Supplement to Comparison of Microarray Designs for Class Comparison and Class Discovery

Dobbin, K.*and Simon, R.

National Cancer Institute
EPN Mailstop 7434
6130 Executive Blvd.
Bethesda, MD 20892
USA

Email: dobbinke@mail.nih.gov
Phone: 301-451-6244

---

*To whom correspondence should be addressed.

1

1 Material Referred to in the Text

1.1 Further Differences Between Our Model and Kerr et al.'s and Wolfinger et al.'s Models

In Wolfinger et al.'s (2001) example, there is an additional AV interaction term which does not appear in our model. The AV effect in their model represents the "channel" effect, which is represented by the AD effect in our model. The reason for the difference is that dye is confounded with variety in their example. Kerr and Churchill (2001) have an additional DG term. A DG interaction term could be introduced into our gene model. A DG effect, even if present, will not bias results for comparisons between samples tagged with the same dye, as is typically done in reference designs. If there is a strong enough DG interaction, then some gene expression may only be detected under one dye. However, in our experience this type of interaction applies to very few genes, and these could be identified in advance from other microarray label experiments. If DG effects exist and samples tagged with different dyes are compared then a reverse labelling of each sample would be required to distinguish DG effects from sample effects. This is important for cluster analysis where contrasts between pairs of samples must be estimated.

1.2 Efficiency Maximized When Replication Is At Sample Level

Consider the reference design shown in Figure 1 of the original paper. In the figure, subscripts denote samples and two samples from each non-reference variety (denoted A and B) are taken; alternatively, the second samples could be replaced with second aliquots of the same samples, so that Array 2 would consist of R_1 and A_1, and Array 4 would consist of R_1 and B_1; but this alternative is inferior to the design shown because it has only one sample from each population of interest instead of two. This lack of replication at the sample level means sample effects are confounded with variety effects, so that no inference about variety effects is possible. Now consider the loop design shown in Figure 2 in the original paper; the figure displays a case with two subsamples from each sample (represented by, for example, the A_1 that appears both on Array 1 and Array 4); we could keep the same number of arrays, and the same loop structure, but eliminate the subsampling, so that each letter would have an unique subscript, requiring 4 samples of each variety; in fact, eliminating the subsampling in Figure 2 would increase the efficiency of the estimates.

1.3 Confounding Of Sample Effects and Spot Effects in Single Aliquot Non-reference Designs

In this case, each sample occurs on exactly one array. If spot effects (AG) are treated as fixed-effects, as we have done, then this design precludes comparison of the gene expression profiles for individual samples on different arrays; specifically, it precludes comparison of the gene-specific

sample effects (FG) for samples on different arrays because these effects will be confounded with the spot effects, i.e. with the blocking factor. Note that in Figure 3 we are forming the loops with respect to *samples* instead of *varieties.* In such a design, it is always more efficient to alternate varieties as well, as in figure 2 in the original paper, and to avoid ever placing two samples of the same variety on an array.[1]

1.4 Figure Displaying Increasing Variance of Sample Contrasts in the Loop Design

Figure 4 shows examples of the variance of contrasts for samples of size 10 and 20, with $\tau_g = \sigma_g$ and with $\tau_g = 2\sigma_g$. As the distance between samples in the loop increases, these plots display the corresponding increase in the variance of contrasts. When there are 20 samples, the proportion of contrast estimates with greater variance under the loop design is larger than when there are 10 samples. Note that as the number of samples increases (beyond 20), this trend will continue. Hence, information about the distance between pairs of samples in a loop design will, on average, decrease as the number of samples increases.

---

[1]Our analytic computations are based on designs which look like one large loop at the sample level, and like a series of repeated loops at the variety level. These designs should have optimal or close to optimal efficiency for a loop design with the small numbers of varieties we consider.

1.5   Details of the Cluster Analysis Monte Carlo Simulation

The "spot" effect is simulated as a random effect with a normal distribution. The gene main effect is represented by a fixed effect. Parameter values for effect sizes were approximated using a subset of a prostate cancer microarray dataset. The base ten log of the intensity levels was the response. We generated data with gene main effect $G_g = 2.5$, and random spot effect AG_{ag} normal with mean 0 and variance 0.61^2; twenty genes were down-regulated in half the non-reference samples and up-regulated in the other half; we looked at simulations where the distance between up-regulated and down-regulated genes was 2.5 and 1.8; for the rest of the $1,000$ genes in the non-reference samples, and for all genes in the reference samples, gene effects were equal, so that $FG_{fg} = 0$; we considered one scenario with inter-sample variance and intra-sample variance 0.43^2, and a second with inter-sample variance $.55^2$ and intra-sample variance $.27^2$. For each Monte Carlo simulation, one-half of the non-reference samples were selected at random to be down-regulated, and the rest were up-regulated.

1.6   More Practical Robustness Issues

Here we discuss some practical robustness issues, including: arrays with multiple spots per gene, array replication, use of blocking factors to reduce variability, and sample size calculation.

If genes are spotted multiple times on each array, then variation in spots will no longer be captured by the array by gene interaction because several spots on a slide will share the same

array and gene labels. Spot effects can then be captured by appending a spot location subscript to the array by gene interaction in the gene model, replacing AG_{ag} with AGL_{agl} where l indexes the location.

Although repeating arrays (i.e. having RNA from the same two sources on more than one array) will generally result in loss of efficiency, it may be desirable in some cases. For instance, some repeated arrays with reversed labelling allow one to estimate the DG interaction term; also, repeated arrays in a BIB design may allow one to estimate the FG interaction term and identify potentially problematic or influential samples.

Similarly, if some factors are not of interest, then one may wish to use these as blocks in a block design to gain efficiency. Creating blocks in a reference design is straightforward. Blocks may be formed in a loop design by creating a separate loop for each level of the blocking variable. Because each RNA sample will be associated with one level of the blocking variable, it will be associated with only one of these unconnected loops. One could not compare samples in different blocks. The loop design locks one into a blocked analysis where the reference design does not.

Correct sample size for a reference design can be calculated in a straightforward way using a log-ratio approach (Simon et al., 2002). With non-reference designs, established methods exist for calculating sample size and power curves for fixed effects and mixed effects models, and can be applied directly to array data (Wolfinger et al., 2001).

