Aust. N. Z. J. Stat. 2017 doi: 10.1111/j.1467-842X.XXX

Multiphase experiments in practice: A look back.

C. J. Brien[1]

University of South Australia and University of Adelaide

Summary

Multiphase experiments are introduced and an overview of their design and analysis as it is currently practised is given via an account of their development since 1955 and a literature survey. Methods that are available for designing and analysing them are outlined, with an emphasis on making explicit the role of the model in their design. The availability of software and its use is detailed. Overall, while multiphase designs have been applied in areas such as plant breeding, plant pathology, greenhouse experimentation, product storage, gene expression studies, and sensory evaluation, their deployment has been limited.

Key words: analysis of variance; laboratory phase; mixed models; multiple randomizations; nonorthogonal design; optimal design; plant breeding experiments; sensory experiments

1. Introduction

McIntyre (1955) originally used the term two-phase for experiments in which there is a single randomization in each phase. The aim 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 involve multiple allocations (Brien & Bailey 2006; Brien et al. 2011), an allocation being the assignment of one set of objects to a second set of objects, to be called the *allocated* and *recipient objects*, respectively. The most common type of allocation is a randomization in which the allocated objects are the treatments and the recipient objects are the units. However, two alternatives are that some aspects of the allocation are systematic or that it is restricted to allow for the efficient estimation of autocorrelation between units. As described in Brien et al. (2011), each phase involves different units (recipient objects) and 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.

Section 2 gives an example that illustrates the features of multiphase experiments and Section 3 explains some of the terminology. Section 4 gives a history of multiphase experiments, Section 5 describes their design and analysis in practice and gives a further example. Section 6 contains a discussion.

[1]Phenomics and Bioinformatics Research Centre, University of South Australia, GPO Box 2471, Adelaide, South Australia. 5001.

Acknowledgments. The author is grateful to the reviewer and the associate editor for suggestions that greatly improved the accessibility of the paper. He also appreciates the helpful comments made by Rosemary Bailey.

Prepared using anzsauth.cls *[Version: 2017/10/03 Version 7]*

2. An expository example

The purpose of Example 1 is to illustrate some of the basic features of multiphase experiments and the techniques used to design, assess and formulate mixed models for them; it also serves as an introduction to the terminology used by Brien & Bailey (2006, 2009), Brien et al. (2011) and Bailey & Brien (2016) in connection with them.

Example 1 (A nonorthogonal two-phase athlete training experiment). This example is a variation on Example 2 from Brien et al. (2011), an orthogonal two-phase experiment in which the effect of nine training conditions for endurance athletes is investigated. As for that experiment, the first phase of the present experiment is a testing phase that extends over four months and the second phase is a laboratory phase. The nine training conditions are the combinations of three surfaces and three intensities of training. In the testing phase three athletes are recruited each month and each will undergo three tests that are separated by seven days. On completion of each test the heart rate of the athlete is measured. In addition to the heart rate taken immediately upon completion of a test, the free haemoglobin is measured using blood specimens taken from the athletes after each test and transported to the laboratory for analysis. The difference is that here it is supposed that the researcher views Intensities and Surfaces as equally important and so a simple lattice design is proposed for allocating the training conditions, this being the optimal design for assigning trios of training condition to tests for an athlete. The systematic design is as follows, where the combinations of the levels of Intensities and Surfaces are numbered from 1 to 9 and the parentheses enclose the subset of the combinations that one athlete will undertake: (1 2 3)(4 5 6)(7 8 9) for athletes in the first month; (1 4 7)(2 5 8)(3 6 9) for athletes in the second month; (1 5 9) (7 2 6) (4 8 3) for athletes in the third month; (1 8 6) (4 2 9) (7 5 3) for athletes in the fourth month (Cochran & Cox 1957, table 9.1). As in the original experiment, because the specimens become available monthly, the batch of specimens for one month is processed before those for the next month are available; the specimens for a month are randomized to the time locations for that month so that the design for this phase is a randomized complete-block design.

This multiphase design has two allocations and a design has been identified for each. Its allocations are exhibited in the factor-allocation diagram in Figure 1. This figure shows that there are three sets of objects: training conditions, tests and locations, each with a panel. The left and middle panels show the randomization of training conditions to tests for the testing phase using one design and the middle and right panels show the allocation of tests to locations for the laboratory-phase using the other design. In each panel, the factors that index the panel's set of objects are listed, along with the nesting between those factors. The factors within a panel are referred to as a tier by Brien (1983) and Brien & Bailey (2006). They are connected by being either (i) allocated in the same allocation, (ii) recipients in the same allocation, or (iii) allocated in one allocation and recipients in another allocation.

