Block designs with very low replication, and other challenges in design of experiments

R. A. Bailey

University of St Andrews

QMUL (emerita)

James Hutton Institute, 30 March 2015

Variety Testing

In breeding trials of new varieties, typically there is very little seed of each of the new varieties.

Traditionally, an experiment has one plot for each new variety and several plots for a well-established “control”: for example, 30 new varieties on one plot each and one control on 8 plots.

Variety Testing

In breeding trials of new varieties, typically there is very little seed of each of the new varieties.

Traditionally, an experiment has one plot for each new variety and several plots for a well-established “control”: for example, 30 new varieties on one plot each and one control on 8 plots.

In the last 10 years, Cullis and colleagues in Australia
(and independently Bueno and Gilmour)
have suggested replacing many occurrences of the the control by double replicates of a small number of new varieties: for example, 24 new varieties with one plot each, 6 new varieties with two plots each, and the control on two further plots.

Variety Testing

In breeding trials of new varieties, typically there is very little seed of each of the new varieties.

Traditionally, an experiment has one plot for each new variety and several plots for a well-established “control”: for example, 30 new varieties on one plot each and one control on 8 plots.

In the last 10 years, Cullis and colleagues in Australia
(and independently Bueno and Gilmour)
have suggested replacing many occurrences of the the control by double replicates of a small number of new varieties: for example, 24 new varieties with one plot each, 6 new varieties with two plots each, and the control on two further plots.

This is an improvement if there are no blocks.

How do we allow for variation between the plots?

How do we allow for variation between the plots?

"…on any given field agricultural operations, at least for centuries, have followed one of two directions, which are usually those of the rows and columns; consequently streaks of fertility, weed infestation, etc., do, in fact, occur predominantly in those two directions."

R. A. Fisher,
letter to H. Jeffreys,
30 May 1938
(selected correspondence edited by J. H. Bennett)

How do we allow for variation between the plots?

"…on any given field agricultural operations, at least for centuries, have followed one of two directions, which are usually those of the rows and columns; consequently streaks of fertility, weed infestation, etc., do, in fact, occur predominantly in those two directions."

R. A. Fisher,
letter to H. Jeffreys,
30 May 1938
(selected correspondence edited by J. H. Bennett)

(This assumption is dubious for field trials in Australia.)

How do we allow for variation between the plots?

"…on any given field agricultural operations, at least for centuries, have followed one of two directions, which are usually those of the rows and columns; consequently streaks of fertility, weed infestation, etc., do, in fact, occur predominantly in those two directions."

R. A. Fisher,
letter to H. Jeffreys,
30 May 1938
(selected correspondence edited by J. H. Bennett)

(This assumption is dubious for field trials in Australia.)

If field operations have been primarily in one direction for a long time, then it is reasonable to divide the fields into blocks whose length runs along that direction.

Blocking in the second phase of a variety trial

The milling phase of a wheat variety trial has 224 varieties to be compared. Only 10 can be milled in any one day. The trial can take place over 28 days, so there are 28 blocks of size 10.

Blocking in the second phase of a variety trial

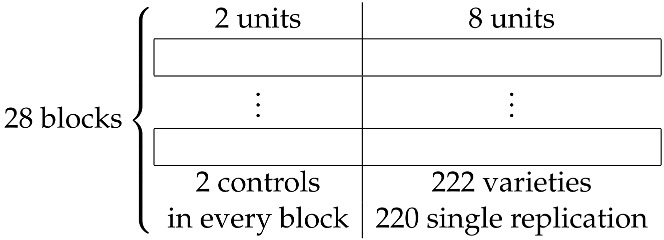
The milling phase of a wheat variety trial has 224 varieties to be compared. Only 10 can be milled in any one day. The trial can take place over 28 days, so there are 28 blocks of size 10.

There are only $280 - 224 = 56$ experimental units "spare" for replication. How should these be allocated?

Blocking in the second phase of a variety trial

The milling phase of a wheat variety trial has 224 varieties to be compared. Only 10 can be milled in any one day. The trial can take place over 28 days, so there are 28 blocks of size 10.

There are only $280 - 224 = 56$ experimental units "spare" for replication. How should these be allocated?

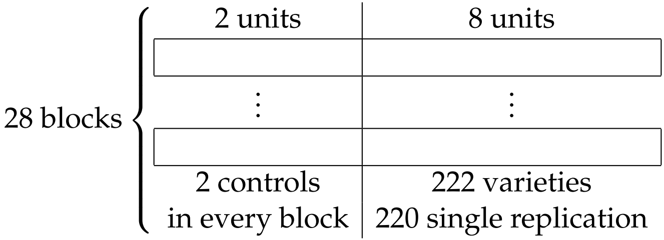


One extreme: 2 "controls" (among the test varieties) in every block.

Blocking in the second phase of a variety trial

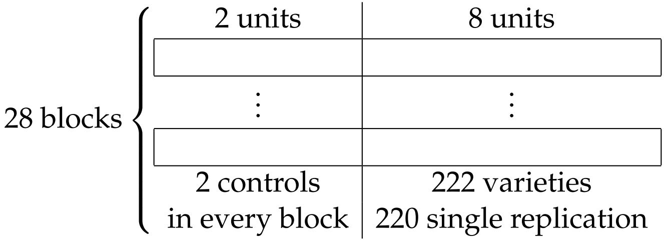
The milling phase of a wheat variety trial has 224 varieties to be compared. Only 10 can be milled in any one day. The trial can take place over 28 days, so there are 28 blocks of size 10.

There are only $280 - 224 = 56$ experimental units "spare" for replication. How should these be allocated?

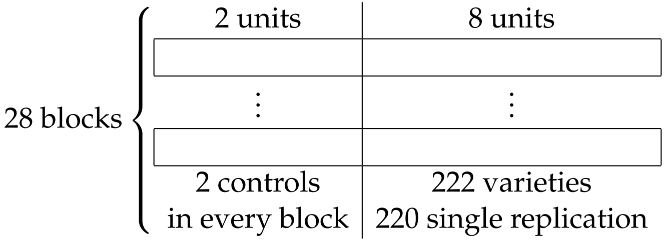


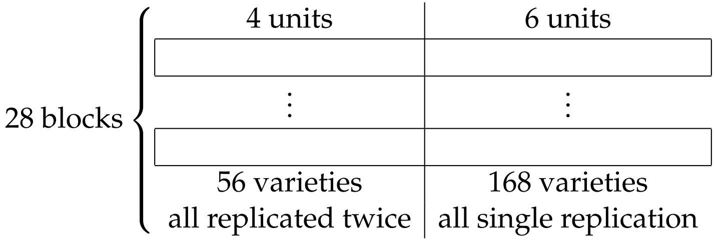
One extreme: 2 "controls" (among the test varieties) in every block.
Even more extreme: 2 uninteresting controls in each block.

Two possible designs for 224 varieties in 28 blocks of 10



Two possible designs for 224 varieties in 28 blocks of 10





The problem

We are given b blocks of size k. We are given v varieties. Assume that

$$\text{average replication} = \bar{r} = \frac{bk}{v} \leq 2.$$

How should we allocate varieties to blocks?

A-optimal designs

We measure the response Y on the plot with variety i in block D, and assume that

$$Y = \tau_i + \beta_D + \text{random noise},$$

where the random noise is $N(0, \sigma^2)$, independently for each plot.

