Multi-part balanced incomplete-block designs

R. A. Bailey
University of St Andrews

Combinatorics Seminar, Shanghai Jiao Tong University,
4 October 2018
Joint work with Peter Cameron (University of St Andrews)

Abstract: I

In order to keep the protocol for a cancer clinical trial simple for each medical centre involved,
it is proposed to limit each medical centre to
only a few of the cancer types and only a few of the drugs.

Let v_1 be the total number of cancer types,
and v_2 the total number of drugs.

At the workshop on
Design and Analysis of Experiments in Healthcare
at the Isaac Newton Institute, Cambridge, UK in 2015,
Valerii Fedorov listed the following desirable properties.

Abstract: II

(a) All medical centres involve the same number, say k_1, of cancer types, where $k_1 < v_1$.
(b) All medical centres use the same number, say k_2, of drugs, where $k_2 < v_2$.
(c) Each pair of distinct cancer types are involved together at the same non-zero number, say λ_{11}, of medical centres.
(d) Each pair of distinct drugs are used together at the same non-zero number, say λ_{22}, of medical centres.
(e) Each drug is used on each type of cancer at the same number, say λ_{12}, of medical centres.

The first four conditions state that,
considered separately, the designs for cancer types and drugs
are balanced incomplete-block designs
(a.k.a. BIBDs or 2-designs)
with the medical centres as blocks. We propose calling a design that satisfies all five properties a 2-*part BIBD* or 2-*part* 2-*design*.

Abstract: III

The parameters of a 2-part 2-design satsify some equations,
and also an inequality that generalizes both Fisher's inequality
and Bose's inequality.

We give several constructions of 2-part 2-designs,
then generalize them to m-part 2-designs.

An example: $v_1 = 6$, $k_1 = 3$, $v_2 = 5$, $k_2 = 2$, $b = 10$

Combinations: 6 Cancer Types and 5 Drugs*

	Cancer					
Block	C1	C2	C3	C4	C5	C6
1	D1,5	D1,5	D1,5			
2	D1,2				D1,2	D1,2
3	D2,3		D2,3	D2,3		
4	D3,4	D3,4				D3,4
5	D4,5			D4,5	D4,5	
6		D1,3		D1,3	D1,3	
7		D2,4	D2,4		D2,4	
8			D3,5		D3,5	D3,5
9			D1,4	D1,4		D1,4
10		D2,5		D2,5		D2,5

Operational constraints for blocks (sub trials):
- No more than 3 cancer types per block
- Only 2 drugs per block

Properties:
- Every pair of drugs at one trial
- Every pair of cancer types at two trials
- Every drug with every cancer type at two trials

Benchmarking: in reality "practical" designs take into account medical knowledge, disease prevalence, differing enrollment rates per cancer type and competing products

9

*Thanks to Prof. Rosemary Bailey

Thanks to Valerii Fedorov for this image.

Comparison with classical factorial designs

Block 1 of our example is shown as

C1	C2	C3
D1, D5	D1, D5	D1, D5

which means that the medical centre which it represents will accept into the trial only patients with cancer types 1, 2 or 3;
patients of each of these types will be randomized
(in approximately equal numbers) to

- drug 1, drug 5 (original idea)
  (placebo may be one of the listed "drugs")
- drug 1, drug 5, and placebo (modified idea)
- drug 1, drug 5, their combination, and placebo
  (further modification).

Contrast this with a classical factorial design in blocks,
which would never have level C1 of factor C occurring in
several combinations in a block while level C4 does not occur
in that block at all.

The concise representation of the design

Block	Cancer types	Drugs
1	C1, C2, C3	D1, D5
2	C1, C5, C6	D1, D2
3	C1, C3, C4	D2, D3
4	C1, C2, C6	D3, D4
5	C1, C4, C5	D4, D5
6	C2, C4, C5	D1, D3
7	C2, C3, C5	D2, D4
8	C3, C5, C6	D3, D5
9	C3, C4, C6	D1, D4
10	C2, C4, C6	D2, D5

Warning! This does not mean that each block has 5 treatments.

Definition of 2-part 2-design

Definition

A 2-part 2-design for v_1 cancer types and v_2 drugs in b medical centres, with further parameters k_1, k_2, λ_{11}, λ_{22} and λ_{12}, is an allocation of cancer types and drugs to medical centres satisfying:

(a) all medical centres involve k_1 cancer types, where $k_1 < v_1$;
(b) all medical centres use k_2 drugs, where $k_2 < v_2$;
(c) each pair of distinct cancer types occur together at λ_{11} medical centres, where $\lambda_{11} > 0$;
(d) each pair of distinct drugs occur together at λ_{22} medical centres, where $\lambda_{22} > 0$;
(e) each drug occurs with each type of cancer at λ_{12} medical centres.