What makes this a multiphase experiment is that there are two allocations involving different sets of recipient objects or units: tests and locations. The outcomes from the first phase are the heart rate and blood specimen for a test, while the outcome from the second phase is the free haemoglobin for a location. The particular features of this example, as compared to multiphase experiments in general, are:

- there is one allocation in each phase — as Brien et al. (2011) describe, if the first phase is an observational study then the only allocation is likely to be the second phase

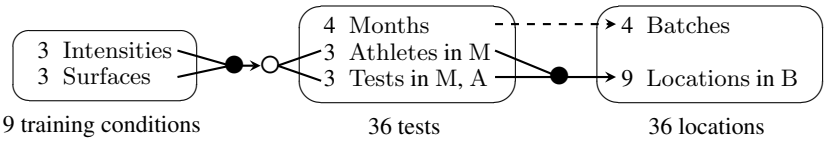

Figure 1. Factor-allocation diagram for Example 1, a nonorthogonal two-phase athlete training experiment: training conditions are randomized to tests and tests are allocated to locations; the left '●' indicates that the combinations of the levels of Intensities and Surfaces are randomized; the '○' indicates that a nonorthogonal design was used in the randomization of the levels of Intensities and Surfaces to the combinations of the levels of Athletes and Tests; the dashed arrow indicates that Months are systematically allocated to Batches; the right '●' indicates that the combinations of the levels of Athletes and Tests are randomized to the Locations; M = Months; A = Athletes; B = Batches.

allocation; on the other hand, there can be more than one allocation in a phase; in this case, the second phase is not a full randomization — it includes a systematic allocation;
- there is a response variable obtained from each phase — often a response variable is obtained from only the last phase;
- the two allocations are in a chain (Brien & Bailey 2009) and are composed, which means that they can be carried out independently — this is not the case for Example 2;
- the two phases are conducted contemporaneously — often the first phase is completed before the second phase is begun;
- the first-phase design is nonorthogonal, but (structure) balanced, and the second-phase design is orthogonal — any combination of orthogonal, nonorthogonal with (structure) balance, and unbalanced designs is possible.

Associated with a factor-allocation diagram is a mixed model that underpins the design chosen. It can be derived from the diagram by forming the generalized (or joint) factors derived from each panel. The set of generalized factors derived from a panel are all factorial combinations of the factors in the panel, subject to the restriction that a nested panel factor cannot occur in a term without its nesting panel factor. A generalized factor is a list of panel factors separated by '∧' when there is more than one panel factor, which is equivalent to ':' in R or '.' in GenStat. Its levels are the observed combinations of the levels of its constituent panel factors. The sets of generalized factors for the example are $\{\text{Batches}, \text{Batches}\wedge\text{Locations}\}$, $\{\text{Months}, \text{Months}\wedge\text{Athletes}, \text{Months}\wedge\text{Athletes}\wedge\text{Tests}\}$ and $\{\text{Intensities}, \text{Surfaces}, \text{Intensities}\wedge\text{Surfaces}\}$.

This leads to the following symbolic mixed model (without the Mean term), which can be used as the basis for the analysis of data from the experiment, whether this be analysis-of-variance (ANOVA) or mixed model fitting:

$$\begin{gathered}\text{Intensities} + \text{Surfaces} + \text{Intensities}\wedge\text{Surfaces} \quad | \\ \text{Batches} + \underline{\text{Batches}\wedge\text{Locations}} + \text{Months} + \text{Months}\wedge\text{Athletes} + \\ \underline{\text{Months}\wedge\text{Athletes}\wedge\text{Tests}},\end{gathered}$$

where terms before the '|' are fixed and those after it are random; the underlined terms, called identity terms, uniquely index the observational units.

To assess the properties of the design, a canonical analysis can be conducted on a systematic or randomized layout for it and displayed in a decomposition table (Brien & Bailey 2009), as Brien et al. (2011, Principle 1) and Bailey & Brien (2016) urge. This

Table 1. Skeleton ANOVA for Example 1, a nonorthogonal two-phase athlete training experiment: B = Batches; M = Months; A = Athletes; I = Intensities; S = Surfaces.

locations tier		tests tier†		training-conditions tier			EMS§					
Source	df	Source	df	eff‡	Source	df	ϕ_{BL}	ϕ_{B}	ϕ_{MAT}	ϕ_{MA}	ϕ_{M}	$q(\cdot)$
Mean	1	Mean	1	1	Mean	1	1	9	1	3	9	$q(\mu)$
Batches	3	Months	3				1	9	1	3	9	
Locations [B]	32	Athletes [M]	8	$\frac{1}{4}$	Intensities	2	1		1	3		$\frac{1}{4}q(\mathrm{I})$
				$\frac{1}{4}$	Surfaces	2	1		1	3		$\frac{1}{4}q(\mathrm{S})$
				$\frac{1}{4}$	I # S	4	1		1	3		$\frac{1}{4}q(\mathrm{IS})$
		Tests [M∧A]	24	$\frac{3}{4}$	Intensities	2	1		1			$\frac{3}{4}q(\mathrm{I})$
				$\frac{3}{4}$	Surfaces	2	1		1			$\frac{3}{4}q(\mathrm{S})$
				$\frac{3}{4}$	I # S	4	1		1			$\frac{3}{4}q(\mathrm{IS})$
					Residual	18	1		1			

