatrix} \eta_1 \\ \vdots \\ \eta_b \end{bmatrix} [\eta_1', \ldots, \eta_b'] \left(\left(\Pi - \frac{1}{b}1_b 1_b' \right) \otimes Q_k \right) \right)$$

$$= \frac{1}{b!} \sum_{\Pi} \left\{ \begin{bmatrix} Q_k \eta_{\Pi^{-1},1} \\ \vdots \\ Q_k \eta_{\Pi^{-1},b} \end{bmatrix} [\eta_{\Pi^{-1},1}' Q_k, \ldots, \eta_{\Pi^{-1},b}' Q_k] \right.$$

$$- \frac{1}{b} \left[1_b \otimes \sum_{i=1}^{b} Q_k \eta_i \right] [\eta_{\Pi^{-1},1}' Q_k, \ldots, \eta_{\Pi^{-1},b}' Q_k]$$

$$- \frac{1}{b} \begin{bmatrix} Q_k \eta_{\Pi^{-1},1} \\ \vdots \\ Q_k \eta_{\Pi^{-1},b} \end{bmatrix} \left[1_b' \otimes \sum_{i=1}^{b} \eta_i' Q_k \right]$$

$$\left. + \frac{1}{b^2} 1_b 1_b' \otimes \sum_{i=1}^{b} Q_k \eta_i \left(\sum_{i=1}^{b} \eta_i' Q_k \right) \right\}.$$

22

Summing over all Π we have

$$\frac{1}{b!}\sum_{\Pi}(\Pi'Q_b \otimes Q_k)\eta\eta'(Q_b\Pi \otimes Q_k)$$

$$= \frac{1}{b!}\begin{bmatrix} (b-1)!\sum_{i=1}^{b}Q_k\eta_i\eta_i'Q_k & (b-2)!\sum_i\sum_{j\neq i}Q_k\eta_i\eta_j'Q_k \\ & \ddots & \\ (b-2)!\sum_i\sum_{j\neq i}Q_k\eta_i\eta_j'Q_k & (b-1)!\sum_{i=1}^{b}Q_k\eta_i\eta_i'Q_k \end{bmatrix}$$

$$-2\frac{(b-1)!}{b\,b!}\left(1_b \otimes \sum_{i=1}^{b}Q_k\eta_i\right)\left(1_b' \otimes \sum_{i=1}^{b}\eta_i'Q_k\right) + \frac{1}{b^2}1_b1_b' \otimes \sum_{i=1}^{b}Q_k\eta_i\left(\sum_{i=1}^{b}\eta_i'Q_k\right)$$

$$= \begin{bmatrix} \frac{b}{b(b-1)}\sum_{i=1}^{b}Q_k\eta_i\eta_i'Q_k & \frac{1}{b(b-1)}\sum_{i=1}^{b}Q_k\eta_i\sum_{j=1}^{b}\eta_j'Q_k \\ & \ddots & \\ \frac{1}{b(b-1)}\sum_{i=1}^{b}Q_k\eta_i\sum_{j=1}^{b}\eta_j'Q_k & \frac{b}{b(b-1)}\sum_{i=1}^{b}Q_k\eta_i\eta_i'Q_k \end{bmatrix}$$

$$-\frac{1}{b(b-1)}1_b1_b' \otimes \sum_{i=1}^{b}(Q_k\eta_i\eta_i'Q_k) - \frac{1}{b^2}1_b1_b' \otimes \sum_{i=1}^{b}Q_k\eta_i\left(\sum_{i=1}^{b}\eta_i'Q_k\right)$$

$$= \left(\frac{b-1}{b}1_b1_b' + (I_b - 1_b1_b')\right) \otimes \frac{-1}{b(b-1)}\left(\sum_{i=1}^{b}Q_k\eta_i\right)\left(\sum_{i=1}^{b}Q_k\eta_i\right)'$$

$$+\left(I_b - \frac{b-1}{b(b-1)}1_b1_b'\right) \otimes \frac{1}{b-1}\sum_{i=1}^{b}Q_k\eta_i\eta_i'Q_k$$

$$= Q_b \otimes S_\eta,$$

which completes the proof of (13).

