BIOINFORMATICS ORIGINAL PAPER Vol. 25 no. 18 2009, pages 2355–2361
doi:10.1093/bioinformatics/btp407

Gene expression

Robustness considerations in selecting efficient two-color

A. H. M. Mahbub Latif[1,2,*], Frank Bretz[3] and Edgar Brunner[4]

[1]Institute of Statistical Research and Training, University of Dhaka, Dhaka 1000, Bangladesh, [2]Queen Mary University of London, School of Mathematical Sciences, London E1 4NS, UK, [3]Novartis Pharma AG, Lichtstrasse 35, 4002 Basel, Switzerland and [4]Abteilung Medizinische Statistik, Universität Göttingen, Humboldtallee 32, D-37073 Göttingen, Germany

Received on February 11, 2009; revised on June 16, 2009; accepted on June 29, 2009

Advance Access publication July 1, 2009

Associate Editor: David Rocke

ABSTRACT

The main goal of microarray experiments is to select a small subset of genes that are differentially expressed among competing *mRNA* samples. For a given set of such *mRNA* samples, it is possible to consider a number of two-color *cDNA* microarray designs with a fixed number of arrays. Appropriate criteria can be used to select an efficient design from such a set of alternative experimental designs. In practice, however, microarray expression data often contain missing observations and the most efficient design (with complete observations) for a specific setup may not be efficient in the presence of missing observations. In this article, we propose two criteria to address the robustness of microarray designs against missing observations. We demonstrate the simultaneous use of efficiency and robustness criteria to select good microarray designs for both one-factor and multi-factor experiments.

Contact: mlatif@isrt.ac.bd

1 INTRODUCTION

The statistical design of microarray experiments plays a vital role in allocating *mRNA* samples under investigation to the available arrays. The application of classical experimental designs to microarray experiments was first investigated by Kerr *et al*. (2000). Microarray experiments can be considered as incomplete block experiments of block size two when comparisons of more than two *mRNA* samples are of interest. For a given set of competing *mRNA* samples, a number of different experimental designs can be considered for a fixed number of arrays. Among the experimental designs used in microarrays, the Common Reference (*CR*) design (Callow *et al*., 2000) is the most commonly used design where competing treatments (i.e. *mRNA* samples) are compared indirectly via a common reference sample. For this design, half of the samples are used for the estimation of the parameters of interest although the information from the reference sample is not of interest itself. Moreover, the indirect comparison inflates the variance of the relevant parameter estimates. Kerr and Churchill (2001) proposed Circular Loop (*CL*) designs instead, which compare the treatments of interest directly by connecting every pair of treatments sequentially. Due to its construction, *CL* designs can be used to estimate the parameters of interest with less variance in comparison with the corresponding *CR* design of the same size (Kerr and Churchill, 2001). Besides *CR* and *CL* designs, the Dye-Swap (*DS*) design, which compares each pair of treatments twice with forward and reverse dye comparisons, is also used in microarray experiments. A general overview of statistical designs in microarray experiments can be found in Landgrebe *et al*. (2006).

The problem of selecting a good experimental design from a set of candidate designs is one of the key questions studied in optimum experimental design theory (Atkinson *et al*., 2007; Pukelsheim, 1993). An experimental design is said to be optimum/efficient if the estimated variance of the relevant parameter estimate is the smallest among the alternative designs. Efficiency of a design is measured by a design criterion (also known as efficiency criterion), which is a function of the information matrix corresponding to the underlying statistical model.

Different design criteria, such as *A*-, *D*- and *E*-optimality, have been proposed in the context of microarray experiments. For one-factor microarray experiments, Kerr and Churchill (2001) reported *A*-optimum designs considering all pairwise treatment comparisons as the effects of interest. Yang and Speed (2002) compared the efficiency of competing loop designs in 2×2 factorial layouts using an *A*-optimality criterion. Landgrebe *et al*. (2006) used a minimax approach based on *E*-optimality to select efficient designs for 2×2 and 3×2 microarray experiments. More recently, Stanzel (2008) provided the theoretical basis for some of the empirical results of Landgrebe *et al*. (2006). Glonek and Solomon (2004) considered a similar model as Yang and Speed (2002) and used an admissibility concept to select efficient multi-factorial microarray designs. Their numerical algorithm is based on a complete search of all possible designs and becomes computationally challenging for a large number of arrays and conditions. To avoid the exhaustive search, Wit *et al*. (2005) used simulated annealing to find near-optimal (*A*- and *D*-optimal) microarray designs in one-factor experiments. Latif (2005) considered a model similar to Landgrebe *et al*. (2006) and used genetic algorithms to search for near-optimal designs in both one-factor and multi-factor experiments. Gupta (2006) discussed a systematic way to find efficient balanced factorial designs for microarray experiments. There are only a few papers on optimum microarray designs based on mixed effects models, although such analysis methods have been used for several years (Wolfinger *et al*,

*To whom correspondence should be addressed.

2001). Tsai *et al*. (2006) considered a statistical model that allows intensities between different hybridization to be correlated and described a heuristic algorithm to obtain *A*-optimal designs for one-factor experiment. Bueno Filho *et al*. (2006) discussed efficient designs of microarray experiments for 3×3 factorial layouts using linear mixed effects models. See also Tempelman (2008) and Passos *et al*. (2009) for recent works on microarray designs using mixed effects models.