Conditions on parameters

Theorem

In a 2-part 2-design with parameters v_1, v_2, b, k_1, k_2, λ_{11}, λ_{22} and λ_{12}, the following hold.

1. *Each cancer type occurs in r_1 blocks, where $v_1 r_1 = bk_1$.*
2. *Each drug occurs in r_2 blocks, where $v_2 r_2 = bk_2$.*
3. $\lambda_{11}(v_1 - 1) = r_1(k_1 - 1)$.
4. $\lambda_{22}(v_2 - 1) = r_2(k_2 - 1)$.
5. $bk_1k_2 = v_1v_2\lambda_{12}$.
6. $b \geq v_1 + v_2 - 1$.

Items 1–5 are obtained by counting something in two different ways.

Item 6 is like Fisher's Inequality (in a 2-design, $b \geq v$).

A generalization of resolvability

Definition

A block design is resolvable if the set of blocks can be partitioned into r replicates of b/r blocks each, in such a way that each treatment occurs once in each replicate.

In general, $r_1 \neq r_2$, so we cannot use the usual definition of resolvable design here.

Definition

A 2-part block design is c-partitionable if the set of blocks can be grouped into c classes of b/c blocks each, in such a way that every cancer type occurs the same number of times in each class and every drug occurs the same number of times in each class.

Theorem

If a 2-part 2-design is c-partitionable then $b \geq v_1 + v_2 + c - 2$.

Notes on the theorem

Theorem

If a 2-part 2-design is c-partitionable then $b \geq v_1 + v_2 + c - 2$.

(a) Every 2-part 2-design is 1-partitionable, so it is always true that

$$b \geq v_1 + v_2 - 1.$$

(b) Bose's Inequality states that, for a resolvable 2-design,

$$b \geq v + r - 1.$$

Our new theorem generalizes that.

Easy construction I: Cartesian product

Let Δ_1 be a BIBD for v_1 treatments in b_1 blocks of size k_1, and let Δ_2 be a BIBD for v_2 treatments in b_2 blocks of size k_2.
Form all b_1b_2 combinations of a block of each sort.
For each block combination, form the Cartesian product of their sets of treatments.

The result is a 2-part 2-design, but it has b_1b_2 blocks, which is often too large.

Easy construction II: Swap

Given a 2-part 2-design, create another one,
interchanging the values of k_1 and $v_1 - k_1$,
by replacing the set of cancer types in each block by the complementary set of cancer types.

The result is also a 2-part 2-design so long as $v_1 - k_1 \geq 2$.

Similarly, swap drugs to interchange k_2 and $v_2 - k_2$.

Easy construction III: Interchange

Given a 2-part 2-design, create another one,
interchanging the values of v_1 and v_2,
and the values of k_1 and k_2,
by interchanging the roles of cancer types and drugs.

Serious construction I: Subcartesian product

Let Δ_1 be a BIBD for v_1 treatments in b_1 blocks of size k_1,
and let Δ_2 be a BIBD for v_2 treatments in b_2 blocks of size k_2.
Suppose that Δ_2 is resolvable with replication r,
and that r divides b_1.

Partition the set of blocks of Δ_1 into r sets of b_1/r blocks,
in any way at all.

Match these sets to the r resolution classes of Δ_2,
in any way at all.

For each matched pair, construct the cartesian product design.

The result is a 2-part 2-design,
and it has $b_1 b_2 / r$ blocks.

An example of a subcartesian product: $v_1 = 3$, $v_2 = 4$

Δ_1 $b = 3$
C1, C2
C1, C3
C2, C3

Block	Cancer types	Drugs	Δ_2 resolvable $r = 3$
1	C1, C2	D1, D3	D1, D3
2	C1, C2	D2, D4	D2, D4
3	C1, C3	D2, D3	D2, D3
4	C1, C3	D1, D4	D1, D4
5	C2, C3	D1, D2	D1, D2
6	C2, C3	D3, D4	D3, D4

Serious construction II: Hadamard matrix

If $v_1 = v_2 = 2k_1 = 2k_2 = 2n$, write down a Hadamard matrix of order $4n$ with all entries $+1$ in the first row.

