Principles in the design of multiphase experiments with a later laboratory phase: orthogonal designs

Chris Brien[1], Bronwyn Harch[2], Ray Correll[2] & Rosemary Bailey[3]

[1]University of South Australia, [2]CSIRO Mathematics, Informatics & Statistics, [3]Queen Mary University of London

http://chris.brien.name/multitier

Chris.brien@unisa.edu.au

Outline

1. Primary experimental design principles
2. Factor-allocation description for standard designs.
3. Principles for simple multiphase experiments.
4. Principles leading to complications, even with orthogonality.
5. Summary

1) Primary experimental design principles

- **Principle 1** (Evaluate designs with skeleton ANOVA tables)
  - Use whether or not data to be analyzed by ANOVA.
- **Principle 2** (Fundamentals): Use randomization, replication and blocking or local control.
- **Principle 3** (Minimize variance): Block entities to form new entities, within new entities being more homogeneous; assign treatments to least variable entity-type.
- **Principle 4** (Split units): confound some treatment sources with more variable sources if some treatment factors:
  i. require larger units than others,
  ii. are expected to have a larger effect, or
  iii. are of less interest than others.

A standard athlete training example

- 9 training conditions — combinations of 3 surfaces and 3 intensities of training — to be investigated.
- Assume the prime interest is in surface differences
  - intensities are only included to observe the surfaces over a range of intensities.
- Testing is to be conducted over 4 Months:
  - In each month, 3 endurance athletes are to be recruited.
  - Each athlete will undergo 3 tests, separated by 7 days, under 3 different training conditions.
- On completion of each test, the heart rate of the athlete will be measured.
- Randomize 3 intensities to 3 athletes in a month and 3 surfaces to 3 tests in an athlete.
  - A split-unit design, employing Principles 2, 3 and 4(iii).

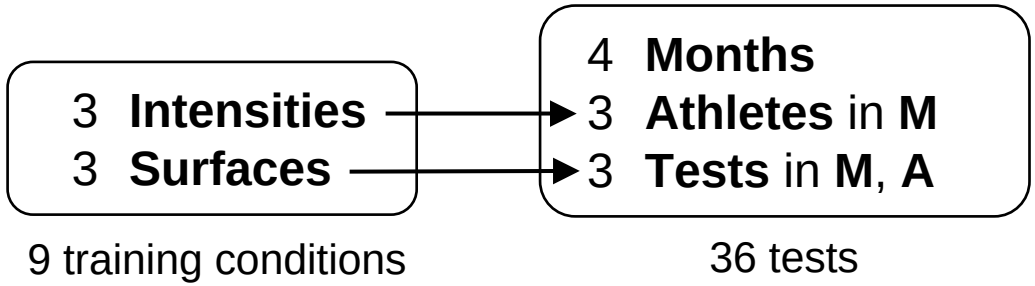
2) Factor-allocation description for standard designs

(Nelder, 1965; Brien, 1983; Brien & Bailey, 2006)

- Standard designs involve a single allocation in which a set of treatments is assigned to a set of units:
  - treatments are whatever are allocated;
  - units are what treatments are allocated to;
  - treatments and units each referred to as a set of objects;
- Often do by randomization using a permutation of the units.
  - More generally treatments are allocated to units e.g. using a spatial design or systematically
- Each set of objects is indexed by a set of factors:
  - Unit or unallocated factors (indexing units);
  - Treatment or allocated factors (indexing treatments).
- Represent the allocation using factor-allocation diagrams that have a **panel** for each set of objects with:
  - a list of the factors; their numbers of levels; their nesting relationships.

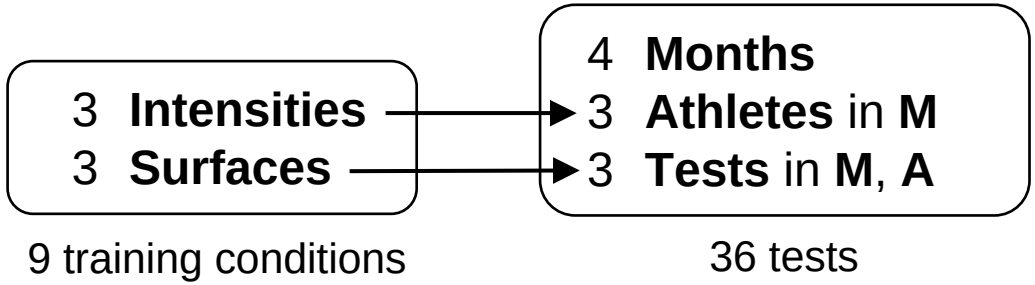
Factor-allocation diagram for the standard athlete training experiment

- One allocation (randomization):
  - a set of training conditions to a set of tests.