1.7   The Model Assumptions

Here we discuss some assumptions of the ANOVA modelling approach, and the impact violations would have on our results, including: inadequacies of the normalization model, correlation of residuals from the normalization model, and inadequacies of the gene model.

Some reviewers expressed concern that the normalization model may be inadequate for two reasons: it makes the unlikely assumption of equal error variance, and it performs the normalization in a linear fashion. We agree that the first assumption is unlikely if the effects in the gene-specific models are non-zero, but that assuming equal variance does produce least squares estimates of normalization model parameters, which we think reasonable. Some researchers believe that making a nonlinear adjustment to each array (Dudoit et al., 2000) is important because it corrects for dye bias associated with spot intensity. Our normalization model is really not essential to the results we have presented, and can be replaced by a different normalization procedure that corrects for the same effects before fitting the gene model. Our analytic and simulation results would still be valid, since they only rely on the structure of the gene model. Our main purpose here is to compare experimental designs, and not necessarily solve the complex question of microarray data normalization.

We have modelled the residuals in the gene model as independent even though they are really correlated from the fitting of the normalization model. We do not think this is problematic. Note that the correlation of the residuals induced by fitting an ANOVA model is a function of the design

matrix, and hence does not depend on effect size. Therefore, we can use matrix algebra to calculate the correlation created by fitting the normalization model for each pair of residuals, regardless of effect size. A computation shows that for any two residuals from the normalization model, the correlation induced by the model falls between $-\frac{1}{G}$ and 0, where G is the number of genes. Since the number of genes is large, the correlation will be practically zero. For instance, with 1000 genes, the correlation of every pair of residuals would fall between -.001 and 0; with 5000 genes, between -.0002 and 0. Such a minor violation of the independence assumption in the gene models is clearly of no practical concern. Also note that any array normalization methodology uses the data to perform the normalization, and will therefore induce correlation among the measured quantities at the gene level; as long as this correlation is very small, the benefits of normalization far outweigh the cost of creating an extremely small violation of independence. As the famous statistician George Box said some years ago, "All models are wrong, but some are useful." The success of microarray technology shows that these are useful models.

We would also emphasize here that systematic relations among the genes and cross-hybridization do not imply that the errors in the gene models will be correlated. For instance, if a collection of genes associated with proliferation have up-regulated expression levels in a subset of the samples, then this may cause the sample effect estimates, $\widehat{FG}$, in the different gene models to be correlated. But it will not induce correlation in the errors terms either within a particular gene model or in different gene models. Similarly, if a target gene hybridizes to two spots on an array, then estimated expression for the two genes associated with the spots may be correlated, but this again will not

induce correlation in the error terms.

Our theoretical and Monte Carlo results rely on ANOVA assumptions and methods in the gene models in order to facilitate the comparisons between the different types of microarray designs. We believe the ANOVA approach to be reasonable, but if one uses this approach, the ANOVA assumptions need checking for specific data. Although ANOVA methods are fairly robust, if these assumptions appear to be seriously violated for a set of data, then this may be problematic for loop or block designs. If, for instance, the residuals for a particular gene appear to violate the assumption of normality, then the ANOVA results may be invalid. If a reference design was used, then more robust non-model based methods can be applied to the data. However, if a loop or block design was used, then the more complex relation between the groups over the arrays may make finding a robust alternative analysis technique difficult.

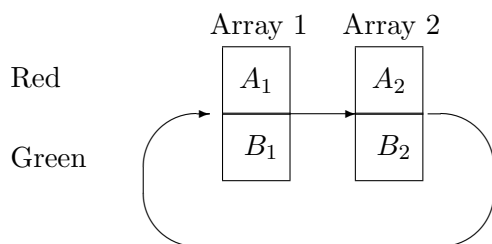## 1.8   Figures and Tables

Figure 1: Loop Design Example: Two samples from each of two varieties. One aliquot from each sample. Single replicate shown.
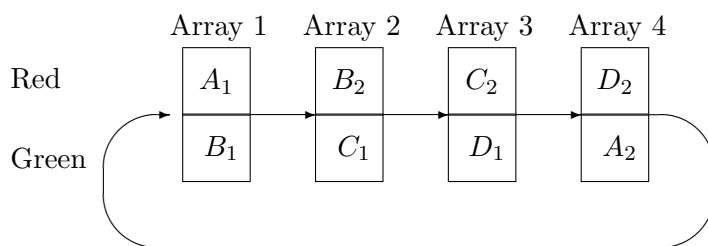
Figure 2: Loop Design Example Figure: Two samples from each of four varieties or phenotypes. One aliquot from each sample.
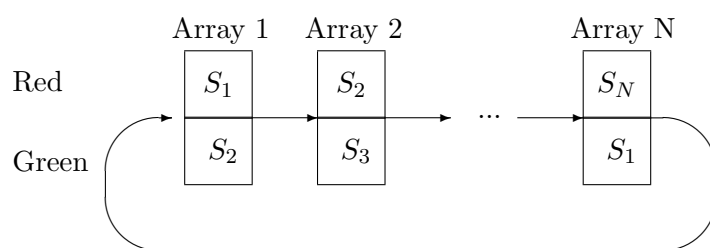
Figure 3: Loop Design for Cluster Analysis: Two aliquots from each of N samples. S_s represents an aliquot from sample s.

Ten Samples　　　　　　　**Twenty Samples**

$$\tau_g^2 = \sigma_g^2 = 1$$

$$\tau_g^2 = 4, \sigma_g^2 = 1$$

Figure 4: Contrast variance of simple reference design and loop design. Distance $= 1 +$ (number of samples between the two samples of interest in the loop, taking shortest route).

