Supplementary materials for this article are available at 10.1007/s13253-011-0060-z.

Multiphase Experiments with at Least One Later Laboratory Phase. I. Orthogonal Designs

C.J. BRIEN, B.D. HARCH, R.L. CORRELL, and R.A. BAILEY

The paper provides a systematic approach to designing the laboratory phase of a multiphase experiment, taking into account previous phases. General principles are outlined for experiments in which orthogonal designs can be employed. Multiphase experiments occur widely, although their multiphase nature is often not recognized. The need to randomize the material produced from the first phase in the laboratory phase is emphasized. Factor-allocation diagrams are used to depict the randomizations in a design and the use of skeleton analysis-of-variance (ANOVA) tables to evaluate their properties discussed. The methods are illustrated using a scenario and a case study. A basis for categorizing designs is suggested. This article has supplementary material online.

Key Words: Analysis of variance; Experimental design; Laboratory experiments; Multiple randomizations; Multi-phase experiments; Multitiered experiments; Two-phase experiments.

1. INTRODUCTION

It is common for the material produced during an experiment to be processed in a laboratory. Reasons for this include the need to measure chemical and physical attributes using equipment such as spectrometers, gas chromatographs, pH meters or wear and strength testers, or to produce processed products such as wine, bread and malt that are subsequently assessed, often by an expert panel. Such experiments consist of two phases (McIntyre 1955), usually with an experimental design required for each phase. Those agricultural experiments that have a laboratory phase after the field phase are two-phase experiments. Clinical trials can also result in two phases, namely clinical treatment and laboratory phases, when specimens from patients are processed in a laboratory. For some experiments

C.J. Brien (✉) is Adjunct Senior Lecturer in Statistics, School of Mathematics and Statistics, University of South Australia, Mawson Lakes Boulevard, Mawson Lakes, SA 5095, Australia (E-mail: *chris.brien@unisa.edu.au*); B.D. Harch is Deputy Director, Ecosciences Precinct, CSIRO Sustainable Agriculture Flagship, GPO Box 2583, Brisbane, Qld 4001, Australia; R.L. Correll is Director, Rho Environmentrics, PO Box 366, Highgate, SA 2063, Australia; R.A. Bailey is Professor of Statistics, School of Mathematical Sciences, Queen Mary University of London, Mile End Road, London E1 4NS, UK.

Journal of Agricultural, Biological, and Environmental Statistics, Volume 16, Number 3, Pages 422–450
DOI: 10.1007/s13253-011-0060-z

both phases occur in the laboratory, such as in food processing when there is a phase in which mixtures are prepared, and a processing phase to produce the final product. More generally experiments may be multiphase. In this paper, the laboratory phase is to be interpreted broadly as a phase in which further processing, measurement, testing and so on are performed, even if, strictly speaking, a laboratory is not involved.

Two-phase experiments were first described by McIntyre (1955), although he considered only designs whose analysis, while performed on second-phase means, is determined by the first-phase design. The crucial feature that McIntyre incorporated into his designs was the use of a randomization in each phase. Cox (1958, p. 83) also pointed out that 'It is frequently not good enough to randomize just one stage [phase] of the experimental procedure and to leave the treatments systematically arrayed at other stages [phases]'. In spite of this early work and even though the use of statistical design principles in the first phase is well-appreciated, the need to employ these principles in laboratory phases is often overlooked. The common practice has been to process in a systematic order, for example to process field produce in 'field order' or, even worse, all samples for each treatment together in the laboratory, or to not consider laboratory processing order at all. More recently Brien (1983) has classified two-phase experiments as being multitiered and Brien and Bailey (2006) have characterized them as involving multiple randomizations. Several authors have recognized the multiphase nature of their experiments. McIntyre (1955) described an experiment with a treatment phase, in which treatments were applied to plant leaves, and an assay phase, in which dilutions from the leaves were applied to assay plants. Brien, May, and Mayo (1987) gave examples of two-phase sensory evaluation experiments. Brien, Harch, and Correll (1998) presented a discussion of the design of and ANOVA for the experiment described in Section 8. Smith et al. (2001) discussed the design and analysis of wheat experiments that involved both field and milling phases. Cullis et al. (2003) described the design and analysis of a three-phase experiment that involved a field phase in which barley lines were grown, a malting phase in which barley malts were produced and a measurement phase in which several traits were assessed. Smith, Lim, and Cullis (2006) developed p/q-rep designs for such experiments. Kerr (2003) and Jarrett and Ruggiero (2008) noted that a microarray experiment can be the measurement phase of a two-phase experiment. Brien and Bailey (2006, Examples 1, 4, 9, 12, 14 and 15, and Figure 7) gave examples involving a first phase followed by a laboratory phase. Clearly, multiphase experiments with a later laboratory phase occur widely.

The purpose of this paper is to provide general principles for designing experiments of this class and so increase awareness of the need to employ design principles in all phases of an experiment. In Section 2, a scenario is introduced, from which different examples are derived to illustrate the principles. Section 3 reviews the Brien and Bailey (2006) approach and recaps the design principles for single-randomization experiments. In Section 4 the nature of multiphase experiments is examined and Section 5 presents a simple two-phase experiment. The attributes of a laboratory-phase design are studied in Section 6. Section 7 outlines some complications and Section 8 applies the principles to a case study. Section 9 summarizes the key characteristics of multiphase experiments, Web Appendix A the principles developed in the paper and Web Appendix B provides definitions of italicized terms. Mixed-model analyses for the examples are sketched in Web Appendix E.

2. THE SCENARIO: ATHLETE TRAINING

The scenario involves research into athlete training and is loosely based on a study reported by Peeling et al. (2009). The effect of training conditions on heart rate in endurance athletes is to be investigated. Twelve athletes are to be recruited and each will undergo three tests, separated by seven days, under different training conditions. On completion of each test, the heart rate of the athlete will be measured.

Experiments for the scenario described so far would employ well-established design principles that are reviewed in Section 3. However, some examples based on it will have blood specimens taken from the athletes for subsequent analysis in the laboratory. This gives rise to questions about the processing order of the specimens in this laboratory phase. Should they be done in treatment order? the order collected? some other order? That is, in addition to a design for testing the athletes, one for the laboratory phase is needed. How do these two designs interrelate?

3. STANDARD DESIGNS

A *standard design* is defined to be the result obtained from allocating a set of *treatments* to a set of *units*. This definition covers virtually all the textbook designs. Often the allocation is at random, in which case it is said to involve a single randomization because it is achievable with a single permutation of the units (Brien and Bailey 2006). The allocation may also be systematic or with other special designs such as spatial designs.

Each unit is an *observational unit*, the unit from which a single value of a response variable is obtained. A *treatment* is a, perhaps conceptual, object that is allocated to one or more units. For convenience we use *object* to refer to either a treatment or a unit. Each set of objects is indexed by a set of factors, termed a *tier* (Brien 1983; Brien and Bailey 2006). The factors indexing the treatments are referred to as the *treatment factors* (or *treatment tier*); they are the factors that are allocated. The factors indexing the observational units are called the *unit factors* (or *unit tier*); they are factors that have another set of factors allocated to them and are sometimes referred to as the block factors (Nelder 1965).

3.1. FACTOR-ALLOCATION DESCRIPTION AND EVALUATION OF AN EXPERIMENT

In this paper, as in Brien and Bailey (2006), an experiment is described in terms of the allocation of multiple sets of objects, along with their associated tiers and the nesting and crossing relations among the factors within a tier. Here, it is termed *factor-allocation description*. This information can be displayed in a *factor-allocation diagram*, an extension of randomization diagrams (Brien and Bailey 2006). It has a panel for each set of objects so that for standard designs it has two panels, one for treatments and the other for units. Each *panel* contains the tier of factors for its objects, and the nesting between the tier's factors is specified; factor crossing is implicit. There are lines and arrows between panels showing how the treatment factors are allocated to the unit factors, these being solid if randomization is employed and dashed if it is systematic. It may also be necessary to add,

between panels, pseudofactors to be used in the allocation. Web Appendix C describes the conventions for such diagrams.

To evaluate designs, after Brien and Bailey (2009), the following principle is utilized in this paper.

Principle 1 (Evaluate designs with skeleton ANOVA tables). Whenever possible, formulate the skeleton ANOVA table using the factor-allocation diagram for an experiment, irrespective of whether its data is to be analysed by ANOVA.

Skeleton ANOVA tables consist of just sources, degrees of freedom and, if applicable, efficiency factors. Optionally, they also include the expected mean squares (E.M.S.). The decomposition tables of Brien and Bailey (2009) are precursors to them, in that they have not had genuine factors substituted for pseudofactors in the sources and do not include the E.M.S. Either table, when based on the factor-allocation diagram, shows the confounding of *treatment sources* with the *unit sources* that results from the design on which it is based. Hence, they are valuable for evaluating designs.

The Brien and Bailey (2009) method starts with a factor-allocation diagram. Then the *set of generalized factors* is derived for each set of objects: it consists of all subsets of the factors within a panel, except that nested factors never occur without the factors that nest them. So, each *generalized factor* is comprised of a (sub)set of factors in a tier and groups the objects; each group is referred to as an *entity* and the type of grouping as the *entity-type* for that generalized factor. We use the notation $F_1 \wedge \cdots \wedge F_n$ to denote the generalized factor whose levels are the combinations of levels of F_1, F_2, ... and F_n, for $n \geq 1$ (Brien and Bailey 2009). For the unit factors, the smallest entity-type is the (observational) unit; its generalized factor consists of all the unit factors and each unit is associated with a unique level of that generalized factor. The other unit generalized factors each define another entity-type and identify one of the ways that the observational units are grouped together.

