Design keys for multi-phase experiments

R. A. Bailey
University of St Andrews
QMUL (emerita)

MODA, June 2016

Abstract

Desmond Patterson introduced the design key in 1965
in the context of rotation experiments in agriculture.

When there are many factors involved,
the design key gives an algorithm for constructing the design
and for keeping track of confounding.

Here I extend the idea to multi-phase experiments,
using one design key for each phase.

H. Desmond Patterson

Introduction of the design key

H. D. Patterson:
The factorial combination of treatments in rotation experiments.
Journal of Agricultural Science, **65** (1965), 171–182.

This paper introduced the design key.

The number of levels of each factor must be a power of a single prime number p. All examples have $p = 2$, but it is mentioned that the method can also be used with $p = 3$.

The examples from this paper are all far too complicated to include in a short talk.
I am amazed that any referee understood it at the time.
Parts of it are almost as if Desmond is thinking aloud.

I shall try to explain it with simpler examples.

Example 1 (Graeco-Latin square): the key

Factors

	Factors	Factors with five levels
Experimental units	Rows	R
	Columns	C
Treatments	Variety of wheat	W
	Quantity of nitrogen	N

Every factor is represented by a single letter.
Levels are integers modulo 5.

Constraints The treatment factors W and N should both be orthogonal to rows and orthogonal to columns.

Design key The design key Φ expresses each treatment factor as a linear combination of factors on the experimental units.

$$\Phi(W) = R + C \qquad \Phi(N) = R + 2C$$

Example 1 (Graeco-Latin square): construction

$$\Phi(W) = R + C \qquad \Phi(N) = R + 2C$$

	C				
R	0	1	2	3	4
0	0,0	1,2	2,4	3,1	4,3
1	1,1	2,3	3,0	4,2	0,4
2	2,2	3,4	4,1	0,3	1,0
3	3,3	4,0	0,2	1,4	2,1
4	4,4	0,1	1,3	2,0	3,2

The experimental units are defined by all combinations of levels of R and C.

The level of W is shown first in each cell.

The level of N is shown second in each cell.

Example 1 (Graeco-Latin square): confounding

$$\Phi(W) = R + C \qquad \Phi(N) = R + 2C$$

Stratum	unit effect	df	tmt factor	tmt effect
Rows	R	4	$W + 2N$	interaction
Columns	C	4	$W + 4N$	interaction
Rows-by-Columns	$R + C$	4	W	variety main
	$R + 2C$	4	N	nitrogen main
	$R + 3C$	4	$W + 3N$	interaction
	$R + 4C$	4	$W + N$	interaction

$$\Phi(W + N) = \Phi(W) + \Phi(N) = 2R + 3C \equiv 6R + 9C = R + 4C$$

Treatment factors (vide R. A. Fisher, 1942)

There is a set $\mathcal{F}$ of treatment factors.

There is one potential treatment for each combination of levels of all the factors in $\mathcal{F}$.

At first, we assume that every factor in $\mathcal{F}$ has p levels, where p is prime. The levels are the integers modulo p. All addition is done modulo p.

Each non-zero linear combination of factors in $\mathcal{F}$ gives a treatment pseudofactor with p levels.
This gives $p - 1$ degrees of freedom for contrasts between treatments, all belonging to the interaction of those genuine treatment factors whose coefficient is non-zero.

If one such linear combination is a non-zero multiple of another, then they correspond to the same df; otherwise the corresponding sets of contrasts are orthogonal to each other.

Factors on the experimental units

There is a set $\mathcal{G}$ of unit factors.

We assume that the real factors on the experimental units form a poset block structure.

This means that they can be defined by a panel diagram, showing

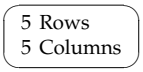
- the list of factors $G_1, \ldots, G_m$ in $\mathcal{G}$
- for each G_i, its number n_i of levels;
- for each G_i, what it is nested in.

There are $n_1 \times \cdots \times n_m$ experimental units, one for each combination of levels of $G_1, \ldots, G_m$.

"G_i is nested in G_j" means
"if two objects have the same level of G_i then this has no significance unless they have the same level of G_j".

The real factors are combinations of levels of none or more of $G_1, \ldots, G_m$ subject to the rule that
if G_i is included and G_i is nested in G_j then G_j must be included.

Poset block structure in Example 1

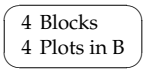


This panel diagram tells us that

- there are factors R and C, each with 5 levels;
- there are 25 experimental units, one for each combination of levels of R and C;
- there is no nesting;
- the real factors on the experimental units are

$\emptyset$	with 1 level;
R	with 5 levels;
C	with 5 levels;
RC	with 25 levels.