Contrast Variances: No Subsampling

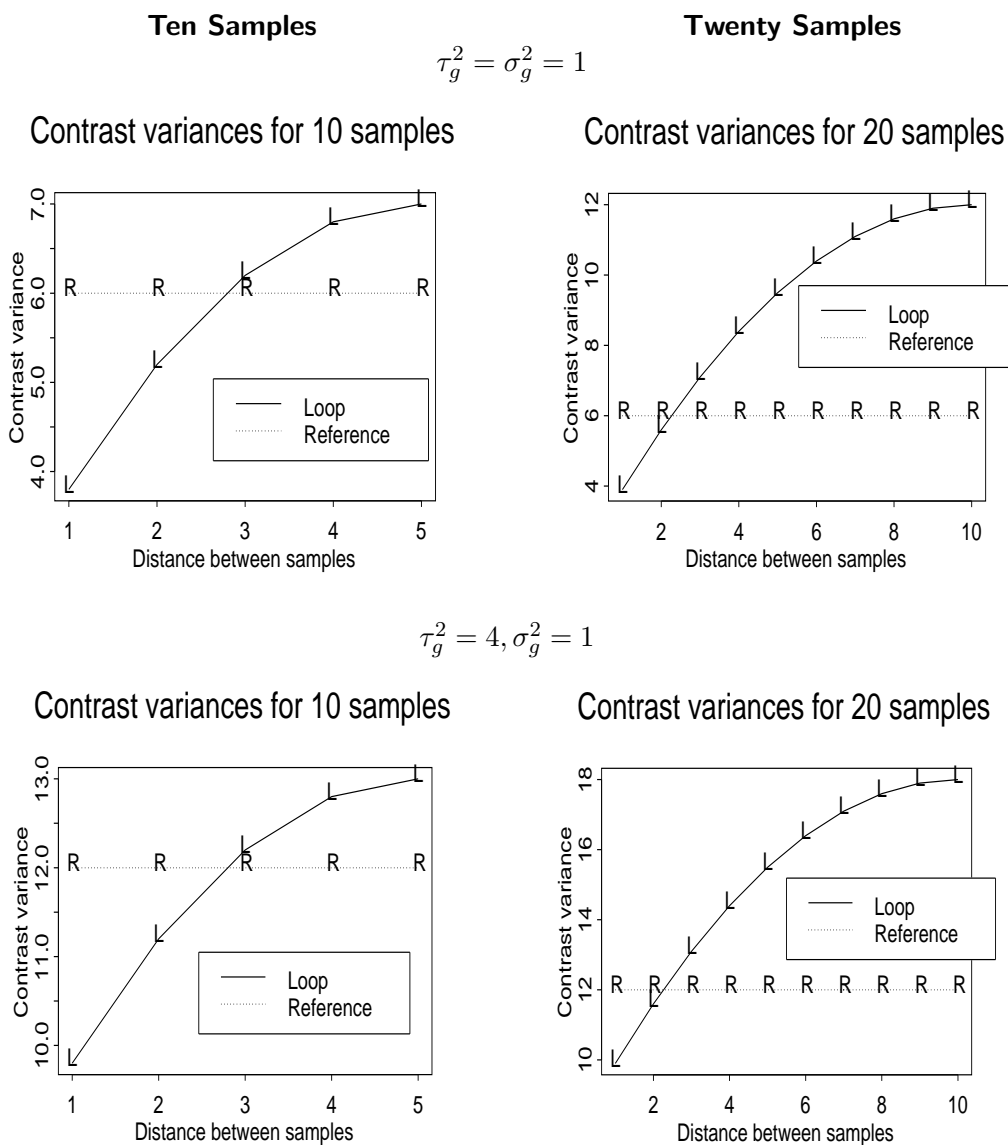| Design	Number of Varieties	Number of Arrays in Single Replicate [a]	Number of Samples in Single Replicate [a]	$Var\left(\widehat{VG}_{ig} - \widehat{VG}_{jg}\right)$
Reference	v	v	v [b]	$\frac{1}{R}\left(2\tau_g^2 + 4\sigma_g^2\right)$
BCB	2	1	2	$\frac{2}{R}\left(\tau_g^2 + \sigma_g^2\right)$
BIB	3	3	6	$\frac{4}{3R}\left(\tau_g^2 + \sigma_g^2\right)$
BIB	4	6	12	$\frac{1}{R}\left(\tau_g^2 + \sigma_g^2\right)$
BIB	5	10	20	$\frac{4}{5R}\left(\tau_g^2 + \sigma_g^2\right)$

Table 1: Variances of contrasts for reference, balanced complete block (BCB), and balanced incomplete block (BIB) designs. One aliquot from each non-reference sample used. Contrast variances are the same for all pairs of (non-reference) samples i and j. In the first row, v is the number of varieties, and the contrast variance formula is the same for all v. For the block designs, the contrast variance formula varies depending on the number of varieties. R is the number of replicates.

---

[a] A single replicate for a reference design is a collection of microarrays arranged so that every array contains one sample from a variety of interest paired with one aliquot from the reference sample, and there are enough arrays so that one sample from each variety of interest is present. A single replicate for a balanced block design is a collection of microarrays arranged so that every array contains two samples, each from a different variety, and there are enough arrays so that every variety appears exactly once on an array with every other variety. (For detailed diagrams of designs, see Supplementary Material at http://linus.nci.nih.gov/ brb/TechReport.htm.)

[b] Does not include reference sample.

Contrast Variances: Two Aliquots Per Sample

Design	Number of Varieties	Number of Arrays in Single Replicate [c]	Number of Samples in Single Replicate [c]	$Var\left(\widehat{VG}_{ig} - \widehat{VG}_{jg}\right)$
Reference	v	2 v	v [d]	$\frac{1}{R}\left(2\tau_g^2 + 2\sigma_g^2\right)$
Loop/BCB	2	2	2	$\frac{1}{R}\left(2\tau_g^2 + \sigma_g^2\right)$
Loop/BIB	3	3	3	$\frac{1}{R}\left(2\tau_g^2 + \frac{4}{3}\sigma_g^2\right)$
Loop (same)	4	4	4	$\frac{1}{R}\left(2\tau_g^2 + \frac{3}{2}\sigma_g^2\right)$
Loop (adjacent)	4	4	4	$\frac{1}{R}\left(2\tau_g^2 + 2\sigma_g^2\right)$

Table 2: Variances of contrasts for reference and loop designs. Two aliquots from each non-reference sample used. In the first row, v is the number of varieties, and the contrast variance formula is the same for all v, although the number of arrays required varies. For the loop designs, the number of arrays required and the contrast variance vary depending on the number of varieties. For 4 or more arrays, the variance of contrasts also depends on the relative positions of the varieties in the loop; "same" indicates varieties which appear together on an array, and "adjacent" varieties which do not. R is the number of replicates.