All of the above investigations of selecting efficient microarray designs used complete observations to calculate the efficiency criteria of the competing microarray designs. However, microarray expression data often contain missing observations due to various reasons including image resolution, image corruption, dust or scratches on the intended array, etc. (Troyanskaya *et al*., 2001). Homemade arrays, which are still used by many laboratories, are more prone to missing observations than commercial arrays. Missing observations can lead to a complete breakdown of an efficient design (i.e. the parameters of interest cannot be estimated unbiasedly anymore) or to a substantial loss in precision of the relevant parameter estimates (Herzberg and Andrews, 1976). In this article, we investigate the robustness of microarray design with respect to missing observations. The importance of robustness in the context of microarray experiments has been stressed in several papers (e.g. Bailey, 2007; Churchill, 2002; Kerr, 2003; Simon *et al*., 2002), but to date no attempts have been made to investigate systematically the robustness considerations in selecting microarray designs. One of the reasons for this could be the absence of any criterion that can quantify the robustness properties of a design. Low *et al*. (1999) presented a method for assessing the robustness of crossover designs to subjects dropping out for studies involving more than two periods. Dey (1993) investigated the robustness of incomplete block design against missing data (John, 1976; Prescott and Mansson, 2001). Tempelman (2005) compared robustness properties of common reference and loop designs, and Tsai *et al*. (2006) mentioned that the optimum designs they constructed for one-factor experiments are robust against one or two missing arrays. None of these approaches suggested any criterion by which robustness properties of a design can be quantified.

The main objective of this article is to formalize different robustness criteria and illustrate their use in selecting good microarray designs for different one-factor and multi-factor experiments. We suggest two robustness criteria, namely, breakdown number and residual efficiency measure for quantifying the robustness properties of a design. Section 2 contains the necessary technical background and the main methodological results. Section 3 illustrates the usefulness of the proposed robustness criteria in selecting good microarray designs from a given set of candidate designs for different one-factor and multi-factor experiments. A short conclusion is given in Section 4.

2 METHODS

2.1 Statistical models

The selection of an efficient microarray design depends on the underlying statistical model and the research questions under investigation. In this section, we describe the statistical model that will be used later to define the efficiency and robustness criteria. Let n denote the number of available arrays from a microarray experiment and let K denote the number of treatments under investigation. For a multi-factor experiment, K can be considered as the total number of treatment combinations (e.g. $K=6$ for a 3×2 experiment).

For a specific gene, we model the n-dimensional vector $\mathbf{Z}$ of normalized log ratios (Huber *et al*., 2002; Yang *et al*., 2002) as

$$\mathbf{Z} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}, \tag{1}$$

where $\mathbf{X}$ denotes the $n \times (K+2)$ design matrix, $\boldsymbol{\beta} = (\delta_1, \delta_2, \tau_1, \ldots, \tau_K)'$ denotes the $(K+2)$-dimensional vector of parameters and $\boldsymbol{\epsilon}$ denotes the n-dimensional vector of independent random errors with mean 0 and variance σ^2. The parameter vector $\boldsymbol{\beta}$ contains the fixed dye effects δ_1 and δ_2 corresponding to the *Cy3* and *Cy5* channels, respectively, and τ_k denotes the k-th treatment mean, $k = 1, \ldots, K$. The model of the form (1) has been proposed by Landgrebe *et al*. (2006) and can be deduced from the global ANOVA model introduced by Kerr and Churchill (2001).

The dye effect is included in our gene-specific model (1) because the standard normalization procedures (Lee *et al*., 2002; Yang *et al*., 2002) can only adjust the overall dye effect, but not the gene-specific dye effect. The gene-specific dye bias is displayed by the genes which do not fall into the overall pattern of the dye effect that characterizes the majority of the genes (Dobbin *et al*., 2003). It was pointed out that even when using normalized data, the dye effects could be significant for some of the genes (Dobbin *et al*., 2003; Kerr, 2003; Landgrebe *et al*., 2006). Thus, we keep the dye effects in the gene-specific model (1).

2.2 Estimability and variance factor

In practice, the research questions of interest can be expressed in terms of a vector of linear functions of the regression parameters $\boldsymbol{\beta}$, e.g. $\mathbf{C}'\boldsymbol{\beta}$, where $\mathbf{C}$ denotes a $(K+2) \times d$ contrast matrix and the value of $d(\geq 1)$ depends on the experimental question. A matrix $\mathbf{C}$ is said to be a contrast matrix if $\mathbf{C}'\mathbf{1}_d = \mathbf{0}_d$, where $\mathbf{1}_d$ and $\mathbf{0}_d$ are d-dimensional vectors with all elements equal to 1 and 0, respectively. Examples for contrast matrices describing relevant experimental questions, such as pairwise treatment comparisons or analysis of interactions, are given in Searle (1971).

Because of the inclusion of the dye effects in the gene-specific model (1) and the fact that the treatment and dye effects are confounded in a single array (Kerr and Churchill, 2001), estimability of the effects of interest becomes an issue. The least squares estimate $\hat{\boldsymbol{\beta}}$ of the regression parameter $\boldsymbol{\beta}$, which is a solution of the consistent system of linear equations $\mathbf{X}'\mathbf{X}\boldsymbol{\beta} = \mathbf{X}'\mathbf{Z}$, is not unique for a non-full rank model similar to (1). However, the estimate of a parametric function $\mathbf{C}'\boldsymbol{\beta}$ is unique, if it is an estimable function. A linear function $\mathbf{C}'\boldsymbol{\beta}$ is said to be estimable, if there exists a linear combination $\mathbf{t}'\mathbf{Z}$ of the response that can be used as an unbiased estimate of $\mathbf{C}'\boldsymbol{\beta}$, i.e. $E(\mathbf{t}'\mathbf{Z}) = \mathbf{C}'\boldsymbol{\beta}$. A necessary and sufficient condition for the estimability of the linear function $\mathbf{C}'\boldsymbol{\beta}$ is