$$\begin{bmatrix}
+1 & +1 & +1 & +1 & +1 & +1 & +1 & +1 & +1 & +1 & +1 & +1 \\
+1 & +1 & +1 & +1 & +1 & +1 & -1 & -1 & -1 & -1 & -1 & -1 \\
+1 & -1 & +1 & -1 & +1 & -1 & +1 & -1 & -1 & +1 & +1 & -1 \\
+1 & -1 & -1 & -1 & +1 & +1 & -1 & -1 & +1 & -1 & +1 & +1 \\
+1 & +1 & +1 & -1 & -1 & -1 & -1 & +1 & +1 & -1 & +1 & -1 \\
+1 & -1 & -1 & +1 & +1 & -1 & +1 & +1 & +1 & -1 & -1 & -1 \\
+1 & -1 & -1 & +1 & -1 & +1 & -1 & +1 & -1 & +1 & +1 & -1 \\
+1 & -1 & +1 & +1 & -1 & -1 & -1 & -1 & +1 & +1 & -1 & +1 \\
+1 & +1 & -1 & -1 & +1 & -1 & -1 & +1 & -1 & +1 & -1 & +1 \\
+1 & +1 & -1 & +1 & -1 & -1 & +1 & -1 & -1 & -1 & +1 & +1 \\
+1 & +1 & -1 & -1 & -1 & +1 & +1 & -1 & +1 & +1 & -1 & -1 \\
+1 & -1 & +1 & -1 & -1 & +1 & +1 & +1 & -1 & -1 & -1 & +1
\end{bmatrix}$$

Serious construction II: Hadamard matrix

If $v_1 = v_2 = 2k_1 = 2k_2 = 2n$, write down a Hadamard matrix of order $4n$ with all entries $+1$ in the first row.
Replace all $\pm$ entries in row 2 with levels of C/D.

$$\begin{bmatrix}
+1 & +1 & +1 & +1 & +1 & +1 & +1 & +1 & +1 & +1 & +1 & +1 \\
\mathrm{C1} & \mathrm{C2} & \mathrm{C3} & \mathrm{C4} & \mathrm{C5} & \mathrm{C6} & \mathrm{D1} & \mathrm{D2} & \mathrm{D3} & \mathrm{D4} & \mathrm{D5} & \mathrm{D6} \\
+1 & -1 & +1 & -1 & +1 & -1 & +1 & -1 & -1 & +1 & +1 & -1 \\
+1 & -1 & -1 & -1 & +1 & +1 & -1 & -1 & +1 & -1 & +1 & +1 \\
+1 & +1 & +1 & -1 & -1 & -1 & -1 & +1 & +1 & -1 & +1 & -1 \\
+1 & -1 & -1 & +1 & +1 & -1 & +1 & +1 & +1 & -1 & -1 & -1 \\
+1 & -1 & -1 & +1 & -1 & +1 & -1 & +1 & -1 & +1 & +1 & -1 \\
+1 & -1 & +1 & +1 & -1 & -1 & -1 & -1 & +1 & +1 & -1 & +1 \\
+1 & +1 & -1 & -1 & +1 & -1 & -1 & +1 & -1 & +1 & -1 & +1 \\
+1 & +1 & -1 & +1 & -1 & -1 & +1 & -1 & -1 & -1 & +1 & +1 \\
+1 & +1 & -1 & -1 & -1 & +1 & +1 & -1 & +1 & +1 & -1 & -1 \\
+1 & -1 & +1 & -1 & -1 & +1 & +1 & +1 & -1 & -1 & -1 & +1
\end{bmatrix}$$

Row 3 $\rightarrow$ {C1,C3,C5||D1,D4,D5} and {C2,C4,C6||D2,D3,D6}.
And so on, so $b = 2(4n - 2) = 8n - 4$.

A good outcome of the Hadamard construction

A Hadamard matrix of order $4n$ leads to a 2-part 2-design with $v_1 = v_2 = 2n$, $k_1 = k_2 = n$ and $b = 8n - 4$.

It is c-partitionable for $c = 4n - 2$.

Often, a subcartesian product can give
a 2-part 2-design with the same parameters,
but this is not usually $(4n - 2)$-partitionable.

Serious construction III: Symmetric BIBD

Start with a BIBD for v treatments in v blocks of size k,
where each pair of blocks have λ treatments in common,
and $\lambda > 1$ and $3 \leq k \leq v - k$.

Choose one block, and identify its treatments with drugs
(so $v_2 = k$).

