A Link between the E-value and the Robustness of Block Designs

J.D.Godolphin

Department of Mathematics, University of Surrey, UK

October 9, 2015

Abstract

Robustness of binary incomplete block designs against giving rise to a disconnected design in the event of observation loss is investigated. A link is established between the E-value of a planned design and the extent of observation loss that can be experienced whilst still guaranteeing an eventual design from which all treatment contrasts can be estimated. Patterns of missing observations covered include loss of entire blocks and loss of individual observations. Simple bounds are provided enabling practitioners to easily assess the robustness of a planned design.

Keywords: breakdown number, connectivity, eigenvalue, information matrix

1 Introduction

Incomplete block designs are used in a wide variety of applications. Typical blocking factors include *batches of raw materials*, *machines*, *fields*, *days*, *sites*. When an experiment is designed involving v treatments, the treatment replications and number and sizes of the blocks depend on resources and practical considerations. Practical limitations may also inform allocation of treatments to blocks: the physical location of a block might cause allocation of specific treatments to be associated with additional cost. The planned design is selected to have good properties within the class of designs satisfying all the constraints.

During experimentation it is not uncommon for some observations to be unavailable: fields of crops can be destroyed by storm; microarray plates can break; an experiment can end prematurely due to funding issues. For some experimental situations, observation loss tends to involve the loss of individual observations at random. For others, the nature of the blocking factor means that observation loss tends to involve entire blocks. In the event of observation loss, the properties of the eventual design can be markedly inferior compared to those of the planned design. As an extreme consequence it will not be possible to test the hypothesis that the treatment effects are equal, or to estimate all treatment contrasts. Two features of the eventual design can give rise to this eventuality. The missing observations may include all replicates of one or more treatments. In this case it may still be possible to make all comparisons between the remaining treatments. Alternatively, the eventual design might contain all v treatments but be disconnected. This occurs when there are consistent partitionings between the v treatments and the blocks of the eventual design.

Dey (1993) introduced two criteria for assessing the robustness of a design against missing data. Following work of Ghosh (1979, 1982), a design is described as being robust against the loss of n observations or t blocks according to Criterion 1, if all treatment contrasts are estimable from every possible eventual design resulting from such observation loss. A planned design is robust according to Criterion 2, if the A-efficiency of any eventual design compares well with that of the planned design.

Much of the work on Criterion 1 robustness is restricted to particular design classes and very specific observation loss. Several authors, including Ghosh (1982) and Bailey *et al.* (2013) demonstrate that various design types, which are universally optimal or D-optimal,

also have optimal Criterion 1 properties. The relevance of designs with blocks of size two in microarray experiments forms a motivation for Mahbub Latif *et al.* (2009), Tsai and Liao (2013) and Bailey *et al.* (2013) to investigate robustness of such designs. Dey *et al.* (2001) focus on the loss of a single observation or a single block from diallel cross designs. Design properties associated with Criterion 1 robustness in the event of the loss of complete blocks from any binary incomplete block design are derived by Baksalary and Tabis (1987), Sathe and Satam (1992) and Godolphin and Warren (2011). Notz *et al.* (1994) investigate the loss of n observations or t blocks according to Criterion 1, Lal *et al.* (2001) give a condition for a block design to be robust against the loss of any n observations and Godolphin and Warren (2014) provide an algorithm for determining robustness of any binary design with all blocks of size k with respect to loss of individual observations.

Notable work on Criterion 2 robustness has been completed by Bhaumik and Whittinghill (1991), Prescott and Mansson (2001) and Morgan and Parvu (2008) who investigate the properties of eventual designs following the loss of a small number of observations or blocks from balanced incomplete block designs. Related work is by Srivastava *et al.* (1991), Das and Kageyama (1992), Ghosh *et al.* (1992) and Bhar (2014).

Criterion 1 robustness is the focus of this paper. The aim is to obtain bounds on the extent of observation loss that a planned design is guaranteed to stand without giving rise to an eventual design from which some treatment contrasts are inestimable. Patterns of observation loss comprising individual observations and complete blocks are considered. Bounds based on the E-value of a design, which improve on the results of Lal *et al.* (2001), are derived. The main results make extensive use of the concepts of treatment and block partitioning. For proper designs, i.e. designs for which all blocks are the same size, lower bounds for the number of blocks, and the number of individual observations, that can be lost are functions of the E-value of the design and of basic design parameters. For designs with blocks of more than one size, the bounds also depend on values obtained via simple algorithms. Thus a straightforward means of using the E-value of a planned incomplete block design as a tool to assess the Criterion 1 robustness of the design is achieved. The results in this paper are not restricted to specific classes of incomplete block design but can be applied to any connected design that is binary, i.e. no treatment occurs more than

once within a block.

The paper is organised as follows. The model is introduced in Section 2 along with preliminary concepts and basic results on connectivity and Criterion 1 robustness. In Section 3, robustness conditions are derived linking treatment partitions and the E-value. Section 4 focuses on proper designs: bounds for the extent of observation loss for which a connected eventual design is guaranteed are established. In Section 5 the results are extended to cover designs with blocks of different sizes. The results are illustrated by examples.

2 Preliminaries

Let D denote a connected binary block design on v treatments applied to experimental units arranged in b blocks. The usual linear model applies with y_{il} , the response obtained by application of treatment with label i to an experimental unit in block l , given by:

$$y_{il} = \mu + \tau_i + \beta_l + \epsilon_{il},$$

where μ is the overall mean, τ_i is the effect of treatment i , β_l is the effect of block l and the ϵ_{il} are uncorrelated random errors with common variance σ^2 . Treatment i is applied to r_i experimental units and the l th block contains k_l experimental units. Allocation of treatments to blocks of D is described by the $v \times b$ incidence matrix $N = (n_{il})$, with $n_{il} = 1$ if treatment i is assigned to an experimental unit in block l , and $n_{il} = 0$ otherwise. The matrix $NK^{-1}N'$ is termed the *weighted concurrence matrix* of D , where N' denotes the transpose of N and K^{-1} is the inverse of the diagonal matrix $K = \text{diag}(k_1, \dots, k_b)$. Weighted treatment concurrences are defined as:

$$\kappa_{ij} = \sum_{l=1}^{l=b} \frac{n_{il}n_{jl}}{k_l}, \text{ for } i, j \in \{1, \dots, v\}, i \neq j$$

and occur as the off-diagonal terms of $NK^{-1}N'$.

Mahbub Latif *et al.* (2009) use the term *breakdown number* for a design D to denote the number of observations that need to be lost before the possibility that not all treatment contrasts are estimable from an eventual design. The concept can be extended to cover

the loss of entire blocks and the term *block breakdown number* is taken to be the number of blocks that need to be lost before the possibility of an eventual design from which not all treatment contrasts are estimable. The notation n_* and t_* will be used for the breakdown and block breakdown numbers of D . An eventual design is denoted by $D_\#$. Classes of eventual designs resulting from the loss of p observations and p blocks, respectively, are given by $D(1, p)$ and $D(2, p)$.

There is some commonality between the situations of the loss of individual observations and of entire blocks. This can be demonstrated by consideration of a design D with blocks all of size k . If $n_* > m(k - 1)$ then all treatment contrasts will be estimable from any $D_\#$ resulting from the loss of all but one observation in each of m blocks. Since blocks containing a single observation do not make any contribution to estimation of treatment contrasts, it follows that $t_* > m$. The converse does not hold in general.