Poset block structure in Example 2



This panel diagram tells us that

- there are factors B and P, each with 4 levels;
- there are 16 experimental units, one for each combination of levels of B and P;
- P is nested in B, so there is no real factor involving P but not B;
- the real factors on the experimental units are

$\emptyset$	with 1 level;
B	with 4 levels;
BP	with 16 levels.

Powers of a prime

If a factor has p^r levels, where $r \geq 2$, then it is represented by r pseudofactors, each with p levels.

The convention is that these pseudofactors are written with the same single letter and subscripts $1, \ldots, r$.

Identification of factorial effects

For a linear combination of factors (and pseudofactors) in $\mathcal{F}$ or a linear combination of factors (and pseudofactors) in $\mathcal{G}$, we need to identify the factorial effect containing the corresponding $p-1$ degrees of freedom.

1. Write down all the letters which occur, ignoring subscripts;
2. if factor C is nested in factor D and letter C occurs then include letter D;
3. remove any duplicate letters.

The set of letters remaining gives the factorial effect.

Example 2 (Factorial design in blocks): the key

Factors

	Factors	(Pseudo-)factors with 2 levels
Experimental units	Blocks (4)	B_1, B_2
	Plots in Blocks (4)	P_1, P_2
Treatments	S (2)	S
	T (2)	T
	U (2)	U
	V (2)	V

The pseudofactors for each factor all have the same letter.
Levels are integers modulo 2.

Constraints All treatment main effects should be orthogonal to blocks. So should as many two-factor interactions as possible.

Design key

$\Phi(S) = P_1$, $\Phi(T) = P_2$, $\Phi(U) = B_1 + P_1 + P_2$, $\Phi(V) = B_2 + P_1 + P_2$

Example 2 (Factorial design in blocks): construction

$\Phi(S) = P_1$, $\Phi(T) = P_2$, $\Phi(U) = B_1 + P_1 + P_2$, $\Phi(V) = B_2 + P_1 + P_2$

The experimental units are defined by all combinations of levels of B_1, B_2, P_1 and P_2.

	Block 1				Block 2				Block 3				Block 4			
B_1	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1
B_2	0	0	0	0	1	1	1	1	0	0	0	0	1	1	1	1
P_1	0	0	1	1	0	0	1	1	0	0	1	1	0	0	1	1
P_2	0	1	0	1	0	1	0	1	0	1	0	1	0	1	0	1
S	0	0	1	1	0	0	1	1	0	0	1	1	0	0	1	1
T	0	1	0	1	0	1	0	1	0	1	0	1	0	1	0	1
U	0	1	1	0	0	1	1	0	1	0	0	1	1	0	0	1
V	0	1	1	0	1	0	0	1	0	1	1	0	1	0	0	1

Example 2 (Factorial design in blocks): confounding

$\Phi(S) = P_1$, $\Phi(T) = P_2$, $\Phi(U) = B_1 + P_1 + P_2$, $\Phi(V) = B_2 + P_1 + P_2$

Stratum	unit effect	df	tmt factor	tmt effect
Blocks (B)	B_1	1	$S+T+U$	3 f.i.
	B_2	1	$S+T+V$	3 f.i.
	B_1+B_2	1	$U+V$	U-by-V intn
Plots in Blocks (BP)	P_1	1	S	main S
	P_2	1	T	main T
	P_1+P_2	1	$S+T$	S-by-T intn
	B_1+P_1	1	$T+U$	T-by-U intn
	B_1+P_2	1	$S+U$	S-by-U intn
	$B_1+P_1+P_2$	1	U	main U
	B_2+P_1	1	$T+V$	T-by-V intn
	B_2+P_2	1	$S+V$	S-by-V intn
	$B_2+P_1+P_2$	1	V	main V
	$B_1+B_2+P_1$	1	$S+U+V$	3 f.i.
	$B_1+B_2+P_2$	1	$T+U+V$	3 f.i.
	$B_1+B_2+P_1+P_2$	1	$S+T+U+V$	4 f.i.

What does the design key do?

A design key is a list
giving an alias for each treatment (pseudo-)factor as
a linear combination of (pseudo-)factors for the experimental units.

This gives

- an algorithm for constructing the design;
- a design that is orthogonal;
- (if it is a fractional replicate) a fraction which is regular;
- an algorithm for identifying the confounding between treatment effects and strata defined by a poset block structure on the experimental units.

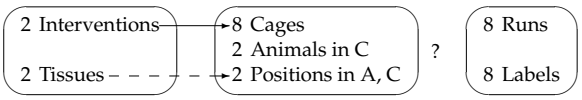
Two-phase experiments

Treatments $\xrightarrow{\Phi}$ Phase I units $\xrightarrow{\Psi}$ Phase II units

Design key Φ allocates treatments to Phase I units.
Design key Ψ allocates Phase I units to Phase II units.