†Canonical efficiency factors for the allocation of tests to locations are all equal to one and so are not shown in the table.
‡Each Eff is the value of all of the nonzero canonical efficiency factors for a training-conditions source confounded with a tests source.
§Each ϕ is a canonical component that, except for ϕ_{BL} and ϕ_{MAT}, is allowed to be negative. Their subscripts are comprised of the first letter of each factor in the corresponding term and the numbers in the table are the coefficients of the canonical components in the EMS. The q-functions are the same quadratic functions of the expectation as the corresponding mean squares are of the data.

exhibits the decomposition of the data space corresponding to the model and the confounding between the sources derived from the different sets of objects. We refer to such a table as the *anatomy* of the design. The expected mean squares (EMS) can be added to this, using the rules outlined in Brien et al. (2011, Web Appendix D), to produce the skeleton-ANOVA table for the experiment. There is a canonical (covariance) component (ϕ) for each random model term; these differ from variance components in that they are allowed to be negative. Table 1 gives this table for the proposed two-phase design; a '#' separates factors that interact and '[...]' contains the generalized factor comprised of the nesting factors. It shows that (i) all three allocated sources, Intensities, Surfaces and Intensities # Surfaces, are confounded with both Athletes [Months] and Tests [Months∧Athletes]. However, the efficiencies show that 75% of the information is confounded with Tests [Months∧Athletes]. All of these sources are in turn confounded with Locations [Batches]. This confounding is reflected in the EMS. From them it is concluded that, if the canonical component for Athletes [Months] is positive, Tests [Months∧Athletes] is likely to be a smaller source of variation than Athletes [Months], because the former includes that canonical component while the latter does not.

It is also apparent from the EMS columns in Table 1, that neither ϕ_{BL} and ϕ_{MAT} nor ϕ_{B} and ϕ_{M} are able to be separately estimated. Thus, attempting to fit the model with mixed-modelling software will be unsuccessful because its variance matrix is singular. To obtain a fit, one of Months∧Athletes∧Tests and Batches∧Locations, and one of Batches and Months,

have to be omitted from the model (Brien & Demétrio 2009, Stage 3(c)). That is, the fitted model will not involve terms for all the potential sources of variation that have been identified for the experiment and so is not the full model. The point is that the canonical component for the omitted term is not assumed to be zero, rather the remaining term is interpreted as having both sources of variation contributing to it. Such models are referred to as 'models of convenience' by Brien & Demétrio (2009) and are a necessary expedient in order to obtain a fit.

3. An explanation of terminology

A glossary of the terminology used in this paper is available from the multitiered experiments web site (Brien et al. 2001). Experiments that involve a single allocation are referred to as *standard experiments*. Often the allocated objects are the treatments, the recipient objects are the units, and the usual method of allocation is randomization. As in Brien & Bailey (2006), a single allocation is one that can be achieved using the following algorithm:

1. A design is chosen and a plan giving a systematic allocation of the allocated factors to the recipient factors is produced.
2. The nesting between the recipient factors for the chosen design is enunciated.
3. A permutation of the levels of the recipient factors that respects the nesting and any systematic aspect of the allocation is randomly selected and applied to them.
4. If required, the combined allocated and permuted recipient factors are rearranged into a convenient order.

Because the randomization of the treatment factors for a split-plot design to its recipient factors can be carried out using this algorithm, for the purpose of counting the number of allocations, it is viewed as a single allocation.

For multiple allocations there are (i) one or more sets of objects that are only ever allocated, (ii) one set of objects that occurs in just one allocation and that are the recipient objects in that allocation, and (iii) all other sets are recipient in one allocation and allocated in one or more of the subsequent allocations. In Example 1, (i) training conditions are only ever allocated, (ii) locations are only involved in the allocation of tests (and training conditions) to locations, where they are the recipient objects, and (iii) tests are the recipient objects when training conditions are allocated, but are those that are allocated to locations.

A factor-allocation diagram (Brien et al. 2011) summarizes a design by representing the allocations in the experiment. It has a panel for each set of objects, with the name of the set of objects below its panel. A panel contains the list of factors indexing the set of objects, along with the nesting (and crossing) of the factors. Normally these relationships reflect the allocations performed. Factor names begin with a capital and object names are with a lower case letter.