$$\mathbf{C}'(\mathbf{X}'\mathbf{X})^{-}(\mathbf{X}'\mathbf{X}) = \mathbf{C}', \tag{2}$$

where $(\mathbf{X}'\mathbf{X})^{-}$ denotes a generalized inverse of the moment matrix $\mathbf{X}'\mathbf{X}$ (Searle, 1971, §5.4). The concept of estimability is crucial: if a linear function $\mathbf{C}'\boldsymbol{\beta}$ is not estimable, the associated experimental question cannot be answered unbiasedly, i.e. any estimate of $\mathbf{C}'\boldsymbol{\beta}$ deviates from the true value by a systematic, unknown quantity. We call an experimental design with associated design matrix $\mathbf{X}$ *connected*, if all the linear functions under investigation are estimable with respect to $\mathbf{X}$.

Following the Gauss–Markov theorem, the best linear unbiased estimator of an estimable linear function $\mathbf{C}'\boldsymbol{\beta}$ is

$$\mathbf{C}'\hat{\boldsymbol{\beta}} = \mathbf{C}'(\mathbf{X}'\mathbf{X})^{-}\mathbf{X}'\mathbf{Z},$$

which is unique, i.e. $\mathbf{C}'\hat{\boldsymbol{\beta}}$ does not depend on the choice of the generalized inverse of $\mathbf{X}'\mathbf{X}$ (Searle, 1971, p. 181). The variance of the estimator $\mathbf{C}'\hat{\boldsymbol{\beta}}$ is

$$\mathrm{Var}(\mathbf{C}'\hat{\boldsymbol{\beta}}) = \sigma^2\{\mathbf{C}'(\mathbf{X}'\mathbf{X})^{-}\mathbf{C}\}, \tag{3}$$

where $\mathbf{C}'(\mathbf{X}'\mathbf{X})^{-}\mathbf{C}$ is called a variance factor which is a non-negative definite square matrix if $d > 1$.

2.3 Efficiency criteria

For a given contrast matrix $\mathbf{C}$, the efficiency of an experimental design can be quantified by considering the variance factor as a function of the design matrix $\mathbf{X}$. To keep the discussion focused, we follow Landgrebe *et al*. (2006) and use E-optimality to introduce several robustness criteria further below. We note, however, that the methods proposed in this article are equally applicable to other optimality criteria than E-optimality. The E-optimality criterion uses the largest eigenvalue of the variance factor $\mathbf{C}'(\mathbf{X}'\mathbf{X})^{-}\mathbf{C}$ after a suitable standardization. Unlike A- or D-optimality, E-optimality does not depend on the dimension of the variance factor. Morgan (2007) discussed some advantages of E-optimality criterion over the A-optimality criterion in constructing optimal incomplete block designs.

In practice, investigators are often interested in more than one research question from a microarray experiment, e.g. all pairwise treatment comparisons, main effects and interactions, etc. Let $\Theta=\{\mathbf{C}_1'\boldsymbol{\beta},\ldots,\mathbf{C}_Q'\boldsymbol{\beta}\}$ denote the set of linear functions corresponding to the Q individual research questions under investigation. Assume that the experimental design corresponding to the design matrix $\mathbf{X}$ is connected with respect to each of the linear functions of the set Θ. For the given set of linear functions Θ, the E-optimality criterion of the design $\mathbf{X}$ is given by the *efficiency measure*

$$\Psi(\mathbf{X},\Theta)=\sum_{q=1}^{Q}Q^{-1}\left\{\frac{\mathrm{tr}(\mathbf{C}_q'\mathbf{C}_q)}{\lambda_{\max}(\mathbf{C}_q'(\mathbf{X}'\mathbf{X})^{-}\mathbf{C}_q)}\right\}, \tag{4}$$

where $\lambda_{\max}(\mathbf{V})$ and $\mathrm{tr}(\mathbf{V})$ denote the largest eigenvalue and the trace of the square matrix $\mathbf{V}$, respectively. The numerator of the expression (4) is used as a normalizing constant to ensure the invariance of the function Ψ under scalar multiplication, i.e. $\Psi(\mathbf{X},r\mathbf{C}_1'\boldsymbol{\beta})=\Psi(\mathbf{X},\mathbf{C}_1'\boldsymbol{\beta})$ for any scalar r. The efficiency measure Ψ is the average of e-efficiencies, defined in Landgrebe *et al*. (2006), calculated for the Q research questions. Note that in (4) all research questions are weighted equally. If some of the questions are more important than others then one can use a weighted average with unequal weights instead of the simple average used in (4).

2.4 Efficient designs and missing observations

For the robustness considerations in this section, we assume that missing expression values occur completely at random, e.g. missing due to technical reasons. That means, the probability of observing a missing expression measurement is equal across all spots of an array and are constant over different arrays. In the sequel, the gene-specific model (1) is assumed and for a specific gene, each array contributes only one datapoint to the analysis.

It has already been mentioned in Section 1 that missing observations may lead to substantially less efficient or even non-estimable estimates of the effects of interest. We now illustrate the loss of efficiency with a numerical example. Consider two different microarray designs for comparing $K=3$ treatments, namely, *2CR* and *DS*, each of which has six arrays. The graphical representations of these two designs are given in Figure 1, where *1*, *2* and *3* denote the treatments to be compared and *R* denotes the common reference sample. The *2CR* design denotes a design consisting of two replications of the basic common reference *CR* design. The *DS* design compares each pair of the treatment conditions twice by reversing the dye label.