Basic results on n_* and t_* depend on the magnitudes of the treatment replications which are arranged in decreasing order as:

$$r_{[1]} \geq r_{[2]} \geq \cdots \geq r_{[v]}. \quad (1)$$

Theorem 1 *Let D be a connected binary design. The breakdown and block breakdown numbers satisfy $t_* \leq n_* \leq r_{[v]}$.*

Proof: Both $D(1, r_{[v]})$ and $D(2, r_{[v]})$ have at least one member which contains fewer than v treatments. Thus $n_* \leq r_{[v]}$ and $t_* \leq r_{[v]}$. Consider a set of n_* observations whose removal gives rise to a $D_\#$ from which not all treatment contrasts are estimable. The observations in such a set must be located in at most n_* blocks. Removal of these entire blocks gives an eventual design from which not all contrasts are estimable. Thus $t_* \leq n_*$.

From Theorem 1, it is evident that a design which is robust against the loss of any p blocks is also robust against the loss of any p observations. Designs with $t_* = r_{[v]}$ are described as being maximally robust with regards to loss of individual observations and loss of entire blocks. Baksalary and Tabis (1987), Sathe and Satam (1992) and Godolphin and Warren (2011) derive conditions for a binary incomplete block design to be maximally robust. However, the concept of maximal robustness has the limitation that the size of $r_{[v]}$ sets the bar at which a design has the status of being maximally robust. Therefore, breakdown numbers can be considered as being more useful measures of Criterion 1 robustness

and are used in this work as a means of assessing design robustness against observation loss. Designs with $r_{[v]} = \text{int}\{\sum_{i=1}^b k_i/v\}$, where $\text{int}\{\cdot\}$ denotes *the integer part of*, which have $n_* = t_* = r_{[v]}$ have optimal robustness against giving rise to eventual designs from which not all treatment contrasts are estimable.

The concept of a treatment partition, denoted by S , is now described. This is an arrangement of the treatments into disjoint non-empty sets, S_1 and S_2 , of sizes h and $v - h$ with $1 \leq h \leq v/2$. A block partition, B , is a partitioning of the blocks of D or $D_\#$ into disjoint non-empty sets B_1 and B_2 . Design D or $D_\#$ is disconnected if and only if a treatment partition exists which is consistent with a block partition so that, all replicates of treatments in S_i are contained in blocks in B_i , for $i = 1, 2$. For a planned experiment to result in a disconnected eventual design is generally disastrous for the aims of the experiment since treatment comparisons can only be made within S_1 and within S_2 . Alternative and equivalent definitions of connectivity in terms of chains between blocks and treatments and in terms of the rank of the information matrix are given by Bose (1963) and Chakrabarti (1963) respectively.

The concepts of connectivity and breakdown numbers are critical in this work and are illustrated by means of Examples 1 and 2. In both cases the design comprises six replications of each of ten treatments arranged in fifteen blocks of size four.

Example 1 Design $D1$ is given by

$$D1 = \begin{array}{cccccccccccccccc} 1 & 1 & 1 & 1 & 2 & 6 & 6 & 6 & 6 & 7 & 1 & 1 & 2 & 2 & 3 \\ 2 & 2 & 2 & 3 & 3 & 7 & 7 & 7 & 8 & 8 & 3 & 4 & 5 & 6 & 4 \\ 3 & 3 & 4 & 4 & 4 & 8 & 8 & 9 & 9 & 9 & 6 & 9 & 8 & 7 & 5 \\ 4 & 5 & 5 & 5 & 5 & 9 & 10 & 10 & 10 & 10 & 7 & 10 & 9 & 8 & 10 \end{array}$$

where, as with all designs displayed in this work, columns show the blocks. By use of the procedure of Godolphin and Warren (2014), it can be established that every member of $D1(1, 5)$ is connected. Thus $n_* = r_{[v]} = 6$ from Theorem 1. Similarly, by considering the loss of all $\binom{15}{4}$ sets of four blocks, every member of $D1(2, 4)$ is found to be connected and so $t_* \geq 5$. However, removal of blocks (columns) 11 to 15, inclusive, gives a disconnected eventual design since the treatment partition $S_1 = \{1, 2, 3, 4, 5\}$, $S_2 = \{6, 7, 8, 9, 10\}$ is consistent with the block partition $B_1 = \{1, 2, 3, 4, 5\}$, $B_2 = \{6, 7, 8, 9, 10\}$. Thus, $t_* = 5$.

Example 2 Design $D2$ is obtained from $D1$ by rearrangement of the treatments in blocks 11 to 15.

$$D2 = \begin{array}{cccccccccccccccc} 1 & 1 & 1 & 1 & 2 & 6 & 6 & 6 & 6 & 7 & 1 & 6 & 1 & 6 & 4 \\ 2 & 2 & 2 & 3 & 3 & 7 & 7 & 7 & 8 & 8 & 2 & 7 & 2 & 7 & 5 \\ 3 & 3 & 4 & 4 & 4 & 8 & 8 & 9 & 9 & 9 & 3 & 8 & 3 & 8 & 9 \\ 4 & 5 & 5 & 5 & 5 & 9 & 10 & 10 & 10 & 10 & 4 & 9 & 10 & 5 & 10 \end{array}$$

Use of Godolphin and Warren (2014) establishes that every member of $D2(1, 3)$ is connected. However, loss of replicates of treatments 10 and 5 from blocks 13 and 14, together with loss of replicates of 4 and 5, or of 9 and 10, from block 15 yields two disconnected members of $D2(1, 4)$, giving $n_* = 4$. The two disconnected eventual designs in $D2(1, 4)$ both have treatment partition $S_1 = \{1, 2, 3, 4, 5\}$, $S_2 = \{6, 7, 8, 9, 10\}$ consistent with a block partition. With the loss of replicates of 4 and 5 from block 15, the block partition is $B_1 = \{1, 2, 3, 4, 5, 11, 13\}$, $B_2 = \{6, 7, 8, 9, 10, 12, 14, 15\}$ and the loss of replicates of 9 and 10 from block 15 corresponds to block partition $B_1 = \{1, 2, 3, 4, 5, 11, 13, 15\}$, $B_2 = \{6, 7, 8, 9, 10, 12, 14\}$. By considering the loss of all sets of two blocks, every member of $D2(2, 2)$ is found to be connected and so $t_* \geq 3$. Removal of blocks 13 to 15 yields a disconnected member of $D2(2, 3)$ with S as above and block partition $B_1 = \{1, 2, 3, 4, 5, 11\}$, $B_2 = \{6, 7, 8, 9, 10, 12\}$. Thus $t_* = 3$. For all three disconnected eventual design described, comparisons can only be made between treatments within S_1 and between treatments in S_2 .

3 General results on design robustness

3.1 Treatment partitions and connectivity

Theorem 1 establishes $r_{[v]}$ as an upper bound for both n_* and t_* . In the remaining part of this work, observation loss is confined to loss from at most $r_{[v]} - 1$ blocks, up to and including the loss of $r_{[v]} - 1$ entire blocks. This restriction ensures that $D_{\#}$ involves v treatments and $b_{\#}$ blocks, where $b - r_{[v]} + 1 \leq b_{\#} \leq b$. Since $D_{\#}$ contains at least one replicate of each treatment in D , all treatment contrasts will be estimable in $D_{\#}$ unless $D_{\#}$ is disconnected.