Next, obtain the labels for *sources* to identify the interaction or nested effects associated with the generalized factors for a panel. We label sources with the notation given in Brien and Demétrio (2009, Table 1). That is, A#B denotes the interaction of A and B, and C [A $\wedge$ B] denotes the differences between C nested within the combinations of levels of A and B. The label for the source for each generalized (pseudo)factor is derived as follows: of the *original factors* in the generalized (pseudo)factor, only those that nest any of the other factors must be in the square brackets joined by '$\wedge$'; the rest are put to the left of the square brackets joined by '#'. Lastly, the label for each source is entered into the ANOVA table, in the column for the tier from which it originates and in rows for the sources from other tiers with which it is confounded. The E.M.S., if required, are obtained using the rules given in Web Appendix D.

Example 1 (A standard athlete training experiment). Suppose that in the scenario (Section 2), three training conditions are to be investigated. Also, the 12 athletes are to be divided into four lots and each lot will undergo the heart-rate testing in a different month. Further, it is proposed that the three training conditions will be randomized to the three

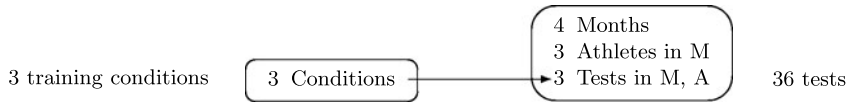

Figure 1. Factor-allocation diagram for the standard athlete training experiment: training conditions are randomized to tests; M = Months; A = Athletes.

Table 1. Skeleton ANOVA table for the standard athlete training experiment[a].

(i) Factor-allocation								(ii) Single-set	
tests tier		training-conditions tier		E.M.S.[b]					
Source	d.f.	Source	d.f.	σ^2_{MAT}	σ^2_{MA}	σ^2_{M}	$q(\cdot)$	source	d.f.
Mean	1	Mean	1					Mean	1
Months	3			1	3	9		Months	3
Athletes [M]	8			1	3			Athletes [M]	8
Tests [M ∧ A]	24	Conditions	2	1			q(C)	Conditions	2
		Residual	22	1				Error	22

[a]The names of sources in skeleton ANOVA tables follow the convention of using only the first letter of a factor, except that its full name is used in the source with the fewest factors, out of all sources in which the factor occurs.
[b]Each σ^2 is a variance component whose subscript is comprised of the first letter of each factor in its term and the numbers in the table are the coefficients of the variance components in the E.M.S. The function q(C) is the same quadratic function of the expectation as is the Conditions mean square.

tests for each athlete. The two sets of objects in this design are the set of training conditions and the set of tests, a test being an (observational) unit. The factor-allocation diagram, in Figure 1, shows the two tiers of factors indexing training conditions and tests. The set of generalized factors for tests is {Months, Months ∧ Athletes, Months ∧ Athletes ∧ Tests}. The entity-types are test, athlete and month. The tests are uniquely indexed by the levels of Months ∧ Athletes ∧ Tests. The athletes are indexed by Months ∧ Athletes; each athlete collects together three tests. The months are indexed by Months; each month collects together nine tests. It is envisaged that athletes will be more variable than tests. The factor-allocation diagram also shows that Conditions are assigned to the Tests within each Athlete within each Month.

Table 1(i) contains the skeleton ANOVA table for this design derived from the factor-allocation diagram. In this table Conditions, from the training-conditions tier, is in the same subtable as Tests [Months ∧ Athletes], from the tests tier. This shows that Conditions is confounded with Tests [Months ∧ Athletes].

3.2. Single-Set Description and Evaluation of an Experiment

Experiments are also commonly described using what can be termed *single-set description* (see for example Searle, Casella, and McCulloch 1992, Section 4.1a). One identifies the smallest set of factors, including the factors of interest to the researcher, that is sufficient to uniquely index the units in the experiment. This may necessitate the incorporation

of a single synthetic factor that is nested within the other factors and whose levels differ for the repeats of the combinations of the other factors. Generally, the factors identified in this approach are a subset of those from the factor-allocation approach, the subset being sufficient to uniquely index the units. Usually unit factors that are 'equivalent' to treatment factors are omitted, this being feasible because it is *impossible* to observe all combinations of these factors.

In a refinement of the single-set approach, Littell et al. (2006) split the single set of factors into treatment and experiment design factors, neither of which uniquely indexes the units. They also recommend the identification of experimental units (EUs). Restating their definition in our terminology, an *experimental unit* is the smallest entity-type to which a treatment generalized factor is independently assigned. They describe EUs using both treatment and experiment design factors. On the other hand, a factor-allocation diagram automatically encapsulates these EUs and identifies them using only unit (experiment design) factors. The entity-type corresponding to the combination of unit factors to which an arrow points is such an EU. Also, the restrictions on randomization are reflected in the factor nesting and crossing relations in the diagram.

In the single-set approach, the identity of the EU is obscured, ANOVA tables do not exhibit the confounding in the experiment and the origins of sources of error are not always obvious. These difficulties becomes more acute when there are multiple randomizations involved. For further discussion see Brien and Demétrio (2009, Section 8) and Brien and Bailey (2010, Section 6).

For Example 1, the single-set approach yields the single set of factors {Months, Athletes, Conditions} and these factors uniquely index the units in the experiment. Conditions is a treatment factor while Months and Athletes are experiment design factors. The crucial aspect of this description, compared to the factor-allocation description, is that Tests is not included. Here, the EUs are defined by the levels of Months $\wedge$ Athletes $\wedge$ Conditions. However, it makes no sense to talk of randomizing Conditions to these levels: they are randomized to the entities that they index. That is, the levels of Months $\wedge$ Athletes $\wedge$ Conditions do not specifically identify the EUs, but merely act as a proxy for them. On the other hand, it is clear from the factor-allocation diagram in Figure 1 that an EU is the test, the same as the observational unit, and that the tests are indexed by the combinations of Months, Athletes and Tests.

Table 1(ii) contains a skeleton ANOVA table derived from the single-set approach. The E.M.S. for it are the same as for the equivalent lines in Table 1(i). Table 1(ii) does not exhibit the confounding in the experiment and so masks the origin of the Error source. One is likely to surmise that it stems from the variation in Conditions differences between Athletes. However, Table 1(i) shows that it comes from the variability of Tests within Athletes.

3.3. PRIMARY EXPERIMENTAL DESIGN PRINCIPLES

The fundamental principles of experimental design, espoused by Fisher and employed in Example 1, are embodied in the following principle (Cox 2009):

Principle 2 (Fundamentals). A good experimental design employs: *replication* to provide a measure of random error and sufficient to achieve adequate precision; *randomization* to avoid systematic effects and other biases; and, where appropriate, *blocking* (or local control) to reduce variation among experimental units.

Next, we focus on blocking in the following principle (Cox 1958, Chapter 7):

Principle 3 (Minimize variance). Block the entities of an entity-type on the units into groups, to form a new entity-type, if it seems that the entities within the new entity-type will be more homogeneous than if they were ungrouped; assign treatments to the least variable entity-type so that the contribution of other entity-types to the variance of the estimates of treatment effects is reduced as far as is possible.

Another view of blocking is that a set of unit factors is identified and their nesting and crossing relationships considered. The relationships are based on the physical setup of the experiment and on what are anticipated to be the substantial sources of variation in the experiment (Brien and Bailey 2006, Section 2.2). As outlined above, the relationships determine the set of generalized factors derived from the unit factors. Next, one determines the manner in which treatment factors are to be assigned to unit factors, so that treatment sources are confounded with unit sources such that Principle 3 (Minimize variance) is achieved. To illustrate, take a rectangular grid of plots, indexed by Rows and Columns; these unit factors are inherently crossed. However, they should only be designated as crossed if substantial row and column differences are envisaged; differences in just the rows direction would result in Columns nested within Rows. For the latter, the design that is blocked in accord with Principle 3 (Minimize variance) is a design with *hierarchical unit factors*: the unit factors are nested one within another, such as in a randomized complete-block design (RCBD). The former situation requires a design with *nonhierarchical unit factors*, such as a Latin square design. For both these designs, Treatments are assigned to Rows $\wedge$ Columns, but in different ways so that the confounding of Treatments is not the same: with Columns [Rows] and Rows # Columns, respectively.

Example 1 has hierarchical unit factors, with Tests nested within Athletes; this implies no order effect in testing an athlete, which may well be justified given the seven days between tests. Otherwise, a Latin square design in each month could be employed as a design with nonhierarchical unit factors. The cost of this is that the degrees of freedom for the Residual would be reduced to 14, which hopefully would be compensated for by a reduced Residual mean square.

The implication of Principle 3 (Minimize variance) is that the general aim should be to have treatment sources estimated solely from, or *confounded* with, the source associated with the smallest entity-type, unit, as these entities are anticipated to have smallest variation. A side-effect is that this usually maximizes the degrees of freedom of the variance for estimated treatment effects, and hence the precision of the estimated variance as well. The proviso 'as far as is possible' is needed for Principle 3 (Minimize variance) because it is not always possible to assign the treatments such that treatment sources are confounded

with just one unit source. As a consequence a nonorthogonal design might need to be employed, in which case minimizing the variance for estimated treatment effects is likely to be a play-off between the amount of information confounded with the smallest entity-type and the size of blocks. However, nonorthogonal design is outside the scope of the present paper.

Example 1 has Conditions confounded with Tests [Months ∧ Athletes]. So, it is confounded with the source for the smallest entity-type, test. In this case, Principle 3 (Minimize variance) is satisfied by grouping the tests according to Athletes within Months and choosing a design in which Conditions is free of differences between Athletes.