In our example, the *DS* design is found to be more efficient than the *2CR* design when the effect of interest is the treatment difference $\tau_1-\tau_2$. The corresponding values of the efficiency measure Ψ, from Equation (4), are

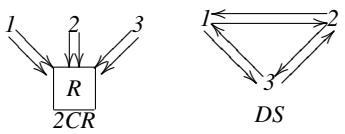

Fig. 1. Graphical representations of *2CR* and DS designs for comparing $K=3$ treatments where *2CR* denotes a design consisting of two replications of the basic *CR* design of the type.

6.0 and 2.0 for the designs *DS* and *2CR*, respectively. That means, in terms of the relative efficiency, the *2CR* design needs to be replicated three times to attain the same efficiency as the *DS* design for estimating $\tau_1-\tau_2$. In other words, the *CR* design requires 18 arrays to attain the same efficiency of the *DS* design which has six arrays.

First, consider the case of one missing observation. For the *2CR* design, assume that an array either of the type $1\longrightarrow R$ or the type $2\longrightarrow R$ is missing. In this case, the efficiency measure Ψ for the comparison $\tau_1-\tau_2$ reduces to 1.3. Note that the efficiency remains unchanged if an array of type $3\longrightarrow R$ is missing. For the *DS* design, Ψ reduces to 3.6 if an array of type $1\longrightarrow 2$ is missing and reduces to 5.1 if the missing observation is associated with the arrays connecting the treatment *3* with any other treatments. This numerical example illustrates that a missing observation may reduce the efficiency of the estimates and the amount of the reduction depends on which array leads to a missing observation. As a side comment, note that these small designs with very low replications will typically have small degrees of freedom and consequently yield poor estimates of the experimental error; and this is before the loss of any arrays.

Consider now the case of more than one missing observation. For two missing observations, the effect $\tau_1-\tau_2$ is not estimable anymore with the *2CR* design if the missing observations come both from either $1\longrightarrow R$ or $2\longrightarrow R$. If one missing observation comes from $1\longrightarrow R$ and another from $2\longrightarrow R$ then the efficiency measure reduces to 1.0. The *2CR* design will not lose any efficiency if both arrays of type $3\longrightarrow R$ are missing. In contrast, the effect $\tau_1-\tau_2$ remains estimable with the *DS* design for any combination of two missing observations (with less efficiency though). If more than three observations are missing, the effect $\tau_1-\tau_2$ is not estimable anymore using the *DS* design, leading to a breakdown of *DS* design.

So far, no attempts have been made to investigate systematically the robustness of microarray designs with respect to missing observations. In the following section, we propose two robustness criteria to measure the robustness property of an experimental design for given experimental questions. All of the following robustness criteria depend on the possible array constellations for a fixed number of missing arrays.

2.5 Robustness criteria

Let $\mathbf{X}_n$ denote the design matrix describing a microarray experiment for which the robustness properties shall be investigated, where n is the size of the design matrix, i.e. the number of arrays used in the experiment. Consider the situation where m out of n observations are missing, i.e. m out of n rows of the design matrix $\mathbf{X}_n$ are missing. Let

$$\mathcal{R}_m(\mathbf{X}_n)=\left\{\mathbf{X}_{n,1}^{(-m)},\mathbf{X}_{n,2}^{(-m)},\ldots,\mathbf{X}_{n,T}^{(-m)}\right\} \tag{5}$$

denotes the set of all possible $T=\binom{n}{m}$ residual design matrices (Dey, 1993) that can be constructed from the design $\mathbf{X}_n$ by leaving m out of n rows. Each residual design has $(n-m)$ arrays, where $\boldsymbol{X}_{n,t}^{(-m)}$ denotes the design matrix corresponding to the t-th residual design, $t=1,\ldots,T$. For example, when $m=0$, the set $\mathcal{R}_0(\mathbf{X}_n)$ contains only the design matrix $\mathbf{X}_n$. Further, define $\theta_q=\mathbf{C}_q'\boldsymbol{\beta}$, $q=1,\ldots,Q$, and let

$$\mathcal{R}_m^{\star}(\mathbf{X}_n,\theta_q)\subseteq\mathcal{R}_m(\mathbf{X}_n) \tag{6}$$

denotes the set of design matrices for which the linear functions θ_q are estimable. In other words, among all the residual designs constructed from the initial design $\mathbf{X}_n$ with m missing arrays, $\mathcal{R}_m^{\star}(\mathbf{X}_n,\theta_q)$ is defined as the set of all connected residual designs with respect to the linear function θ_q. For notational simplicity, we denote the sets $\mathcal{R}_m^{\star}(\mathbf{X}_n,\theta_q)$ and $\mathcal{R}_m(\mathbf{X}_n)$ by $\mathcal{R}_m^{\star}$ and $\mathcal{R}_m$, respectively. Also, we let $T^{\star}(\leq T)$ denote the cardinality of $\mathcal{R}_m^{\star}(\mathbf{X}_n,\theta_q)$.