Put

$$V_{ij}\,\sigma^2 \quad = \quad \text{variance of the best linear unbiased estimator for } \tau_i - \tau_j;$$

$$V_T = \sum_{i=1}^{v-1} \sum_{j=i+1}^{v} V_{ij} \quad \propto \quad \text{sum of variances of variety differences.}$$

A block design is A-optimal if it minimizes V_T.

Silly names just for this talk

Definition

Call a variety a

a drone if it has replication 1;

Silly names just for this talk

Definition

Call a variety a

a drone if it has replication 1;

a queen-bee if it occurs in every block;

Silly names just for this talk

Definition

Call a variety a

- a drone if it has replication 1;
- a queen-bee if it occurs in every block;
- a worker otherwise.

Silly names just for this talk

Definition

Call a variety a

- a drone if it has replication 1;
- a queen-bee if it occurs in every block;
- a worker otherwise.

Silly names just for this talk

Definition

Call a variety a

- a drone if it has replication 1;
- a queen-bee if it occurs in every block;
- a worker otherwise.

Is it better to put all the drones into one block (or a few blocks), or are they better distributed equally among all the blocks?

How should we distribute the drones?

Block A	Block B
n drones	m drones

If we move all the drones in block B into block A
then we reduce nm variances from $2 + V_{AB}$ to 2,
where V_{AB} is the variance of the estimator of the difference between the block effects of A and B in the design obtained by ignoring the drones.

How should we distribute the drones?

Block A	Block B
n drones	m drones

If we move all the drones in block B into block A
then we reduce nm variances from $2 + V_{AB}$ to 2,
where V_{AB} is the variance of the estimator of the difference between the block effects of A and B in the design obtained by ignoring the drones.

Then we have to remove m non-drones from block A,
and this increases the variances between these $n + m$ drones and the remaining $v - n - m$ varieties.

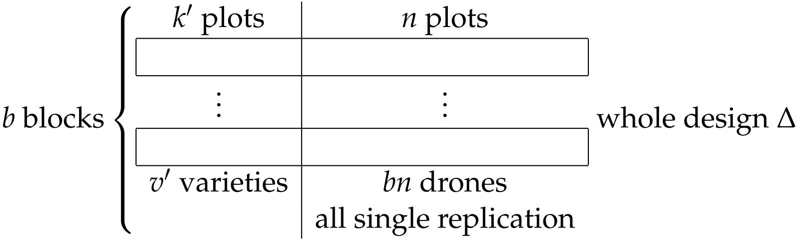
How should we distribute the drones?

Block A	Block B
n drones	m drones

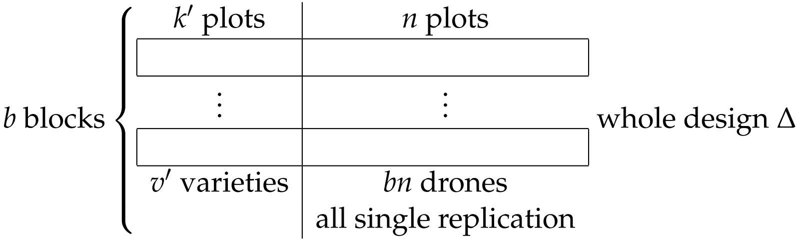
If we move all the drones in block B into block A
then we reduce nm variances from $2 + V_{AB}$ to 2,
where V_{AB} is the variance of the estimator of the difference between the block effects of A and B in the design obtained by ignoring the drones.

Then we have to remove m non-drones from block A,
and this increases the variances between these $n + m$ drones and the remaining $v - n - m$ varieties. This more than compensates for the original reduction in variance.

From now on, distribute drones as equally as possible

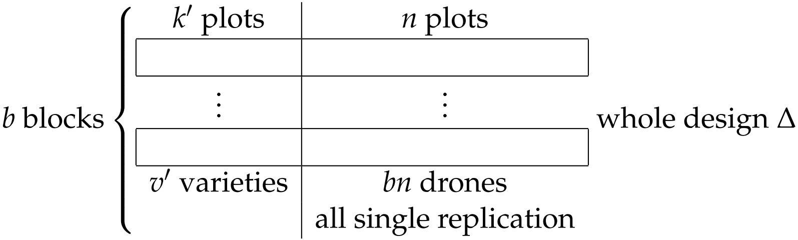


From now on, distribute drones as equally as possible



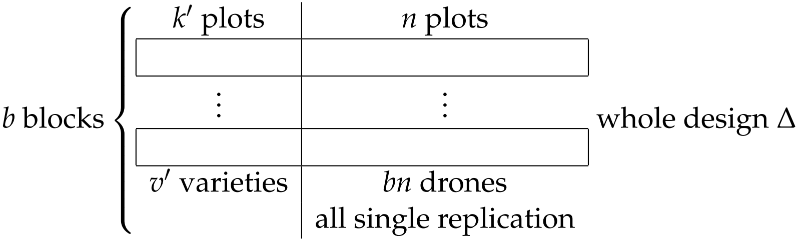
Whole design Δ has v treatments in b blocks of size $k = k' + n$;

From now on, distribute drones as equally as possible



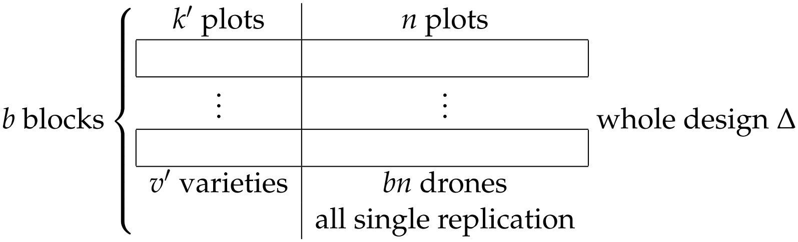
Whole design Δ has v treatments in b blocks of size $k = k' + n$;
the subdesign Γ has v' core varieties in b blocks of size k'.

From now on, distribute drones as equally as possible



Whole design Δ has v treatments in b blocks of size $k = k' + n$;
the subdesign Γ has v' core varieties in b blocks of size k'.
(The core varieties may include extra drones.)

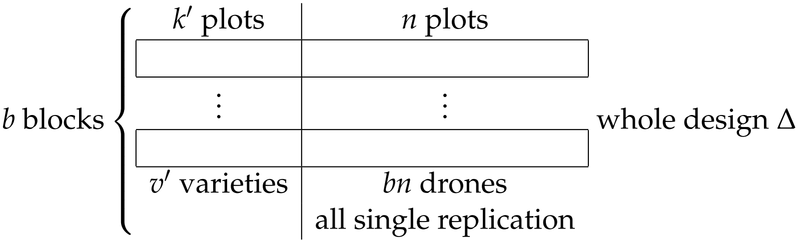
From now on, distribute drones as equally as possible



Whole design Δ has v treatments in b blocks of size $k = k' + n$;
the subdesign Γ has v' core varieties in b blocks of size k'.
(The core varieties may include extra drones.)

$$n \geq n_0 = \left\lfloor \frac{2v - bk}{b} \right\rfloor$$

From now on, distribute drones as equally as possible



Whole design Δ has v treatments in b blocks of size $k = k' + n$;
the subdesign Γ has v' core varieties in b blocks of size k'.
(The core varieties may include extra drones.)