Combining these allows us to keep track of confounding all the way through,
which helps us to choose suitable design keys in the first place.

Example 3 (Proteomics): constraints

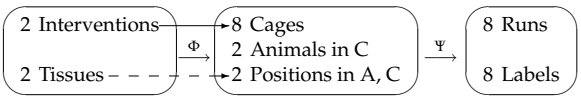


- Interventions probably has the biggest variance from Phase I, so try to confound this with a low-variance term in Phase II.
- If possible, confound the rest of Cages with the same term, to avoid losing degrees of freedom for the residual.
- If possible, make Tissues and I#T orthogonal to Runs and orthogonal to Labels.

Design key

I	$\Phi(I) = C_1$	$C_1,\ C_2,\ C_3$	$\Psi(C_i) = R_i + L_i$	$R_1,\ R_2,\ R_3$
		A	$\Psi(A) = R_1$	
T	$\Phi(T) = P + C_3$	P	$\Psi(P) = L_2$	$L_1,\ L_2,\ L_3$

Example 3 (Proteomics): confounding



I	$\Phi(I) = C_1$	$C_1,\ C_2,\ C_3$	$\Psi(C_i) = R_i + L_i$	$R_1,\ R_2,\ R_3$
		A	$\Psi(A) = R_1$	
T	$\Phi(T) = P + C_3$	P	$\Psi(P) = L_2$	$L_1,\ L_2,\ L_3$

$\Psi(P + C_2) = R_2$ Positions[A,C]; Runs

$\Psi(\Phi(T)) = \Psi(P + C_3) = L_2 + R_3 + L_3$ T; P[A,C]; R#L

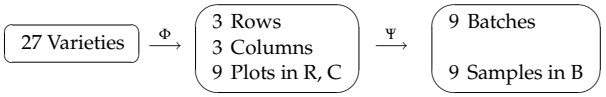
$\Psi(\Phi(I + T)) = \Psi(C_1 + P + C_3) = R_1 + L_1 + L_2 + R_3 + L_3$

I#T; P[A,C]; R#L

Example 3 (Proteomics): skeleton anova

units		animal-bits		treatments		
source	df	source	df	source	df	EMS
Mean	1	Mean	1	Mean	1	$\xi_0 + 2\eta_0 + q_0$
Runs	7	Animals[C]$_1$	1			$\xi_R + 2\eta_{CA}$
		Positions[A,C]$_1$	2			$\xi_R + 2\eta_{CAP}$
		Residual	4			ξ_R
Labels	7	Animals[C]$_2$	1			$\xi_L + 2\eta_{CA}$
		Positions[A,C]$_2$	2			$\xi_L + 2\eta_{CAP}$
		Residual	4			ξ_L
R#L	49	Cages	7	Interventions	1	$\xi_{RL} + 2\eta_C + q(\mathrm{I})$
				Residual	6	$\xi_{RL} + 2\eta_C$
		Animals[C]$_3$	6			$\xi_{RL} + 2\eta_{CA}$
		Positions[A,C]$_3$	12	Tissues	1	$\xi_{RL} + 2\eta_{CAP} + q(\mathrm{T})$
				I#T	1	$\xi_{RL} + 2\eta_{CAP} + q(\mathrm{IT})$
				Residual	10	$\xi_{RL} + 2\eta_{CAP}$
		Residual	24			ξ_{RL}

Example 4 (Field then laboratory): constraints

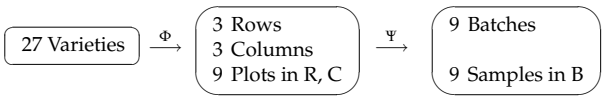


$V_1,\ V_2,\ V_3$	$\Phi(V_3) = R + C$	R	$B_1,\ B_2$
	$\Phi(V_1) = P_1$	C	
	$\Phi(V_2) = P_2$	$P_1,\ P_2$	$S_1,\ S_2$

Constraints All Variety effects should be orthogonal to Rows and orthogonal to Columns in Phase I.
Then at least 2df for Varieties must be confounded with R#C, so there is no loss of generality in taking this design key Φ for the first phase.

Question What should we do in the second phase, given that at least 2df for Varieties must be confounded with Batches?