2.6 Breakdown number

For a given linear function θ_q, the *breakdown number* of a design denotes the minimum number of missing observations that lead to at least one disconnected residual design. More specifically, the breakdown number m_0

(say) states that the effect of interest is estimable for all the residual designs with (m_0-1) missing observations, but there exists at least one residual design with m_0 missing observations for which the effect is no longer estimable. Formally, the breakdown number of a design $\mathbf{X}_n$ for a given linear function θ_q is defined as

$$BdN(\mathbf{X}_n,\theta_q)=\min_m\left\{T^\star<T\right\}, \tag{7}$$

where $T^\star$ and T are as defined above. A large value of the breakdown number indicates larger robustness. For example, to estimate the effect $\tau_1-\tau_2$ in the example from Section 2.4, the breakdown number of the *2CR* design is 2, since two missing observations may already lead to non-estimable comparisons, whereas the breakdown number of the *DS* design is 4. This indicates a larger robustness of the *DS* design as compared with the *2CR* design to estimate $\tau_1-\tau_2$. If more than one linear function is of interest, the minimum of the corresponding breakdown numbers can be used as a robustness criterion, i.e.

$$BdN(\mathbf{X}_n,\Theta)=\min\left\{BdN(\mathbf{X}_n,\theta_q),q=1,\ldots,Q)\right\}, \tag{8}$$

where Θ is already defined for Equation (4). Note that the breakdown number does not depend on the optimality criterion but depends on the estimability of the linear function under investigation. Table 1 gives the pseudo code to compute the breakdown number for a given design and research question.

2.7 Residual efficiency measure

For a given design, the breakdown number provides the minimum number of missing observations that lead to at least one disconnected residual design. In practice, one may need to select a good design among several designs having the same breakdown number. In such a situation, the average efficiency for all residual designs with a fixed number of missing observations can be used to select a good design and it can be used as a robustness criterion. For a given linear function θ_q, the *residual efficiency measure* of the design $\mathbf{X}_n$ with m missing arrays is defined as

$$\Psi_m(\mathbf{X}_n,\theta_q)=\begin{cases}\displaystyle\sum_{\mathbf{X}\in\mathcal{R}_m} T^{-1}\Psi(\mathbf{X},\theta_q), & \text{if } m<BdN\\ \infty, & \text{otherwise.}\end{cases} \tag{9}$$

Note that the residual efficiency measure can only be computed if the number of missing arrays is less than the corresponding breakdown number. Similar to the efficiency measure Ψ, larger values of the residual efficiency measure indicate larger robustness. For a fixed number m of missing arrays, the residual efficiency measure can be used to compare the robustness characteristics of two or more designs provided their individual breakdown numbers are all greater than m. For more than one linear function, the average of the corresponding residual efficiency measures (9) can be used as a robustness criterion. Note that if there are no missing observations, $\Psi=\Psi_0$, i.e. the efficiency measure and the residual efficiency measure are equal.

Table 1. Pseudo code to compute breakdown number and residual efficiency measure for a given design and research question (the residual efficiency measure will be introduced in the following section)

- Check that the linear function $\theta=\mathbf{C}'\boldsymbol{\beta}$ is estimable for the design matrix $\mathbf{X}_n$
- Set $m=1$

REPEAT {
- generate the set of possible $T=\binom{n}{m}$ residual design matrices $\mathcal{R}_m(\mathbf{X}_n)$ as shown in (5)
- check estimability of θ for each $\mathbf{X}\in\mathcal{R}_m$ using (2)
- **IF** θ is estimable for each $\mathbf{X}\in\mathcal{R}_m$ **THEN** {
  - compute $\Psi_m(\mathbf{X}_n,\theta)$ using (9)
  - $m=m+1$

} **ELSE** $BdN=m$ and **STOP**

} **UNTIL** $m<n$

If the number of missing arrays is greater than or equal to the corresponding breakdown number, the residual efficiency measure cannot be used as robustness criterion. In such situations, the proportion of connected designs (ratio of $T^\star$ to T) can be used as a robustness criterion [see Latif (2005) for further details].

Table 1 also gives the pseudo code to compute the residual efficiency measure for a given design and research question. A related R Development Core Team (2008) package that implements the pseudo code from Table 1 for *A*-, *D*- and E-optimality is available from the first author upon request.

Biological replications are often used in microarray experiments to enable researchers making inferences about treatment effects for the populations included in the experiment. To analyse microarray data with biological replications, linear mixed effects models need to be used instead of fixed effects models of the type (1). Note, however, that the robustness criteria considered in this section can be extended to include also mixed effects models. For instance, the breakdown number can be defined under mixed effects models using the definition of estimable linear function given in Passos *et al*. (2009). In principle, the methods from this article can thus be extended to microarray experiments including biological replications. However, more theoretical work is required to formalize optimality considerations for mixed effects models.

3 RESULTS

In this section, we illustrate the methods from Section 2 by comparing the efficiency and robustness of several designs for one- and two-factor experiments. The basic designs for each of the experimental layouts are defined first and replications or combinations thereof are used to construct composite designs. By replication, we mean here technical replications.

We have already used the notation *2CR* for the design that consists of two replications of the basic *CR* design. In general, we will use the following notation:

(i) b*XX* denotes a design that consists of b $(=2,3,\ldots)$ replications of the *XX* design.
(ii) $\overline{XX}$ denotes the design that uses the reverse dye labelling of the *XX* design.
(iii) b$XX/\overline{XX}$ denotes the design that consists of the arrays of the b*XX* and $\overline{XX}$ designs.

3.1 One-factor experiments

We now investigate the robustness of some microarray designs to compare all pairwise combinations of $K=4$ treatments. The basic microarray designs considered here are *CR* and *CL*, each of which consists of four arrays. The corresponding *DS* design has 12 arrays, see Figure 2. The $\overline{CL}$ ($\overline{CR}$) design, which uses the opposite dye labelling protocol of the *CL* (*CR*) design, is considered here to examine the effect of having different dye labelling protocols in different replications of the *CL* (*CR*) design.