$$n \geq n_0 = \left\lfloor \frac{2v - bk}{b} \right\rfloor \qquad k' \leq k_0 = k - n_0$$

Sum of the pairwise variances

Theorem (cf. Herzberg and Jarrett, 2007)

If there are n drones in each block of Δ,
and the core design Γ has v' varieties in b blocks of size k'
then the sum of the variances of variety differences in Δ

$$= V_T(\Delta) = bn(bn + v' - 1) + V_T(\Gamma) + nV_{BT}(\Gamma) + n^2 V_B(\Gamma),$$

where

$$\begin{aligned} V_T(\Gamma) &= \text{the sum of the variances of variety differences in } \Gamma \\ V_B(\Gamma) &= \text{the sum of the variances of block differences in } \Gamma \\ V_{BT}(\Gamma) &= \text{the sum of the variances of sums of} \\ &\quad \text{one treatment and one block in } \Gamma. \end{aligned}$$

Sum of variances in whole design if Γ is equi-replicate

$$V_T(\Delta) = bn(bn + v' - 1) + V_T(\Gamma) + nV_{BT}(\Gamma) + n^2 V_B(\Gamma)$$

$$\begin{aligned} V_T(\Gamma) &= \text{the sum of the variances of variety differences in } \Gamma \\ V_B(\Gamma) &= \text{the sum of the variances of block differences in } \Gamma \\ V_{BT}(\Gamma) &= \text{the sum of the variances of sums of} \\ &\quad \text{one treatment and one block in } \Gamma. \end{aligned}$$

Sum of variances in whole design if Γ is equi-replicate

$$V_T(\Delta) = bn(bn + v' - 1) + V_T(\Gamma) + nV_{BT}(\Gamma) + n^2 V_B(\Gamma)$$

$$\begin{aligned} V_T(\Gamma) &= \text{the sum of the variances of variety differences in } \Gamma \\ V_B(\Gamma) &= \text{the sum of the variances of block differences in } \Gamma \\ V_{BT}(\Gamma) &= \text{the sum of the variances of sums of one treatment and one block in } \Gamma. \end{aligned}$$

If Γ is equi-replicate with replication r' then

$$\begin{aligned} \frac{k'}{b} V_B(\Gamma) - b &= \frac{r'}{v'} V_T(\Gamma) - v'; \\ V_{BT}(\Gamma) &= \frac{2b}{v'} V_T(\Gamma) + \frac{v'}{k'}(b - v' - 1), \end{aligned}$$

so $V_B(\Gamma)$ and $V_{BT}(\Gamma)$ are both increasing functions of $V_T(\Gamma)$.

Sum of variances in whole design if Γ is equi-replicate

$$V_T(\Delta) = bn(bn + v' - 1) + V_T(\Gamma) + nV_{BT}(\Gamma) + n^2 V_B(\Gamma)$$

$$\begin{aligned} V_T(\Gamma) &= \text{the sum of the variances of variety differences in } \Gamma \\ V_B(\Gamma) &= \text{the sum of the variances of block differences in } \Gamma \\ V_{BT}(\Gamma) &= \text{the sum of the variances of sums of one treatment and one block in } \Gamma. \end{aligned}$$

If Γ is equi-replicate with replication r' then

$$\begin{aligned} \frac{k'}{b} V_B(\Gamma) - b &= \frac{r'}{v'} V_T(\Gamma) - v'; \\ V_{BT}(\Gamma) &= \frac{2b}{v'} V_T(\Gamma) + \frac{v'}{k'}(b - v' - 1), \end{aligned}$$

so $V_B(\Gamma)$ and $V_{BT}(\Gamma)$ are both increasing functions of $V_T(\Gamma)$.

Consequence

For a given choice of k', use the core design Γ which minimizes $V_T(\Gamma)$.

Sum of variances in whole design if there are many drones

$$V_T(\Delta) = bn(bn + v' - 1) + V_T(\Gamma) + nV_{BT}(\Gamma) + n^2 V_B(\Gamma)$$

$$\begin{aligned} V_T(\Gamma) &= \text{the sum of the variances of variety differences in } \Gamma \\ V_B(\Gamma) &= \text{the sum of the variances of block differences in } \Gamma \\ V_{BT}(\Gamma) &= \text{the sum of the variances of sums of} \\ &\quad \text{one treatment and one block in } \Gamma. \end{aligned}$$

Sum of variances in whole design if there are many drones

$$V_T(\Delta) = bn(bn + v' - 1) + V_T(\Gamma) + nV_{BT}(\Gamma) + n^2 V_B(\Gamma)$$

$$\begin{aligned} V_T(\Gamma) &= \text{the sum of the variances of variety differences in } \Gamma \\ V_B(\Gamma) &= \text{the sum of the variances of block differences in } \Gamma \\ V_{BT}(\Gamma) &= \text{the sum of the variances of sums of} \\ &\quad\ \text{one treatment and one block in } \Gamma. \end{aligned}$$

Consequence

If v is large then n is large, so we need to focus on reducing $V_B(\Gamma)$,
so it may be best to increase the number of drones
and decrease k' (the size of blocks in the core design Γ),
so that average replication within Γ is more than 2.

An example of this non-intuitive result

If there are $4(2+n)$ varieties in 4 blocks of size $4+n$, the design on the left is A-better than the design on the right if and only if $n < 50$.

1	2	3	4	n drones
1	2	5	6	n drones
3	6	7	8	n drones
4	5	7	8	n drones

1	2	3	$n+1$ drones
1	2	4	$n+1$ drones
1	3	4	$n+1$ drones
2	3	4	$n+1$ drones

A definite result

Theorem

Suppose that we are given b blocks of size k, and v varieties.
For $i = 1, 2$, let design Δ_i have core subdesign Γ_i with block size k_i.
If Γ_1 is the dual of a balanced incomplete block design and $k_1 > k_2$ then Δ_2 is worse than Δ_1 on the A criterion,
no matter how big v is.

An example of the good result

If there are $4n + 6$ varieties in 4 blocks of size $3 + n$,
the design on the left is A-better than the design on the right,
for all values of n.

1	2	3	n drones
1	4	5	n drones
2	4	6	n drones
3	5	6	n drones

1	2	$n + 1$ drones
1	2	$n + 1$ drones
1	2	$n + 1$ drones
1	2	$n + 1$ drones

Strategy

Given b, v and k, how do we find an A-optimal design for v varieties in b blocks of size k when

$$\frac{bk}{2} \leq v \leq b(k-1)+1?$$

Strategy

Given b, v and k, how do we find an A-optimal design for v varieties in b blocks of size k when

$$\frac{bk}{2} \leq v \leq b(k-1)+1?$$

Average replication ≤ 2

Strategy

Given b, v and k, how do we find an A-optimal design for v varieties in b blocks of size k when

$$\frac{bk}{2} \leq v \leq b(k-1)+1?$$

Average replication ≤ 2 Maximum v for estimability

Strategy

Given b, v and k, how do we find an A-optimal design for v varieties in b blocks of size k when

$$\frac{bk}{2} \leq v \leq b(k-1)+1?$$

Average replication ≤ 2 | Maximum v for estimability

Case 1. $b = 2$ or $b = 3$ (very small b).