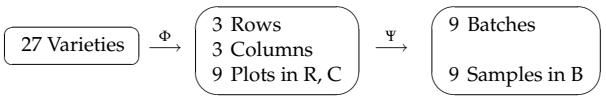
Example 4 (Field then laboratory): option 1



$V_1,\ V_2,\ V_3$	$\Phi(V_3) = R + C$	R	$\Psi(R) = B_1$ $B_1,\ B_2$
	$\Phi(V_1) = P_1$	C	$\Psi(C) = B_2$
	$\Phi(V_2) = P_2$	$P_1,\ P_2$	$\Psi(P_i) = S_i$ $S_1,\ S_2$

samples		plots		varieties		
source	df	source	df	source	df	EMS
Batches	8	Rows	2			$\xi_B + \eta_R$
		Columns	2			$\xi_B + \eta_C$
		R#C	4	V_3	2	$\xi_B + \eta_{RC} + q(V_3)$
				Residual	2	$\xi_B + \eta_{RC}$

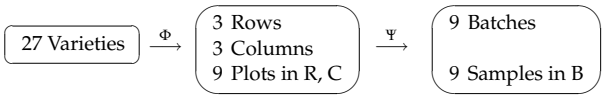
Example 4 (Field then laboratory): option 2



$V_1,\ V_2,\ V_3$	$\Phi(V_3) = R + C$	R	$\Psi(R) = B_1$	$B_1,\ B_2$
	$\Phi(V_1) = P_1$	C	$\Psi(C) = S_1$	
	$\Phi(V_2) = P_2$	$P_1,\ P_2$	$\Psi(P_1) = B_2$	$S_1,\ S_2$
			$\Psi(P_2) = S_2$	

samples		plots		varieties		
source	df	source	df	source	df	EMS
Batches	8	Rows	2			$\xi_B + \eta_R$
		Plots[R,C]$_1$	6	V_1	2	$\xi_B + \eta_{RCP} + q(V_1)$
				Residual	4	$\xi_B + \eta_{RCP}$

Example 4 (Field then laboratory): option 3



$V_1,\ V_2,\ V_3$	$\Phi(V_3) = R + C$	R	$\Psi(R) = B_1 + S_2$	$B_1,\ B_2$
	$\Phi(V_1) = P_1$	C	$\Psi(C) = S_1$	
	$\Phi(V_2) = P_2$	$P_1,\ P_2$	$\Psi(P_1) = B_2$	$S_1,\ S_2$
			$\Psi(P_2) = S_2$	

samples		plots		varieties		
source	df	source	df	source	df	EMS
Batches	8	Plots[R,C]$_1$	8	V_1	2	$\xi_B + \eta_{RCP} + q(V_1)$
				Residual	6	$\xi_B + \eta_{RCP}$

Possible generalizations

More phases Treatments are applied to first-phase units. For $i > 1$, in each phase i, the material from the units in phase $i - 1$ is applied to units in phase i using another design key, but no further treatments are applied.

The foregoing ideas can be applied recursively, and no new concepts are involved.

Treatments in the second phase Keep two phases. Apply one set of treatments in the first phase, and another set of treatments in the second phase. There may be interactions between the two sets of treatments.

Everything works. Any confounding between the two sets of treatments is easily discovered. See example in paper.

More than one prime Use a separate design key for each prime. See next slide.

Generalization to more than one prime

Suppose that more than one prime is involved.

If any factor has a composite number of levels, express it as a product of pseudofactors, each with a prime number of levels.

For each prime p_i separately, consider only the treatment (pseudo-)factors and unit (pseudo-)factors which have p_i levels, and make a design key for them, using arithmetic modulo p_i.

Suppose that $p_1, \ldots, p_k$ are among the primes involved, and that, for $i = 1, \ldots, k$, T_i is a linear combination of treatment factors or pseudofactors with p_i levels.
Then T_i belongs to an effect defined by a subset $\mathcal{S}_i$ of the initial letters of the genuine treatment factors.
It can be shown that the $\prod_{i=1}^{k}(p_i - 1)$ df for the interaction between $T_1, \ldots, T_k$ all belong to the effect defined by the subset $\mathcal{S}_1 \cup \cdots \cup \mathcal{S}_k$. (You don't get the cancellation that can happen with a single prime.)

The same result holds for factors on the experimental units.

Summary

- For simple cases with a single prime, the design key is just another method of getting designs that you know how to get anyway.
- The design key gives an algorithm for construction and an algorithm for identifying confounding. In particular, you can check the confounding of proposed designs without having to construct them in their entirety.
- These advantages become more evident as the number of factors increases, as the number of phases increases, as the number of primes increases.
- For advice on choosing a design key for a single phase, see Patterson and Bailey (1978), Kobilinsky (1985), Cheng and Tsai (2013), Cheng's book (2014), Kobilinsky, Bouvier and Monod (2015), Kobilinsky, Monod and Bailey (2016).