3.4. SPLIT-UNIT PRINCIPLE

In factorial experiments, which have more than one treatment factor, another possibility is to use the split-unit (also split-plot) principle and confound different treatment sources with the sources for different unit generalized factors. The principle employed in this is as follows:

Principle 4 (Split-units). Confound some treatment sources with unit sources for which greater variation is expected 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.

Situation (i) applies in agricultural experiments when some factors, such as irrigation treatments, must be applied to larger units and in industrial experiments when the levels of some treatment factors, such as process temperature, are difficult to set. The first phase of Example 2 illustrates situation (iii).

4. MULTIPHASE EXPERIMENTS

McIntyre (1955) originally used the term 'two-phase' for experiments in which there is a single randomization in each phase; these are referred to as *normal*. The object of the second phase is to evaluate the material produced in the first phase and a response variable is measured at the end of the second phase. More generally, *multiphase experiments* are possible in which there is a phase for each set of units that produces an outcome. The outcome can be material for processing in the next phase, or values for response variables, or both. The *phase* is the period of time during which a set of units are engaged in producing their outcome. Only the final phase need have a response variable. Also, one phase might overlap another phase, as in Example 2.

The two randomizations in normal two-phase experiments form a *chain* (Brien and Bailey 2009) and are either *composed* or *randomized inclusive* (Brien and Bailey 2006, Figure 7 and Example 9). However, the number of randomizations in two-phase experiments varies. There is always be least one: the randomization of material from the first phase in the laboratory phase. When the first phase is an observational study, then this is the only randomization. For example, suppose that tissue is taken from animals that differ in some characteristic, such as sex or genetic make-up, for which the expression of the character

is predetermined for each animal. That is, the outcome of this phase is tissue specimens. These are then subject to a microarray analysis, with specimens randomized to arrays. The outcome of the microarray phase is an intensity measurement. More than two randomizations are also possible. One case is that the first phase involves multiple randomizations: for example, *independent randomizations* in a plant phase (Brien and Bailey 2006, Example 5), composed randomizations in a grazing phase (Brien and Bailey 2006, Example 3) or *unrandomized-inclusive randomizations* in a superimposed phase (Brien and Bailey 2006, Example 10). Another possibility is that the second phase involves more than one randomization, such as when treatments are introduced in the laboratory phase. These will be *two-to-one randomizations* (Brien and Bailey 2010). In general, the number of tiers, and hence panels, in a factor-allocation diagram is related to the number of randomizations.

From here on we concentrate on normal two-phase experiments in which the second phase is a laboratory phase, although sometimes experiments have the extra randomization of laboratory treatments. The extension to other multiphase experiments is straightforward.

5. KEEPING IT SIMPLE

In the spirit of keeping it simple, the following principle is proposed.

Principle 5 (Simplicity desirable). Whenever possible, in choosing a design to assign first-phase units to laboratory units, randomize first-phase unit factors that have treatments assigned to them so that sources associated with these factors are confounded with a single laboratory-unit source.

This is advocating composed randomizations, with an orthogonal laboratory design. Simplicity is achieved in that each of two composed randomizations can be done ignoring the other, because the result of one is not needed to do the other. Consequently, the randomizations can be done in either order. Even more importantly, there is no degradation of the properties of first-phase sources with such designs. Also, in this section, we confine our attention to experiments that do not use laboratory replications and treatments, to avoid the complications that come with them (Sections 7.3 and 7.4). With no replication of the first-phase units in a laboratory phase, the number of first-phase and laboratory units must be equal and so separate Residuals for laboratory sources do not exist. It is said that each laboratory source is *exhausted* by the first-phase sources. One consequence of this is that some variance components may not be estimable. In particular, only the sum of the two variance components for the variation arising from first-phase and laboratory units is estimable. If, in addition, the first-phase unit sources are not split using pseudofactors, then the full decomposition for the experiment is equivalent to the decomposition for the first phase.

The simplest laboratory phase has first-phase units completely randomized to laboratory units. Another uncomplicated type is when the blocking in the laboratory phase conforms to the blocking in the first phase: for example, an RCBD in both phases with blocks and

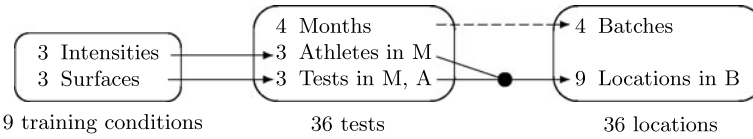

Figure 2. Factor-allocation diagram for the simple two-phase athlete training experiment: training conditions are randomized to tests and tests are allocated to locations; the '●' indicates that the combinations of the levels of Athletes and Tests are randomized to the Locations; the dashed arrow indicates that Months are systematically allocated to Batches; M = Months; A = Athletes; B = Batches.

units from the first phase randomized to blocks, of the same size, and units from the laboratory phase, respectively. Example 2 is of this type and Brien and Demétrio (2009) give a multiphase example.

Example 2 (A simple two-phase athlete training experiment). Suppose that in the scenario (Section 2) nine training conditions are to be investigated and these are the combinations of three surfaces and three intensities of training. Also, assume that the prime interest is in surface differences, with intensities included to observe the surfaces over a range of intensities. Further, 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 transported to the laboratory for analysis. The experiment involves a testing and a laboratory phase, with the product of the first phase being the blood specimen. As the specimens become available monthly, the batch of specimens for one month are to be processed, in a random order, before those for the next month are available.

In designing the first phase, part (iii) of Principle 4 (Split-units) is invoked and a split-unit design used, with (a) Intensities randomized to Athletes within Months and (b) Surfaces randomized to Tests within Athletes and Months. That is, for the single-set approach, the EUs in this randomization are athletes and tests. Its factor-allocation diagram is given in the two left panels of Figure 2. Its skeleton ANOVA table, which will be the basis for the analysis of heart rate, is obtained from Table 2 by deleting the locations tier and its variance components. This shows that the design is orthogonal, each training-conditions source being confounded with just one tests source. Intensities is confounded with the potentially more variable Athletes [Months]; Surfaces and Intensities # Surfaces are confounded with Tests [Months $\wedge$ Athletes].

For the laboratory phase, a second standard design is needed for allocating the tests to locations. Assume that there is nothing to suggest that the between-locations variation will be reduced by processing less than nine specimens as an entity. Hence the application of Principle 3 (Minimize variance) suggests that Batches is the only basis for grouping locations in the laboratory analysis. The factor-allocation diagram for this standard design, which is also orthogonal, is given in the two right panels of Figure 2. Note the different units for each phase: tests and locations. In this case Months are systematically assigned to Batches because the order of processing the months is the same order in which they are produced. This design conforms to Principle 5 (Simplicity desirable) because, in the allocation of tests to locations, the Athletes and Tests are randomized only to Locations so that their sources are confounded with just the Locations [Batches] source. Hence, in

Table 2. Skeleton ANOVA table for the simple two-phase athlete training experiment.

locations tier		tests tier		training-conditions tier		E.M.S.[a]					
Source	d.f.	Source	d.f.	Source	d.f.	σ^2_{BL}	σ^2_{B}	σ^2_{MAT}	σ^2_{MA}	σ^2_{M}	$q(\cdot)$
Mean	1	Mean	1	Mean	1						
Batches	3	Months	3			1	9	1	3	9	
Locations [B]	32	Athletes [M]	8	Intensities	2	1		1	3		$q(\mathrm{I})$
				Residual	6	1		1	3		
		Tests [M ∧ A]	24	Surfaces	2	1		1			$q(\mathrm{S})$
				I#S	4	1		1			$q(\mathrm{IS})$
				Residual	18	1		1			

[a]Each σ^2 is a variance component whose subscript is comprised of the first letter of each factor in its term and the numbers in the table are the coefficients of the variance components in the E.M.S. The q-functions are the same quadratic functions of the expectation as are the corresponding mean squares.

allocating tests to locations, the randomization of training conditions factors to tests can be ignored and the design involves two composed randomizations, which are in a chain. The single-set-approach EUs are batch and location.

Table 2 gives the skeleton ANOVA table for the combined phases, which will be the basis for the analysis of free haemoglobin. The most important impact of the laboratory phase is that it adds extra sources of variability for Batches and Batches ∧ Locations. The value of the skeleton ANOVA table is that it shows that, as the numbers of tests and locations are equal and whole tests sources are confounded with single locations sources, the two decompositions are equivalent, as predicted. The result is that Batches and Locations [Batches] are exhausted by tests sources. The variance components for Batches, Months, Batches ∧ Locations and Months ∧ Athletes ∧ Tests are not estimable, but the sum of the first pair and that of the second pair are.

6. THE LABORATORY-PHASE DESIGN

A normal two-phase experiment, like Example 2, involves a first phase in which treatments are allocated to units using a *first-phase design* and a second phase in which first-phase units are allocated to second-phase units using a *second-phase design*. Thus, the combined *two-phase design* for a normal two-phase experiment is comprised of the equivalent of at least two standard designs and there are three sets of objects. When the second phase is a laboratory phase, the second-phase design is referred to as the *laboratory-phase design* and the three sets of objects as (i) first-phase treatments, (ii) first-phase units, and (iii) laboratory units. Now there are two types of units; the laboratory units are the observational units. There are three tiers, or sets of factors, each indexing one of these sets of objects. Additional phases and laboratory treatments add additional sets of objects and associated tiers; the observational units are the units for the last phase.

A second principle that is specific to multiphase experiments is

Principle 6 (Preplan all). If possible, plan all phases of an experiment before commencing it.

This principle is needed because there are situations in which limitations in the laboratory phase need to be taken into account. In other cases, it is not possible to apply this principle. In Example 2, although both phases are being planned together, this is not crucial.