Strategy

Given b, v and k, how do we find an A-optimal design for v varieties in b blocks of size k when

$$\frac{bk}{2} \leq v \leq b(k-1)+1?$$

Average replication ≤ 2 Maximum v for estimability

Case 1. $b=2$ or $b=3$ (very small b).

Case 2. $v=b(k-1)+1$ or $v=b(k-1)$ (very large v).

Strategy

Given b, v and k, how do we find an A-optimal design for v varieties in b blocks of size k when

$$\frac{bk}{2} \leq v \leq b(k-1)+1?$$

Average replication ≤ 2 | Maximum v for estimability

Case 1. $b = 2$ or $b = 3$ (very small b).

Case 2. $v = b(k-1)+1$ or $v = b(k-1)$ (very large v).

Case 3. $k_0 \geq b-1$.

$$k_0 = k - \left\lfloor \frac{2v-bk}{b} \right\rfloor = \text{biggest space per block for non-drones.}$$

Strategy

Given b, v and k, how do we find an A-optimal design for v varieties in b blocks of size k when

$$\frac{bk}{2} \leq v \leq b(k-1)+1?$$

Average replication ≤ 2 Maximum v for estimability

Case 1. $b=2$ or $b=3$ (very small b).

Case 2. $v=b(k-1)+1$ or $v=b(k-1)$ (very large v).

Case 3. $k_0 \geq b-1$.

Case 4. $2 < k_0 < b-1$ (small k_0 but not Case 2).

$$k_0 = k - \left\lfloor \frac{2v-bk}{b} \right\rfloor = \text{biggest space per block for non-drones.}$$

Case 1. Only 2 blocks, of size k

Morgan and Jin (2007) showed that the A-optimal designs are those with $2n$ drones and q queen bees,
where $n = n_0 = v - k$ and $q = k' = k_0 = k - n_0 = 2k - v$.

1	2	3	4	$\ldots$	q	A_1	A_2	A_3	$\ldots$	A_n
1	2	3	4	$\ldots$	q	B_1	B_2	B_3	$\ldots$	B_n
queens						drones				

Case 1 continued. 3 blocks of size k

Using the nice theorem, RAB has shown that
the A-optimal designs are as follows when v is divisible by 3
(and presumably small changes deal with the other cases).
There are $3w$ workers and $3n$ drones,
where $3w = 3k - v$ and $n = n_0 = k - 2w$ and $k' = k_0 = 2w$.

1	2	4	5	...	$3w-2$	$3w-1$	A_1	A_2	A_3	...	A_n
1	3	4	6	...	$3w-2$	$3w$	B_1	B_2	B_3	...	B_n
2	3	5	6	...	$3w-1$	$3w$	C_1	C_2	C_3	...	C_n

w copies of design using all pairs from 3

drones

Case 2. $v = b(k-1)+1$

This is the maximum number of varieties that can be tested in b blocks of size k with all comparisons estimable.

Mandal, Shah and Sinha (1991), for $k=2$,
and Bailey and Cameron (2013), for general block size,
showed that, no matter how many blocks there are,
the A-optimal design has the following form.

1	A_1	A_2	A_3	$\dots$	A_{k-1}
1	B_1	B_2	B_3	$\dots$	B_{k-1}
1	C_1	C_2	C_3	$\dots$	C_{k-1}
1	D_1	D_2	D_3	$\dots$	D_{k-1}
1	E_1	E_2	E_3	$\dots$	E_{k-1}
1 queen	$v-1$ drones				

Case 2 continued. $v = b(k-1)$

The A-optimal designs were found for all cases
by Krafft and Schaefer (1997).

Case 2 continued. $v = b(k-1)$

The A-optimal designs were found for all cases
by Krafft and Schaefer (1997).

small k and b

1	2	A_1
2	3	B_1
3	4	C_1
4	5	D_1
5	6	E_1
6	1	F_1

chain

Case 2 continued. $v = b(k-1)$

The A-optimal designs were found for all cases by Krafft and Schaefer (1997).

small k and b — increase k if $b \geq 5$

1	2	A_1
2	3	B_1
3	4	C_1
4	5	D_1
5	6	E_1
6	1	F_1

chain

1	2	A_1	A_2
2	3	B_1	B_2
3	1	C_1	C_2
1	D_1	D_2	D_3
1	E_1	E_2	E_3
1	F_1	F_2	F_3

smaller chain

Case 2 continued. $v = b(k-1)$

The A-optimal designs were found for all cases by Krafft and Schaefer (1997).

small k and b

1	2	A_1
2	3	B_1
3	4	C_1
4	5	D_1
5	6	E_1
6	1	F_1

chain

increase k if $b \geq 5$

1	2	A_1	A_2
2	3	B_1	B_2
3	1	C_1	C_2
1	D_1	D_2	D_3
1	E_1	E_2	E_3
1	F_1	F_2	F_3

smaller chain

then increase b

1	2	A_1	A_2
1	2	B_1	B_2
1	C_1	C_2	C_3
1	D_1	D_2	D_3
1	E_1	E_2	E_3
1	F_1	F_2	F_3
1	G_1	G_2	G_3

1 queen

Case 2 continued. $v = b(k-1)$

The A-optimal designs were found for all cases
by Krafft and Schaefer (1997).

small k and b — increase k if $b \geq 5$ — then increase b

1	2	A_1
2	3	B_1
3	4	C_1
4	5	D_1
5	6	E_1
6	1	F_1

chain

1	2	A_1	A_2
2	3	B_1	B_2
3	1	C_1	C_2
1	D_1	D_2	D_3
1	E_1	E_2	E_3
1	F_1	F_2	F_3

smaller chain

1	2	A_1	A_2
1	2	B_1	B_2
1	C_1	C_2	C_3
1	D_1	D_2	D_3
1	E_1	E_2	E_3
1	F_1	F_2	F_3
1	G_1	G_2	G_3

1 queen

Youden and Connor (1953) had recommended chain designs.

Case 3. $k \geq k_0 \geq b - 1$

For simplicity, assume that b divides $2v$, so that

$$n_0 = \frac{2v - bk}{b} = \text{minimum number of drones per block.}$$

Then

$$\frac{b(2k - b + 1)}{2} \geq v \geq \frac{bk}{2} \geq \frac{b(b-1)}{2}.$$

Case 3. $k \geq k_0 \geq b - 1$

For simplicity, assume that b divides $2v$, so that

$$n_0 = \frac{2v - bk}{b} = \text{minimum number of drones per block.}$$

Then

$$\frac{b(2k - b + 1)}{2} \geq v \geq \frac{bk}{2} \geq \frac{b(b-1)}{2}.$$

Let Γ_0 be the design for $b(b-1)/2$ varieties
replicated twice in b blocks of size $b - 1$
in such a way that
there is one variety in common to each pair of blocks.
This is A-optimal for these numbers.

Case 3 continued. $k \geq k_0 \geq b - 1$

If $k_0 = s(b-1)$ then take Γ to be s copies of Γ_0.
This is always A-optimal.

Case 3 continued. $k \geq k_0 \geq b - 1$

If $k_0 > b - 1$ but k_0 is not a multiple of $b - 1$,
then the following strategy seems likely to be good
(but it is not A-optimal when $b = k_0 = 4$ and v is very large).

Case 3 continued. $k \geq k_0 \geq b-1$

If $k_0 > b-1$ but k_0 is not a multiple of $b-1$,
then the following strategy seems likely to be good
(but it is not A-optimal when $b = k_0 = 4$ and v is very large).
$n_0 =$ minimal number of drones per block.