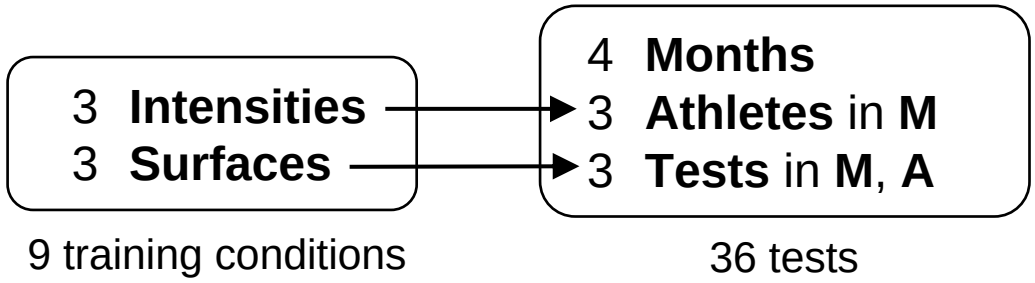
- The set of factors belonging to a set of objects forms a **tier**:
  - they have the same status in the allocation (randomization):
  - {Intensities, Surfaces} or {Months, Athletes, Tests}
  - Textbook experiments are two-tiered.
- A crucial feature is that diagram automatically shows EU and restrictions on randomization/allocation.

Some derived items



- Sets of generalized factors (terms in the mixed model):
  - Months, Months∧Athletes, Months∧Athletes∧Tests;
  - Intensities, Surfaces, Intensities∧Surfaces.
- Corresponding types of entities (groupings of objects):
  - month, athlete, test (last two are Eus);
  - intensity, surface, training condition (intensity-surface combination).
- Corresponding sources (in an ANOVA):
  - Months, Athletes[M], Tests[M∧A];
  - Intensities, Surfaces, Intensities#Surfaces.

Skeleton ANOVA



tests tier		training conditions tier		E[MSq]		
source	df	source	df	σ^2_{MAT}	σ^2_{MA}	σ^2_{M}
Months	3			1	3	9
Athletes[M]	8	Intensities	2	1	3	$q_I(\mu)$
		Residual	6	1	3	
Tests[M∧A]	24	Surfaces	2	1		$q_S(\mu)$
		I#S	4	1		$q_{IS}(\mu)$
		Residual	18	1		

- Intensities is confounded with the more-variable Athletes[M] & Surfaces with Tests[M^A].

3) Principles for simple multiphase experiments

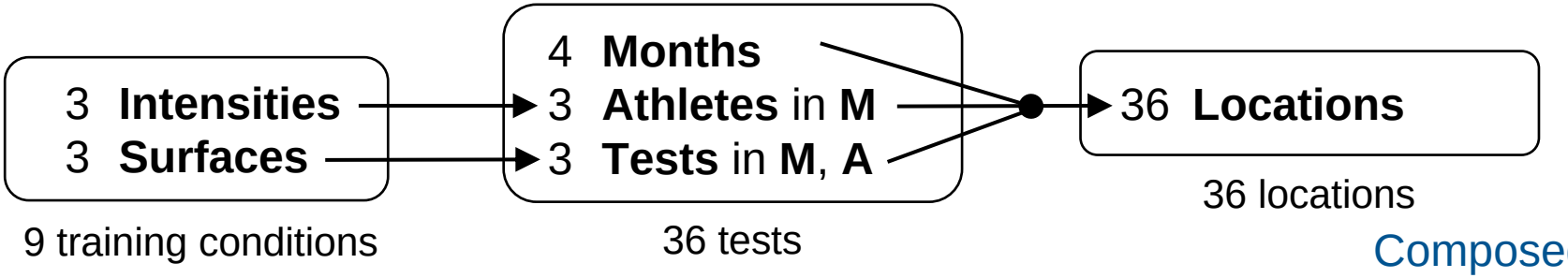
- Suppose in the athlete training experiment:
  - in addition to heart rate taken immediately upon completion of a test,
  - the free haemoglobin is to be measured using blood specimens taken from the athletes after each test, and
  - the specimens are transported to the laboratory for analysis.
- The experiment is two phase: testing and laboratory phases.
  - The outcome of the testing phase is heart rate and a blood specimen.
  - The outcome of the laboratory phase is the free haemoglobin.
- How to process the specimens from the first phase in the laboratory phase?

Some principles

- **Principle 5** (Simplicity desirable): assign first-phase units to laboratory units so that each first-phase source is confounded with a single laboratory source.
  - Use composed randomizations with an orthogonal design.
- **Principle 6** (Preplan all): if possible.
- **Principle 7** (Allocate all and randomize in laboratory): always allocate all treatment and unit factors and randomize first-phase units and lab treatments.
- **Principle 8** (Big with big):
  - Confound big first-phase sources with big laboratory sources, provided no confounding of treatment with first-phase sources.

A simple two-phase athlete training experiment

- Simplest is to randomize specimens from a test to locations (in time or space) during the laboratory phase.