An eventual design, $D_\#$, will have $v \times b_\#$ incidence matrix $N_\#$. Define $K_\#$ as the $b_\# \times b_\#$ diagonal matrix containing the sizes of each block of $D_\#$. The relationship between the weighted concurrence matrices of D and $D_\#$ is given by:

$$NK^{-1}N' = N_\#K_\#^{-1}N'_\# + \Psi. \quad (2)$$

The off-diagonal terms of $\Psi = (\psi_{ij})$ give the change in each weighted concurrence as a result of the observation loss. Some weighted concurrences can be larger in $D_\#$ than in D , but terms in $N_\#K_\#^{-1}N'_\#$ are all non-negative. Consider any treatment partition, S . The rows of N and $N_\#$ can be arranged as $[N'_1 \ N'_2]'$ and $[N'_{\#1} \ N'_{\#2}]'$ where rows of N_i and $N_{\#i}$ correspond to treatments of S_i in the same order, for $i = 1, 2$. Then (2) is represented in a manner consistent with S as:

$$\begin{bmatrix} N_1K^{-1}N'_1 & N_1K^{-1}N'_2 \\ N_2K^{-1}N'_1 & N_2K^{-1}N'_2 \end{bmatrix} = \begin{bmatrix} N_{\#1}K_\#^{-1}N'_{\#1} & N_{\#1}K_\#^{-1}N'_{\#2} \\ N_{\#2}K_\#^{-1}N'_{\#1} & N_{\#2}K_\#^{-1}N'_{\#2} \end{bmatrix} + \begin{bmatrix} \Psi_1 & \Psi_{12} \\ \Psi'_{12} & \Psi_2 \end{bmatrix}. \quad (3)$$

Entries in Ψ_{12} give the change in each weighted concurrence, κ_{ij} , involving a treatment in S_1 and one in S_2 . Denote the sum of these terms by $\Sigma_S = 1'_h \Psi_{12} 1_{v-h}$, where 1_a denotes an $a \times 1$ vector with every element unity.

Theorem 2 *Let D be a binary connected design and let $D_\#$ be an eventual design. The following inequality holds for every treatment partition, $S = \{S_1, S_2\}$, with equality if and only if $D_\#$ is disconnected and S is consistent with a block partition in $D_\#$:*

$$\Sigma_S \leq \sum_{i \in S_1} \sum_{j \in S_2} \kappa_{ij}. \quad (4)$$

Proof: Consider any treatment partition S . All entries in $N_\#K_\#^{-1}N'_\#$ are non-negative, so from (3), $1'_h \Psi_{12} 1_{v-h} \leq 1'_h N_1 K^{-1} N'_2 1_{v-h}$. Noting $1'_h N_1 K^{-1} N'_2 1_{v-h} = \sum_{i \in S_1} \sum_{j \in S_2} \kappa_{ij}$, this gives $\Sigma_S \leq \sum_{i \in S_1} \sum_{j \in S_2} \kappa_{ij}$.

Design $D_\#$ is disconnected if and only if consistent treatment and block partitions S and B exist so that, in $D_\#$, all replicates of treatments in S_i are contained in blocks in B_i , for $i = 1, 2$. This is equivalent to $N_\#K_\#^{-1}N'_\#$ being block diagonal with $N_{\#1}K_\#^{-1}N'_{\#2} = 0_{h,v-h}$, where $0_{h,v-h}$ denotes the $h \times (v-h)$ matrix with all entries zero. From (3):

$$\begin{aligned} N_{\#1}K_\#^{-1}N'_{\#2} = 0_{h,v-h} &\iff 1'_h \Psi_{12} 1_{v-h} = 1'_h N_1 K^{-1} N'_2 1_{v-h} \\ &\iff \Sigma_S = \sum_{i \in S_1} \sum_{j \in S_2} \kappa_{ij}. \end{aligned}$$

3.2 The E-value and connectivity

The information matrix for estimation of treatment contrasts is:

$$C = R - NK^{-1}N', \quad (5)$$

where $R = \text{diag}(r_1, \dots, r_v)$ is a diagonal matrix of the treatment replications. It is well known that the symmetric matrix C is positive semi-definite and that each row has zero sum. For a connected design, D , all treatment contrasts are estimable and C has rank $v - 1$. Denote the non-zero eigenvalues of C by $\mu_1 \leq \mu_2 \leq \dots \leq \mu_{v-1}$. Various functions of the μ_i are used as measures of design quality: details are given in Shah and Sinha (1989). In particular, μ_1 is the E-value of D . A design is E-optimal, within a class of designs, if no member of the class has larger μ_1 . Within a design class, an E-optimal design is one for which the largest variance of a normalised treatment contrast is minimised. An upper bound for μ_1 is provided by the following result, given by Jacroux (1991) and generalised by Morgan (2007).

Theorem 3 *Partition C as*

$$C = \begin{bmatrix} C_{11} & C_{12} \\ C'_{12} & C_{22} \end{bmatrix},$$

where C_{11} is $g \times g$ and $1 \leq g \leq v - 1$. Then,

$$\mu_1 \leq \frac{v1'_g C_{11} 1_g}{g(v - g)}. \quad (6)$$

Theorem 3 relates to the treatment partition with sets $\{1, \dots, g\}$ and $\{g + 1, \dots, v\}$. By rearranging the rows of N and R , Theorem 3 can be applied to any treatment partition.

Theorem 4 *Let D be a binary connected design and let $D_\#$ be an eventual design. If $h_0 \leq v/2$ and*

$$\Sigma_S < \frac{\mu_1 h_0 (v - h_0)}{v}, \quad (7)$$

for every S with $h_0 \leq h \leq v/2$ then, either $D_\#$ is connected or any treatment partition which is consistent with a block partition has $h < h_0$.

Proof: Let S be any treatment partition with $h \geq h_0$ and arrange the rows of N as $[N'_1 N'_2]'$ and order the diagonal entries of R to be consistent with this partition. The information

matrix calculated from (5), with the rearranged N and R , is consistent with S . Use of Theorem 3 gives an upper bound for μ_1 that is specific to S :

$$\mu_1 \leq \frac{v 1'_h C_{11} 1_h}{h(v-h)}. \quad (8)$$

From the property that each row sum of C is zero:

$$C_{11} 1_h = -C_{12} 1_{v-h} = N_1 K^{-1} N'_2 1_{v-h},$$

giving

$$1'_h C_{11} 1_h = 1'_h N_1 K^{-1} N'_2 1_{v-h} = \sum_{i \in S_1} \sum_{j \in S_2} \kappa_{ij}$$

and an alternative representation of (8):

$$\frac{\mu_1 h(v-h)}{v} \leq \sum_{i \in S_1} \sum_{j \in S_2} \kappa_{ij}. \quad (9)$$

Let S be a treatment partition with $h_0 \leq h \leq v/2$. If $\Sigma_S < \mu_1 h_0(v-h_0)/v$ for every such S then $\Sigma_S < \mu_1 h(v-h)/v$ and, from (9), $\Sigma_S < \sum_{i \in S_1} \sum_{j \in S_2} \kappa_{ij}$, for every such partition. Then, by Theorem 2, no S with $h \geq h_0$ is consistent with a block partition in $D_\#$. This establishes that, either $D_\#$ is connected or any treatment partition which is consistent with a block partition in $D_\#$ has $h < h_0$.