This implies

$$\text{var}_\delta(\ell'\hat\tau) = \frac{1}{t!c_1^2}\,\ell'\sum_{\Gamma}\Gamma'H_d'(Q_b \otimes S_\eta)H_d\Gamma\ell.$$

Condition (GYD1) implies that there are $t \times t$-permutation matrices $\Delta_1,\ldots,\Delta_b$ with

$$H_d = \begin{bmatrix} H_{d1} \\ \vdots \\ H_{db} \end{bmatrix} = \begin{bmatrix} [I_k, 0]\Delta_1 \\ \vdots \\ [I_k, 0]\Delta_b \end{bmatrix},$$

see Kunert and Utzig (1993). Since d is uniform on the periods if follows that

$$\sum_{i=1}^{b}[I_k, 0]\Delta_i = \frac{b}{t}1_k1_t'.$$

23

Therefore,

$$
\begin{aligned}
\mathrm{var}_\delta(\ell'\hat\tau) &= \frac{1}{t!c_1^2}\,\ell'\sum_\Gamma\Gamma'\bigg\{\sum_{i=1}^b\Delta_i'\begin{bmatrix}I_k\\0\end{bmatrix}S_\eta\,[I_k,0]\,\Delta_i-\frac{1}{b}\left(\sum_{i=1}^b\Delta_i'\begin{bmatrix}I_k\\0\end{bmatrix}\right)S_\eta\left(\sum_{i=1}^b[I_k,0]\,\Delta_i\right)\bigg\}\Gamma\ell\\
&= \frac{1}{t!c_1^2}\,\ell'\sum_\Gamma\Gamma'\bigg\{\sum_{i=1}^b\Delta_i'\begin{bmatrix}I_k\\0\end{bmatrix}S_\eta\,[I_k,0]\,\Delta_i\bigg\}\Gamma\ell.
\end{aligned}
$$

For an arbitrary matrix M with row and column sums zero it holds

$$
\frac{1}{t!}\sum_\Gamma\Gamma'M\Gamma=\frac{\mathrm{tr}\,M}{t-1}Q_t.
$$

Note that S_η has row and column sums zero. Together with (12) this implies

$$
\begin{aligned}
\mathrm{var}_\delta(\ell'\hat\tau) &= \frac{1}{c_1^2(t-1)}\,\mathrm{tr}\left(\sum_{i=1}^b\Delta_i'\begin{bmatrix}I_k\\0\end{bmatrix}S_\eta\,[I_k,0]\,\Delta_i\right)\times\ell'Q_t\ell\\
&= \frac{1}{k-1}\,\mathrm{tr}\,S_\eta\times\ell'C_\delta^+\ell\\
&= \frac{1}{(b-1)(k-1)}\,\eta'(Q_b\otimes Q_k)\eta\times\ell'C_\delta^+\ell. \qquad(14)
\end{aligned}
$$

It only remains to show that

$$
\mathrm{E}_\eta(\hat\sigma^2)=\frac{1}{(b-1)(k-1)}\,\eta'(Q_b\otimes Q_k)\eta.
$$

For this calculation we can assume without loss of generality that $\tau=0$. Then

$$
\mathrm{E}_\eta(\hat\sigma^2)=\frac{1}{(b-1)(k-1)-t+1}\,\left\{\eta'(Q_b\otimes Q_k)\eta-\mathrm{E}_\eta\hat\tau'C_\delta\hat\tau\right\}.
$$