Composed randomizations

locations tier		tests tier		training conditions tier		E[MSq]			
source	df	source	df	source	df	σ^2_L	σ^2_{MAT}	σ^2_{MA}	σ^2_M	
Locations	35	Months	3			**1**	1	3	9	
		Athletes[M]	8	Intensities	2	**1**	1	3		$q_I(\mu)$
				Residual	6	**1**	1	3		
		Tests[M∧A]	24	Surfaces	2	**1**	1			$q_S(\mu)$
				I#S	4	**1**	1			$q_{IS}(\mu)$
				Residual	18	**1**	1			

A simple two-phase athlete training experiment (cont'd)

locations tier		tests tier		training conditions tier		E[MSq]				
source	df	source	df	source	df	σ^2_L	σ^2_{MAT}	σ^2_{MA}	σ^2_M	
Locations	35	Months	3			1	1	3	9	
		Athletes[M]	8	Intensities	2	1	1	3		$q_I(\mu)$
				Residual	6	1	1	3		
		Tests[M∧A]	24	Surfaces	2	1	1			$q_S(\mu)$
				I#S	4	1	1			$q_{IS}(\mu)$
				Residual	18	1	1			

- No. tests = no. locations = 36 and so tests sources exhaust the locations source.
- Cannot separately estimate locations and tests variability, but can estimate their sum.
- But do not want to hold blood specimens for 4 months.

A simple two-phase athlete training experiment (cont'd)

- Simplest is to align lab-phase and first-phase blocking.

Composed randomizations

locations tier		tests tier		training conditions tier		E[MSq]				
source	df	source	df	source	df	σ^2_{BL}	σ^2_{B}	σ^2_{MAT}	σ^2_{MA}	σ^2_{M}	
Batches	**3**	Months	3			1	**9**	1	3	9	
Locations[B]	32	Athletes[M]	8	Intensities	2	1		1	3		$q_{I}(\mu)$
				Residual	6	1		1	3		
		Tests[M∧A]	24	Surfaces	2	1		1			$q_{S}(\mu)$
				I#S	4	1		1			$q_{IS}(\mu)$
				Residual	18	1		1			

- Note Months confounded with Batches (i.e. Big with Big).

The multiphase law

- DF for sources from a previous phase can never be increased as a result of the laboratory-phase design.
- However, it is possible that first-phase sources are split into two or more sources, each with fewer degrees of freedom than the original source.

locations tier		tests tier		training conditions tier		E[MSq]					
source	df	source	df	source	df	σ^2_{BL}	σ^2_{B}	σ^2_{MAT}	σ^2_{MA}	σ^2_{M}	
Batches	3	Months	3			1	9	1	3	9	
Locations[B]	32	Athletes[M]	8	Intensities	2	1		1	3		$q_I(\mu)$
				Residual	6	1		1	3		
		Tests[M∧A]	24	Surfaces	2	1		1			$q_S(\mu)$
				I#S	4	1		1			$q_{IS}(\mu)$
				Residual	18	1		1			

- DF for first phase sources unaffected.

4) Principles leading to complications, even with orthogonality

- **Principle 9** (Use pseudofactors):
  - An elegant way to split sources (as opposed to introducing grouping factors unconnected to real sources of variability).
- **Principle 10** (Compensating across phases):
  - Sometimes, if something is confounded with more variable first-phase source, can confound with less variable lab source.
- **Principle 11** (Laboratory replication):
  - Replicate laboratory analysis of first-phase units if lab variability much greater than 1st-phase variation;
  - Often involves splitting **product** from the first phase into **portions** (e.g. batches of harvested crop, wines, blood specimens into aliquots, drops, lots, samples and fractions).
- **Principle 12** (Laboratory treatments):
  - Sometimes treatments are introduced in the laboratory phase and this involves extra randomization.

5) Summary

- Have provided 4 standard principles and 8 principles specific to orthogonal, multiphase designs.
- In practice, will be important to have some idea of likely sources of laboratory variation.
- Are laboratory treatments to be incorporated?
- Will laboratory replicates be necessary?

References

- Brien, C. J. (1983). Analysis of variance tables based on experimental structure. *Biometrics,* **39**, 53-59.
- Brien, C.J., and Bailey, R.A. (2006) Multiple randomizations (with discussion). *J. Roy. Statist. Soc., Ser. B*, **68**, 571–609.
- Brien, C.J., Harch, B.D., Correll, R.L. and Bailey, R.A. (2011) Multiphase experiments with laboratory phases subsequent to the initial phase. I. Orthogonal designs. *Journal of Agricultural, Biological and Environmental Statistics*, available online.
- Littell, R. C., G. A. Milliken, et al. (2006). *SAS for Mixed Models*. Cary, N.C., SAS Press.
- Peeling, P., B. Dawson, et al. (2009). Training Surface and Intensity: Inflammation, Hemolysis, and Hepcidin Expression. *Medicine & Science in Sports & Exercise,* **41**, 1138-1145.
- Searle, S. R., G. Casella, et al. (1992). *Variance components*. New York, Wiley.

Web address for link to Multitiered experiments site: http://chris.brien.name/multitier