Concentrating on the laboratory phase, ultimately its design is in many ways the same as for standard designs: at least one set of objects is to be allocated to another set of objects. Unsurprisingly then, the Principles outlined in Section 3 remain applicable. However, there are differences in their application that are now explored. In particular, while Principle 2 (Fundamentals) applies to the second-phase design, there are different emphases. For example, replication of the factors being allocated is not mandatory (Section 7.3). For instance, they are not replicated in Example 2.

With respect to randomization, a principal tenet of this paper is that there should be a *laboratory randomization*, to avoid systematic trends and other biases clouding effects of interest. That is, wherever possible, first-phase units should be randomized to laboratory units, and any laboratory treatments randomized too. An alternative is to process in order of first-phase units ('field order'), but this relies on the first-phase blocking being appropriate for the laboratory phase and misses the opportunity for further randomization to increase the robustness of inferences about treatments or compensate for a poor first-phase randomization. However, laboratory-phase designs differ from first-phase designs.

Firstly, there are generally more factors, the first-phase unit factors, to be allocated to units in the laboratory phase than in a first phase. So, a more complicated design, often with pseudofactors, is likely to be needed (Section 7.1).

Secondly, except when the first phase is an observational study, an essential difference is that there are at least two tiers to be allocated in the laboratory phase and, as noted for single-set description, it is impossible to observe all combinations of the levels of their factors. Using the single-set approach, one would take the subset that uniquely indexes the first-phase units and ignore some first-phase unit factors; this is not because the ignored factors are without effect. For instance, in Example 2, it is impossible to observe all 36×9 combinations of Months, Athletes, Tests, Intensities and Surfaces. However, the observations are uniquely indexed by just Months, Intensities and Surfaces. Consequently, to simplify the laboratory design, just these factors could be allocated and Athletes and Tests simply ignored. The difficulty with this is that one can easily lose track of how sources of variation in the first phase affect the response (see Brien and Bailey 2010, Section 6). To avoid this, all first-phase unit factors must be allocated in the laboratory phase, as expressed in the following principle.

Principle 7 (Allocate all and randomize in laboratory). The laboratory-phase design should *always* allocate *all* the first-phase unit factors, as well as any laboratory treatments, to the laboratory units, using randomization wherever possible.

Generally, Principles 3 (Minimize variance) and 4 (Split-units) apply in carrying out this latest principle. Consequently, the designer will look to confound the first-phase unit sources that have treatments confounded with them, with the smallest sources of laboratory variation.

Example 2 conforms to Principle 7 (Allocate all and randomize in laboratory) with the combinations of Months, Athletes and Tests randomized to the laboratory units. Also, while a split-plot design is used in the first phase, Principle 3 (Minimize variance) is satisfied without needing to invoke Principle 4 (Split-units) in the laboratory phase. The design for the case study (Section 8) does not conform to Principle 7 (Allocate all and randomize in laboratory).

A third difference is that, whenever blocking has been employed in the first phase, there are unit factors that are purely nuisance factors to be randomized to the laboratory units. To cover this situation, part (ii) of Principle 4 (Split-unit) is extended to cover unit sources as follows:

Principle 8 (Big with big). Confound big first-phase unit sources that have no treatment sources confounded with them, with potentially big second-phase unit sources.

This principle complements Principle 3 (Minimize variance). It differs from part (iii) of Principle 4 (Split-unit) because the precision of a factor of interest is not being sacrificed to gain precision for other factors of greater interest. It means, for example, that block main effects from the two phases should be confounded, provided there is no treatment source confounded with them. In Example 2, Months is a first-phase nuisance factor and it is allocated to Batches, which satisfies Principle 8 (Big with big).

A final feature of laboratory-phase designs is that, in cases like the case study (Section 8), there are also laboratory treatments to be randomized, as advocated in Principle 7 (Allocate all and randomize in laboratory). This increases the number of tiers and randomizations.

The third aspect of Principle 2 (Fundamentals) is blocking. Laboratory units are often the times at which an analysis is performed, or positions in a machine each time a set of specimens are processed together. They may be considered, generically, as *locations* but, for convenience, a contextually appropriate name will be used. A commonly occurring source of heterogeneity, to guard against in designing a laboratory phase, is smooth or nonsmooth trend across locations such as results from equipment drift. Blocking locations to minimize the variation affecting treatments, as embodied in Principle 3 (Minimize variance), involves forming groups of homogeneous locations. An obvious way to do this, in time, is to form laboratory blocks from consecutive times or time periods. But what then are the relationships between the factors indexing the blocks and those indexing the locations? While, frequently, the locations factors are treated as nested in the blocks factors, more often they are inherently crossed: the first locations in all of the blocks share the property that they are in the same relative locations in all blocks. Hence, a design with nonhierarchical unit factors, in which blocks and locations are crossed, would seem to be dictated, and this would result in the elimination of both smooth and nonsmooth trends

over the locations, provided these trends are reasonably consistent across blocks. However, as discussed in Section 3, the factor relationships are not determined solely by the inherent relationships. So, if consistent differences between locations across blocks are not expected, then a design with hierarchical laboratory-unit factors would be appropriate. In Example 2, the laboratory units are locations. The laboratory-unit factors are hierarchical, with Locations nested in Batches. This requires that there be no systematic trend across locations that is consistent between batches.

According to Principle 3 (Minimize variance), the objective is to confound first-phase, treatment sources with the smallest source of laboratory variation possible. It may be necessary to invoke Principle 4 (Split-unit) and confound different treatment sources with different laboratory sources.

In designing all phases of an experiment, except the first, the following important law applies:

Multiphase law 1. The degrees of freedom for sources from a previous phase can never be increased as a result of the design for a subsequent phase. However, it is possible that the design splits a source from a previous phase into two or more sources, each with fewer degrees of freedom than the original source.

In Example 2, the first-phase and combined decompositions are equivalent and so there is no change in the degrees of freedom for the first-phase sources in the combined decomposition.

7. COMPLICATIONS, EVEN WITH ORTHOGONALITY

This section explores several ways in which experiments deviate from the simple situations described in Section 5, even though their designs are orthogonal. Extra principles are developed as needed.

7.1. PSEUDOFACTORS FOR SPLITTING SOURCES BEING RANDOMIZED

Possibly the most common complication with multiphase experiments is that it is useful or necessary to deploy *pseudofactors*, which group together levels of some generalized factor, usually one that is being randomized. The difference between pseudofactors and other factors is that changes in the response variable between the levels of a pseudofactor are ascribed to the genuine generalized factor from which they are formed. As outlined in Brien and Bailey (2006, Section 8.2), there are different ways in which the need for pseudofactors arises. Whatever the way, as Brien and Bailey (2009) describe, the result is that they split the source corresponding to the generalized factor from which they are formed. Pseudofactors may occur with both composed and randomized-inclusive randomizations. At times the use of pseudofactors can be avoided by dropping factors from previous phases, but this contravenes Principle 7 (Allocate all and randomize in laboratory). As a result the following principle is advanced.

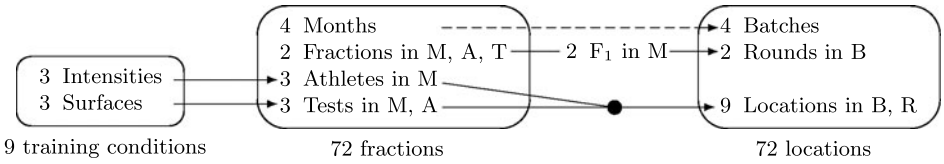

Figure 3. Factor-allocation diagram for the replicated two-phase athlete training experiment: training conditions are randomized to fractions and fractions are allocated to locations; the '●' indicates that the combinations of the levels of Athletes and Tests are randomized to the Locations; the dashed arrow indicates that Months are systematically allocated to Batches; M = Months; A = Athletes; T = Tests; B = Batches; R = Rounds; F_1 is a pseudofactor for Fractions that groups fractions that are to be assigned to the same Rounds level.

Principle 9 (Use pseudofactors). Use pseudofactors to split sources, when necessary, to keep track of all factors in the experiment or to produce structure-balanced designs.

The resultant splitting of sources may result in the degradation of the properties of the first phase, but does not when the split source has no first-phase, treatment source confounded with it. The following example uses pseudofactors with composed randomizations without degradation of properties. A simpler example of this is Example 4 of Brien and Bailey (2006), which is analysed by Bailey and Brien (2011, Example 2). Web Appendix Example 1 (Web Appendix F), in which pseudofactors are used with randomized-inclusive randomizations, displays some degradation.

The use of pseudofactors for structure-balanced designs (Brien and Bailey 2009) is included in Principle 9 (Use pseudofactors) for completeness, these being nonorthogonal designs in general.

Example 3 (A replicated two-phase athlete training experiment). Suppose that in Example 2 the analysis of the free haemoglobin is to be duplicated for each specimen. To do this two fractions are to be taken from each blood specimen and one fraction from all specimens processed together in a single round as before. Then, in a second round, the second set of fractions is to be processed together. The factor-allocation diagram for this experiment is in Figure 3.

While the unrandomized objects in the first phase have now become fractions, the first phase randomization of treatments to fractions is essentially the same as in Example 2. On the other hand, the laboratory-phase allocation of fractions to locations now involves a pseudofactor F_1 because the randomization of Fractions to Rounds is not *consonant* (Brien and Bailey 2006): there are in total 72 levels of Months $\wedge$ Athletes $\wedge$ Tests $\wedge$ Fractions to be assigned to the eight levels of Batches $\wedge$ Rounds. To deal with this, the two-level pseudofactor F_1, which is nested within Months, is introduced to group one level of Fractions from all tests within a month. To avoid confounding Rounds with any systematic difference between the fractions, it is necessary either to label randomly the levels of Fractions within each test or to assign randomly the fractions from each test to the two groups indexed by F_1. The levels of F_1 within a month are randomized to the levels of Rounds within a batch. As part of this randomization, the tests within each level of F_1 are randomized to the locations within each round.