Corollary 1 *Consider a design for which any disconnected $D_\#$ has $h \geq h_0$ for any treatment partition, S , consistent with a block partition. Then,*

$$\Sigma_S \geq \frac{\mu_1 h_0(v-h_0)}{v}. \quad (10)$$

Theorem 4 and Corollary 1 will be used in Sections 4 and 5 to provide an upper bound for Σ_S that is independent of h and is valid for any treatment partition that could be consistent with a block partition in a disconnected $D_\#$.

4 Results for proper designs

In this section, the planned design, D , has k common for all blocks. Restriction of the observation loss to a maximum of $r_{[v]} - 1$ blocks enables the focus to be limited to a subset of treatment partitions when investigating design robustness. This is achieved via the following result:

Theorem 5 *Let D be a binary design with blocks of size k and let $D_{\#}$ be an eventual design. If $D_{\#}$ is disconnected then any S which is consistent with a block partition of $D_{\#}$ has $h \geq k$.*

Proof: From the restriction on the extent of the observation loss, $D_{\#}$ contains at least one replicate of every treatment in D in a block of size k . Assume $D_{\#}$ is disconnected. Then there is a treatment partition, $S = \{S_1, S_2\}$, with S_1 containing $h \leq v/2$ treatments, such that each block in $D_{\#}$ contains treatments from only one of S_1, S_2 . Consider any treatment in S_1 . This treatment occurs in a block in $D_{\#}$ with $k - 1$ other treatments. Thus, S_1 contains at least k treatments, i.e. $h \geq k$.

From Theorem 5 it is evident that for any treatment partition in D or $D_{\#}$ which is consistent with a block partition, sets S_1 and S_2 each have cardinalities at least k . Thus such a partition can only exist if $v \geq 2k$, giving:

Corollary 2 *Let D be a binary design with all blocks of size k and with $v < 2k$. Then D is connected and $t_* = n_* = r_{[v]}$.*

Since any design with $v < 2k$ has known optimal robustness properties from Corollary 2, investigation on robustness conditions for proper designs can be confined to designs with $v \geq 2k$. From Theorem 5, any disconnected $D_{\#}$ arising from a design with $v \geq 2k$ has $k \leq h \leq v/2$ for any treatment partition that is consistent with a block partition.

Upper bounds for Σ_S are now established for specific configurations of observation loss. Patterns of observation loss considered comprise the loss of p individual observations and the loss of p entire blocks, where $p \leq r_{[v]} - 1$.

Lemma 1 *Let D be a connected binary design on v treatments in b blocks of size k and let S be a treatment partition. Then:*

- (i) *every member of $D(1, p)$ has $\Sigma_S \leq p(k - 1)/k$;*
- (ii) *every member of $D(2, p)$ has $\Sigma_S \leq p(2k^2 - 1 + (-1)^k)/8k$.*

Proof: (i) The maximum contribution to Σ_S from the loss of a single observation arises if the observation comprises the only treatment from S_1 or S_2 in a block. Such an observation

contributes $1/k$ to each of the weighted concurrences with the other $k - 1$ treatments in the block and thus contributes $1/k$ to each of $k - 1$ elements of $N_1 K^{-1} N_2'$ and Ψ_{12} , giving a total contribution of $(k - 1)/k$ to Σ_S . The maximum possible value of Σ_S from the loss of p observations occurs if all p observations are from different blocks and if each is the only treatment from S_1 or S_2 in the block. The result follows.

(ii) For even k , the maximum contribution to Σ_S from a single block occurs for a block in which exactly half the treatments are from each of S_1 and S_2 . In this case, the contribution to Σ_S is $(k/2)^2 \times 1/k = k/4$. For odd k , the maximum occurs for a block with $(k + 1)/2$ treatments from S_1 or S_2 and $(k - 1)/2$ treatments from the other set. In this case, the contribution is $(k + 1)(k - 1)/4 \times 1/k$. Cases with k even and k odd are both covered by $(2k^2 - 1 + (-1)^k)/8k$. The result follows for p blocks.

The general results from §3 together with the results now obtained for proper designs can be combined to obtain lower bounds for n_* and t_* :

Theorem 6 *Let D be a connected binary block design with blocks of size k , on $v \geq 2k$ treatments. Lower bounds for the breakdown and block breakdown numbers are given by $\mathcal{N}$ and $\mathcal{T}$:*

$$(i) \quad n_* \geq \mathcal{N} = \min \left\{ \left\lceil \frac{\mu_1 k^2 (v - k)}{v(k - 1)} \right\rceil, r_{[v]} \right\}; \quad (11)$$

$$(ii) \quad t_* \geq \mathcal{T} = \min \left\{ \left\lceil \frac{8\mu_1 k^2 (v - k)}{v(2k^2 - 1 + (-1)^k)} \right\rceil, r_{[v]} \right\}, \quad (12)$$

where $[x]$ denotes the ceiling of x .

Proof: (i) From Theorem 1, $n_* \leq r_{[v]}$. Assume at least one member of $D(1, p)$ is disconnected, $D_\#$ say, for some $p < r_{[v]}$. By Theorem 2, a treatment partition S exists which is consistent with a block partition in $D_\#$ with $\Sigma_S = \sum_{i \in S_1} \sum_{j \in S_2} \kappa_{ij}$. Further, from Theorem 5, any such partition has $k \leq h \leq v/2$ and by Corollary 1,

$$\Sigma_S \geq \frac{\mu_1 k (v - k)}{v}. \quad (13)$$

From Lemma 1,

$$\Sigma_S \leq \frac{p(k-1)}{k}. \quad (14)$$

Combining (13) and (14),

$$\frac{\mu_1 k(v-k)}{v} \leq \Sigma_S \leq \frac{p(k-1)}{k},$$

giving:

$$p \geq \left\lceil \frac{\mu_1 k^2(v-k)}{v(k-1)} \right\rceil.$$

The inequality for n_* follows. Result (ii) is obtained in a similar way.

Theorem 6 gives lower bounds for the breakdown numbers of a given design. The innate simplicity of the bounds, being functions of only v, k, μ_1 and $r_{[v]}$, mean that they provide a useful tool at the planning stage. The bound for t_* improves on that of Lal *et al.*, who establish that $t_* \geq \lceil \mu_1 \rceil$. Furthermore, the bounds of Theorem 6 can be rearranged to give lower bounds for μ_1 that correspond to designs with guaranteed robustness in the event of chosen levels and configurations of observation loss.

Corollary 3 *Let D be a connected binary block design on v treatments with blocks of size k and let $p \in \{1, \dots, r_{[v]} - 1\}$.*

(i) *D is robust to the loss of any p observations, if*

$$\mu_1 > \frac{vp(k-1)}{k^2(v-k)}. \quad (15)$$

(ii) *D is robust to observation loss from any p blocks, including loss of p entire blocks, if*

$$\mu_1 > \frac{vp(2k^2 - 1 + (-1)^k)}{8k^2(v-k)}. \quad (16)$$

Corollary 3 gives conditions on μ_1 for the breakdown numbers, n_* and t_* , to be at least $p+1$. With knowledge of likely patterns and extent of observation loss, a practitioner can restrict design search to designs with sufficiently large μ_1 to ensure that a selected design has desirable robustness properties.