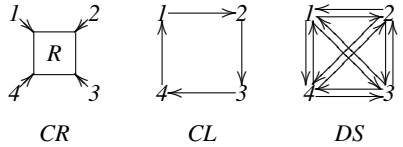

Fig. 2. Graphical representations of the microarray designs with four arrays (*CR* and *CL*) and 12 arrays (*DS*) for comparing $K=4$ treatments.

Table 2. Robustness and efficiency criteria Ψ_m for selected designs with eight arrays for all pairwise comparisons of $K=4$ treatments

Design	*BdN*	Ψ_0	Ψ_1	Ψ_2	Ψ_3
2CR	2	2.00	1.67	–	–
CR/$\overline{CR}$	2	2.00	1.60	–	–
2CL	2	**4.89**	**4.04**	–	–
CL/$\overline{CL}$	**4**	**4.89**	**4.04**	**3.19**	**2.33**

Table 3. Robustness and efficiency criteria Ψ_m for selected designs with 12 arrays for all pairwise comparisons of $K=4$ treatments

Design	*BdN*	Ψ_0	Ψ_1	Ψ_2	Ψ_3	Ψ_4	Ψ_5
3CR	3	3.00	2.70	2.40	–	–	–
3CL	3	7.33	6.59	5.84	–	–	–
2CL/$\overline{CL}$	**6**	7.33	6.59	5.84	5.08	4.32	3.55
DS	**6**	**8.00**	**7.18**	**6.36**	**5.54**	**4.72**	3.06

3.1.1 Scenario I Consider the situation where we are only interested in the designs with eight arrays. Four composite designs with eight arrays are constructed from the basic *CR* and *CL* designs (Fig. 2): *2CR*, *CR*/$\overline{CR}$, *2CL* and *CL*/$\overline{CL}$.

Table 2 shows the corresponding *BdN* values and residual efficiency measures Ψ_m, $m<BdN$. Among the designs with eight arrays, the loop designs *2CL* and *CL*/$\overline{CL}$ are found to be more efficient ($\Psi_0=4.89$) than the common reference designs *2CR* and *CR*/$\overline{CR}$ ($\Psi_0=2.00$) for $m=0$. The *CL*/$\overline{CL}$ design is found to be more robust ($BdN=4$) than the other three designs ($BdN=2$). Clearly, reverse dye labelling for different replications can improve substantially the robustness for some designs, such as the loop design. This example also demonstrates the usefulness of the breakdown number in comparing the two designs *2CL* and *CL*/$\overline{CL}$ which are equally efficient if there are no missing observations.

3.1.2 Scenario II Assume now that we are interested in the four designs *3CR*, *3CL*, *2CL*/$\overline{CL}$ and *DS*, where each of them uses 12 arrays. As seen from Table 3, the *3CR* design ($\Psi_0=3.00$) is less efficient than the loop designs ($\Psi_0=7.33$) if no observation is missing (i.e. $m=0$). Among the competing designs, the *2CL*/$\overline{CL}$ and *DS* designs are more robust ($BdN=6$) as compared with the *3CL* and *3CR* designs ($BdN=3$). For this setup, the *DS* design is recommended because it is slightly more efficient than the *2CL*/$\overline{CL}$ design, even after accounting for the possibility of missing observations.

3.2 Two-factor experiments

In this section, we consider two-factor experiments. To be more precise, we consider the efficiency and robustness of several designs for a 3×2 experiment. In a 3×2 experiment, one factor (say, *A*) has three levels and the other (say, *B*) has two levels. We assume that the effects of interest are the two main effects (*A*, *B*) and the interaction ($A\times B$). Landgrebe *et al*. (2006) discussed some basic types of the microarray designs, namely, circular loop (*CL*), cross–loop (*XL*), triangular loop (*TL*), *A*-loop (*AL*), *B*-swap (*BS*) and star-swap (*RS*) (Fig. 3) and reported efficient composite designs for different combinations of the effects of interest.

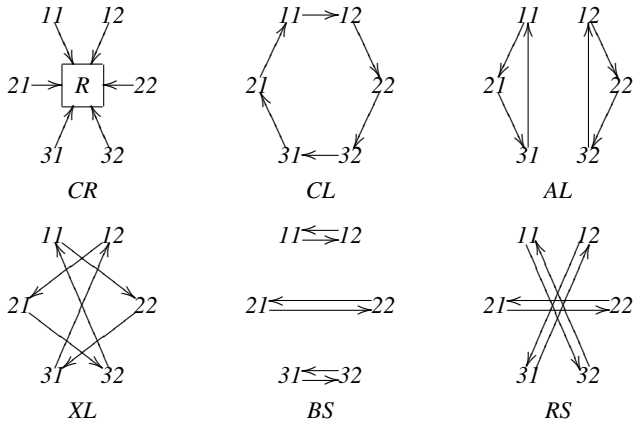

Fig. 3. Graphical representation of basic microarray designs for the 3×2 experimental layout, each of which has six arrays. Treatment combinations are specified by a pair of the treatment labels corresponding to the factors *A* and *B*.

3.2.1 Scenario I Among the basic designs with six arrays, the main effects (*A* and *B*) and interaction ($A\times B$) are only found estimable with the designs *CR*, *XL* and *CL*. The *XL* design is found to be more efficient than the *CR* design for estimating the main effects *A* ($\Psi=6.00$ versus $\Psi=2.00$) and *B* ($\Psi=4.00$ versus $\Psi=1.00$). The *CL* design is more efficient than the *CR* design for estimating the main effect *B* ($\Psi=1.10$ versus $\Psi=1.00$) and interaction $A\times B$ ($\Psi=6.00$ versus $\Psi=2.00$). If all three effects are of equal interest, the *XL* design is found to be the most efficient design. The *CL* design is found to be the best design if only the interaction is of interest. For the *CR*, *XL* and *CL* designs the breakdown number is 1 irrespective of the effects *A*, *B* and $A\times B$.