Observe that

$$
\begin{aligned}
\mathrm{E}_\eta(\hat\tau'C_\delta\hat\tau) &= \frac{1}{t!b!c_1}\sum_\Gamma\sum_\Pi\eta'(Q_b\Pi\otimes Q_k)H_d\Gamma\Gamma'H_d'(\Pi'Q_b\otimes Q_k)\eta\\
&= \frac{1}{t!b!c_1}\,\mathrm{tr}\sum_\Gamma\sum_\Pi\{\Gamma'H_d'(\Pi'Q_b\otimes Q_k)\eta\eta'(Q_b\Pi\otimes Q_k)H_d\Gamma\}\\
&= \frac{1}{c_1}\,\mathrm{tr}\left(\sum_{i=1}^b\Delta_i'\begin{bmatrix}I_k\\0\end{bmatrix}S_\eta\,[I_k,0]\,\Delta_i\right)\\
&= \frac{t-1}{(b-1)(k-1)}\,\eta'(Q_b\otimes Q_k)\eta
\end{aligned}
$$

24

and get

$$
\begin{aligned}
\mathrm{E}_\eta(\hat{\sigma}^2) &= \frac{1}{(b-1)(k-1)-t+1}\left\{\eta'(Q_b \otimes Q_k)\eta - \frac{t-1}{(b-1)(k-1)}\eta'(Q_b \otimes Q_k)\eta\right\} \\
&= \frac{1}{(b-1)(k-1)}\eta'(Q_b \otimes Q_k)\eta.
\end{aligned}
$$

This completes the proof.

B Proof of Proposition 2

The proof of proposition 2 follows that of proposition 1. Note that for a Generalized Youden Design with $k = t$ we have $C_d = c_2 Q_t$, where

$$
c_2 = b + \frac{(b-ta)\{b-t(a+1)\}}{b(t-1)},
$$

where a is the integer part of b/t. After randomizing treatment labels and assessors we still have a Generalized Youden Design with the same information matrix as before randomization, i.e. $C_\delta = c_2 Q_t$.

To show that $\mathrm{E}(\ell'\hat{\tau}) = \ell'\tau$, it suffices to replace c_1 by c_2 in the proof of proposition 1 since the calculations up to (11) do not require d to be a Generalized Youden Design.

It remains to show that $\mathrm{E}(\hat{\mathrm{var}}_\delta(\ell'\hat{\tau})) = \mathrm{var}_\delta(\ell'\hat{\tau})$. Again, we replace c_1 by c_2 in the proof of proposition 1. Since $k = t$,

$$
H_d = \begin{bmatrix} \Delta_1 \\ \vdots \\ \Delta_b \end{bmatrix},
$$

and $\sum_{i=1}^b \Delta_i = a 1_t 1_t' + Q$ where the elements q_{ij} of Q are 1 if treatment j appears in period i exactly $a+1$ times and 0 else. Note that

$$
QQ' = \begin{bmatrix} b-ta & & \frac{(b-ta)(b-ta-1)}{t-1} \\ & \ddots & \\ \frac{(b-ta)(b-ta-1)}{t-1} & & b-ta \end{bmatrix}
$$

25

is completely symmetric. It follows

$$\operatorname{tr}\left(\sum_i \Delta_i\right)' S_\eta \left(\sum_i \Delta_i\right) = \operatorname{tr} S_\eta \left(\sum_i \Delta_i\right)\left(\sum_i \Delta_i\right)'$$

$$= \operatorname{tr}\left\{ c_3 S_\eta 1_t 1_t' + S_\eta \begin{bmatrix} -\frac{(b-ta)\{b-t(a+1)\}}{t-1} & & 0 \\ & \ddots & \\ 0 & & -\frac{(b-ta)\{b-t(a+1)\}}{t-1} \end{bmatrix} \right\}$$

$$= -\frac{(b-ta)\{b-t(a+1)\}}{t-1} \operatorname{tr} S_\eta$$

for some constant c_3.