[c]A single replicate for reference design is a collection of microarrays arranged so that every array contains one sample from a variety of interest paired with one aliquot from the reference sample, and there are enough arrays so that one sample from each variety and two aliquots from each sample are represented. A single replicate for a loop design is a collection of microarrays arranged so that every array contains two aliquots, each from a different sample, there are enough arrays so that two aliquots from each sample are represented, and the samples are arranged in a loop design (see Figure 3).

[d]Does not include reference sample.

2 Application: A Comparison of an ANOVA and Log-ratio Analysis

We applied these analyses to the publicly available apo AI dataset described in Dudoit et al. (2000). This was a reference design experiment with two non-reference varieties (control mice and mice with the apolipoprotein gene knocked-out). None of the non-reference samples were subsampled, so that sample effects were confounded with the error term. Therefore, sample effects and interactions were removed from the models. The resulting models were:

$$log\left(Y_{advfgr}\right) = \mu + A_a + D_d + AD_{ad} + V_v + \epsilon_{advfgr}$$

$$r_{advfgr} = G_g + AG_{ag} + VG_{vg} + \gamma_{advfgr}.$$

We compared our results to those of Dudoit et al. (2000), who used a t-test approach based on log ratios to compare the 8 knocked out mice to the 8 control mice. Dudoit et al. identified 8 genes significantly differentially expressed using a permutation analysis. Table 3 shows the nine most significant genes found using the ANOVA approach. The top 8 most significant genes match exactly the 8 found by Yang et al, and have highly significant Bonferroni adjusted p-values. No other genes appeared significant under the ANOVA approach. Indeed, the ninth most significant gene using the ANOVA approach had an adjusted Bonferroni p-value of 1, and appears in the table.

Reference

Dudoit, S., Yang, Y.H., Callow, M.J., and Speed, T.P. (2000) Statistical methods for identifying differentially expressed genes in replicated cDNA microarray experiments. Technical Report,

Gene Name	Reported by Dudoit et al.	Gene Row Number	Rank	Bonferroni F-test P-value
Apo AI	yes	2149	1	$3x10^{-8}$
Apo CIII	yes	1739	2	$7x10^{-7}$
Apo AI	yes	5356	3	$8x10^{-5}$
Sterol C5 desaturase	yes	4139	4	$2x10^{-4}$
Apo AI	yes	540	5	$4x10^{-4}$
EST AA080005	yes	1496	6	$1x10^{-3}$
Sterol C5 desaturase	yes	4941	7	$2x10^{-3}$
Apo CIII	yes	2537	8	$4x10^{-3}$
cDNA	no	2106	9	1

Table 3: Comparison of results for log-ratio analysis and ANOVA analysis based on log intensities.

http://www.stat.berkeley.edu/users/terry/zarray/TechReport/578.pdf.

3 Efficiency Computations

3.1 Notation

r_{avf} indicates the residual (from the normalization model) corresponding to sample f from variety v on array a for a particular gene (subscript omitted). Inter-sample variance for a gene g is denoted τ_g^2, and intra-sample variance for gene g is denoted σ_g^2.

The usual constraints are used to estimate parameters, namely,

$$\sum_a \widehat{AG} = 0$$
$$\sum_v n_v \widehat{VG} = 0,$$

where n_v is the number of aliquots associated with variety v that appear in the design.

3.2   Efficiency Calculations for the Reference Design

Step 1: Suppose we have N non-reference varieties and one sample from each of these varieties and a simple reference design.

Array 1	Array 2	...	Array N
r_{100}	r_{200}	...	r_{N00}
r_{111}	r_{222}	...	r_{NNN}

The estimate of a variety effect is

$$
\begin{aligned}
\widehat{VG}_{1g} &= r_{111} - (r_{100} - \bar{r}_{\cdot 00}) - \bar{r}_{\cdots} \\[2mm]
\widehat{VG}_{2g} &= r_{222} - (r_{200} - \bar{r}_{\cdot 00}) - \bar{r}_{\cdots} \\[2mm]
\widehat{VG}_{1g} - \widehat{VG}_{2g} &= r_{111} - r_{222} - (r_{100} - r_{200}) \\[2mm]
&= (V_1 - V_2) + (F_1 - F_2) + (\gamma_{111} + \gamma_{222} - \gamma_{100} - \gamma_{200}) \\[2mm]
var\left(\widehat{VG}_{1g} - \widehat{VG}_{2g}\right) &= 2\tau_g^2 + 4\sigma_g^2
\end{aligned}
$$

Step 2: If we repeat this experiment by adding a R new batches of N samples and pairing each with another aliquot from the reference variety, then we have

$$
var\left(\widehat{VG}_{1g} - \widehat{VG}_{2g}\right) \;=\; \frac{2}{R}\tau_g^2 + \frac{4}{R}\sigma_g^2
$$

For any reference design without subsampling of non-reference varieties, this can be used to find the variance of a contrast.

3.3   Reference design with two aliquots per sample

Array 1	Array 2	...	Array N	Array N+1	...	Array 2N
r_{100}	r_{200}	...	r_{N00}	$r_{(N+1)00}$	...	$r_{(2N)00}$
r_{111}	r_{222}	...	r_{NNN}	$r_{(N+1)11}$	...	$r_{(2N)NN}$

Suppose we have one replication of the reference design in the previous subsection, which is not a true replication but a replication with from the same set of samples. Add a fourth subscript for the replicates.

$$
\widehat{VG}_{1g} \;=\; \bar{r}_{111\cdot} - \left(\bar{r}_{100\cdot} - \bar{r}_{\cdot 00\cdot}\right) - \bar{r}_{\cdots\cdot}
$$

$$\widehat{VG}_{2g} = \bar{r}_{222\cdot} - (\bar{r}_{200\cdot} - \bar{r}_{\cdot 00\cdot}) - \bar{r}_{\cdots}$$

$$\widehat{VG}_{1g} - \widehat{VG}_{2g} = \bar{r}_{111\cdot} - \bar{r}_{222\cdot} - (r_{100\cdot} - r_{200\cdot})$$

$$= (V_1 - V_2) + (F_1 - F_2) + (\bar{\gamma}_{111\cdot} + \bar{\gamma}_{222\cdot} - \bar{\gamma}_{100\cdot} - \bar{\gamma}_{200\cdot})$$

$$var\left(\widehat{VG}_{1g} - \widehat{VG}_{2g}\right) = 2\tau_g^2 + 2\sigma_g^2$$

3.4 Efficiency Calculations for Balanced Incomplete Block Designs

3.4.1 Balanced block design with two varieties, no subsampling

Array 1	Array 2	...
r_{111}	r_{223}	...
r_{122}	r_{214}	...