Extending Brien & Bailey (2009), the allocations in Figure 1 could be described as two allocations in a chain. Following them, the randomizations in each allocation can be classified as either composed or randomized-inclusive. This distinction has to do with whether knowledge about the first allocation is specifically used in making the second allocation: while it is for randomized-inclusive randomizations, it is not for composed randomizations. The two randomizations in Example 1 are composed. Knowledge of the allocation of training

conditions to tests is not needed in making the allocation of tests to locations: the nine combinations of athletes and tests for a month are simply randomized to the nine locations for a batch, irrespective of which training conditions are assigned to the athletes and tests. This is not the case for Example 2, where in assigning tests to locations, one has to ensure that the tests with certain training conditions are assigned to particular batches and locations. This difference is reflected in the scripts provided in the Supplementary Materials.

For Example 1, a nonorthogonal design with structure balance is proposed for the first phase. Structure balance is defined by Brien (1992) and Brien & Bailey (2009) and relates to the number of different nonzero canonical efficiency factors; it is also called complete or general balance (Nelder 1965; Preece 1982; Payne & Tobias 1992). Canonical efficiency factors take values between zero and one, inclusive, and reflect the amount of information available for allocated sources (Williams & Piepho 2015). Structure-balanced designs include all orthogonal and some nonorthogonal designs. Three types of designs are recognized here:

Orthogonal: For a design to be orthogonal, (i) all the nonzero canonical efficiency factors between any allocated and any recipient source must equal one, and (ii) a set of orthogonal subspaces must result from combining the subspace for a recipient source with the subspace for each allocated source confounded with it; this second condition is a form of adjusted orthogonality (Eccleston & Russell 1975).

Nonorthogonal with structure balance: To be nonorthogonal with structure balance, a design must meet the conditions for an orthogonal design, except that, while all nonzero canonical efficiency factors between an allocated and a recipient source must be equal, there must be at least one such pair of sources for which the value is not one.

Unbalanced: Unbalanced designs are those that do not meet the conditions for nonorthogonality with structure balance.

Brien & Bailey (2009) proved, for randomizations in a chain, that if each design is structure balanced then the composite design is also structure balanced. Qualifying balance with structure is done to emphasize that the balance applies to a particular structure, i.e. to particular orthogonal decompositions of the sample space. Consequently, for block designs and the structure corresponding to decompositions for (i) between and within blocks and (ii) between treatments, only balanced incomplete-block designs are structure balanced; partially balanced incomplete-block designs are unbalanced. On the other hand, Houtman & Speed (1983) give a definition of balance that enables them to claim that all block designs with equal block size exhibit general balance, there being some, possibly uninterpretable, structure on the treatments for which this is true.

In Example 1, the recipient structure for the allocation of training conditions to tests is the decomposition specified by the recipient sources for the tests. The design is nonorthogonal with structure-balance: (i) for all pairs of an allocated source and a recipient source, the nonzero canonical efficiency factors for each pair are all equal and for at least one pair the value is not one (e.g. the nonzero canonical efficiency factors for the allocated source $\mathrm{Intensities}$ and the recipient source $\mathrm{Months}\,[\mathrm{Athletes}]$ are all equal to 0.25), and (ii) when the three allocated sources are combined with $\mathrm{Months}\,[\mathrm{Athletes}]$, they remain orthogonal to each other, as they do when combined with $\mathrm{Tests}\,[\mathrm{Months}{\wedge}\mathrm{Athletes}]$.

The practical importance of these different design types relates to (i) the variances of prediction differences based on them and (ii) the complexity of the analysis and its

interpretation. Firstly, given a single treatment factor with a set number of equally replicated levels, the same number of replicates and the same sources of variation in the experiment, the average variance of prediction differences is likely to be less for an orthogonal design than for a design with structure balance, which, in turn, is likely to be less than for an unbalanced design, which leads Mead, Gilmour & Mead (2012) to recommend that designs be as balanced as possible. For example, suppose that an experiment for six treatments each replicated five times is to be run and that the block canonical component is equal to the residual variance component (σ^2). It can be shown that the average variance of pairwise differences between treatments for a randomized complete-block design will be $0.4\sigma^2$. For a balanced incomplete-block design with three plots per block, six treatments and five replicates, it will be $0.5\sigma^2$ when based on intrablock estimates and $0.47\sigma^2$ when inter- and intra-block treatment information is combined. Now consider a partially balanced incomplete block design with the same numbers of treatments, blocks, plots per block and replicates as for the balanced incomplete block, but with one treatment occurring in every block and each pair of the other five treatments occurring in one block. The average variance of pairwise differences between treatments will be $0.583\sigma^2$ when based on intrablock treatment estimates and $0.542\sigma^2$ when based on combined treatment estimates.