The result of the laboratory allocation is that the sources Athletes [Months] and Tests [Months ∧ Athletes] are only confounded with Locations [Batches ∧ Rounds], as shown in the skeleton ANOVA table in Table 3. So the two fractions sources that have training-conditions sources confounded with them are not split. This means that the allocation of fractions to locations can be done ignoring the result of the assignment of training-conditions to fractions and so the randomizations are composed. However, the source Fractions [Months ∧ Athletes ∧ Tests] is split; part is confounded with Rounds [Batches] and the rest with Locations [Batches ∧ Rounds]. The need for this can be predicted from the nonconsonant randomization of Fractions to Rounds, which required the inclusion of the two-level pseudofactor F_1 that effects the split. In this example, in contrast to Web Appendix Example 1, the splitting of a source does not result in randomized-inclusive randomizations, because nothing is confounded with Fractions [Months ∧ Athletes ∧ Tests] and so what happens with this source is irrelevant to the type of multiple randomizations. All this could be avoided if Fractions is omitted, but that would breach Principles 7 (Allocate all and randomize in laboratory) and 9 (Use pseudofactors). An advantage of retaining Fractions and its pseudofactor is that it draws attention to the need to randomize the fractions to the different rounds within batches to ensure an unbiased estimate of the sum of the components for Batches ∧ Rounds ∧ Locations and Months ∧ Athletes ∧ Tests ∧ Fractions.

7.2. BALANCING PRECISION ACROSS PHASES

While it was noted in Section 6 that Principles 3 (Minimize variance) and 4 (Split-units) apply in carrying out Principle 7 (Allocate all and randomize in laboratory), a possibility that is unique to multiphase experiments is the subject of the following principle:

Principle 10 (Compensating across phases). If treatments are confounded with a large source of unit variation in the first phase, then consider confounding this source with a smaller source of variation in the laboratory phase.

Example 4 (A compensating two-phase athlete training experiment). Suppose that, contrary to the suggestion made in Example 2, Surfaces and Intensities are of similar interest. Also, it is anticipated that there will be an interaction between Surfaces and Intensities and maximum precision in estimating these interaction effects is desired. Further suppose that, unlike in previous examples, the researcher believes that there will be a common order effect between tests in the same month, but that the order effect will be less than the differences between athletes. It is decided to employ a strip-unit design (Cochran and Cox 1957, Section 7.32) to allocate Surfaces to Athletes within Months and Intensities to Tests within Months.

In the laboratory phase, unlike in Example 2, the researcher believes that it will be advantageous to group sets of three consecutive locations into blocks, called Periods. The laboratory-unit factors remain hierarchical. To compensate for Surfaces being allocated to the more variable Athletes in the first phase, it is decided to employ Principle 10 (Compensating across phases): Intensities are be allocated to the more variable Periods and Surfaces to Locations within Periods. Figure 4 shows the factor allocations.

Table 3. Skeleton ANOVA table for the replicated two-phase athlete training experiment.

locations tier		fractions tier[a]		training-conditions tier		E.M.S.[b]							
Source	d.f.	Source	d.f.	Source	d.f.	σ^2_{BRL}	σ^2_{BR}	σ^2_{B}	σ^2_{MATF}	σ^2_{MAT}	σ^2_{MA}	σ^2_{M}	$q(\cdot)$
Mean	1	Mean	1	Mean	1								
Batches	3	Months	3			1	9	18	1	2	6	18	
Rounds [B]	4	Fractions $[M \wedge A \wedge T]_1$	4			1	9		1				
Locations $[B \wedge R]$	64	Athletes [M]	8	Intensities	2	1			1	2	6		$q(I)$
				Residual	6	1			1	2	6		
		Tests $[M \wedge A]$	24	Surfaces	2	1			1	2			$q(S)$
				I#S	4	1			1	2			$q(IS)$
				Residual	18	1			1	2			
		Fractions $[M \wedge A \wedge T]_{\vdash}$	32			1			1				

[a]Fractions $[M \wedge A \wedge T]_1$ is the part of Fractions $[M \wedge A \wedge T]$ corresponding to the pseudofactor F_1, which is nested within M, and Fractions $[M \wedge A \wedge T]_{\vdash}$ is the part orthogonal to Fractions $[M \wedge A \wedge T]_1$.

[b]Each σ^2 is a variance component whose subscript is comprised of the first letter of each factor in its term and the numbers in the table are the coefficients of the variance components in the E.M.S. The q-functions are the same quadratic functions of the expectation as are the corresponding mean squares.

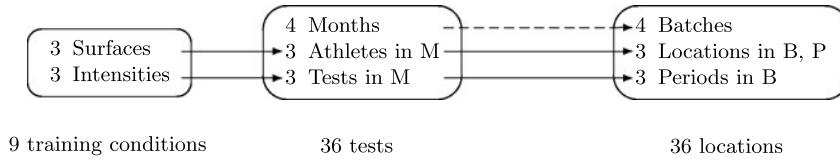

Figure 4. Factor-allocation diagram for the compensating two-phase athlete training experiment: training conditions are randomized to tests and tests are allocated to locations; the dashed arrow indicates that Months are systematically allocated to Batches; M = Months; B = Batches; P = Periods.

The skeleton ANOVA table for the experiment is given in Table 4. It shows that the desired balancing of precision is achieved, although ultimately judging the success of the approach requires knowledge of the relative magnitudes of the sources of variation from both phases. Again, the numbers of units for the two phases are equal and so, as can be deduced from Table 4, all locations sources are exhausted by tests sources. As in Example 2, for some pairs of variance components, only their sums are estimable.

7.3. LABORATORY REPLICATION

For laboratory replication, the following principle, based on McIntyre (1955), applies:

Principle 11 (Laboratory replication). Replicated measurement of first-phase units is not required, but is highly desirable when uncontrolled variation in the laboratory phase is large relative to the first phase. It is also needed if the relative magnitudes of field and laboratory variation are to be assessed.

With the possibility of laboratory replication comes the need to distinguish between the actual *products* of the first phase (batches of harvested crop, wines, blood specimens) and *portions* of them (aliquots, drops, lots, samples and fractions), which are not required if there is no laboratory phase. Portions are necessary in situations such as when measurements are to be replicated and the process is destructive; they are not necessary if the process is nondestructive. In designing the first phase, usually treatments are randomized only to products, even when there are portions that are the units. Indeed, without a laboratory phase, the first-phase units would be its products.

Of course, the inclusion of laboratory replicates does not increase the Residual degrees of freedom for treatment sources from the first phase (Multiphase law 1). Their inclusion serves to decrease the variance of the estimates of the treatment effects, provided there is variability in the laboratory replicates. Smith, Lim, and Cullis (2006) discuss the design of multiphase plant breeding experiments. Their experience is that the laboratory variation is greater than field variation in flour yield data and they want separate estimates of the magnitude of laboratory and field variation. Hence, following Principle 11 (Laboratory replication), laboratory replicates are required: they propose the use of partial replication that replicates only some of the field units that correspond to test varieties unreplicated in the field phase. More generally, the use of partial replication in the laboratory phase, by employing staggered nested designs (Ojima 2000) for this phase, is desirable when

Table 4. Skeleton ANOVA table for the compensating two-phase athlete training experiment.

locations tier		tests tier		training-conditions tier		E.M.S.[a]							
Source	d.f.	Source	d.f.	Source	d.f.	σ^2_{BPL}	σ^2_{BP}	σ^2_{B}	σ^2_{MAT}	σ^2_{MA}	σ^2_{MT}	σ^2_{M}	$q(\cdot)$
Mean	1	Mean	1	Mean	1								
Batches	3	Months	3			1	3	9	1	3	3	9	
Periods [B]	8	Tests [M]	8	Intensities	2	1	3		1		3		$q(\mathrm{I})$
				Residual	6	1	3		1		3		
Locations [B $\wedge$ P]	24	Athletes [M]	8	Surfaces	2	1			1	3			$q(\mathrm{S})$
				Residual	6	1			1	3			
		T#A [M]	16	I#S	4	1			1				$q(\mathrm{IS})$
				Residual	12	1			1				

[a]Each σ^2 is a variance component whose subscript is comprised of the first letter of each factor in its term and the numbers in the table are the coefficients of the variance components in the E.M.S. The q-functions are the same quadratic functions of the expectation as are the corresponding mean squares.

laboratory testing is expensive. Cox and Solomon (2003, Section 3.7) consider relative costs in designing them.

For Example 3, which involves laboratory replicates, a product is the blood specimen from a test and a portion is a fraction of a blood specimen. The degrees of freedom of the Residuals for estimating the variance for training-conditions sources are unchanged between Tables 2 and 3. However, Example 3 allows one to estimate the variability associated with laboratory replicates: it is given by $\sigma^2_{\mathrm{BIL}} + \sigma^2_{\mathrm{MATF}}$, since both of these components are an essential part of it. It is clear from Table 3 that the variability associated with location-fraction combinations can be separated from that for tests, as presaged in Principle 11 (Laboratory replication). This is not the case for Example 2, as can be seen from Table 2. However, the inclusion of laboratory replicates is only worthwhile if there is variability between them. To see this, consider the variance of Surface mean differences in Example 3:

$$2\left(\frac{\sigma^2_{\mathrm{BIL}} + \sigma^2_{\mathrm{MATF}}}{24} + \frac{\sigma^2_{\mathrm{MAT}}}{12}\right).$$

If there are no laboratory replicates then the denominator of the first term is reduced to 12 and so the variance is increased. It is noted that while Fractions would not be included in the analysis, it would still contribute to the Residual for Tests [Months $\wedge$ Athletes] because each test would still involve a fraction. That is, it would be an unidentified contributor to the test variability, as in Table 2. On the other hand, if there is no variability between laboratory replicates then the first term would be zero and the variance of the mean differences would be the same whether or not laboratory replicates are included.