Example 1 Revisited For $D1$, the value of μ_1 is 1.7456. Using Theorem 6:

$$\left\lceil \frac{\mu_1 k^2(v-k)}{v(k-1)} \right\rceil = \lceil 5.5859 \rceil = 6 \quad \text{and} \quad \left\lceil \frac{8\mu_1 k^2(v-k)}{v(2k^2 - 1 + (-1)^k)} \right\rceil = \lceil 4.1894 \rceil = 5,$$

giving, $n_* \geq \min\{6, 6\}$ and $t_* \geq \min\{5, 6\}$. Since n_* and t_* are bounded above by $r_{[v]} = 6$, $D1$ has $n_* = 6$ and the lower bound for t_* is 5, consistent with earlier conclusion. From Corollary 3 (i), for any design, D , in ten treatments, with blocks of size four and minimum treatment replication six, a value of at least 1.5625 for μ_1 ensures that D is robust against the loss of any five observations. Part (ii) gives a lower bound for μ_1 of 2.0833 to ensure that D is robust against loss from up to five blocks, including the loss of any five entire blocks.

Example 3 Balanced incomplete block designs The robustness properties, established by Ghosh (1982), for the class of balanced incomplete block designs have an alternative derivation via Corollary 2 and Theorem 6. Let D be a balanced incomplete block design on v treatments in blocks of size k with common treatment replication r . If $v < 2k$ then $n_* = t_* = r$ by Corollary 2. Consider the situation with $v \geq 2k$. All non-zero eigenvalues of C have the value $v(k-1)/(k(v-1))$. Using Theorem 6(ii):

$$\begin{aligned} \frac{8\mu_1 k^2(v-k)}{v(2k^2-1+(-1)^k)} &= \frac{8vrk^2(k-1)(v-k)}{kv(v-1)(2k^2-1+(-1)^k)} \\ &\geq \frac{4r(k-1)(v-k)}{k(v-1)} \\ &= 4r \left(\frac{k-1}{k} \right) \left(\frac{v-k}{v-1} \right) \\ &> 4r \left(\frac{1}{2} \right) \left(\frac{1}{2} \right) = r. \end{aligned}$$

Thus, $t_* = n_* = r$.

A final illustration of the capabilities of the results of §4 is given by investigation of the robustness properties of equireplicate designs constructed from a set of component designs. The designs considered range from ones which are very robust to ones which are very vulnerable to observation loss. Further, the nature of the designs means that it is relatively straightforward to determine the actual values of n_* and t_* , thus providing a useful comparison of bounds $\mathcal{N}$, $\mathcal{T}$ and actual n_* , t_* for designs with a wide range of properties.

Example 4 Component designs $B1, B2, B3, C1, C2$ each comprise four replicates of 14 treatments arranged in 14 blocks of size four. Component design $B1$ consists of a pair of balanced incomplete block designs, obtained by cycling initial blocks within the subsets of

treatments $\{1, \dots, 7\}$ and $\{8, \dots, 14\}$ respectively:

$$B1 = \begin{array}{cccccccccccccccc} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 & 11 & 12 & 13 & 14 \\ 2 & 3 & 4 & 5 & 6 & 7 & 1 & 9 & 10 & 11 & 12 & 13 & 14 & 1 \\ 3 & 4 & 5 & 6 & 7 & 1 & 2 & 10 & 11 & 12 & 13 & 14 & 8 & 9 \\ 6 & 7 & 1 & 2 & 3 & 4 & 5 & 13 & 14 & 8 & 9 & 10 & 11 & 12 \end{array}.$$

Design $B2$ is obtained from $B1$ by interchanging treatments 5 and 12 in blocks 7 and 14, and $B3$ is obtained from $B2$ by interchanging treatments 2 and 9 in blocks 7 and 14. Design $C1$ is a regular group divisible design given as $R112$ by Clatworthy (1973) and constructed cyclically from the initial block containing 1, 2, 5 and 7. Finally, $C2$ is similarly obtained from initial block with treatments 1, 2, 3 and 4, but does not have the properties of balance associated with group divisible designs. Tables 1 and 2 give all designs based on the five component designs for $r = 4, 8, 12$, with the exception of designs constructed entirely from copies of $B1$ which are disconnected. Designs comprising copies of $C1$ and $B1$ are examples of group divisible designs constructed by Lemma 3(c) of Hurd and Sarvate (2007). Within each value of r the designs cover a wide range of E-value. In the cases where the bounds $\mathcal{N}$ and $\mathcal{T}$ are lower than r , comparison of the bounds and actual breakdown numbers demonstrate that the bounds provide a useful measure of the robustness of the corresponding design.

5 Results for designs with blocks of different sizes

The approach to determine lower bounds for n_* and t_* when k_i is not common for all blocks in D parallels that of §4, but is more complicated since the upper bounds on Σ_S from Lemma 1 depend on blocks being of common size and so are not available. The block sizes are ordered and labelled in the same manner as that used for the treatment replications in (1), with:

$$k_{[1]} \geq k_{[2]} \geq \dots \geq k_{[b]} \geq 2. \quad (17)$$

Godolphin and Warren (2011) provide a straightforward algorithm for evaluating a lower bound, y_p , for the number of treatments in S_1 for any treatment partition which could be

Table 1: Bounds for n_* , t_* from Theorem 6 and actual values for Example 4 with $r = 4, 8$.
 $\dagger$ denotes a group divisible design.

$r = 4$					
Design	μ_1	$\mathcal{N}$	$\mathcal{T}$	n_*	t_*
$C1^\dagger$	3.0000	4	4	4	4
$C2$	0.9138	4	3	4	4
$B3$	0.5000	2	2	4	2
$B2$	0.3517	2	2	2	2

$r = 8$					
Design	μ_1	$\mathcal{N}$	$\mathcal{T}$	n_*	t_*
$B1 C1^\dagger$	3.0000	8	8	8	8
$C2 C2$	1.8276	7	6	8	8
$B3 C2$	1.5863	7	5	8	8
$B2 C2$	1.5662	6	5	8	8
$B1 C2$	1.2073	5	4	8	6
$B3 B3$	1.0000	4	3	8	4
$B2 B3$	0.8739	4	3	6	4
$B2 B2$	0.7035	3	3	4	4
$B1 B3$	0.5359	3	2	4	2
$B1 B2$	0.3920	2	2	2	2

Table 2: Bounds for n_* , t_* from Theorem 6 and actual values for Example 4 with $r = 12$.

† denotes a group divisible design.

$r = 12$						$r = 12$					
Design	μ_1	$\mathcal{N}$	$\mathcal{T}$	n_*	t_*	Design	μ_1	$\mathcal{N}$	$\mathcal{T}$	n_*	t_*
$C1C1C1^\dagger$	9.0000	12	12	12	12	$B3C2C2$	2.5497	10	8	12	12
$C1C1C2$	6.9138	12	12	12	12	$B1C2C2$	2.2298	9	7	12	12
$B3C1C1$	6.5000	12	12	12	12	$B3B3C2$	2.1510	9	7	12	10
$B2C1C1$	6.3517	12	12	12	12	$B2B3C2$	2.1018	9	7	12	10
$B1C1C1^\dagger$	6.0000	12	12	12	12	$B2B2C2$	2.0246	8	6	12	10
$C1C2C2$	4.8276	12	12	12	12	$B1B3C2$	1.7497	7	5	12	8
$B3C1C2$	4.5863	12	12	12	12	$B1B2C2$	1.6799	7	5	10	8
$B2C1C2$	4.5662	12	12	12	12	$B3B3B3$	1.5000	6	5	12	6
$B1C1C2$	4.2073	12	12	12	12	$B2B3B3$	1.3805	6	4	10	6
$B3B3C1$	4.0000	12	12	12	12	$B1B1C2$	1.2909	5	4	8	6
$B2B3C1$	3.8739	12	12	12	12	$B2B2B3$	1.2338	5	4	8	6
$B2B2C1$	3.7035	12	11	12	12	$B2B2B2$	1.0552	5	4	6	6
$B1B3C1$	3.5359	12	11	12	12	$B1B3B3$	1.0478	4	3	8	4
$B1B2C1$	3.3920	12	10	12	12	$B1B2B3$	0.9165	4	3	6	4
$B1B1C1^\dagger$	3.0000	12	9	12	12	$B1B2B2$	0.7581	3	3	4	4
$C2C2C2$	2.7414	11	8	12	12	$B1B1B3$	0.5478	3	2	4	2
$B2C2C2$	2.5691	10	8	12	12	$B1B1B2$	0.4046	2	2	2	2

associated with a disconnected eventual design arising from observation loss from up to p blocks. The algorithm and a property of the sequence $y_1, \dots, y_{r_{[v]}-1}$ follow.