3.2.2 Scenario II The designs with 12 arrays for the 3×2 experiment are constructed from combinations and/or replications of the basic designs as shown in Figure 3. In this case, the concept of the admissible designs (Glonek and Solomon, 2004) is used to reduce the number of candidate designs. Among the 21 possible designs with 12 arrays, six designs are found to be admissible. The efficiency measures and breakdown numbers of the six admissible designs are reported in Table 4. For comparison purposes, the values of the efficiency measures and robustness criteria corresponding to the designs *2CR* and *XL*/$\overline{XL}$ are also reported in Table 4.

If all three effects are of equal interest, the *2XL*, *AL/XL*, *AL/BS*, *XL*/$\overline{XL}$ and *XL/BS* designs are most efficient ($\Psi_0=8.00$) and among these designs, the *2XL* design is found to be less robust as compared with the other four designs in terms of the minimum breakdown numbers. Table 5 displays a comparison between these four equally efficient ($\Psi_0=8.00$) and robust ($BdN=4$) designs on the basis of the residual efficiency measure Ψ_m, $m<4$. Table 5 shows that the design *XL/BS* outperforms the other three designs.

If only the effect $A\times B$ is of interest, the design *AL/BS* is found to be the most efficient ($\Psi=14.00$) and robust ($BdN=4$). If the effects *A* and $A\times B$ are of interest, the *CL/AL*, *AL/XL* and *AL/BS* designs are most efficient ones ($\Psi=10.00$), however, the *CL/AL*

Table 4. Efficiency measure and breakdown number for designs with 12 arrays in a 3×2 experiment

Design	Ψ				*BdN*			
	A	*B*	$A \times B$	average	*A*	*B*	$A \times B$	min
2CR	4.00	2.00	4.00	3.33	2	2	2	2
CL/AL	8.00	1.10	12.00	7.03	4	2	4	2
CL/XL	8.00	5.14	8.00	7.05	4	4	4	4
2XL	**12.00**	**8.00**	4.00	8.00	2	2	2	2
AL/XL	**12.00**	4.00	8.00	8.00	4	4	4	4
XL/$\overline{XL}$	**12.00**	**8.00**	4.00	8.00	4	4	4	4
XL/BS	6.00	**8.00**	10.00	8.00	4	4	4	4
AL/BS	6.00	4.00	**14.00**	8.00	4	4	4	4

Table 5. Residual efficiency measures for the selected designs from a 3×2 experiment when the main effects and interaction are of equal interest

Design	Ψ_0	Ψ_1	Ψ_2	Ψ_3
XL/BS	8.00	**6.42**	**5.20**	**4.08**
XL/$\overline{XL}$	8.00	6.29	5.10	3.98
AL/BS	8.00	6.25	5.10	3.97
AL/XL	8.00	6.20	5.00	3.81

Table 6. For different combinations of effects, the best designs for 3×2 experiment with 6 and 12 of arrays

Number of arrays (*n*)	Effect combinations of interest			
	$A \times B$	$A \times B$, *A*	$A \times B$, *B*	$A \times B$, *A*, *B*
6	*CL*	*CL, XL*	*CL*	*XL*
12	*AL/BS*	*AL/XL, AL/BS*	*AL/BS, XL/BS*	*AL/BS*, *XL/*$\overline{XL}$ *AL/XL, XL/BS*

design is less robust ($BdN = 2$) compared with the other two designs ($BdN = 4$). The *AL/BS* and *XL/BS* designs are found to be the best designs ($\Psi = 9.00$) when the effects $A \times B$ and *B* are of interest and both of these designs are equally robust ($BdN = 4$).

Table 6 shows the efficient and robust designs for 3×2 experiments with 6 and 12 arrays for different combinations of effects of interest. This result shows that one could find more efficient and robust designs than the *CR* design for 3×2 experiments.

4 CONCLUSIONS

Many papers have been published in selecting good/efficient microarray designs for both one-factor and multi-factor experiments. Though missing values are often observed in microarray experiments, so far no attempts have been made to include the possibility of having missing observations in the selection procedure of good microarray designs. In this article, two robustness criteria (breakdown number and residual efficiency measure) are proposed that can be used to quantify the robustness characteristics of a design. The use of the proposed robustness criteria in selecting good microarray designs from a set of candidate designs has been illustrated with designs for both the one- and two-factor experiments.

For pairwise treatment comparisons, the common reference designs are less efficient as compared with loop designs in one-factor experiments. However, the common reference design is equally robust to the loop design only if the number of arrays is equal to the number of treatments. Reverse dye labelling improves the robustness of the loop designs, but does not do the same for the corresponding common reference design.

Selection of good designs depends on both the research questions under investigation and the number of arrays the experimenter intend to use. In this study, we only consider designs that are replicates of initially considered basic designs. A more general procedure is required to find good designs with any number of arrays. The novel approach of using both efficiency and robustness criteria in selecting good microarray designs can be easily tailored to such general questions.

ACKNOWLEDGEMENTS

The authors are grateful to three anonymous referees and Ben Parker, Queen Mary, University of London, UK, for their helpful comments, which improved the presentation of the article. The first author would like to acknowledge the support he received for this work during participating DoE program at the Isaac Newton Institute at the University of Cambridge, UK.