Step 1: Suppose we have two samples for each variety.

$$\widehat{VG}_{1g} - \widehat{VG}_{2g} = \bar{r}_{\cdot 1} - \bar{r}_{\cdot 2}$$

$$= \frac{1}{2}(r_{111} + r_{214} - r_{122} - r_{223})$$

$$= \frac{1}{2}(2V_1 + F_1 + F_4 + \gamma_{111} + \gamma_{214} - 2V_2 - F_2 - F_3 - \gamma_{122} - \gamma_{223})$$

$$= V_1 - V_2 + \frac{1}{2}(F_1 + F_4 - F_2 - F_3) +$$

$$\frac{1}{2}(\gamma_{111} + \gamma_{214} - \gamma_{122} - \gamma_{223})$$

$$var\left(\widehat{VG}_{1g} - \widehat{VG}_{2g}\right) = \tau_g^2 + \sigma_g^2$$

Step 2: If we repeat the process by adding R new pairs of blocks, each with varieties laid out in the same pattern and new samples, then

$$var(\widehat{VG}_{1g} - \widehat{VG}_{2g}) = \frac{1}{R}\tau_g^2 + \frac{1}{R}\sigma_g^2$$

3.4.2 Balanced incomplete block (BIB) design with 3 varieties, no subsampling

Array 1	Array 2	Array 3	...
r_{111}	r_{223}	r_{335}	...
r_{122}	r_{234}	r_{316}	...

Step 1: Suppose we have two samples from each variety.

A computation shows that

$$\widehat{VG}_{1g} = \frac{1}{3}(r_{111} - r_{122} + r_{316} - r_{335})$$

$$\widehat{VG}_{2g} = \frac{1}{3}(r_{122} - r_{111} + r_{223} - r_{234})$$

$$\widehat{VG}_{1g} - \widehat{VG}_{2g} \;=\; \frac{2}{3}(r_{111} - r_{122})\frac{1}{3}(r_{316} - r_{335} - r_{223} + r_{234})$$

$$var\left(\widehat{VG}_{1g} - \widehat{VG}_{2g}\right) \;=\; \frac{8}{9}(\tau_g^2 + \sigma_g^2) + \frac{4}{9}(\tau_g^2 + \sigma_g^2)$$

$$=\; \frac{4}{3}(\tau_g^2 + \sigma_g^2)$$

The variance is the same for every pair of samples.

Step 2: If we repeat the process by adding R new triplets of blocks, each with varieties laid out in the same patterns and new samples, then

$$var\left(\widehat{VG}_{1g} - \widehat{VG}_{2g}\right) \;=\; \frac{4}{3R}(\tau_g^2 + \sigma_g^2)$$

3.4.3   BIB design with 4 varieties, no subsampling

Array 1	Array 2	Array 3	Array 4	Array 5	Array 6	...
r_{111}	r_{223}	r_{335}	r_{437}	r_{519}	$r_{64(11)}$	...
r_{122}	r_{234}	r_{346}	r_{418}	$r_{54(10)}$	$r_{62(12)}$	...

Step 1: Suppose we have three samples per variety. A computation shows that

$$\widehat{VG}_{1g} = \frac{1}{4}(r_{111} - r_{122} - r_{447} + r_{418} + r_{519} - r_{5,3,10})$$

$$\widehat{VG}_{2g} = \frac{1}{4}(r_{122} - r_{111} + r_{223} - r_{234} + r_{6,2,11} - r_{6,4,12})$$

$$var\left(\widehat{VG}_{1g} - \widehat{VG}_{2g}\right) = \tau_g^2 + \sigma_g^2$$

The variance is the same for every pair of varieties.

Step 2: Repeat the balanced incomplete block design R times.

$$var\left(\widehat{VG}_{1g} - \widehat{VG}_{2g}\right) = \frac{1}{R}(\tau_g^2 + \sigma_g^2)$$

3.4.4 BIB design with 5 varieties, no subsampling

Array 1	Array 2	Array 3	Array 4	Array 5
r_{111}	r_{223}	r_{335}	r_{447}	r_{519}
r_{122}	r_{234}	r_{346}	r_{458}	$r_{55(10)}$

Array 6	Array 7	Array 8	Array 9	Array 10	...
$r_{63(11)}$	$r_{74(13)}$	$r_{82(15)}$	$r_{95(17)}$	$r_{(10)5(19)}$	...
$r_{61(12)}$	$r_{71(14)}$	$r_{84(16)}$	$r_{92(18)}$	$r_{(10)3(20)}$	...

Step 1: Suppose we have four samples per variety. A computation shows that

$$\widehat{VG}_{1g} = \frac{1}{5}(r_{111} - r_{122} + r_{519} - r_{55(10)} + r_{61(12)} - r_{63(11)} + r_{71(14)} - r_{74(13)})$$

$$\widehat{VG}_{2g} = \frac{1}{5}(r_{122} - r_{111} + r_{223} - r_{234} + r_{82(15)} - r_{84(16)} + r_{92(18)} - r_{95(17)})$$

$$\widehat{VG}_{1g} - \widehat{VG}_{2g} = \frac{2}{5}(r_{111} - r_{122}) +$$

$$\frac{1}{5}\left(r_{519} - r_{55(10)} + r_{61(12)} - r_{63(11)} + r_{71(14)} - r_{74(13)}\right.$$

$$\left. -r_{223} + r_{234} - r_{82(15)} + r_{84(16)} - r_{92(18)} + r_{95(17)}\right)$$

$$var\left(\widehat{VG}_{1g} - \widehat{VG}_{2g}\right) = \frac{4}{5}(\tau_g^2 + \sigma_g^2)$$

Step 2: Repeat the balanced incomplete block R times.

$$var\left(\widehat{VG}_{1g} - \widehat{VG}_{2g}\right) = \frac{4}{5R}(\tau_g^2 + \sigma_g^2)$$

3.4.5 Two varieties in a balanced complete block design, with subsampling

Array 1	Array 2	...
r_{111}	r_{222}	...
r_{122}	r_{211}	...