Lemma 2 *For $p \in \{1, \dots, r_{[v]} - 1\}$, let $x_{0,p} = r_{[v]} - p$ and for $m = 1, 2, \dots$ define*

$$x_{m,p} = r_{[v-y_{m,p}+1]} - p, \quad \text{and} \quad y_{m,p} = k_{[b-x_{m-1,p}+1]}. \quad (18)$$

(i) *Then $x_{0,p}, x_{1,p}, x_{2,p}, \dots$ and $y_{1,p}, y_{2,p}, \dots$ are monotonically nondecreasing sequences of integers with stop values x_p and y_p , where $1 \leq x_p \leq r_{[1]} - p$ and $k_{[b]} \leq y_p \leq k_{[1]}$.*

(ii) *The sequence $y_1, y_2, \dots, y_{r_{[v]}-1}$ is monotonically decreasing.*

Proof: (i) This is proved in Lemma 1 of Godolphin and Warren (2011) for the case $p = r_{[v]} - 1$ and the general result is obtained in the same way.

(ii) Let p and q be integers such that $1 \leq p < q \leq r_{[v]} - 1$ and assume that $x_{m,p} > x_{m,q}$ for some integer m . Then, from the definitions of $y_{m,p}$ and $x_{m,p}$ in (18), $y_{m,p} \geq y_{m,q}$ and $x_{m+1,p} > x_{m+1,q}$. Continuing in this way gives $y_p \geq y_q$. Now consider $m = 1$. Using (18) again, $y_{1,p} = y_{1,q} = k_{[b]}$ and $x_{1,p} > x_{1,q}$. Thus, (ii) follows by an inductive argument.

Lemma 2 is now used to obtain information on treatment partitions that could correspond to a disconnected eventual design $D_{\#}$. The statement and approach of Theorem 7 parallel those of Theorem 5.

Theorem 7 *Let D be a connected binary design with blocks of size $k_{[1]} \geq k_{[2]} \geq \dots \geq k_{[b]}$ and let $D_{\#}$ be an eventual design resulting from the loss of observations from up to $p \leq r_{[v]} - 1$ blocks of D . If $D_{\#}$ is disconnected then any treatment partition, S , which is consistent with a block partition of $D_{\#}$ has $h \geq y_p$.*

Proof: Assume $D_{\#}$ is disconnected. A treatment partition $S = \{S_1, S_2\}$ exists which is consistent with a block partition in $D_{\#}$. Consider any treatment in S_1 . This treatment appears in at least $r_{[v]} - p = x_{0,p}$ blocks of $D_{\#}$ from which no observations are missing. One or more of these blocks has size at least $k_{[b-x_{0,p}+1]} = y_{1,p}$. Thus, S_1 contains at least $y_{1,p}$ treatments. Consider any set of $y_{1,p}$ treatments in S_1 . One or more of these treatments occurs in at least $r_{[v-y_{1,p}+1]} - p = x_{1,p}$ blocks of $D_{\#}$ from which no observations are missing and one or more of these blocks has size at least $k_{[b-x_{1,p}+1]} = y_{2,p}$. Thus, S_1 contains at

least $y_{2,p}$ treatments. Argument continues in this manner to conclude that S_1 contains at least y_p treatments.

Corollary 4 mirrors Corollary 2 and is given without proof.

Corollary 4 *Let D be a binary design with $v < 2y_p$ for some value of p with $1 \leq p \leq r_{[v]} - 1$. Then D is connected and the breakdown numbers satisfy $p < t_* \leq n_* \leq r_{[v]}$.*

Following the approach of §4, but necessarily working with the weighted concurrences rather than the ordinary concurrences, upper bounds are obtained for Σ_S which correspond to the loss of p individual observations and to any loss from p blocks, including the loss of p entire blocks, where $p \leq r_{[v]} - 1$. A preliminary result is required.

Lemma 3 *Let $z_{0,l} = 0$ for $l = 1, \dots, b$. To update from $z_{i,l}$ to $z_{i+1,l}$, let q be the smallest integer which satisfies*

$$\frac{2z_{i,q} + 1}{k_{[q]}} \leq \frac{2z_{i,l} + 1}{k_{[l]}}, \quad \text{for } l = q + 1, \dots, b. \quad (19)$$

Define $z_{i+1,l} = z_{i,l}$ for $l \neq q$, and $z_{i+1,q} = z_{i,q} + 1$. The sequence $z_{p,1}, z_{p,2}, \dots, z_{p,b}$ is monotonically decreasing.

Proof: Suppose the sequence is not monotonically decreasing. Then, for some l , $z_{p,l} < z_{p,l+1}$. Since $z_{0,l} = z_{0,l+1}$, there is a step at which, $z_{i,l} = z_{i,l+1}$ and $z_{i+1,l} < z_{i+1,l+1}$. This implies that

$$\frac{2z_{i,l} + 1}{k_{[l]}} > \frac{2z_{i,l+1} + 1}{k_{[l+1]}}. \quad (20)$$

But this provides a contradiction since $z_{i,l} = z_{i,l+1}$ and $k_{[l]} \geq k_{[l+1]}$.

Lemma 4 *Let D be a connected binary design on v treatments arranged in b blocks of sizes $k_{[1]} \geq k_{[2]} \geq \dots \geq k_{[b]}$ and let $S = \{S_1, S_2\}$ be a treatment partition. Then, for $p \leq r_{[v]} - 1$:*

(i) *every member of $D(1, p)$ has*

$$\Sigma_S \leq \sum_{l=1}^b \frac{z_{p,l}(k_{[l]} - z_{p,l})}{k_{[l]}} = Z_p; \quad (21)$$

(ii) *every member of $D(2, p)$ has*

$$\Sigma_S \leq \sum_{i=1}^p \frac{2k_{[i]}^2 - 1 + (-1)^{k_{[i]}}}{8k_{[i]}} = K_p. \quad (22)$$

Proof: (i) Assume the maximum possible value of Σ_S for the loss of i observations from a design with blocks of sizes $k_{[1]} \geq k_{[2]} \geq \dots \geq k_{[b]}$ corresponds to the loss of $z_{i,l}$ observations from the l th ordered block, where $\sum_{l=1}^b z_{i,l} = i$. This maximum is achieved if the $z_{i,l}$ observations comprise all the treatments from either S_1 or S_2 in the block, for $l = 1, \dots, b$. In this situation, the contribution to Σ_S from the l th ordered block is $z_{i,l}(k_{[l]} - z_{i,l})/k_{[l]}$ and the upper bound for Σ_S is achieved by summing over blocks. Now consider the loss of one additional observation. The change in maximum contribution to Σ_S between a situation in which $z_{i,l}$ observations are lost from a block of size $k_{[l]}$ and these comprise the only treatments from S_1 or S_2 in the block and the loss of $z_{i,l} + 1$ observations which comprise the only treatments from S_1 or S_2 is $(k_{[l]} - 2z_{i,l} - 1)/k_{[l]} = 1 - (2z_{i,l} + 1)/k_{[l]}$. Thus, an upper bound for the contribution to Σ_S for the loss of $i + 1$ observations corresponds to the loss of $z_{i+1,l}$ observations comprising the only treatments in S_1 or S_2 from the l th ordered block. When $i = 0$, $z_{0,l}$ observations are lost from each block and the condition follows by inductive reasoning.