Funding: UK Engineering and Physical Sciences Research Council (grant number EP/C541715/1 to A.H.M.M.L., in part); BMBF project SKAVOE (to F.B., in part)

Conflict of Interest: none declared.

REFERENCES

Atkinson,A.C. *et al.* (2007) *Optimum Experimental Designs, with SAS*. Oxford: Oxford University Press.

Bailey,R.A. (2007) Designs for two-color microarray experiments. *Appl. Stat.*, **54**, 365–394.

Bueno Filho,J.S.S. *et al.* (2006) Design of microarray experiments for genetical genomic studies. *Genetics*, **174**, 945–957.

Callow,M.J. *et al.* (2000) Microarray expression profiling identifies genes with altered expression in HDL-deficient mice. *Genome Res.*, **10**, 2022–2029.

Churchill,G.A. (2002) Fundamentals of experimental design for cDNA microarrays. *Nat. Genet.*, **32** (Suppl. S), 490–495.

Dey,A. (1993) Robustness of block designs against missing data. *Stat. Sin.*, **3**, 219–231.

Dobbin,K. *et al.* (2003) Questions and answers on design of dual-label microarrays for identifying differentially expressed genes. *J. Natl Cancer Inst.*, **95**, 1362–1369.

Glonek,G.F.V. and Solomon,P.J. (2004) Factorial and time course designs for cDNA microarray experiments. *Biostatistics*, **5**, 89–111.

Gupta,S. (2006) Balanced factorial designs for cDNA microarray experiments. *Commun. Stat. Theory Methods*, **35**, 1469–1476.

Herzberg,A.M. and Andrews,D.F. (1976) Some considerations in the optimal design of experiments in non-optimal situations. *J. R. Stat. Soc. B*, **38**, 284–289.

Huber,W. *et al.* (2002) Variance stabilization applied to microarray data calibration and to the quantification of differential expression. *Bioinformatics*, **18** (Suppl. 1), 96–104.

John,W.M.J. (1976) Robustness of balanced incomplete block designs. *Ann. Stat.*, **4**, 960–962.

Kerr,M.K. (2003) Design considerations for efficient and effective microarray studies. *Biometrics*, **59**, 822–828.

Kerr,M.K. and Churchill,G.A. (2001) Experimental design for gene expression microarrays. *Biostatistics*, **2**, 183–201.

Kerr,M.K. *et al.* (2000) Analysis of variance for gene expression microarray data. *J. Comput. Biol.*, **7**, 819–837.

Landgrebe,J. *et al.* (2006) Efficient design and analysis of two-color factorial microarray design. *Comput. Stat. Data Anal.*,**50**, 499–517.

Latif,A.H.M.M. (2005) Efficiency and Robustness Issues in Complex Statistical Designs for Two-color Microarray Experiments. PhD Thesis, Center for Statistics, University of Goettingen, Germany.

Lee,M.-L.T. *et al.* (2002) Models for microarray gene expression data. *J. Biopharm. Stat.*, **12**, 1–19.

Low,J.L. *et al.* (1999) Assessing robustness of crossover designs to subjects dropping out. *Stat. Comput.*, **9**, 219–227.

Morgan,J.P. (2007) Optimal incomplete block designs. *J. Am. Stat. Assoc.*, **102**, 655–663.

Passos,V.L. *et al.* (2009) Optimal designs for one- and two-color microarrays using mixed models: comparative valuation of their efficiencies. *J. Comput. Biol.*, **16**, 67–83.

Prescott,P. and Mansson,R. (2001) Robustness of balanced incomplete block designs to randomly missing observations. *J Stat. Plan. Inference*, **92**, 283–296.

Pukelsheim,F. (1993) *Optimal Designs of Experiments*. John Wiley & sons, New York.

R Development Core Team (2008) *R : A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

Searle,S.R. (1971) *Linear Models*. John Wiley & sons, New York.

Simon,R. *et al.* (2002) Design of studies using DNA microarrays. *Genet. Epidemiol.*, **23**, 21–36.

Stanzel,S. (2008) Optimale statistische Versuchsplanung dreifaktorieller Zwei–Farben cDNA–Microarray–Experimente. PhD Thesis, Fakultät für Statistik, Technische Universität Dortmund, Germany.

Tempelman,R.J. (2005) Assessing statistical precision, power, and robustness of alternative experimental designs for two-color microarray platform based on mixed effects models. *Vet. Immunol. Immunopathol.*, **105**, 175–186.

Tempelman,R.J. (2008) Statistical analysis of efficient unbalanced factorial designs for two-color microarray experiments. *Int. J. Plant Genomics*, **2008**, 1–16.

Troyanskaya,O. *et al.* (2001) Missing value estimation methods for DNA microarrays. *Bioinformatics*, **17**, 520–525.

Tsai,S.-F. *et al.* (2006) Statistical designs for two-color microarray experiments involving technical replications. *Comput. Stat. Data Anal.*, **51**, 2078–2090.

Wit,E. *et al.* (2005) Near-optimal designs for dual channel microarray studies. *Appl. Stat.*, **54**, 817–830.

Wolfinger,R.D. *et al.* (2001) Assessing gene significance from cDNA microarray expression data via mixed models . *J. Comput. Biol.*, **8**, 625–637.

Yang,Y. and Speed,T. (2002). Design issues for cDNA microarray experiments. *Nat. Rev. Genet.*, **3**, 579–588.

Yang,Y.H. *et al.* (2002). Normalization for cDNA microarray data: a robust composite method addressing single and multiple slide systematic variation. *Nucleic Acids Res.*, **30**, e15.