7.4. LABORATORY TREATMENTS

The inclusion of laboratory treatments is obviously at the behest of the researcher. They are rather like superimposed treatments, being applied in a later stage of the experiment. Portions are needed when laboratory treatments are to be tested on the same product and different samples are needed for each treatment. However, while Principle 3 (Minimize variance) still applies, there are now both first-phase and laboratory sources to consider because laboratory treatments can be assigned to either first-phase or laboratory generalized factors. In either case, if they are to be assigned to a generalized factor that no first-phase treatment sources have been assigned to, directly or indirectly, then independent or coincident randomizations can be used to assign them, like Brien and Bailey (2006, Examples 6 and 12). Otherwise, unrandomized-inclusive randomizations (Brien and Bailey 2006) may be needed. Note that confounding laboratory treatments solely with laboratory variation is impossible when the numbers of field-phase and laboratory-phase units are equal, although this may be of little consequence when they can be allocated to first-phase portions. The following principle pertains to their randomization.

Principle 12 (Laboratory treatments). To minimize the variance of the estimates of laboratory treatment effects, confound them with sources to which only small components of laboratory variation contribute. When also confounding with first-phase unit sources, they too should be as small as possible.

7.5. DESIGN KEYS CAN BE USEFUL

Design keys (Patterson and Bailey 1978) are more useful in the laboratory phase than the first phase, particularly when there are several factors in each tier of the randomization, an orthogonal design is required and randomized-inclusive randomizations requiring pseudofactors for first-phase units are needed. Then design keys, using these pseudofactors, can facilitate the design of the laboratory phase because they allow the designer to specify the confounding between first-phase unit pseudofactors and laboratory pseudofactors. Their use is demonstrated in Example 6 for the case study (Section 8).

8. CASE STUDY: A BIODIVERSITY EXPERIMENT

A two-phase experiment, akin to that described by Harch et al. (1997), consisted of field and laboratory phases. The field experiment used an RCBD with four blocks to look at the effect of two tillage treatments on bacterial and fungal diversity. For each plot, soil samples were taken at the one place; two samples were taken at each of two different depths (0–5 cm and 5–10 cm). The resulting 32 soil samples were taken to the laboratory for analysis with a gas chromatograph. In this laboratory phase there were 64 runs and these were divided into two occasions of 32 runs each. During the first occasion, fractions from each of the 32 soil samples were analysed in a systematic order and then, during the second occasion, another 32 fractions were analysed in the same sample order. The two samples taken at each depth were preprocessed using two different methods (ground versus sieved). Each occasion was divided into two intervals, during each of which the fractions from 16 samples from two of the four blocks were assayed. The order of processing of the 16 samples from two blocks, A and B say, is shown in Table 5. Also given are the values for the response variable, a Gini coefficient computed from readings from BIOLOG^{TM} plates taken at selected incubation times during a run.

The arrangement used has the obvious defect that Methods and, to a lesser extent, Depths were confounded with systematic trends across the Runs arising from problems such as equipment drift, operator learning and fatigue, and changes in the laboratory ambience. Furthermore, the seemingly most important, and possibly smallest, treatment effect of Tillage was not confounded with potentially the smallest source of random variation, that between pairs of consecutive runs. Given these defects, the question that arises is how one might improve on the laboratory-phase design.

The first step is to adopt Principle 7 (Allocate all and randomize in laboratory) and to require that all factors from the first phase and the laboratory treatments are randomized in the laboratory phase. A product from the first phase is the soil for a Blocks $\wedge$ Plots $\wedge$ Depths combination, and Tillage is randomized to Plots within Blocks. In the laboratory phase, methods are laboratory treatments, which must be applied to different portions of soil and so two samples are to be taken from each product. Also, there are to be laboratory duplicates and so, since measurement is destructive, two fractions are to be taken from each sample. It is decided to divide the laboratory phase into two occasions, during each of which one of the fractions from all 32 Blocks $\wedge$ Plots $\wedge$ Depths $\wedge$ Samples combinations are processed, the processing order differing between occasions. Also, it is assumed that:

Table 5. Laboratory-phase order and observed Gini coefficients for the biodiversity experiment.

							Gini coefficient (×100)			
						Occasion	1		2	
						Interval	1	2	1	2
Run	Method	Block[a]	Plot	Tillage[b]		Depth				
1	ground	A	1	CC	CC	0–5	66.54	66.14	65.32	63.46
2	ground	A	1	CC	CC	5–10	71.45	67.24	68.64	64.34
3	ground	A	2	DD	DD	0–5	66.22	63.26	64.46	63.36
4	ground	A	2	DD	DD	5–10	67.00	63.95	68.37	63.96
5	ground	B	1	DD	CC	0–5	63.90	63.53	63.91	64.11
6	ground	B	1	DD	CC	5–10	69.17	65.33	67.37	65.44
7	ground	B	2	CC	DD	0–5	64.42	61.36	63.49	62.62
8	ground	B	2	CC	DD	5–10	64.02	63.36	64.84	64.03
9	sieved	A	1	CC	CC	0–5	66.44	69.01	66.44	68.64
10	sieved	A	1	CC	CC	5–10	72.04	71.04	72.42	66.89
11	sieved	A	2	DD	DD	0–5	64.90	65.72	68.24	63.81
12	sieved	A	2	DD	DD	5–10	70.48	70.88	71.52	70.88
13	sieved	B	1	DD	CC	0–5	68.18	64.97	66.86	65.09
14	sieved	B	1	DD	CC	5–10	73.05	66.89	71.86	67.74
15	sieved	B	2	CC	DD	0–5	65.56	62.54	64.15	65.06
16	sieved	B	2	CC	DD	5–10	67.61	65.85	70.46	67.48

[a]The letters A and B refer to Blocks 1 and 2, respectively, for the first and third columns of the Gini coefficients and Blocks 3 and 4, respectively, for the other two columns.
[b]CC and DD stand for Conventional Cultivation and Direct Drilling, respectively. The first column of Tillage refers to those applied in Blocks 1 and 2 and the second column to those applied in Blocks 3 and 4.

Table 6. Laboratory framework for the runs in each occasion for the biodiversity experiment.

(i) Initial

Run / Interval	1	2	3	4	5	6	7	8
1								
2								
3								
4								

(ii) After blocking

Cluster	1		2		3		4	
Run / Interval	1	2	1	2	1	2	1	2
1	1	2	1	2	1	2	1	2
2	3	4	3	4	3	4	3	4
3	5	6	5	6	5	6	5	6
4	7	8	7	8	7	8	7	8

NOTE: Numbers in the cells of (ii) are the levels of Analyses within Clusters and Occasions.

(i) the equipment has to be recalibrated after every eight runs; and

(ii) in terms of consecutive runs, a pair is less variable than four, which in turn are less variable than eight.

That this is plausible is established by an analysis of the data from this experiment, reported in Web Appendix G.1. As the equipment is recalibrated after every eight runs, a set of eight consecutive runs forms an interval. Thus the basic set-up of the laboratory phase can be described as two Occasions by four Intervals by eight Runs, as illustrated in Table 6(i).

Example 5 (Hierarchical laboratory-unit factors for the biodiversity experiment). The factors Tillage, Methods and Depths, yielding eight combinations, are of particular interest

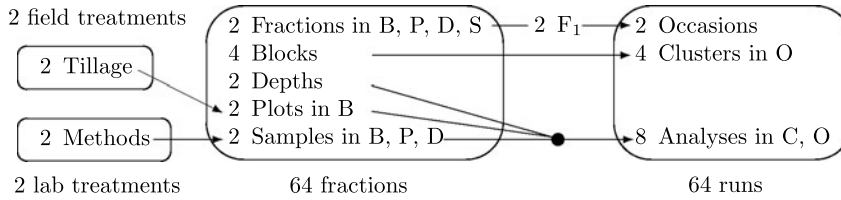

Figure 5. Factor-allocation diagram for the design with hierarchical laboratory-unit factors for the biodiversity experiment: field and lab treatments are randomized to fractions, and fractions to runs; the '●' indicates that the combinations of the levels of Depths, Plots and Samples are randomized to Analyses; B = Blocks; P = Plots; D = Depths; S = Samples; O = Occasions; C = Clusters; F_1 is a pseudofactor for Fractions that groups fractions that are to be assigned to the same Occasions level.

to the researchers and need to be randomized to the runs within an occasion. It would be ideal if laboratory blocks of eight homogeneous runs could be identified to accommodate the two Plots by two Samples by two Depths, corresponding to Tillage ∧ Methods ∧ Depths combinations, in a field block. This would comply with Principle 5 (Simplicity desirable), because composed randomizations would result, and with Principle 3 (Minimize variance). Also, Principle 12 (Laboratory treatments) would be satisfied because laboratory treatments would be confounded with the smallest source of laboratory variation, in addition to being confounded with portions (the samples) of the first-phase product. Our case-study assumptions mean that groups of eight runs within an interval will not accomplish our aim. They do imply that the four intervals within an occasion will be similar, as will be pairs of runs, so that a group of two consecutive runs across four intervals would be homogeneous. Label these groups using the 4-level factor Clusters, with an 8-level factor Analyses to index the runs within each level of Clusters: see Table 6(ii); the processing sequence is still runs within an interval. The factor-allocation diagram for this design is in Figure 5. The randomization of field treatments to fractions and the latter to runs is composed. However, as in Example 4, the randomization of Fractions to Occasions is not consonant and so a pseudofactor is needed to group the fractions taken from the Blocks ∧ Plots ∧ Depths ∧ Samples combinations for each Occasion. Also, the randomizations of field and lab treatments to fractions are independent, as in Brien and Bailey (2006, Example 6). When the laboratory phase is planned after the first phase has been completed, the two randomizations must be done separately. This handling of the laboratory treatment randomization is a simpler alternative to that in Brien and Bailey (2006, Example 13).