Identify the other treatments with cancer types (so $v_1 = v - k$).

Each remaining block gives a block of our 2-part 2-design, so

$$\begin{aligned} b &= v - 1 \\ k_2 &= \lambda \\ k_1 &= k - \lambda \\ \lambda_{11} &= \lambda \\ \lambda_{12} &= \lambda \\ \lambda_{22} &= \lambda - 1. \end{aligned}$$

An example from a symmetric BIBD: $v_1 = 6$, $v_2 = 5$

rows are blocks				
1	5	3	4	9
2	6	4	5	10
3	7	5	6	0
4	8	6	7	1
5	9	7	8	2
6	10	8	9	3
7	0	9	10	4
8	1	10	0	5
9	2	0	1	6
10	3	1	2	7
0	4	2	3	8

2-part 2-design				
drugs		cancer types		
D2	D4	C2	C3	C5
D2	D3	C1	C3	C4
D1	D4	C3	C4	C6
D2	D5	C2	C4	C6
D3	D5	C3	C5	C6
D4	D5	C1	C4	C5
D1	D2	C1	C5	C6
D1	D5	C1	C2	C3
D1	D3	C2	C4	C5
D3	D4	C1	C2	C6

1	5	3	4	9
D1	D2	D3	D4	D5

0	2	6	7	8	10
C1	C2	C3	C4	C6	C5

This is exactly the first 2-part 2-design that I showed you.

Serious construction IV: Augmentation

Given a 2-part 2-design with $v_2 = 2k_2 + 1$, add an extra drug,
increasing v_2 to $v_2 + 1$, k_2 to $k_2 + 1$ and b to $2b$.

Replace each previous block by two new blocks,
both with the original subset of cancer types.

One of these has the same drugs as before, plus the new drug.
The other has all the remaining drugs.

Easy construction IV: Group-divisible designs

If $v_1 = v_2$ and $k_1 = k_2$ then the concise form of a 2-part 2-design
is a "semi-regular group-divisible incomplete block-design for
two groups of treatments".

Look these up in Clatworthy's
Tables of Two-Associate Class Partially Balanced Designs.

Serious construction V: Permutation groups

If there is a group G which
acts doubly transitively on the set of cancer types
and also acts doubly transitively on the set of drugs,
then choose an initial block
and then get the remaining blocks by
applying the permutations in G to it.

Interesting examples are too large to fit on a slide!

Extending the problem

On 28 March 2016, Valerii sent me the png file of the first design in this talk. When I thanked him, he emailed back the next day with

Dear Rosemary,
It can be never ending story ....
For instance, can we extend the table below and add another factor: oncogenes (biomarker)? ...

3-part 2-designs

In a 3-part 2-design, we also have a set of v_3 biomarkers, such that

(a) all medical centres involve k_1 cancer types, where $k_1 < v_1$;
(b) all medical centres use k_2 drugs, where $k_2 < v_2$;
(c) each pair of distinct cancer types occur together at λ_{11} medical centres, where $\lambda_{11} > 0$;
(d) each pair of distinct drugs occur together at λ_{22} medical centres, where $\lambda_{12} > 0$;
(e) each drug occurs with each type of cancer at λ_{12} medical centres;
(f) all medical centres use k_3 biomarkers, where $k_3 < v_3$;
(g) each pair of distinct biomarkers occur together at λ_{33} medical centres, where $\lambda_{33} > 0$;
(h) each biomarker occurs with each type of cancer at λ_{13} medical centres;
(i) each biomarker occurs with each drug at λ_{23} medical centres.

Serious new construction: Orthogonal array

Let Δ_1 be a BIBD for v_1 treatments in b_1 blocks of size k_1,
Δ_2 a BIBD for v_2 treatments in b_2 blocks of size k_2,
and Δ_3 a BIBD for v_3 treatments in b_3 blocks of size k_3.

Use an orthogonal array of strength 2, with three columns, where column i has b_i symbols.

For each row of the orthogonal array, construct the cartesian product of the three blocks, one in each of Δ_1, Δ_2 and Δ_3.