Secondly a structure-balanced design, orthogonal or not, has a single variance for the estimated effects for each source; unbalanced designs have variable variances for the estimated effects for at least one source. Orthogonal designs can be easily analysed using ANOVA or mixed-model analysis, while the analysis of nonorthogonal designs with structure balance may not be achievable with ANOVA; unbalanced designs usually require a mixed model analysis. Further, because the allocated sources in a structure-balanced design remain orthogonal to each other when combined with recipient sources in all cases, there is not (partial) aliasing between any sources from the same tier to complicate the interpretation of the results and the decomposition is unique.

Consequently, all other things being equal, it is preferable that the design be orthogonal; if that is not possible then a nonorthogonal design with structure balance is preferred; if even that is not achievable then an optimal unbalanced design should be sought.

Bailey & Brien (2016) distinguish between analyses that are ANOVA-applicable and those that are not. The importance of ANOVA-applicability again relates to the properties of the predictions. To be ANOVA-applicable, not only must the design be structure balanced, but the estimates of the fixed effects for each source must be obtained from a single source: the combination of information from different sources is not required. The multiphase analysis for Example 1, with $\mathrm{Athletes}$ and Months regarded as random, is not ANOVA-applicable because the predictions are obtained by combining information from $\mathrm{Months}\,[\mathrm{Athletes}]$ and $\mathrm{Tests}\,[\mathrm{Months}\wedge\mathrm{Athletes}]$, both merged with $\mathrm{Locations}\,[\mathrm{Batches}]$. The analysis would be ANOVA-applicable if Months and $\mathrm{Athletes}$ were assumed fixed so that predictions for $\mathrm{Intensities}$ and $\mathrm{Surfaces}$ were obtained using only information from $\mathrm{Tests}\,[\mathrm{Months}\wedge\mathrm{Athletes}]$ merged with $\mathrm{Locations}\,[\mathrm{Batches}]$. However, ANOVA-applicability is less important than the balance properties of a multiphase design, particularly if a mixed-model analysis is to be undertaken anyway.

4. Some history

It is now more than 60 years since the publication of McIntyre's seminal 1955 paper on multiphase experiments. As is outlined here and is evident from the bibliography of 101 items concerned with multiphase experiments provided in the Supplementary Materials, much activity has ensued since, particularly this century with 83 publications so far. The history is split into that connected with the development of structure-balanced designs and that connected with those that are not.

4.1. Structure-balanced designs

The first known description of multiphase experiments was in McIntyre (1955), which was followed by two related articles, McIntyre (1956) and Curnow (1959). Cox (1958, section 5.5) mentioned that randomizations might have to be done in phases, although he termed them stages. Brien, May & Mayo (1987, examples 2–5) reported four wine-evaluation experiments that were designed and run in the period 1972–76. In the absence of any knowledge of McIntyre (1955) they were labelled as multitiered experiments, but by the end of 1976 W. B. Hall had drawn the author's attention to McIntyre's paper and it became clear that the wine-evaluation experiments are also multiphase experiments.

Brien (1979a,b) included a multiphase sensory example and described a multitiered approach to multiphase experiments. The work was discussed and shared with T. P. Speed in 1980 and was published as Brien (1983). In these papers the term tier, which can be characterized as a set of factors in a panel of a diagram such as Figure 1, was defined and multitiered was coined to refer to experiments with more than two tiers. These papers gave symbolic mixed models and skeleton-ANOVA tables for the examples. Brien (1979c) extended Wilkinson (1970) to produce an algorithm for computing the ANOVA for data from a multitiered experiment in which the designs in both phases are nonorthogonal. Houtman & Speed (1983, section 5.5) noted that, as a class, two-phase experiments do not satisfy general balance. Wood, Williams & Speed (1988) discussed the ANOVA for the subclass of multiphase experiments in which the first-phase design is orthogonal and gave expressions for the combined estimates of treatment effects along with their variances. This was followed by Brien (1992), who derived the ANOVA with EMS for multiphase experiments, based on mixed models, in which both designs display structure balance. Also in 1992 collaboration between R. A. Bailey and the author began. Brien & Payne (1999) is the first published multiphase example in which the designs for both phases, while nonorthogonal, are structure balanced, and they showed how to compute the ANOVA for it.

Brien, Harch & Correll (1998) and Smith et al. (1999) independently realized that a common and important type of multiphase experiment is one in which there is a later laboratory phase, although the latter were unaware at the time that their experiments could be described as multiphase.

A web site for multitiered experiments was established in 2001 (Brien et al. 2001) to promulgate unpublished material on multitiered experiments. It remains available, although in the interim it has been revised and much of the content has been published. Brien & Bailey (2006) used many examples to describe different types of multiple randomizations and several of these are multiphase experiments. Brien & Bailey (2009) and Brien & Bailey (2010) characterized these types as either (i) randomizations in a chain, in which one set of

objects is randomized to another and then the second set of objects to a third and so on, or (ii) two-to-one randomizations, in which two sets of objects are randomized to a third set.