In constructing the skeleton ANOVA table, given in Table 7, the intertier interactions of Depths with Tillage and Methods are included. The table shows that all the sources of prime interest are confounded with the laboratory source anticipated to have the smallest variability. Importantly, it also shows that maximum Residual degrees of freedom are available for testing sources involving Tillage, Methods and Depths, which, although they are limited by the field-phase arrangement, are not reduced by the laboratory-phase design. For example, the Residual under Tillage has the three degrees of freedom from the field-phase RCBD. Blocks variation from the first phase is likely to be large and so, applying Principle 8 (Big with big), Blocks are confounded with Clusters, a larger source of random variation from the laboratory phase.

Table 7. Skeleton ANOVA table for the design with hierarchical laboratory-unit factors for the biodiversity experiment.

runs tier		fractions tier[a]		treatments tiers		E.M.S.[b]									
Source	d.f.	Source	d.f.	Source	d.f.	σ^2_{OCA}	σ^2_{OC}	σ^2_{O}	σ^2_{BPDSF}	σ^2_{BPDS}	σ^2_{BPD}	σ^2_{BP}	σ^2_{BD}	σ^2_{B}	$q(\cdot)$
Mean	1	Mean	1	Mean	1										
Occasions	1	Fractions $[B \wedge P \wedge D \wedge S]_O$	1			1	8	32	1						
Clusters [O]	6	Blocks	3			1	8		1	2	4	8	8	16	
		Fractions $[B \wedge P \wedge D \wedge S]_B$	3			1	8		1						
Analyses $[O \wedge C]$	56	Depths	1			1			1	2	4		8		$q(D)$
		B#D	3			1			1	2	4		8		
		Plots [B]	4	Tillage	1	1			1	2	4	8			$q(T)$
				Residual	3	1			1	2	4	8			
		D#P [B]	4	T#D	1	1			1	2	4				$q(TD)$
				Residual	3	1			1	2	4				
		Samples $[B \wedge P \wedge D]$	16	Methods	1	1			1	2					$q(M)$
				T#M	1	1			1	2					$q(TM)$
				M#D	1	1			1	2					$q(MD)$
				T#M#D	1	1			1	2					$q(TMD)$
				Residual	12	1			1	2					
		Fractions $[B \wedge P \wedge D \wedge S]_{\vdash}$	28			1			1						

[a]Fractions $[B \wedge P \wedge D \wedge S]_O$ is the part of Fractions $[B \wedge P \wedge D \wedge S]$ corresponding to F_1, Fractions $[B \wedge P \wedge D \wedge S]_B$ is the part corresponding to $B\#F_1$ and Fractions $[B \wedge P \wedge D \wedge S]_{\vdash}$ is the part orthogonal to the other two Fractions sources.

[b]Each σ^2 is a variance component whose subscript is comprised of the first letter of each factor in its term and the numbers in the table are the coefficients of the variance components in the E.M.S. The q-functions are the same quadratic functions of the expectation as are the corresponding mean squares.

The success of this, perhaps not so obvious, arrangement does rely on the similarities between different intervals within an occasion. If there was substantial variation between intervals, then groups of runs within intervals would be beneficial. Groups of size 2, 4 or 8 are possible, of which 2 or 4 seem preferable. Thus, Principle 4 (Split-unit) would need to be exploited and at least one source confounded with a larger source of variation. Perhaps Depths has a large effect and so is a candidate.

The design does not fully satisfy Principle 3 (Minimize variance) because laboratory treatments are not randomized to the smallest first-phase entity-type. This would be rectified if Methods were randomized to four samples that replace the two fractions in two samples. Even if there is little difference in the variability of Samples and Fractions, it would also have the advantage that the Residual degrees of freedom for Methods, and its interactions, would increase from 12 to 40. This potential improvement is likely to be overlooked if the factor Fractions is omitted, because then it is not obvious that a sample is not the smallest first-phase entity-type.

Example 6 (Nonhierarchical laboratory-unit factors for the biodiversity experiment). The basic framework for the laboratory phase of the biodiversity experiment is two Occasions by four Intervals by eight Runs [see Table 6(i)], and all three factors are inherently crossed. Suppose that, not only are differences between runs expected to be consistent across intervals and occasions, but that there are also substantial differences between intervals, in contrast to Example 5, and that these are consistent across runs but not occasions. Thus a design that conforms to these expectations, and so to Principle 3 (Minimize variance) also, would have two rectangles of four rows by eight columns, with columns latinized (Williams 1986) across rectangles. Such a design is now produced. It is taken also that each sample should occur just once in a rectangle. Applying Principle 8 (Big with big) sees Blocks assigned to Intervals. Then, consistent with Principle 3 (Minimize variance), the eight Plots ∧ Depths ∧ Samples combinations within Blocks are assigned to four Intervals by eight Runs within Occasions, confounding as much information as possible with Intervals # Runs [Occasions]. The factor-allocation diagram for such a design is given in Figure 6 and shows that the randomizations of the two treatments tiers to fractions and then fractions to runs are randomized inclusive. They are not composed because some sources involving Plots or Samples, with which Tillage and Methods sources are confounded, must be confounded with more than one runs source in randomizing fractions to runs.

The design key method is used to obtain an orthogonal design for randomizing fractions to runs, by assigning multiple pseudofactors, each with two levels, to each of the nonprime factors. Giving just the first letter of the genuine factor names and subscripting this letter for pseudofactors, the sets of two-level genuine factors and pseudofactors are O, I_1, I_2, R_1, R_2 and R_3 for the runs tier and B_1, B_2, P, D, S and F for the fractions tier. However, because the randomizations are randomized inclusive, we have to keep track of Tillage and Methods and ensure that the associated sources are confounded with appropriate laboratory sources. To do this, the two-level factors Plots and Samples are replaced by pseudofactors P_1 and S_1, respectively. As shown in Figure 6, P_1 identifies plots with the same tillage; similarly with S_1 for samples and methods. That is, P_1 and S_1 amount to a relabelling of

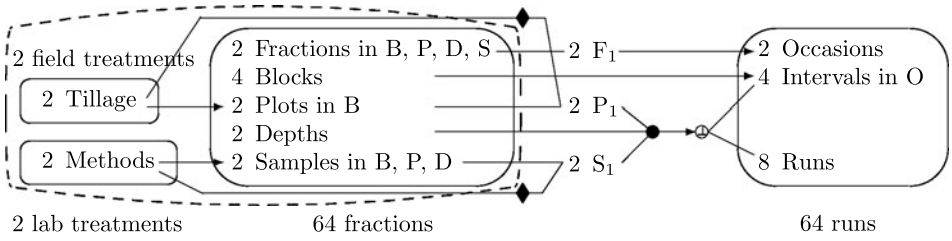

Figure 6. Factor-allocation diagram for the design with nonhierarchical laboratory-unit factors for the biodiversity experiment: field and lab treatments are randomized to fractions, and field and lab treatments and fractions to runs, the latter dependent on the allocation of treatments to samples; the '●' indicates that the levels combinations of P_1, Depths and S_1 are randomized to the levels combinations of Intervals and Runs; the ⏀ indicates that an orthogonal design is used; P_1 is a pseudofactor on Plots determined by Tillage, as indicated by the upper '♦'; S_1 is a pseudofactor on Samples determined by Methods, as indicated by the lower '♦'; the dashed oval indicates that the factors in the enclosed panels are combined into a pseudotier for randomizing them to runs; B = Blocks; P = Plots; D = Depths; S = Samples; O = Occasions; F_1 is a pseudofactor for Fractions that groups fractions that are to be assigned to the same Occasions level.

their factors, which is captured in a second design key for this randomization. However, the origins of P_1 and S_1 are in first-phase unit, not treatment, differences. One could expediently replace P_1 with T and S_1 with M and use a single design key, but this is not done so as to retain Plots and Samples in the randomization and analysis, as advocated by Principles 7 (Allocate all and randomize in laboratory) and 9 (Use pseudofactors). As for Example 5, the pseudofactor F_1 is used in assigning Fractions. The selected design keys are

(Pseudo)factor	B_1	B_2	P_1	D	S_1	F_1	T	M
Alias	I_1	I_2	I_1R_1	I_2R_2	$OI_1I_2R_3$	O	P_1	S_1

The aliases for P_1 and S_1 in the first design key have been chosen so that they, along with all two-factor interactions involving them and D, are confounded with the Intervals # Runs [Occasions] source and their three-factor interaction confounded with a two-factor source. Consequently, so are the corresponding sources with Tillage, Methods and Depths. Further details about the generation of randomized layouts and the aliases for the effects are available in Web Appendix G.2.