Step 1: Suppose we have one sample from each variety, and two aliquots per sample.

$$\widehat{VG}_{1g} - \widehat{VG}_{2g} = \bar{r}_{\cdot 1} - \bar{r}_{\cdot 2}$$

$$= \frac{1}{2}(r_{111} + r_{211} - r_{122} - r_{222})$$

$$= \frac{1}{2}(2V_1 + 2F_1 + \gamma_{111} + \gamma_{211} - 2V_2 - 2F_2 - \gamma_{122} - \gamma_{222})$$

$$= V_1 - V_2 + \frac{1}{2}(2F_1 - 2F_2) +$$

$$\frac{1}{2}(\gamma_{111} + \gamma_{211} - \gamma_{122} - \gamma_{222})$$

$$var\left(\widehat{VG}_{1g} - \widehat{VG}_{2g}\right) = 2\tau_g^2 + \sigma_g^2$$

Step 2: Repeat this pattern R times using new samples.

$$var\left(\widehat{VG}_{1g} - \widehat{VG}_{2g}\right) = \frac{2}{R}\tau_g^2 + \frac{1}{R}\sigma_g^2$$

3.4.6 Three varieties in a BIB design, with subsampling

Array 1	Array 2	Array 3	...
r_{111}	r_{222}	r_{333}	...
r_{122}	r_{233}	r_{311}	...

Step 1: Suppose we have one sample from each variety, with two aliquots per sample.

$$\widehat{VG}_{1g} = \frac{1}{3}(r_{111} - r_{122} + r_{316} - r_{335})$$

$$\widehat{VG}_{2g} = \frac{1}{3}(r_{122} - r_{111} + r_{222} - r_{233})$$

$$\widehat{VG}_{1g} - \widehat{VG}_{2g} = \frac{2}{3}(r_{111} - r_{122}) + \frac{1}{3}(r_{311} - r_{333} - r_{222} + r_{233})$$

$$= \frac{2}{3}(V_1 - V_2 + F_1 - F_2 + \gamma_{111} - \gamma_{122})$$

$$+ \frac{1}{3}(V_1 - V_2 + F_1 - F_2 + \gamma_{311} - \gamma_{333} - \gamma_{222} + \gamma_{233})$$

$$var\left(\widehat{VG}_{1g} - \widehat{VG}_{2g}\right) = 2\tau_g^2 + \frac{4}{3}\sigma_g^2$$

Step 2: Repeat this pattern R times with new samples.

$$var\left(\widehat{VG}_{1g} - \widehat{VG}_{2g}\right) = \frac{2}{R}\tau_g^2 + \frac{4}{3R}\sigma_g^2$$

3.4.7 Four varieties in a BIB design, with subsampling

Array 1	Array 2	Array 3	Array 4	Array 5	Array 6	...
r_{111}	r_{222}	r_{333}	r_{433}	r_{511}	r_{644}	...
r_{122}	r_{233}	r_{344}	r_{411}	r_{544}	r_{622}	...

Step 1: Suppose we have one sample from each variety, and three aliquots from each sample.

$$
\begin{aligned}
\widehat{VG}_{1g} &= \frac{1}{4}(r_{111} - r_{122} - r_{444} + r_{411} + r_{511} - r_{5,3,3}) \\
\widehat{VG}_{2g} &= \frac{1}{4}(r_{122} - r_{111} + r_{222} - r_{233} + r_{6,2,2} - r_{6,4,4}) \\
\widehat{VG}_{1g} - \widehat{VG}_{2g} &= \frac{1}{2}(r_{111} - r_{122}) + \\
&\quad \frac{1}{4}(r_{411} - r_{444} + r_{511} - r_{533} + r_{233} - r_{222} + r_{644} - r_{622}) \\
var\left(\widehat{VG}_{1g} - \widehat{VG}_{2g}\right) &= 2\tau_g^2 + \sigma_g^2
\end{aligned}
$$

Step 2: Repeat the pattern R times with new samples.

$$
var\left(\widehat{VG}_{1g} - \widehat{VG}_{2g}\right) = \frac{2}{R}\tau_g^2 + \frac{1}{R}\sigma_g^2
$$

3.4.8 Five varieties in a BIB design, with subsampling

Array 1	Array 2	Array 3	Array 4	Array 5
r_{111}	r_{222}	r_{333}	r_{444}	r_{511}
r_{122}	r_{233}	r_{344}	r_{455}	r_{555}

Array 6	Array 7	Array 8	Array 9	Array 10	...
r_{633}	r_{744}	r_{822}	r_{955}	$r_{(10)55}$	...
r_{611}	r_{711}	r_{844}	r_{922}	$r_{(10)33}$	...

Step 1:

$$\widehat{VG}_{1g} = \frac{1}{5}(r_{111} - r_{122} + r_{511} - r_{555} + r_{611} - r_{633} + r_{711} - r_{744})$$

$$\widehat{VG}_{2g} = \frac{1}{5}(r_{122} - r_{111} + r_{222} - r_{233} + r_{822} - r_{844} + r_{922} - r_{955})$$

$$\widehat{VG}_{1g} - \widehat{VG}_{2g} = \frac{2}{5}(r_{111} - r_{122}) +$$

$$\frac{1}{5}(r_{511} - r_{555} + r_{611} - r_{633} + r_{711} - r_{744} - r_{222} + r_{233} - r_{822} + r_{844} - r_{922} + r_{955})$$

$$var\left(\widehat{VG}_{1g} - \widehat{VG}_{2g}\right) = 2\tau_g^2 + \frac{4}{5}\sigma_g^2$$

Step 2

$$var\left(\widehat{VG}_{1g} - \widehat{VG}_{2g}\right) = \frac{2}{R}\tau_g^2 + \frac{4}{5R}\sigma_g^2$$

3.5 Efficiency Calculations for Loop Designs

3.5.1 Two or three varieties in a loop

The loop design is the same as the BIB design when there are two or three varieties.

3.5.2 Four varieties in a loop (w.r.t. varieties), without subsampling

Array 1	Array 2	Array 3	Array 4	...
r_{111}	r_{223}	r_{335}	r_{447}	...
r_{122}	r_{234}	r_{346}	r_{418}	...