ANOVA for the full range of structure-balanced experiments with randomizations in a chain only became available with the publication of both Brien & Bailey (2009), who gave their decomposition tables, and Bailey & Brien (2016), who gave skeleton-ANOVA tables with EMS and the estimators of treatment effects and their standard errors. The corresponding analysis for experiments that involve two-to-one randomizations is yet to be published, although Brien & Bailey (2010) gave their decomposition tables.

On the other hand, an approach to the analysis of the data from multitiered experiments that is based on the fitting of randomization-based mixed models was described by Brien & Bailey (2006), it being derived from the approach in Brien (1992). Brien & Demétrio (2009) applied this approach to longitudinal data. Bailey & Brien (2016) gave the derivation of randomization-based mixed models and discussed their position in the general class of mixed models. Piepho, Büchse & Emrich (2003) also formulated a mixed model for a two-phase experiment. Bate & Chatfield (2016) proposed an approach to formulating mixed models for experiments that includes multiphase experiments. While their approach often leads to the same models as those of the Brien-Bailey approach, they include some terms and omit others compared to the Brien-Bailey approach (see their Section 8); one reason for this is that they initially include all interactions between factors from different tiers and the rules for removing terms do not necessarily omit them.

A particular innovation in Brien & Bailey (2006) was the randomization diagram, introduced to assist in visualizing multiple randomizations. These were extended to factor-allocation diagrams by Brien et al. (2011) and the conventions used in them were detailed in their Web Appendix C and on the multitiered web site (Brien et al. 2001). Brien et al. (2011) also catalogued a set of principles for designing orthogonal multiphase experiments. They suggested that design keys could be useful in designing multiphase experiment, especially in the later phases. Bailey (2016) developed this idea further. Brien (2017c) extended Brien et al. (2011) to cover nonorthogonal multiphase experiments..

4.2. Designs not structure balanced

A. B. Smith, B. R. Cullis and their collaborators have developed a distinct strand of multiphase experiments. They have been concerned with plant, mostly cereal, early-generation, breeding experiments where large numbers of lines are involved. An early public discourse on multiphase experiments by them was a presentation in 1999 (Smith et al. 1999), which was followed by Smith et al. (2001) that included analyses of multiphase experiments to detect quantitative trait loci (QTL). The experiments described were each based on a single trial. They involved spatially-optimized designs for the field phase; in the laboratory phase, replication was recommended to guard against laboratory variation. It was demonstrated that there was variation between and within days in the laboratory in respect of results on the milling of wheat; Snell et al. (2002) also demonstrated day-to-day variation in the assessment of rice traits. Cullis et al. (2003) extended the analysis to a series of trials, establishing that substantial trends can occur in the laboratory phase when barley malting quality is being assessed.

The design of single multiphase experiments to investigate quality traits from grain experiments was advanced by Smith, Lim & Cullis (2006) with the introduction of p/q-rep

designs for them. These designs are based on the p-rep (partially replicated) designs of Cullis, Smith & Coombes (2006) in which $p\%$ of the lines involved in the experiment are duplicated. In the case of p/q-rep designs, $p\%$ of the lines are replicated in the field phase and $q\%$ of the plots are duplicated in the laboratory phase. Also, the efficacy of using composite samples from the field phase was investigated in a simulation study.

Butler (2013) developed a method for design generation of multiphase experiments using optimal-design techniques. Given a mixed model for a multiphase experiment, the method is capable of developing an optimal design with partial replication in each phase.

Smith et al. (2011) incorporated the use of composite samples into single-phase experiments and Smith et al. (2015) extended the approach to suit the first phase of a multiphase experiment.

Given that the designs they used were unbalanced and often optimized for autocorrelation between units, it is no surprise that this group were the first to advocate the analysis of multiphase experiments using mixed-model software in which REML estimation is employed. Consequently, their focus has been on formulating mixed models for experiments rather than ANOVAs and Smith et al. (2001), Cullis et al. (2003), Smith, Lim & Cullis (2006) and Smith et al. (2015) developed an approach similar to that of Brien (1992), Brien & Bailey (2006) and Brien & Demétrio (2009). In particular, Cullis et al. (2003) proposed tier-based mixed models.

4.3. Application areas for multiphase experiments

The publications in which multiphase experiments have been used are listed for several application areas. For each area, the following aspects of the design are described in more detail for one example: the nature of the phases; the units for each phase; the outcomes of each phase; the designs employed in each phase; and the types of allocations used.

4.3.1. Plant pathology

The main example in McIntyre (1955) was a plant pathology experiment, while Cowley et al. (2012), and subsequent papers from this group, advocated for multiphase experiments in plant pathology. Peacock et al. (2003) described a two-phase plant pathology experiment, although it was not recognized as such at the time of publication. In general, one phase might be in a greenhouse and the second phase in a laboratory or both phases might occur in laboratories.