From this, we get the variance of our estimate

$$
\begin{aligned}
\operatorname{var}_\delta(\ell'\hat\tau) &= \frac{1}{t!c_2^2} \ell' \sum_\Gamma \Gamma'\left\{ \sum_{i=1}^b \Delta_i' S_\eta \Delta_i - \frac{1}{b}\left(\sum_{i=1}^b \Delta_i'\right) S_\eta \left(\sum_{i=1}^b \Delta_i\right)\right\} \Gamma \ell \\
&= \frac{1}{t!c_2^2} \ell' \frac{c_2 \operatorname{tr} S_\eta}{t-1} Q_t \\
&= \frac{1}{(b-1)(t-1)} \eta'(Q_b \otimes Q_t)\eta \times \ell' C^+ \ell.
\end{aligned}
$$

Since $\operatorname{var}_\delta(\ell'\hat\tau) = \sigma^2 \ell' C_\delta^+ \ell$, we need to show

$$\mathrm{E}_\eta(\hat\sigma^2) = \frac{1}{(b-1)(t-1)} \eta'(Q_b \otimes Q_t)\eta.$$

With the same methods as in the proof of proposition 1 we get

$$\mathrm{E}_\eta(\hat\sigma^2) = \frac{1}{(b-2)(t-1)} \left\{\eta'(Q_b \otimes Q_t)\eta - \mathrm{E}(\hat\tau' C_\delta \hat\tau)\right\}.$$

Here,

$$
\begin{aligned}
\mathrm{E}(\hat\tau' C_\delta \hat\tau) &= \frac{1}{b!t!c_2} \sum_\Gamma \sum_\Pi \eta'(Q_b \Pi \otimes Q_t) H_d \Gamma Q_t \Gamma' H_d'(\Pi' Q_b \otimes Q_t)\eta \\
&= \frac{1}{b!t!c_2} \operatorname{tr}\left\{ \sum_\Gamma \sum_\Pi \Gamma' H_d'(\Pi' Q_b \otimes Q_t)\eta\eta'(Q_b \Pi \otimes Q_t) H_d \Gamma \right\} \\
&= \operatorname{tr} S_\eta.
\end{aligned}
$$

Finally

$$
\begin{aligned}
\mathrm{E}_\eta(\hat\sigma^2) &= \frac{1}{(b-2)(t-1)} \left\{\eta'(Q_b \otimes Q_t)\eta - \frac{1}{b-1}\eta'(Q_b \otimes Q_t)\eta\right\} \\
&= \frac{1}{(b-1)(t-1)} \eta'(Q_b \otimes Q_t)\eta.
\end{aligned}
$$

This completes the proof of proposition 2.

C Proof of Proposition 3

We use the randomization distribution of δ to consider the performance of the estimate $\ell'\hat{\tau}$ under model (5). Then

$$\mathrm{E}_\delta(\ell'\hat{\tau}) = \mathrm{E}_\delta\left\{\frac{1}{c_1}\ell'T_\delta'(Q_b\otimes Q_k)(T_\delta\tau + F_\delta\rho + \eta)\right\}$$

and from (10) and (11) we have

$$\mathrm{E}_\delta\left\{\frac{1}{c_1}\ell'T_\delta'(Q_b\otimes Q_k)(T_\delta\tau + \eta)\right\} = \ell'\tau.$$

The bias equals

$$\mathrm{E}_\delta\left\{\frac{1}{c_1}\ell'T_\delta'(Q_b\otimes Q_k)(I_b\otimes W)T_\delta\rho\right\} = \frac{1}{t!c_1}\ell'\sum_\Gamma\Gamma'H_d'(Q_b\otimes Q_kW)H\Gamma\rho.$$

Note that

$$H_d'(Q_b\otimes Q_kW)H_d = H_d'(I_b\otimes Q_kW)H_d - \frac{1}{b}H_d'(1_b1_b'\otimes Q_kW)H_d.$$

Since d is uniform on the periods, we have

$$H_d'(1_b1_b'\otimes Q_kW)H_d = \left(\sum_{i=1}^b H_{di}'\right)Q_kW\left(\sum_{i=1}^b H_{di}\right) = 0.$$

Let $D_{d12} = H_d'(I_b\otimes Q_kW)H_d$. Then

$$D_{d12} = M_d - \frac{1}{k}\sum_{i=1}^b\begin{bmatrix} n_{di1}\tilde{n}_{di1} & & n_{di1}\tilde{n}_{dit} \\ & \ddots & \\ n_{dit}\tilde{n}_{d11} & & n_{dit}\tilde{n}_{dit} \end{bmatrix}. \tag{15}$$

Here M_d is the matrix whose entries m_{dij} indicate how often in d treatment i is preceded by treatment j and n_{dij} indicates how often assessor i receives treatment j. $\tilde{n}_{dij}$ indicates how often assessor i receives treatment j in periods $1,\ldots,k-1$. Note that there is no run-in period and no product can be preceded by itself.