Step 1: Suppose we have two samples from each variety. A calculation shows that:

$$\widehat{VG}_{1g} = \frac{3}{8}(r_{111} - r_{122} + r_{418} - r_{447}) - \frac{1}{8}(r_{234} - r_{223} + r_{335} - r_{346})$$

$$\widehat{VG}_{2g} = \frac{3}{8}(r_{122} - r_{111} + r_{223} - r_{234}) - \frac{1}{8}(r_{346} - r_{335} + r_{447} - r_{418})$$

$$\widehat{VG}_{1g} - \widehat{VG}_{2g} = \frac{3}{4}(r_{111} - r_{122}) + \frac{1}{4}(r_{418} - r_{447} + r_{234} - r_{223} + r_{346} - r_{335})$$

$$var\left(\widehat{VG}_{1g} - \widehat{VG}_{2g}\right) = \frac{3}{2}(\tau_g^2 + \sigma_g^2)$$

$$\widehat{VG}_{3g} = \frac{3}{8}(r_{234} - r_{223} + r_{335} - r_{346}) - \frac{1}{8}(r_{111} - r_{122} + r_{418} - r_{447})$$

$$\widehat{VG}_{1g} - \widehat{VG}_{3g} = \frac{1}{2}(r_{111} - r_{122} + r_{418} - r_{447} - r_{234} + r_{223} - r_{335} + r_{346})$$

$$var\left(\widehat{VG}_{1g} - \widehat{VG}_{3g}\right) = 2(\tau_g^2 + \sigma_g^2)$$

Step 2: Suppose there are R replications of this design with new samples.

$$var\left(\widehat{VG}_{1g} - \widehat{VG}_{2g}\right) = \frac{3}{2R}(\tau_g^2 + \sigma_g^2)$$

$$var\left(\widehat{VG}_{1g} - \widehat{VG}_{3g}\right) = \frac{2}{R}(\tau_g^2 + \sigma_g^2)$$

3.5.3 Four varieties in a loop (w.r.t. samples), with two aliquots per sample and the samples arranged in a loop design to allow for clustering

Array 1	Array 2	Array 3	Array 4	...
r_{111}	r_{222}	r_{333}	r_{444}	...
r_{122}	r_{233}	r_{344}	r_{411}	...

Step 1: Suppose there is one sample from each variety and two aliquots from each sample.

$$\widehat{VG}_{1g} = \frac{3}{8}(r_{111} - r_{122} + r_{411} - r_{444}) - \frac{1}{8}(r_{233} - r_{222} + r_{333} - r_{344})$$

$$\widehat{VG}_{2g} = \frac{3}{8}(r_{122} - r_{111} + r_{222} - r_{233}) - \frac{1}{8}(r_{344} - r_{333} + r_{444} - r_{411})$$

$$\widehat{VG}_{1g} - \widehat{VG}_{2g} = \frac{3}{4}(r_{111} - r_{122}) + \frac{1}{4}(r_{411} - r_{444} + r_{233} - r_{222} + r_{344} - r_{333})$$

$$var(\widehat{VG}_{1g} - \widehat{VG}_{2g}) = 2\tau_g^2 + \frac{3}{2}\sigma_g^2$$

$$\widehat{VG}_{1g} - \widehat{VG}_{3g} = \frac{1}{2}(r_{111} - r_{122} + r_{411} - r_{444} - r_{233} + r_{222} - r_{333} + r_{344})$$

$$var\left(\widehat{VG}_{1g} - \widehat{VG}_{3g}\right) = 2\tau_g^2 + 2\sigma_g^2$$

Step 2: Suppose there are R replications of the design with new samples.

$$var\left(\widehat{VG}_{1g} - \widehat{VG}_{2g}\right) = \frac{2}{R}\tau_g^2 + \frac{3}{2R}\sigma_g^2$$

$$var\left(\widehat{VG}_{1g} - \widehat{VG}_{3g}\right) = \frac{2}{R}\tau_g^2 + \frac{2}{R}\sigma_g^2$$

Step 3: Note that looping with respect to samples (instead of varieties) will result in a slightly different pattern with respect to the sample index than performing replications. However, because this change does not effect the array or variety indexes, the variance of variety contrasts will be the same as the variance calculated in step 2.

3.5.4 Five varieties in a loop (w.r.t. varieties), without subsampling

Array 1	Array 2	Array 3	Array 4	Array 5	...
r_{111}	r_{223}	r_{335}	r_{447}	r_{519}	...
r_{122}	r_{234}	r_{346}	r_{458}	$r_{55(10)}$	...

Step 1: Suppose there are four samples from each variety.

$$\widehat{VG}_{1g} = \frac{2}{5}(r_{111} - r_{122} + r_{51,10} - r_{459}) - \frac{1}{5}(r_{234} - r_{223} + r_{447} - r_{458})$$

$$\widehat{VG}_{2g} = \frac{2}{5}(r_{122} - r_{111} + r_{223} - r_{234}) - \frac{1}{5}(r_{346} - r_{335} + r_{459} - r_{51,10})$$

$$\widehat{VG}_{1g} - \widehat{VG}_{2g} = \frac{4}{5}(r_{111} - r_{122}) + \frac{1}{5}(r_{51,10} - r_{459} + r_{234} - r_{223} - r_{447} + r_{458} + r_{346} - r_{335})$$

$$var\left(\widehat{VG}_{1g} - \widehat{VG}_{2g}\right) = \frac{8}{5}(\tau_g^2 + \sigma_g^2)$$

$$\widehat{VG}_{3g} = \frac{2}{5}(r_{234} - r_{223} + r_{335} - r_{346}) - \frac{1}{5}(r_{111} - r_{122} + r_{458} - r_{447})$$

$$\widehat{VG}_{1g} - \widehat{VG}_{3g} = \frac{3}{5}(r_{111} - r_{122} + r_{223} - r_{234}) + \frac{2}{5}(r_{458} - r_{447} + r_{346} - r_{335} + r_{51,10} - r_{459})$$

$$var\left(\widehat{VG}_{1g} - \widehat{VG}_{3g}\right) = \frac{12}{5}(\tau_g^2 + \sigma_g^2)$$

Step 2: Suppose this pattern is repeated R times.

$$var\left(\widehat{VG}_{1g} - \widehat{VG}_{2g}\right) = \frac{8}{5R}(\tau_g^2 + \sigma_g^2)$$

$$var\left(\widehat{VG}_{1g} - \widehat{VG}_{3g}\right) = \frac{12}{5R}(\tau_g^2 + \sigma_g^2)$$