Experiments 2–5 in Cowley et al. (2012) investigated the resistance of 32 lupin lines to leaf and pod blight. The units in the first phase were pots in a greenhouse and, for the detached-leaf assay in the second phase, were petri dishes, each containing a leaf from the first phase. The petri dishes were placed into a shelf-unit in a culture room, the shelf-units consisting of shelves stacked vertically and containing a rectangular grid of dishes. The outcome of the first phase was leaves and that of the second-phase was a set of responses from the assessment of the leaves for disease symptoms. The first-phase design was a randomized complete-block design and is orthogonal. The second phase used a resolved row-column design that was most likely unbalanced and involved replication of the units from the first phase in that several leaves were taken from each pot. Both allocations were randomized and composed.

4.3.2. Greenhouse/growth chamber experimentation:

Preece (1991) and Brien et al. (2013) discussed two-phase experiments in which both phases occur in greenhouses: the first phase might be an establishment phase after which plants are moved to a new location for a growth phase. Sometimes a treatment, such as exposure to heat, applied to some of the plants in a greenhouse experiment might involve their spending a period in a growth chamber. Molenaar et al. (2017) reported an experiment in which plants were cultivated in the first phase and cuttings from that phase were rooted in the second phase, while Boxriker et al. (2017b) and Boxriker et al. (2017a) described two-phase experiments in which the first phase occurred in a greenhouse and the second phase in a laboratory.

Two-phase experiment I in Molenaar et al. (2017) involved 500 Pelargonium lines and the two phases were conducted in different greenhouses. The first-phase units were pairs of plants for a line and the second-phase units were the cuttings from a pair of plants. The outcomes of the first phase were cuttings and several longitudinal response variables for the pairs of plants; the outcome of the second phase was the response variable "root formation" measured on three dates. The first phase design was a resolved incomplete-block (α-)design and the second-phase design was a pragmatic design that involved systematic placement of the replicates from the first phase and the randomization of pairs of plants from the first phase to areas within trays, the trays being arranged on tables. It was a matter of biological necessity that the arrangement of lines into areas on trays and tables was not optimized. Both designs were unbalanced and the allocations composed.

4.3.3. Product storage

Wood, Williams & Speed (1988, section 4.2) described a milk-storage experiment, Brien & Bailey (2006, example 13) described a potato storage experiment and Wilkinson et al. (2008) and Hardner et al. (2016) both analysed apple storage experiments with four phases. Generally, the first phase was a production phase and subsequent phases were storage phases.

Experiment 2 from Wilkinson et al. (2008) investigated the effect of the chemical 1-MCP on apples in storage. The four phases in the experiment were orchard, treatment, storage and export phases. The first-phase units were bags of 50 apples and the second-phase units were treatment cartons of 50 apples; the third-phase units were storage positions and the fourth-phase units were export positions. The outcome of the first phase was bags of harvested fruit, of the second phase was cartons of treated and non-treated fruit, of the third phase was cartons of stored fruit and of the final phase was (i) cartons of exported fruit and (ii) two response variables: fruit firmness and browning. The design for the first phase was an observational study in which apples were sampled from 12 orchards. The second phase involved two independent randomizations: (i) a randomized complete-block design with subsampling was used to assign the 1-MCP and Control to treatment tubs; and (ii) an orthogonal design involving nested randomizations was used to allocate the bags of fruit to the treatment cartons. Orthogonal designs with nested randomizations were then used to assign treatment cartons to storage positions and storage positions to export positions. While the randomizations in the treatment phase were independent, the randomizations in the storage and export phases were composed.

4.3.4. Sensory evaluation

Kempthorne (1952, pp.579–581) described a sensory evaluation experiment that was a two-phase experiment, albeit unrecognized as such. The first phase was a food storage experiment and the second phase was the evaluation phase. There was no mention of what occurred in the food production stage that would have preceded the storage experiment. Presumably it did not involve a designed experiment, which would have contributed another phase to make the experiment three-phase. Brien (1983), Brien, May & Mayo (1987), Brien (1992), Smith et al. (2003), Brien & Bailey (2006) and Brien et al. (2011) all recognized that sensory experiments can be two-phase. In these examples the first phase was a field trial and the second phase was an evaluation phase in which produce from the field-phase plots was evaluated. The fourth stage of the storage experiment reported by Hardner et al. (2016) was a sensory evaluation phase. In many sensory experiments the first phase is a production phase that is not subject to an experiment and variability in that phase is largely ignored.