(ii) Using a similar argument to that of Lemma 1, the maximum contribution to Σ_S from a block of size k is $(2k^2 - 1 + (-1)^k)/8k$. For $i > j$:

$$\frac{2k_{[i]}^2 - 1 + (-1)^{k_{[i]}}}{8k_{[i]}} \leq \frac{2k_{[j]}^2 - 1 + (-1)^{k_{[j]}}}{8k_{[j]}}.$$

Thus, for the loss of p blocks, Σ_S has the upper bound specified.

Information on the upper bound of Σ_S , from Lemma 4, for patterns of observation loss considered, can now be combined with information on the lower bound of h for treatment partitions which could potentially correspond to block partitions in eventual designs, from Theorem 7, and the associated lower bound of Σ_S from Corollary 1, to establish lower bounds for the breakdown numbers.

Theorem 8 *Let D be a connected binary block design on v treatments in b blocks of sizes $k_{[1]} \geq k_{[2]} \geq \dots \geq k_{[b]} \geq 2$, with minimum treatment replication $r_{[v]}$ and E -value μ_1 . For $p \in \{1, \dots, r_{[v]} - 1\}$, if any of conditions (i) to (iii) are satisfied then $n_* > p$, and if condition (i) or condition (iii) is satisfied then $t_* > p$.*

(i)

$$y_p > v/2; \tag{23}$$

(ii)

$$\mu_1 > \frac{Z_p v}{y_p(v - y_p)}; \quad (24)$$

(iii)

$$\mu_1 > \frac{K_p v}{y_p(v - y_p)}. \quad (25)$$

Proof: Let p be a member of $\{1, \dots, r_{[v]} - 1\}$. From Corollary 4, if $y_p > v/2$ then $n_* \geq t_* > p$. Assume $y_p \leq v/2$. Consider the loss of p individual observations from D . Assume and that $n_* \leq p$ so that at least one member of $D(1, p)$ is disconnected and let $D_\#$ denote such a member. By Theorem 2, a treatment partition $S = \{S_1, S_2\}$ exists, which is consistent with a block partition in $D_\#$, with $\Sigma_S = \sum_{i \in S_1} \sum_{j \in S_2} \kappa_{ij}$. From Theorem 7 any such partition has $y_p \leq h \leq v/2$. Thus by Corollary 1,

$$\Sigma_S \geq \frac{\mu_1 y_p (v - y_p)}{v}. \quad (26)$$

An upper bound for Σ_S is Z_p from condition (i) of Lemma 4. Combining this bound and the lower bound of (26) gives:

$$\begin{aligned} \frac{\mu_1 y_p (v - y_p)}{v} &\leq \Sigma_S \leq Z_p \\ \mu_1 &\leq \frac{Z_p v}{y_p (v - y_p)}. \end{aligned}$$

Thus, if $y_p \leq v/2$, condition (ii) is sufficient to confirm that $n_* > p$. That condition (iii) is sufficient to confirm $t_* \geq n_* > p$ is established in a similar way, by considering any extent of observation loss from p blocks.

Since y_p decreases monotonically as p increases, and Z_p and K_p are strictly increasing, the lower bounds given by conditions (ii) and (iii) of Theorem 8 increase monotonically with p . To identify the best lowest bounds, the suggested approach is to start with $p = r_{[v]} - 1$ and then, if necessary, to decrease p until bounds for n_* and t_* are identified.

Example 4. Design $D3$ is pairwise balanced and variance balanced and is given by Hedayat and Stufken (1989). The design involves seven treatments arranged in six blocks of size six, ten blocks of size three and seven blocks of size two. Treatments 1 to 6 are replicated

eleven times and treatment 7 has replication twelve.

$$\begin{array}{cccccccccccccccccccc}
1 & 1 & 1 & 1 & 1 & 2 & 1 & 1 & 1 & 1 & 1 & 2 & 2 & 2 & 3 & 3 & 1 & 2 & 3 & 4 & 5 & 6 \\
2 & 2 & 2 & 2 & 3 & 3 & 2 & 2 & 3 & 4 & 5 & 3 & 4 & 5 & 4 & 4 & 7 & 7 & 7 & 7 & 7 & 7 \\
3 & 3 & 3 & 4 & 4 & 4 & 3 & 4 & 5 & 6 & 6 & 6 & 5 & 6 & 5 & 6 & & & & & & \\
D3 = & 4 & 4 & 5 & 5 & 5 & 5 & & & & & & & & & & & & & & & \\
& 5 & 6 & 6 & 6 & 6 & 6 & & & & & & & & & & & & & & & \\
& 7 & 7 & 7 & 7 & 7 & 7 & & & & & & & & & & & & & & &
\end{array}$$

From Theorem 1, $t_* \leq n_* \leq 11$. Consider observation loss restricted to a maximum of 10 blocks, i.e. $p = 10$. Use of Lemma 2 gives $y_{10} = 2$ and Lemma 4 gives $K_{10} = 6 \times 3/2 + 4 \times 2/3 = 11.6667$. Thus the right hand side of (25) is $K_{10}v/(y_{10}(v - y_{10})) = 8.1667$. The value of μ_1 for $D3$ of 9.3333 means that condition (iii) of Theorem 8 is satisfied. It is concluded that $D3$ has breakdown and block breakdown numbers both equal to 11.

The right hand side terms of the conditions of Theorem 8 depend only on parameters of the experiment, i.e. $v, r_{[i]}, k_{[j]}$, and do not depend on the allocation of treatments to blocks that identifies a specific design. When selecting from a class of designs with common parameters, bounds for μ_1 can be calculated using Theorem 8, according to the level and pattern of observation loss considered plausible, and these bounds can be used as one of the factors to inform design choice.

Example 5. The class of group divisible designs with unequal block sizes (GDUBs) is introduced by Lee and Jacroux (1987) to provide a source of designs with good optimality properties. An example of a GDUB design, $D4$ say, considered by Lee and Jacroux has $v = 10, b = 65$ and is equireplicate with $r = 18$. The design is constructed from two group divisible sub-designs $R36$ and $R108$ of Clatworthy (1973). Both $R36$ and $R108$ have $v = 10$. Sub-design $R36$ has 40 blocks of size 2 and $R108$ has 25 blocks of size 4. Design $D4$ has $\mu_1 = 12.5$. From Theorem 1, $t_* \leq n_* \leq 18$. For $i = 1, \dots, 17$, Lemma 2 gives $y_i = 2$ and Lemma 4 gives $Z_i = 3i/4, K_i = i$. For $i = 17$, the right hand side of (25) is $K_{17}v/(y_{17}(v - y_{17})) = 10.625$. Thus by condition (iii) of Theorem 8, $D4$ has $n_* = t_* = 18$. It is interesting to consider the extent of adjustment that can be made to $D4$, to accommodate any practical constraints, whilst retaining a planned design with good robustness properties. For example, if knowledge of the experimental situation indicates

that any observation loss is likely to involve individual observations at random then, for $p = 17$, the right hand side of (24) is $Z_{17} v / (y_{17}(v - y_{17})) = 7.9688$. Thus the treatment allocation to the experimental units can be adjusted provided that the new design is binary and has $\mu_1 > 7.9688$ and will still ensure a planned design with $n_* = 18$.