3.6 Relative Efficiency Tables

Sample Contrast Relative Efficiencies: 10 Arrays

Distance	1	2	3	4	5
Rel Eff	$\frac{2\tau_g^2+4\sigma_g^2}{2\tau_g^2+1.8\sigma_g^2}$	$\frac{2\tau_g^2+4\sigma_g^2}{2\tau_g^2+3.2\sigma_g^2}$	$\frac{2\tau_g^2+4\sigma_g^2}{2\tau_g^2+4.2\sigma_g^2}$	$\frac{2\tau_g^2+4\sigma_g^2}{2\tau_g^2+4.8\sigma_g^2}$	$\frac{2\tau_g^2+4\sigma_g^2}{2\tau_g^2+5\sigma_g^2}$

Table 4: Ten Arrays: Loop versus reference design Relative efficiencies for individual sample contrasts $= \frac{var_{reference}(\widehat{FG}_{1g}-\widehat{FG}_{2g})}{var_{loop}(\widehat{FG}_{1g}-\widehat{FG}_{2g})}$. Distance $= 1 +$ (number of samples between the two samples of interest in the loop, taking shortest route).

Sample Contrast Relative Efficiencies: 20 Arrays

Distance	1	2	3	4	5
Rel Eff	$\frac{2\tau_g^2+4\sigma_g^2}{2\tau_g^2+1.9\sigma_g^2}$	$\frac{2\tau_g^2+4\sigma_g^2}{2\tau_g^2+3.6\sigma_g^2}$	$\frac{2\tau_g^2+4\sigma_g^2}{2\tau_g^2+5.1\sigma_g^2}$	$\frac{2\tau_g^2+4\sigma_g^2}{2\tau_g^2+6.4\sigma_g^2}$	$\frac{2\tau_g^2+4\sigma_g^2}{2\tau_g^2+7.5\sigma_g^2}$

Distance	6	7	8	9	10
Rel Eff	$\frac{2\tau_g^2+4\sigma_g^2}{2\tau_g^2+8.4\sigma_g^2}$	$\frac{2\tau_g^2+4\sigma_g^2}{2\tau_g^2+9.1\sigma_g^2}$	$\frac{2\tau_g^2+4\sigma_g^2}{2\tau_g^2+9.6\sigma_g^2}$	$\frac{2\tau_g^2+4\sigma_g^2}{2\tau_g^2+9.9\sigma_g^2}$	$\frac{2\tau_g^2+4\sigma_g^2}{2\tau_g^2+10.0\sigma_g^2}$

Table 5: Twenty Arrays: Loop versus reference design Relative efficiencies for individual sample contrasts $= \frac{var_{reference}(\widehat{FG}_{1g}-\widehat{FG}_{2g})}{var_{loop}(\widehat{FG}_{1g}-\widehat{FG}_{2g})}$. Distance $= 1 +$ (number of samples between the two samples of interest in the loop, taking shortest route).

Reference vs. BIB Design

Number of Varieties	Same Number of Arrays Used	Same Non-Reference Aliquots Used
2	$\dfrac{2\tau_g^2 + 4\sigma_g^2}{\tau_g^2 + \sigma_g^2}$	$\dfrac{(2\tau_g^2 + 4\sigma_g^2)/2}{\tau_g^2 + \sigma_g^2}$
3	$\dfrac{2\tau_g^2 + 4\sigma_g^2}{\frac{4}{3}\tau_g^2 + \frac{4}{3}\sigma_g^2}$	$\dfrac{(2\tau_g^2 + 4\sigma_g^2)/2}{\frac{4}{3}\tau_g^2 + \frac{4}{3}\sigma_g^2}$
4	$\dfrac{(2\tau_g^2 + 4\sigma_g^2)/3}{(\tau_g^2 + \sigma_g^2)/2}$	$\dfrac{(2\tau_g^2 + 4\sigma_g^2)/3}{(\tau_g^2 + \sigma_g^2)/1}$
5	$\dfrac{(2\tau_g^2 + 4\sigma_g^2)/2}{(\frac{4}{5}\tau_g^2 + \frac{4}{5}\sigma_g^2)/1}$	$\dfrac{(2\tau_g^2 + 4\sigma_g^2)/4}{(\frac{4}{5}\tau_g^2 + \frac{4}{5}\sigma_g^2)/1}$

Table 6: Relative efficiencies of variety contrasts $= \dfrac{var_{reference}\left(\widehat{VG}_{1g} - \widehat{VG}_{2g}\right)}{var_{bib}\left(\widehat{VG}_{1g} - \widehat{VG}_{2g}\right)}$. Balanced block designs use one aliquot per sample (to maximize efficiency).

Reference vs. Loop Design

Number of Varieties	Same Number of Arrays Used	Same Non-Reference Aliquots Used
2	$\dfrac{2\tau_g^2 + 4\sigma_g^2}{2\tau_g^2 + \sigma_g^2}$	$\dfrac{2\tau_g^2 + 2\sigma_g^2}{2\tau_g^2 + \sigma_g^2}$
3	$\dfrac{2\tau_g^2 + 4\sigma_g^2}{2\tau_g^2 + \frac{4}{3}\sigma_g^2}$	$\dfrac{2\tau_g^2 + 2\sigma_g^2}{2\tau_g^2 + \frac{4}{3}\sigma_g^2}$
4 (same)	$\dfrac{2\tau_g^2 + 4\sigma_g^2}{2\tau_g^2 + \frac{3}{2}\sigma_g^2}$	$\dfrac{2\tau_g^2 + 2\sigma_g^2}{2\tau_g^2 + \frac{3}{2}\sigma_g^2}$
4 (adjacent)	$\dfrac{2\tau_g^2 + 4\sigma_g^2}{2\tau_g^2 + 2\sigma_g^2}$	$\dfrac{2\tau_g^2 + 2\sigma_g^2}{2\tau_g^2 + 2\sigma_g^2}$

Table 7: Relative efficiencies of variety contrasts $= \dfrac{var_{reference}\left(\widehat{VG}_{1g} - \widehat{VG}_{2g}\right)}{var_{loop}\left(\widehat{VG}_{1g} - \widehat{VG}_{2g}\right)}$. Loop designs use 2 aliquots per sample (to allow for clustering of samples).