Construction Method

1. *put n_0 drones in each block;*

Case 3 continued. $k \geq k_0 \geq b - 1$

If $k_0 > b - 1$ but k_0 is not a multiple of $b - 1$,
then the following strategy seems likely to be good
(but it is not A-optimal when $b = k_0 = 4$ and v is very large).
$n_0 =$ minimal number of drones per block.

Construction Method

1. *put n_0 drones in each block;*
2. *put in one copy of Γ_0;*

Case 3 continued. $k \geq k_0 \geq b - 1$

If $k_0 > b - 1$ but k_0 is not a multiple of $b - 1$,
then the following strategy seems likely to be good
(but it is not A-optimal when $b = k_0 = 4$ and v is very large).
$n_0 =$ minimal number of drones per block.

Construction Method

1. *put n_0 drones in each block;*
2. *put in one copy of Γ_0;*
3. *put in as many further copies of Γ_0 as possible;*

Case 3 continued. $k \geq k_0 \geq b-1$

If $k_0 > b-1$ but k_0 is not a multiple of $b-1$,
then the following strategy seems likely to be good
(but it is not A-optimal when $b = k_0 = 4$ and v is very large).
$n_0 =$ minimal number of drones per block.

Construction Method

1. *put n_0 drones in each block;*
2. *put in one copy of Γ_0;*
3. *put in as many further copies of Γ_0 as possible;*
4. *in any remaining space,*
   use a good design for workers with replication 2
   (so long as there is at least one copy of Γ_0,
   it probably doesn't make much difference which one is used).

Case 3. Example: $b = 8$ and $k = 15$ (so $60 \leq v \leq 92$)

60 varieties: all workers ($n_0 = 0$)

1	2	3	4	5	6	7	29	30	31	32	33	34	35	57
1	8	9	10	11	12	13	29	36	37	38	39	40	41	57
2	8	14	15	16	17	18	30	36	42	43	44	45	46	58
3	9	14	19	20	21	22	31	37	42	47	48	49	50	58
4	10	15	19	23	24	25	32	38	43	47	51	52	53	59
5	11	16	20	23	26	27	33	39	44	48	51	54	55	59
6	12	17	21	24	26	28	34	40	45	49	52	54	56	60
7	13	18	22	25	27	28	35	41	46	50	53	55	56	60
one copy of Γ_0							another copy of Γ_0							

Case 3. Example: $b = 8$ and $k = 15$ (so $60 \leq v \leq 92$)

76 varieties: 44 workers, 32 drones ($n_0 = 4$)

1	2	3	4	5	6	7	29	30	31	32	A_1	A_2	A_3	A_4
1	8	9	10	11	12	13	33	34	35	36	B_1	B_2	B_3	B_4
2	8	14	15	16	17	18	37	38	39	40	C_1	C_2	C_3	C_4
3	9	14	19	20	21	22	41	42	43	44	D_1	D_2	D_3	D_4
4	10	15	19	23	24	25	29	33	37	41	E_1	E_2	E_3	E_4
5	11	16	20	23	26	27	30	34	38	42	F_1	F_2	F_3	F_4
6	12	17	21	24	26	28	31	35	39	43	G_1	G_2	G_3	G_4
7	13	18	22	25	27	28	32	36	40	44	H_1	H_2	H_3	H_4
Γ_0							16 workers replication 2				drones			

Case 3. Example: $b = 8$ and $k = 15$ (so $60 \leq v \leq 92$)

92 varieties: 28 workers, 64 drones ($n_0 = 8$)

1	2	3	4	5	6	7	A_1	A_2	A_3	A_4	A_5	A_6	A_7	A_8
1	8	9	10	11	12	13	B_1	B_2	B_3	B_4	B_5	B_6	B_7	B_8
2	8	14	15	16	17	18	C_1	C_2	C_3	C_4	C_5	C_6	C_7	C_8
3	9	14	19	20	21	22	D_1	D_2	D_3	D_4	D_5	D_6	D_7	D_8
4	10	15	19	23	24	25	E_1	E_2	E_3	E_4	E_5	E_6	E_7	E_8
5	11	16	20	23	26	27	F_1	F_2	F_3	F_4	F_5	F_6	F_7	F_8
6	12	17	21	24	26	28	G_1	G_2	G_3	G_4	G_5	G_6	G_7	G_8
7	13	18	22	25	27	28	H_1	H_2	H_3	H_4	H_5	H_6	H_7	H_8

Γ_0 | drones

Case 4. $2 < k_0 < b - 1$

For various values of $k_i \leq k_0$,
find the best core subdesign Γ_i for v_i' varieties in b blocks of size k_i.

Case 4. $2 < k_0 < b - 1$

For various values of $k_i \leq k_0$,
find the best core subdesign Γ_i for v_i' varieties in b blocks of size k_i. (For equi-replicate core subdesigns,
it is often easier to find the best dual design, which is obtained by interchanging the roles of blocks and varieties.)

Case 4. $2 < k_0 < b - 1$

For various values of $k_i \leq k_0$,
find the best core subdesign Γ_i for v_i' varieties in b blocks of size k_i. (For equi-replicate core subdesigns,
it is often easier to find the best dual design, which is obtained by interchanging the roles of blocks and varieties.)

$$
\begin{aligned}
V_T(\Gamma_i) &= \text{the sum of the variances of variety differences in } \Gamma_i \\
V_B(\Gamma_i) &= \text{the sum of the variances of block differences in } \Gamma_i \\
V_{BT}(\Gamma_i) &= \text{the sum of the variances of sums of} \\
&\quad\ \text{one treatment and one block in } \Gamma_i.
\end{aligned}
$$

If there are n_i drones in each block then, in the whole design Δ,

$$V_T(\Delta) = bn_i(bn_i + v_i' - 1) + V_T(\Gamma_i) + n_i V_{BT}(\Gamma_i) + n_i^2 V_B(\Gamma_i).$$

Case 4. $2 < k_0 < b - 1$

For various values of $k_i \leq k_0$,
find the best core subdesign Γ_i for v_i' varieties in b blocks of size k_i. (For equi-replicate core subdesigns, it is often easier to find the best dual design, which is obtained by interchanging the roles of blocks and varieties.)

$$\begin{aligned} V_T(\Gamma_i) &= \text{the sum of the variances of variety differences in } \Gamma_i \\ V_B(\Gamma_i) &= \text{the sum of the variances of block differences in } \Gamma_i \\ V_{BT}(\Gamma_i) &= \text{the sum of the variances of sums of} \\ &\quad \text{one treatment and one block in } \Gamma_i. \end{aligned}$$

If there are n_i drones in each block then, in the whole design Δ,

$$V_T(\Delta) = bn_i(bn_i + v_i' - 1) + V_T(\Gamma_i) + n_i V_{BT}(\Gamma_i) + n_i^2 V_B(\Gamma_i).$$

Use this formula to find the core subdesign which gives the smallest $V_T(\Delta)$.