An example using an orthogonal array: $v_1 = v_2 = v_3 = 3$

design Δ_1		design Δ_2		design Δ_1	
Block 1	C1, C2	Block 1	D1, D2	Block 1	B1, B2
Block 2	C1, C3	Block 2	D1, D3	Block 2	B1, B3
Block 3	C2, C3	Block 3	D2, D3	Block 3	B2, B3

Orthogonal array			Block	Cancer types	Drugs	Bio-markers
1	1	1	1	C1, C2	D1, D2	B1, B2
2	2	2	2	C1, C3	D1, D3	B1, B3
3	3	3	3	C2, C3	D2, D3	B2, B3
1	3	2	4	C1, C2	D2, D3	B1, B3
2	1	3	5	C1, C3	D1, D2	B2, B3
3	2	1	6	C2, C3	D1, D3	B1, B2
1	2	3	7	C1, C2	D1, D3	B2, B3
2	3	1	8	C1, C3	D2, D3	B1, B2
3	1	2	9	C2, C3	D1, D2	B1, B3

General multi-part BIBDs

The foregoing definition extends to m different types of thing.

Theorem
Let Δ be an m-part 2-design with v_i things of type i, for $i = 1, \ldots, m$. If the parameters are b, v_i, k_i, λ_{ii} and λ_{ij} for $1 \le i < j \le m$, then the following hold.

1. *For $i = 1, \ldots, m$, each thing of type i occurs in r_i blocks, where $v_i r_i = bk_i$.*
2. *For $i = 1, \ldots, m$, $\lambda_{ii}(v_i - 1) = r_i(k_i - 1)$.*
3. *For $1 \le i < j \le m$, $bk_ik_j = v_iv_j\lambda_{ij}$.*
4. *If Δ is c-partitionable then $b \ge v_1 + \cdots + v_m + c - m$.*
5. *In particular, $b \ge v_1 + \cdots + v_m - m + 1$.*

Nothing new under the sun

In May 2018 we learnt that designs like these had already been proposed by Randy Sitter (*Biometrika*, 1993) and Rahul Mukerjee (*Journal of Statistical Planning and Inference*, 1998).

Mukerjee's main construction is the general orthogonal array (possibly trivial, i.e. all possible rows) applied to parts of m component BIBDs which are all c-partitionable (possibly with $c = 1$).

There are three main differences between their approach and ours.

Different applications: I

Sitter is concerned with sampling.
The population consists of m distinct strata of sizes $v_1, \ldots, v_m$.
He wants to draw a random sample of size k_i from stratum i,
for $i = 1, \ldots, m$, measure something on each element sampled,
and hence estimate something about the population.

Different applications: II

Mukerjee's proposed application is to designed experiments,
but it is different from ours.

For him, each of our "blocks" is a single experimental unit,
to which subsets of different types of treatment are applied.

Block	Cancer types	Drugs
1	C1, C2, C3	D1, D5

Experimental unit	Chemical mixture applied to the soil	Variety of wheat
1	C1, C2, C3	D1, D5
one measurement	a mixture of three chemicals	variety obtained by cross-breeding varieties D1 and D5

Different constraints

Sitter's and Mukerjee's applications
do not need the constraint that $\lambda_{ii} > 0$.
So their designs allow $k_i = 1$, which ours do not.

Different terminology

Sitter and Mukerjee called their designs
balanced orthogonal multi-arrays,
following Brickell (*Congressus Numerantium*, 1984).

Brickell does not require the separate designs to be balanced,
but he does require $\lambda_{ij} = 1$ if $i \neq j$.
Sitter and Mukerjee explicitly dropped this last condition.

Orthogonal multi-arrays (in Brickell's original definition)
have been, and are still, used widely in coding theory and in
statistics (where they are also called semi-Latin squares).

So we think that it is better not to call these designs
orthogonal multi-arrays.

A very general construction

m c-partitionable 2-designs
(c may be 1)

orthogonal array with m columns
(this may have all possible rows)

The ingredients can be
c-partitionable
multi-part 2-designs.

It suffices to have
all but one
c-partitionable,
so long as c divides
the number of blocks
in the other one.

A final example

The Hadamard construction gives a design for 6 cancer types
and 6 drugs, with 3 cancer types and 3 drugs in each block.
The design has 20 blocks,
and can be partitioned into 10 classes of 2 blocks,
each of which is a single replicate of cancer types and of drugs.

Suppose that there are 5 biomarkers,
and we want 2 in each block.
There are 10 pairs of biomarkers. Match pairs to classes,
and put those two biomarkers in both blocks in that class.

Suppose that there are 6 biomarkers,
and we want 3 in each block.
There is a BIBD for 6 biomarkers in 10 blocks of size 3.
Match these blocks to the original classes.

So our new designs for $m = 2$ lead to
new designs for larger values of m.