6 Concluding remarks

The bounds on the breakdown and block breakdown numbers established in §4 and §5 can be used to check robustness properties of a design of interest and are intended to provide a tool for practitioners planning an experiment. When the bounds do not confirm that the design is robust against the loss of $r_{[v]} - 1$ observations or blocks, investigation of a large number of examples indicates that the bounds provided tend to be ‘close enough’ to the true values of n_* and t_* to provide valuable information. Future work to further investigate the quality of the bounds, especially for different values of v and k , is planned. Initial investigation suggests that the quality of the bounds in Theorem 6 may be better for larger k .

The results of §4 and §5 indicate that, within a design class, E-optimal designs will be amongst those designs having the highest lower bounds for robustness. Therefore, in general E-optimal designs can be expected to have good robustness properties within a class and are recommended for use in situations where observation loss is considered likely. When the parameters are such that E-optimal designs are not known, the results lead to an alternative strategy: a lower bound for the E-value which corresponds to a desirable level of robustness can be calculated. Restricting candidate designs to those which at least match this E-value ensures a predetermined minimum degree of robustness against observation loss.

The majority of incomplete block designs used in practice are binary, making a robustness study of such designs particularly relevant. The results and approaches suggested have three notable strengths. First, the methods can be applied to any binary incomplete block design and are not restricted to specific design classes or to designs with particular characteristics, such as equal replication or equal block size. Second, being based only on basic design parameters such as the E-value, treatment replications and block sizes, calculations

are straightforward, allowing easy investigation of robustness of alternative designs at the planning stage. Finally, the versatility of the approach in handling configurations of observation loss involving entire blocks and individual observations is an additional strength.

Only binary designs have been considered throughout this work. The lower bounds for the cardinality of S_1 , of k for proper designs and from Theorem 7 for designs with blocks of different sizes, depend on the binary nature of the design. With a non-binary design the cardinalities can be smaller and the results on n_* and t_* from Theorems 6 and 8 do not necessarily hold. Work is in progress to adapt the results to accommodate non-binary designs.

The author is grateful to suggestions from two referees and an associate editor which have improved the clarity and final presentation of the paper.

References

- Bailey, R.A. 2007. Designs for two-colour microarray experiments. *Appl. Statist.* 56, 365-394.
- Bailey, R.A., Schiffl, K. and Hilgers, R-D. 2013. A note on robustness of D-optimal block designs for two-colour microarray experiments. *J. Statist. Plann. Inference* 143, 1195-1202.
- Baksalary, J.K., Tabis, Z. 1987. Conditions for the robustness of block designs against the unavailability of data. *J. Statist. Plann. Inference* 16, 49-54.
- Bhar, L. 2014. Robustness of variance balanced designs. *Sankhya B* 76, 305-316.
- Bhaumik, D.K., Whittinghill, D.C. 1991. Optimality and robustness to the unavailability of blocks in block designs. *J. R. Statist. Soc. B* 53, 399-407.
- Bose, R.C. 1963. Combinatorial properties of partially balanced designs and association schemes. *Sankya Ser. A*, 25, 109-136.
- Chakrabarti, M.C. 1963. On the C -matrix in design of experiments. *J. Indian Statist. Assoc.* 1, 8-23.
- Clatworthy, W.H. 1973. Tables of two-associate-class partially balanced designs. *Nat. Bureau Stand. Appl. Math. Ser.* 63.
- Das, A., Kageyama, S. 1992. Robustness of BIB and extended BIB designs against the nonavailability of any number of observations in a block. *Comput. Statist. Data Anal.* 14, 343-358.

- Dey, A. 1993. Robustness of block designs against missing data. *Statist. Sinica* 3, 219-331.
- Dey, A., Srivastava, R., Parsad, R. 2001. Robustness of block designs for diallel crosses against missing observations. *J. Ind. Soc. Ag. Statist.* 54, 376-384.
- Foody, W., Hedayat, A. 1977. On theory and applications of BIB designs with repeated blocks. *Annals of Statistics* 5, 932-945.
- Ghosh, S. 1979. On robustness of designs against incomplete data. *Sankhya B* 40, 204-208.
- Ghosh, S. 1982. Robustness of BIBD against the unavailability of data. *J. Statist. Plann. Inference* 6, 29-32.
- Ghosh, S., Rao, S.B. and Singhi, N.M. 1983. On a robust property of PBIB designs. *J. Statist. Plann. Inference* 8, 355-364.
- Godolphin, J.D., Warren, H.R. 2011. Improved conditions for the robustness of binary block designs against the loss of entire blocks. *J. Statist. Plann. Inference* 141, 3498-3505.
- Godolphin, J.D., Warren, H.R. 2014. An Efficient Procedure for the Avoidance of Disconnected Incomplete Block Designs. *Comput. Statist. Data Anal.* 71, 1134-1146.
- Hedayat, A., Stufken, J. 1989. A relation between pairwise balanced and variance balanced block designs. *J. American Statistical Association* 84, 753-755.
- Hurd, S.P., Sarvate, D.G. 2008. Group divisible designs with block size four and two groups. *Discrete Mathematics* 308, 2663-2673.
- Jacroux, M. 1991. A note on the E-optimality of block designs. *Metrika* 38, 203-214.
- Lal, K., Gupta, V.K. and Bhar, L. 2001. Robustness of designed experiments against missing data. *J. Appl. Statist.* 28, 63-79.
- Lee, K.Y. and Jacroux, M. 1987. Some sufficient conditions for the E - and MV - Optimality of block designs having blocks of unequal size. *Ann. Inst. Statist. Math.* 39, 385-397.
- Mahbub Latif, A.H.M., Bretz, F. and Brunner, E. 2009. Robustness considerations in selecting efficient two-color microarray designs. *Bioinformatics* 25, 2355-2361.
- Morgan, J.P., Parvu, V. 2008. Most robust BIBDs. *Statist. Sinica* 18, 689-707.
- Morgan, J.P. 2007. Optimal incomplete block designs. *J. American Statistical Association* 102, 655-663.
- Notz, W.I., Whittinghill, D.C. and Zhu, Y. 1994. Robustness to the unavailability of data in block designs. *Metrika* 41, 263-275.

- Prescott, P., Mansson, R. 2001. Robustness of balanced incomplete block designs to randomly missing observations. J. Statist. Plann. Inference 92, 283-296.
- Sathe, Y.S. and Satam, M.R. 1992. Some more robust block designs against the unavailability of data. J. Statist. Plann. Inference 30, 93-98.
- Shah, K.R. and Sinha, B.K. 1989. Theory of Optimal Designs. New York: Springer-Verlag.
- Srivastava, R., Gupta, V.K. and Dey, A. 1991. Robustness of some designs against missing data. J. Appl. Statist. 18, 313-318.
- Tsai, S-F., Liao, C-T. 2013. Minimum breakdown designs in blocks of size two. J. Statist. Plann. Inference 143, 202-208.