Case 4. $2 < k_0 < b - 1$

For various values of $k_i \leq k_0$,
find the best core subdesign Γ_i for v_i' varieties in b blocks of size k_i. (For equi-replicate core subdesigns, it is often easier to find the best dual design, which is obtained by interchanging the roles of blocks and varieties.)

$$\begin{aligned} V_T(\Gamma_i) &= \text{the sum of the variances of variety differences in } \Gamma_i \\ V_B(\Gamma_i) &= \text{the sum of the variances of block differences in } \Gamma_i \\ V_{BT}(\Gamma_i) &= \text{the sum of the variances of sums of} \\ &\quad\ \text{one treatment and one block in } \Gamma_i. \end{aligned}$$

If there are n_i drones in each block then, in the whole design Δ,

$$V_T(\Delta) = bn_i(bn_i + v_i' - 1) + V_T(\Gamma_i) + n_i V_{BT}(\Gamma_i) + n_i^2 V_B(\Gamma_i).$$

Use this formula to find the core subdesign which gives the smallest $V_T(\Delta)$.
As the number of varieties increases, it becomes more important to choose Γ_i with a small value of $V_B(\Gamma_i)$.

Case 4 continued. $k_0 = 4 < b - 1$, $V_B(\Gamma_i) \div b(b-1)/2$

Best design for b blocks known to RAB

	Γ_1	Γ_2	Γ_3	Γ_4
k_i	2	3	3	4
	2 queens, both boring	2 queens, 2 workers (rep 2)	b workers rep 3	$2b$ workers rep 2
$b = 6$	1	1^-	0.85	0.87
$b = 7$	1	1^-	0.86	0.92
$b = 8$	1	1^-	0.89	0.93
$b = 9$	1	1^-	0.92	
$b = 10$	1	1^-		
$b = 11$	1	1^-		
$b = 12$	1	1^-	0.98	
$b = 13$	1	1^-	1	1.07
$b = 14$	1	1^-		
$b = 15$	1	1^-	1.01	1.08

Case 4 continued. $k_0 = 4 < b - 1$, $V_B(\Gamma_i) \div b(b-1)/2$

	Best design for b blocks known to RAB			
	Γ_1	Γ_2	Γ_3	Γ_4
k_i	2	3	3	4
	2 queens, both boring	2 queens, 2 workers (rep 2)	b workers rep 3	$2b$ workers rep 2
$b = 6$	1	1^-	0.85	0.87
$b = 7$	1	1^-	0.86	0.92
$b = 8$	1	1^-	0.89	0.93
$b = 9$	1	1^-	0.92	
$b = 10$	1	1^-		
$b = 11$	1	1^-		
$b = 12$	1	1^-	0.98	
$b = 13$	1	1^-	1	1.07
$b = 14$	1	1^-		
$b = 15$	1	1^-	1.01	1.08

As v increases, Γ_3 becomes better than Γ_4.

Case 4 continued. $k_0 = 4 < b-1$, $V_B(\Gamma_i) \div b(b-1)/2$

Best design for b blocks known to RAB

	Γ_1	Γ_2	Γ_3	Γ_4
k_i	2	3	3	4
	2 queens, both boring	2 queens, 2 workers (rep 2)	b workers rep 3	$2b$ workers rep 2
$b = 6$	1	1^-	0.85	0.87
$b = 7$	1	1^-	0.86	0.92
$b = 8$	1	1^-	0.89	0.93
$b = 9$	1	1^-	0.92	
$b = 10$	1	1^-		
$b = 11$	1	1^-		
$b = 12$	1	1^-	0.98	
$b = 13$	1	1^-	1	1.07
$b = 14$	1	1^-		
$b = 15$	1	1^-	1.01	1.08

As v increases, Γ_3 becomes better than Γ_4.

If $b \geq 14$, then, as v increases, Γ_1 and Γ_2 become better than Γ_3.

Case 4 continued. $2 < k_0 < b - 1$ when $b = 8$: $k_0 = 6$

$k = k_0 = 6$, and 24 varieties, all workers, all replicated twice.

1	2	3	4	5	6
7	8	9	10	11	12
1	7	13	14	15	16
2	8	17	18	19	20
3	9	13	17	21	22
4	10	14	18	23	24
5	11	15	19	21	23
6	12	16	20	22	24

(One worker for each pair of blocks
except for $\{A, B\}$, $\{C, D\}$, $\{E, F\}$ and $\{G, H\}$.)

Case 4 continued. $k = 5$ and $k = 6$ when $b = 8$: $k_0 = 5$

$k = 5$
20 varieties:
20 workers, no drones

1	2	3	4	5
6	7	8	9	10
1	11	12	13	14
2	6	15	16	17
3	7	11	18	19
4	8	12	15	20
5	9	13	16	18
10	14	17	19	20

$k = 6$
28 varieties:
20 workers, 8 drones

1	2	3	4	5	A_1
6	7	8	9	10	B_1
1	11	12	13	14	C_1
2	6	15	16	17	D_1
3	7	11	18	19	E_1
4	8	12	15	20	F_1
5	9	13	16	18	G_1
10	14	17	19	20	H_1

Case 4 continued. $k = 5$ and $k = 6$ when $b = 8$: $k_0 = 4$

$k = 5$
24 varieties:
16 workers, 8 drones

1	2	3	4	A_1
5	6	7	8	B_1
9	10	11	12	C_1
13	14	15	16	D_1
1	5	9	13	E_1
2	6	10	14	F_1
3	7	11	15	G_1
4	8	12	16	H_1

$k' = 4$
rep = 2

$k = 6$
32 varieties:
8 workers, 24 drones

1	2	4	A_1	A_2	A_3
2	3	5	B_1	B_2	B_3
3	4	6	C_1	C_2	C_3
4	5	7	D_1	D_2	D_3
5	6	8	E_1	E_2	E_3
6	7	1	F_1	F_2	F_3
7	8	2	G_1	G_2	G_3
8	1	3	H_1	H_2	H_3

$k' = 3$
rep 3

Case 4 continued. $k = 5$ and $k = 6$ when $b = 8$: $k_0 = 3$

$k = 5$
28 varieties:
12 workers, 16 drones

1	2	3	A_1	A_2
1	4	5	B_1	B_2
4	6	7	C_1	C_2
6	8	9	D_1	D_2
2	8	10	E_1	E_2
5	10	11	F_1	F_2
7	11	12	G_1	G_2
3	9	12	H_1	H_2

$k = 6$
36 varieties:
12 workers, 24 drones

1	2	3	A_1	A_2	A_3
1	4	5	B_1	B_2	B_3
4	6	7	C_1	C_2	C_3
6	8	9	D_1	D_2	D_3
2	8	10	E_1	E_2	E_3
5	10	11	F_1	F_2	F_3
7	11	12	G_1	G_2	G_3
3	9	12	H_1	H_2	H_3

Health Warnings

The overall message is that there can be phase changes as the spare capacity for replication ($bk - v$) decreases.
Therefore it is necessary to compare core subdesigns Γ_i with different block size k_i.

Health Warnings

The overall message is that there can be phase changes as the spare capacity for replication $(bk - v)$ decreases.
Therefore it is necessary to compare core subdesigns Γ_i with different block size k_i.

Although this overall message is correct,
no one has checked the arithmetic in the examples presented,
so individual cases may be wrong.

Health Warnings

The overall message is that there can be phase changes as the spare capacity for replication $(bk - v)$ decreases.
Therefore it is necessary to compare core subdesigns Γ_i with different block size k_i.