We have $\mathrm{tr}\, D_{d12} = -b(k-1)/k$ and for $r = 1,\ldots,k$ it holds that

$$\sum_{i=1}^b n_{dir}\tilde{n}_{di1} + \ldots + n_{dir}\tilde{n}_{dit} = k(k-1)\frac{b}{t}\quad\text{and}$$

$$\sum_{i=1}^b n_{di1}\tilde{n}_{dir} + \ldots + n_{dit}\tilde{n}_{dir} = k(k-1)\frac{b}{t}.$$

27

This implies that D_{d12} has row and column sums zero.

Thus the bias equals

$$\frac{1}{t!c_1}\ell'\sum_{\Gamma}\Gamma'D_{d12}\Gamma\rho = -\frac{1}{k}\ell'\rho.$$

To calculate the variance we can once again assume without loss of generality that $\tau = 0$. We get

$$
\begin{aligned}
\mathrm{E}_\delta(\ell'\hat{\tau})^2 &= \frac{1}{c_1^2}\mathrm{E}_\delta\left\{\ell'T_\delta'(Q_b\otimes Q_k)(F_\delta\rho+\eta)(F_\delta\rho+\eta)'(Q_b\otimes Q_k)T_\delta\ell\right\}\\
&= \frac{1}{c_1^2}\mathrm{E}_\delta\ell'\Big\{D_{\delta12}\rho\rho'D_{\delta12}' + D_{\delta12}\rho\eta'(Q_b\otimes Q_k)T_\delta\\
&\quad + T_\delta'(Q_b\otimes Q_k)\eta\rho'D_{\delta12}' + T_\delta'(Q_b\otimes Q_k)\eta\eta'(Q_b\otimes Q_k)T_\delta\Big\}\ell,
\end{aligned}
$$

where $D_{\delta12} = T_\delta'(Q_b\otimes Q_k)F_\delta = \Gamma'D_{d12}\Gamma$.

From (14) we obtain that

$$\frac{1}{c_1^2}\mathrm{E}_\delta\left\{\ell'T_\delta'(Q_b\otimes Q_k)\eta\right\}^2 = \frac{1}{(b-1)(k-1)}\eta'(Q_b\otimes Q_k)\eta\times\ell'C_\delta^+\ell.$$

The middle terms vanish because

$$\mathrm{E}_\delta\left\{D_{\delta12}\rho\eta'(Q_b\otimes Q_k)T_\delta\right\} = \frac{1}{b!t!}\sum_{\Gamma}\sum_{\Pi}\Gamma'D_{d12}\Gamma\rho\eta'(Q_b\Pi\otimes Q_k)H_d\Gamma$$

and

$$\sum_{\Pi}Q_b\Pi = (b-1)!Q_b1_b1_b' = 0.$$

For the first term we get

$$\mathrm{E}_\delta(\ell'D_{\delta12}\rho\rho'D_{\delta12}'\ell) = \frac{1}{t!}\sum_{\Gamma}(\ell'\Gamma'D_{d12}\Gamma\rho)^2.$$

If we let $\tilde{N}$ such that

$$D_{d12} = -\frac{c_1}{k}Q_t + \tilde{N},$$

then $\tilde{N}$ has row and column sums zero and $\operatorname{tr}\tilde{N} = 0$.