From the skeleton ANOVA table in Table 8, which includes intertier interactions, it is clear that the design is orthogonal, provided pseudofactors are used to identify the subspaces of fractions sources that are confounded with different runs sources. Table 8 also shows that variance components for Occasions ∧ Intervals ∧ Runs and Blocks ∧ Plots ∧ Depths ∧ Samples ∧ Fractions are not estimable and that two estimates of the variance component for Blocks ∧ Plots ∧ Depth ∧ Samples are obtained from equating observed and expected mean squares. Further, the design used in the laboratory phase sacrifices (i) one Residual degree of freedom for Plots [Blocks], and (ii) Residual degrees of freedom for Samples [Blocks ∧ Plots ∧ Depths]. More information about the sources in the skeleton ANOVA table and how to generate a skeleton ANOVA table in GenStat are available in Web Appendix G.2.

Table 8. Skeleton ANOVA table for the design with nonhierarchical laboratory-unit factors for the biodiversity experiment.

runs tier		fractions tier[a]		treatments tiers		E.M.S.[b]											
Source	d.f.	Source	d.f.	Source	d.f.	σ^2_{OIR}	σ^2_{OR}	σ^2_{OI}	σ^2_{R}	σ^2_{O}	σ^2_{BPDSF}	σ^2_{BPDS}	σ^2_{BPD}	σ^2_{BP}	σ^2_{BD}	σ^2_{B}	$q(\cdot)$
Mean	1	Mean	1	Mean	1												
Occasions	1	Fractions $[B \wedge P \wedge D \wedge S]_O$	1			1	4	8		32	1						
Runs	7	Plots $[B]_R$	1			1	4		8		1	2	4	8			
		$B\#D_R$	1			1	4		8		1	2	4		8		
		$D\#P\,[B]_R$	1			1	4		8		1	2	4				
		Fractions $[B \wedge P \wedge D \wedge S]_R$	4			1	4		8		1						
Intervals [O]	6	Blocks	3			1		8			1	2	4	8	8	16	
		Fractions $[B \wedge P \wedge D \wedge S]_{OI}$	3			1		8			1						
O#R	7	Samples $[B \wedge P \wedge D]_{OR}$	4	T#M#D	1	1	4				1	2					q(TMD)
				Residual	3	1	4				1	2					
		Fractions $[B \wedge P \wedge D \wedge S]_{OR}$	3			1	4				1						
I#R [O]	42	Depths	1			1					1	2	4		8		q(D)
		$B\#D_{\vdash}$	2			1					1	2	4		8		
		Plots $[B]_{\vdash}$	3	Tillage	1	1					1	2	4	8			q(T)
				Residual	2	1					1	2	4	8			
		$D\#P\,[B]_{\vdash}$	3	T#D	1	1					1	2	4				q(TD)
				Residual	2	1					1	2	4				
		Samples $[B \wedge P \wedge D]_{\vdash}$	12	Methods	1	1					1	2					q(M)
				T#M	1	1					1	2					q(TM)
				M#D	1	1					1	2					q(MD)
				Residual	9	1					1	2					
		Fractions $[B \wedge P \wedge D \wedge S]_{\vdash}$	21			1					1						

[a]Fractions $[B \wedge P \wedge D \wedge S]_O$ is the part of Fractions $[B \wedge P \wedge D \wedge S]$ corresponding to F_1; the subscript letters on sources indicate that they are the parts of the source estimated from the corresponding runs source, for example $D\#P\,[B]_R$ is the part of $D\#P\,[B]$ estimated from the source Runs; the subscript '$\vdash$' on a source indicates that this source is the part of the source orthogonal to all other parts of the same source, for example $B\#D_{\vdash}$ is the part of $B\#D$ orthogonal to $B\#D_R$.

[b]Each σ^2 is a variance component whose subscript is comprised of the first letter of each factor in its term and the numbers in the table are the coefficients of the variance components in the E.M.S. The q-functions are the same quadratic functions of the expectation as are the corresponding mean squares.

9. DISCUSSION

This paper has demonstrated that, while the laboratory-phase design is often a standard design, there are several principles specific to it that need to be employed in using these designs for multiphase experiments. In addition to employing these principles, it is useful for the designer to characterize prospective designs according to the following features:

1. Relationships between laboratory-unit factors: *hierarchical* or *nonhierarchical*.
2. Number of phases: *two-phase* or *more than two phases*.
3. Number of randomizations within phases: *one in each* or *none in some* and/or *multiple in some*.
4. The experiment's randomization: *single*, *two composed*, *two randomized-inclusive* or *three or more*.
5. Laboratory-phase features: *laboratory treatments* or *not*; *laboratory replicates* or *not*.
6. Nature of standard designs used: *orthogonal*, *nonorthogonal but structure balanced* or *unbalanced*.
7. Type of variance structure: *orthogonal* or *nonorthogonal variance structure*.

While this paper has covered the first five features, the last two will be dealt with in a second paper.

SUPPLEMENTARY MATERIALS

Web Appendices referenced in this paper are available from *supplementary data*. Also, there is a multitiered experiments web site at *http://chris.brien.name/multitier/*.

ACKNOWLEDGEMENTS

We are very grateful to the referees of earlier versions of the paper for suggestions that led to substantial improvements. We also thank Clive Pankhurst and Bruce Hawke who provided us with the data for the Biodiversity example in Section 8.

[Received March 2011. Accepted May 2011. Published Online June 2011.]

REFERENCES

Bailey, R. A., and Brien, C. J. (2011), "Data Analysis for Multitiered Experiments Using Randomization Models: A Chain of Randomizations," Unpublished manuscript.

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)," *Journal of the Royal Statistical Society, Series B*, 68, 571–609.

——— (2009), "Decomposition Tables for Experiments. I. A Chain of Randomizations," *The Annals of Statistics*, 37, 4184–4213.

——— (2010), "Decomposition Tables for Experiments. II. Two-One Randomizations," *The Annals of Statistics*, 38, 3164–3190.

Brien, C. J., and Demétrio, C. G. B. (2009), "Formulating Mixed Models for Experiments, Including Longitudinal Experiments," *The Journal of Agricultural, Biological and Environmental Statistics*, 14, 253–280.

Brien, C. J., Harch, B. D., and Correll, R. L. (1998), "Design and ANOVA for Experiments Involving a Field Trial and Laboratory Analyses," Paper presented to The Ninth International Conference on Quantitative Methods for the Environmental Sciences, Gold Coast, Australia.

Brien, C. J., May, P., and Mayo, O. (1987), "Analysis of Judge Performance in Wine-Quality Evaluations," *Journal of Food Science*, 52, 1273–1279.

Cochran, W. G., and Cox, G. M. (1957), *Experimental Designs* (2nd ed.), New York: Wiley.

Cox, D. R. (1958), *Planning of Experiments*, New York: Wiley.

——— (2009), "Randomization in the Design of Experiments," *International Statistical Review*, 77, 415–429.

Cox, D. R., and Solomon, P. (2003), *Components of Variance*, Boca Raton: Chapman and Hall/CRC.

Cullis, B. R., Smith, A. B., Panozzo, J. F., and Lim, P. (2003), "Barley Malting Quality: Are We Selecting the Best?" *Australian Journal of Agricultural Research*, 54, 1261–1275.

Harch, B. D., Correll, R. L., Meech, W., Kirkby, C. A., and Pankhurst, C. E. (1997), "Using the Gini Coefficient with BIOLOG Substrate Utilisation Data to Provide an Alternative Quantitative Measure for Comparing Bacterial Soil Communities," *Journal of Microbial Methods*, 30, 91–101.

Jarrett, R. G., and Ruggiero, K. (2008), "Design and Analysis of Two-Phase Experiments for Gene-Expression Microarrays—Part I," *Biometrics*, 64, 208–216.

Kerr, M. K. (2003), "Design Considerations for Efficient and Effective Microarray Studies," *Biometrics*, 59, 822–828.

Littell, R., Milliken, G., Stroup, W., Wolfinger, R., and Schabenberger, O. (2006), *SAS for Mixed Models* (2nd ed.), Cary: SAS Press.

McIntyre, G. A. (1955), "Design and Analysis of Two Phase Experiments," *Biometrics*, 11, 324–334.

Nelder, J. A. (1965), "The Analysis of Randomized Experiments with Orthogonal Block Structure. I. Block Structure and the Null Analysis of Variance," *Proceedings of the Royal Society, Series A*, 283, 147–161.

Ojima, Y. (2000), "Generalized Staggered Nested Designs for Variance Components Estimation," *Journal of Applied Statistics*, 27, 541–553.

Patterson, H. D., and Bailey, R. A. (1978), "Design Keys for Factorial Experiments," *Applied Statistics*, 27, 335–343.

Peeling, P., Dawson, B., Goodman, C., Landers, G., Wiegerinck, E. T., Swinkels, D. W., and Trinder, D. (2009), "Training Surface and Intensity: Inflammation, Hemolysis, and Hepcidin Expression," *Medicine and Science in Sports and Exercise*, 41, 1138–1145.

Searle, S. R., Casella, G., and McCulloch, C. E. (1992), *Variance Components*, New York: Wiley.

Smith, A. B., Cullis, B. R., Appels, R., Campbell, A. W., Cornish, G. B., Martin, D., and Allen, H. M. (2001), "The Statistical Analysis of Quality Traits in Plant Improvement Programs with Application to the Mapping of Milling Yield in Wheat," *Australian Journal of Agricultural Research*, 52, 1207–1219.

Smith, A. B., Lim, P., and Cullis, B. R. (2006), "The Design and Analysis of Multi-Phase Plant Breeding Experiments," *The Journal of Agricultural Science*, 144, 393–409.

Williams, E. R. (1986), "Row and Column Designs with Contiguous Replicates," *Australian Journal of Statistics*, 28, 154–163.