Although this overall message is correct,
no one has checked the arithmetic in the examples presented,
so individual cases may be wrong.

This is work is progress, not a finished project.

The cornerstones of experimental design

- Replication.
- Blocking.
- Randomization.

The cornerstones of experimental design

- Replication.
- Blocking.
- Randomization.

What is old?
What is new?

Replication: the old. I

In 2012 the UK's Food and Environment Research Agency conducted an experiment to find out "the effects of neonicotinoid seed treatments on bumble bee colonies under field conditions" (from a DEFRA report available on the web, Crown copyright 2013).

`fera.co.uk/ccss/documents/defraBumbleBeereportPS2371V4A.pdf`

Replication: the old. I

In 2012 the UK's Food and Environment Research Agency conducted an experiment to find out "the effects of neonicotinoid seed treatments on bumble bee colonies under field conditions" (from a DEFRA report available on the web, Crown copyright 2013).

`fera.co.uk/ccss/documents/defraBumbleBeereportPS2371V4A.pdf`

Site	Treatment of oilseed rape seeds
Site A, near Lincoln	no treatment
Site B, near York	Modesto^{TM}
Site C, near Scunthorpe	Chinook^{TM}

Replication: the old. I

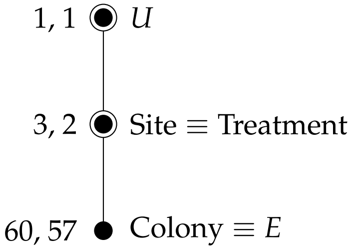
In 2012 the UK's Food and Environment Research Agency conducted an experiment to find out "the effects of neonicotinoid seed treatments on bumble bee colonies under field conditions" (from a DEFRA report available on the web, Crown copyright 2013).

`fera.co.uk/ccss/documents/defraBumbleBeereportPS2371V4A.pdf`

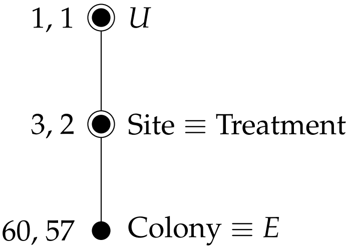
Site	Treatment of oilseed rape seeds
Site A, near Lincoln	no treatment
Site B, near York	Modesto™
Site C, near Scunthorpe	Chinook™

Twenty colonies of bumble bees were placed at each site. Various outcomes were measured on each colony.

Replication: the old. II



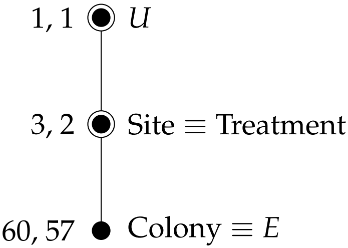
Replication: the old. II



Skeleton analysis of variance

Stratum	Source	df
U	Mean	1
Sites	Treatments	2
Colonies		57

Replication: the old. II

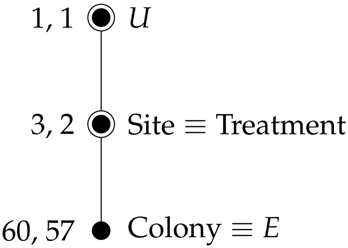


Skeleton analysis of variance

Stratum	Source	df
U	Mean	1
Sites	Treatments	2
Colonies		57

There is no residual mean square in the stratum containing Treatments, so we cannot tell if observed differences are caused by differences between treatments or differences between sites.

Replication: the old. II

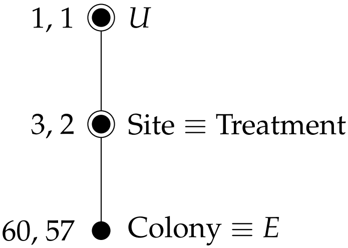


Skeleton analysis of variance

Stratum	Source	df
U	Mean	1
Sites	Treatments	2
Colonies		57

There is no residual mean square in the stratum containing Treatments, so we cannot tell if observed differences are caused by differences between treatments or differences between sites. Therefore, there is no way of giving confidence intervals for the estimates of treatment differences, or of giving P values for testing the hypothesis of no treatment difference.

Replication: the old. II

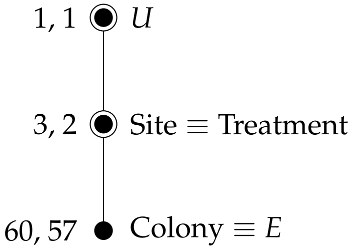


Skeleton analysis of variance

Stratum	Source	df
U	Mean	1
Sites	Treatments	2
Colonies		57

There is no residual mean square in the stratum containing Treatments, so we cannot tell if observed differences are caused by differences between treatments or differences between sites. Therefore, there is no way of giving confidence intervals for the estimates of treatment differences, or of giving P values for testing the hypothesis of no treatment difference. The official report does claim to give confidence intervals and P values.

Replication: the old. II



Skeleton analysis of variance

Stratum	Source	df
U	Mean	1
Sites	Treatments	2
Colonies		57

There is no residual mean square in the stratum containing Treatments, so we cannot tell if observed differences are caused by differences between treatments or differences between sites. Therefore, there is no way of giving confidence intervals for the estimates of treatment differences, or of giving P values for testing the hypothesis of no treatment difference. The official report does claim to give confidence intervals and P values.

The Hasse diagram can clearly show such false replication before the experiment is carried out.

Replication: the old. III

False replication is one of the oldest and most common mistakes in design of experiments.

- S. H. Hurlbert: Pseudoreplication and the design of ecological field experiments. *Ecological Monographs* **54** (1984), 187–211.
- I gave advice to the Ministry of Agriculture, Fisheries and Foods about this problem in the 1980s. See Example 1.1 of my 2008 book *Design of Comparative Experiments*.
- T. H. Sparks, R. A. Bailey & D. A. Elston: Pseudoreplication: common (mal)practice. *SETAC (Society of Environmental Toxicology and Chemistry) News*, **17(3)** (1997), 12–13.
- S. H. Hurlbert: The ancient black art and transdisciplinary extent of pseudoreplication. *Journal of Comparative Psychology* **123** (2009), 436–443.

Replication: the old. III

False replication is one of the oldest and most common mistakes in design of experiments.

- S. H. Hurlbert: Pseudoreplication and the design of ecological field experiments. *Ecological Monographs* **54** (1984), 187–211.
- I gave advice to the Ministry of Agriculture, Fisheries and Foods about this problem in the 1980s. See Example 1.1 of my 2008 book *Design of Comparative Experiments*.
- T. H. Sparks, R. A. Bailey & D. A. Elston: Pseudoreplication: common (mal)practice. *SETAC (Society of Environmental Toxicology and Chemistry) News*, **17(3)** (1997), 12–13.
- S. H. Hurlbert: The ancient black art and transdisciplinary extent of pseudoreplication. *Journal of Comparative Psychology* **123** (2009), 436–443.

Why is this still happening?

Replication: the old. III

False replication is one of the oldest and most common mistakes in design of experiments.

- S. H. Hurlbert: Pseudoreplication and the design of ecological field experiments. *Ecological Monographs* **54** (1984), 187–211.
- I gave advice to the Ministry of Agriculture, Fisheries and Foods about this problem in the 1980s. See Example 1.1 of my 2008 book *Design of Comparative Experiments*.
- T. H. Sparks, R. A. Bailey & D. A. Elston: Pseudoreplication: common (mal)practice. *SETAC (Society of Environmental Toxicology and Chemistry) News*, **17(3)** (1997), 12–13.
- S. H. Hurlbert: The ancient black art and transdisciplinary extent of pseudoreplication. *Journal of Comparative Psychology* **123** (2009), 436–443.