This implies

$$
\frac{1}{t!} \sum_\Gamma (\ell'\Gamma'D_{d12}\Gamma\rho)^2
$$

$$
= \frac{1}{t!} \left\{ \sum_\Gamma \left(-\frac{c_1}{k}\,\ell'\Gamma'Q_t\Gamma\rho\right)^2 - 2\frac{c_1}{k}\sum_\Gamma \ell'\Gamma'Q_t\Gamma\rho\,\ell'\Gamma'\tilde{N}\Gamma\rho + \sum_\Gamma \left(\ell'\Gamma'\tilde{N}\Gamma\rho\right)^2 \right\}
$$

$$
\geq \left(-\frac{c_1}{k}\,\ell'\rho\right)^2 - 2\frac{c_1}{k}\,\ell'\rho\,\ell' \sum_\Gamma \Gamma'\tilde{N}\Gamma\rho
$$

$$
= \left(-\frac{c_1}{k}\,\ell'\rho\right)^2. \tag{16}
$$

If both the starting design d and the design d^* which consists of the first $k-1$ periods of d are Generalized Youden Designs and uniform on the periods and d is also balanced for carryover, then the off-diagonal elements of M_d in (15) equal $b(k-1)/(t(t-1))$ and $\sum_i n_{dir}\tilde{n}_{dir} = b(k-1)/t$ for $r = 1, \ldots, t$. Finally, for $r \neq s$ we have

$$
\sum_i n_{dir}\tilde{n}_{dis} = \frac{b(k-1)^2}{t(t-1)}.
$$

Hence $D_{d12} = -\frac{b(k-1)}{k(t-1)}Q_t$, $\tilde{N} = 0$ and equality holds in (16).

Combining the results above we have

$$
\mathrm{var}_\delta(\ell'\hat{\tau}) = \mathrm{E}_\delta\left\{(\ell'\hat{\tau})^2\right\} - \left\{\mathrm{E}_\delta(\ell'\hat{\tau})\right\}^2
$$

$$
\geq \left(\frac{1}{c_1}\right)^2 \left(\frac{-c_1}{k}\right)^2 (\ell'\rho)^2 + \frac{1}{(b-1)(k-1)}\,\eta'(Q_b \otimes Q_k)\eta \times \ell'C_\delta^+\ell - \left(-\frac{1}{k}\,\ell'\rho\right)^2
$$

$$
= \frac{1}{(b-1)(k-1)}\,\eta'(Q_b \otimes Q_k)\eta \times \ell'C_\delta^+\ell
$$

with equality for a balanced design d with both d and d^* GYD. As in Proposition 3, this shows that the variance of the case without carryover effects is a lower bound, and that the balanced design where d^* is a GYD has the same variance with or without carryover effects.

D   Data

Periods	Assessors														
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1	0.18	0.35	0.32	0.06	0.49	0.55	0.21	0.48	0.21	0.67	0.88	0.90	0.38	0.23	0.50
2	0.49	0.71	0.69	0.43	0.72	0.86	0.52	0.67	0.57	1.04	1.16	1.02	0.67	0.47	0.52
3	1.24	1.23	0.77	0.84	1.56	1.06	1.07	0.73	1.91	1.65	1.23	1.05	1.41	0.48	2.60
4	2.42	2.53	1.99	3.01	1.91	1.36	1.67	1.07	1.96	2.06	2.18	1.16	1.66	1.46	5.29
5	3.68	4.43	3.16	10.00	4.88	1.42	3.70	1.50	6.08	3.82	10.00	2.34	3.40	3.38	6.75

Periods	Assessors														
	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30
1	0.27	0.37	0.24	0.71	0.45	0.20	1.18	0.37	0.44	0.30	0.71	0.49	0.51	0.33	0.53
2	0.28	0.87	0.47	1.35	0.80	0.34	1.93	0.41	0.53	0.52	0.83	0.87	0.84	0.52	1.23
3	0.98	1.25	0.92	3.63	1.20	0.61	2.61	1.70	1.95	1.12	1.50	1.32	0.94	1.12	1.25
4	2.62	3.10	1.03	5.23	1.78	2.04	6.37	3.49	2.50	2.73	1.81	1.37	2.39	1.75	1.28
5	2.64	3.53	1.23	5.82	5.84	3.14	8.49	4.70	4.05	2.91	2.29	1.73	3.14	1.87	4.00

The data that was used in the analysis, rounded to two digits.