Why is this still happening?

Why is it still happening in experiments undertaken or commissioned by publicly funded bodies?

Replication: the new

For a long time, we have used high replication as a surrogate condition for low variance.

Replication: the new

For a long time, we have used high replication as
a surrogate condition for low variance.

My work on block designs with low average replication
shows that, when you are close to the wire,
the surrogate may not be a good guide.

Replication: the new

For a long time, we have used high replication as a surrogate condition for low variance.

My work on block designs with low average replication shows that, when you are close to the wire, the surrogate may not be a good guide.

At the Design and Analysis of Experiments conference in Cary, North Carolina, 4–6 March 2015, there were several talks (in different sessions, with no link planned in advance) where the message was that a surrogate measure may not work when you are close to the wire.

Replication: the new

For a long time, we have used high replication as
a surrogate condition for low variance.

My work on block designs with low average replication shows that, when you are close to the wire, the surrogate may not be a good guide.

At the Design and Analysis of Experiments conference in Cary, North Carolina, 4–6 March 2015, there were several talks (in different sessions, with no link planned in advance) where the message was that a surrogate measure may not work when you are close to the wire.

For example, Ching-Shui Cheng discussed supersaturated designs (for m 2-level factors in n experimental units, with $n < m$). One of the classical surrogates for model identifability is equal replication ("balance") of the levels of each factor. He showed that this is not a good guide when the models have high dimension.

Blocking: the old. I

The old paradigm is to group the experimental units into *blocks* of alike units,
and use the blocks in both the design and the analysis,
in order to remove bias from the estimates of treatment differences AND from the estimate of experimental error
in order to increase power.

Blocking: the old. I

The old paradigm is to group the experimental units into *blocks* of alike units,
and use the blocks in both the design and the analysis,
in order to remove bias from the estimates of treatment differences AND from the estimate of experimental error in order to increase power.

We know how to do this for several systems of blocks, under orthogonality.

Blocking: the old. II

In experiments in human-computer interaction, it is common to ask each participant to undertake a certain task under several different scenarios. With four scenarios, the experiment might use 20 participants once each on four days, and assign scenarios using Latin squares.

A	*B*	*C*	*D*	*A*	*B*	*C*	*D*	*A*	*B*	*C*	*D*	*A*	*B*	*C*	*D*	*A*	*B*	*C*	*D*
B	*C*	*D*	*A*	*B*	*C*	*D*	*A*	*B*	*C*	*D*	*A*	*B*	*C*	*D*	*A*	*B*	*C*	*D*	*A*
C	*D*	*A*	*B*	*C*	*D*	*A*	*B*	*C*	*D*	*A*	*B*	*C*	*D*	*A*	*B*	*C*	*D*	*A*	*B*
D	*A*	*B*	*C*	*D*	*A*	*B*	*C*	*D*	*A*	*B*	*C*	*D*	*A*	*B*	*C*	*D*	*A*	*B*	*C*

Blocking: the old. III

Skeleton analysis of variance

Stratum	Source	df
U	Mean	1
Days		3
Participants		19
Days#Participants	Scenarios	3
	residual	54

Blocking: the old. III

Skeleton analysis of variance

Stratum	Source	df
U	Mean	1
Days		3
Participants		19
Days#Participants	Scenarios	3
	residual	54

… the ANOVA produces three F-values … risk of over-testing the data …

so do not include Days in the ANOVA

- P. Cairns: HCL… Not as it should be: inferential statistics in HCI research. *Proceedings of HCI 2007*, 195–201.

Blocking: the old. III

Skeleton analysis of variance

Stratum	Source	df
U	Mean	1
Days		3
Participants		19
Days#Participants	Scenarios	3
	residual	54

... the ANOVA produces three F-values ... risk of over-testing the data ...

so do not include Days in the ANOVA

- P. Cairns: HCL... Not as it should be: inferential statistics in HCI research. *Proceedings of HCI 2007*, 195–201.

This is wrong on two counts:

(i) We do not test for differences between days and between participants: we expect such differences and fit them in the model.

(ii) If we do not remove these differences, we decrease the power for detecting differences between scenarios.

Blocking: the new. I

As new technologies are used in experimentation
(for example, in genomics)
there may be systems of blocks which the experimenters do not recognise because they are not called ‘blocks’.
How do we ensure that known methods of construction, randomization and data analysis are not lost?

Blocking: the new. I

As new technologies are used in experimentation
(for example, in genomics)
there may be systems of blocks which the experimenters do not recognise because they are not called ‘blocks’.
How do we ensure that known methods of construction, randomization and data analysis are not lost?

In agricultural fields in some countries,
it is more realistic to model spatial variation by spatial correlation than by discrete blocks.
How should we construct and randomize such experiments, and analyse data from them?

Blocking: the new. II

In some recent experiments, the experimental units can be thought of as the nodes of a graph, with edges between some nodes. For example, Gerry Humphris of the University of St Andrews is experimenting with non-medical interventions such as sending a text message to known binge drinkers on Friday afternoons. One drinker (a node) may alter his behaviour and thus affect the behaviour of other people that he knows (the nodes joined to him in the graph).
How should we construct and randomize such experiments, and analyse data from them?

Randomization: the old. I

There is a large body of theory about how to randomize experiments in *simple orthogonal block structures* in such a way that estimators of treatment effects and of their variances are both unbiased over the randomization.

Randomization: the old. II

If you randomize an experiment, do not like the outcome, throw it away and re-randomize, what happens?

Randomization: the old. II

If you randomize an experiment, do not like the outcome, throw it away and re-randomize, what happens?

More variablity will be assigned to experimental error than to treatment differences, so the experimenter is less likely to detect genuine differences between treatments.

Randomization: the old. II

If you randomize an experiment, do not like the outcome, throw it away and re-randomize, what happens?

More variablity will be assigned to experimental error than to treatment differences, so the experimenter is less likely to detect genuine differences between treatments. See:

- R. A. Fisher: *Design of Experiments*. Oliver and Boyd, Edinburgh, 1935.
- W. J. Youden: Randomization and experimentation. *Technometrics* **14**, (1972), 13–22.

Randomization: the old. II

If you randomize an experiment, do not like the outcome, throw it away and re-randomize, what happens?

More variablity will be assigned to experimental error than to treatment differences, so the experimenter is less likely to detect genuine differences between treatments. See:

- R. A. Fisher: *Design of Experiments*. Oliver and Boyd, Edinburgh, 1935.
- W. J. Youden: Randomization and experimentation. *Technometrics* **14**, (1972), 13–22.

This has not stopped people from developing software for throwing away "undesirable" outcomes of randomization.

- D. T. Bowman: TFPlan: software for restricted randomization in field plot design. *Agronomy Journal* **92**, (2000), 1276–1278.
- A talk presented at the Tenth Working Seminar on Statistical Methods in Variety Testing at Będlewo, Poland, in July 2014.

Randomization: the new

For spatial correlation, and for designs on the nodes of graphs,

- how should we randomize?
- what criterion of validity should we use?