Consider the two-phase sensory experiment first published in Brien & Payne (1999) and discussed in Brien & Bailey (2006, 2009) and Bailey & Brien (2016). It investigated the effects of different trellis and pruning levels on grape vines and involved a field and an evaluation phase. The units in the field phase were half-plots of vines and in the evaluation phase were table positions of glasses of wine. The outcomes of the first phase were the grapes from a half-plot and their weight and the outcome of the second phase was the score for a wine made from grapes from the field phase. The designs for both phases were nonorthogonal with structure balance and the two randomizations were composed. The production phase, in which wine was produced from the grapes, did not employ an experimental design, although one can imagine that order effects in this phase may well occur and so a design for it could be beneficial.

4.3.5. Gene-expression studies

In 2002 C. M. Triggs (pers. comm.) first pointed out to the author that microarray experiments are multiphase. Kerr (2003) recognized that microarray experiments are two-phase and Jarrett & Ruggiero (2008) addressed their two-phase nature. Milliken, Garrett & Travers (2007) reported on a two-phase microarray experiment, without realizing that it was two-phase; the general approach of Brien & Bailey (2006) and Brien et al. (2011) was used by Brien (2008, 2011a) to formulate its analysis and those of other gene-expression experiments. The first phase is usually a production phase and the second phase entails the preparation and measurement of the arrays. Brien (2011b), recognizing that the preparation and measurement of arrays involves several stages that can be subject to batch effects, outlined the potential to divide this process into several phases and block for batch effects. Nettleton (2012, 2014) also discussed the design of gene-expression experiments, including next-generation sequencing experiments and the possibility of multiphase experiments.

The example in Jarrett & Ruggiero (2008, section 2) is a two-phase, two-channel microarray experiment in which the multiphase design is not structure balanced. The first phase was a field phase in which plants are grown in plots and the second phase was a laboratory phase in which two samples from each plant were processed and allocated to a dye on a microarray. The outcome of the first phase was a plant sample and that of the second phase is a $\log_2(\text{Fluorescence Intensity})$ value for a dye from an array. The design for the first phase was a balanced incomplete-block design, which is nonorthogonal with structure

balance. For the second phase a Youden square design was used to assign plots to array-dye combinations and the resulting design is nonorthogonal with just first-order balance, not structure balance. It is true that each allocated source in the samples to plants allocation had a single value for the nonzero canonical efficiency factors with each recipient source. However, the allocated sources $\mathrm{Plants}\,[\mathrm{Blocks}]$ and $\mathrm{Samples}\,[\mathrm{Blocks}\wedge\mathrm{Plants}]$ are not orthogonal to each other when estimated from the recipient sources arising from Arrays and Dyes, making the allocation unbalanced (for more details see Brien 2017c, example 3). Nonetheless, the two randomizations are composed.

4.3.6. Human

Many medical, psychological and other experiments involving humans have been multiphase experiments, but unrecognized as such. The athlete training experiments, used in some of the examples in Brien et al. (2011) and in the examples in this paper, illustrate the manner in which multiphase experiments can arise. Walwyn & Roberts (2010) described a situation in psychotherapy experiments that required a multiphase experiment, although their distinction between composed and randomized-inclusive randomizations was not quite right (Brien 2012). See Example 1 for a detailed example.

4.3.7. Laboratory analysis of crops

Plant breeding experiments have been prominent in the application of multiphase experiments that involve a laboratory phase to analyse the produce from a field phase. Many of the published examples stem from the work described in Section 4.2, as is evidenced by the following references: Smith et al. (1999, 2001); Eckermann et al. (2001); Cullis et al. (2003); Lever et al. (2005); Smith, Lim & Cullis (2006); Osborne et al. (2007); Raman et al. (2009); Cavanagh et al. (2010); Diepeveen et al. (2012); Fox et al. (2013); Oakey et al. (2013); Foito et al. (2015); Smith et al. (2015); Elias et al. (2016). Plant breeding experiments can involve the phenotyping of lines for traits that require a multiphase experiment. The first phase is a field trial from which yield and other traits can be measured. This phase generally exhibits spatial variation such that neighbouring plots are correlated and so the design needs to be optimized for the estimation of such variation. Then follows one or more laboratory phases in which quality traits of the harvested grain are to be measured. For example, for wheat, there may be a milling phase in which the grain is milled to produce flour; subsequently it may be made into dough and perhaps into bread; for barley, the laboratory phase may involve malting the grain. It has been established that there is often considerable variation in the laboratory processes involved and Smith et al. (2001), Cullis et al. (2003), Smith, Lim & Cullis (2006) and Smith et al. (2015) have argued that this needs to be taken into account in the design and analysis of experiments to maximize their precision. Brien (2017c) gave a detailed treatment of a two-phase plant breeding experiment that involves field and milling phases.

Flakelar et al. (2017) described a four-phase experiment to compare the bioactive compounds in the oil from 64 canola genotypes at two sites. The first phase comprised a field trial at each of two sites and the subsequent phases were stages in the chemical analysis to determine the amounts of the compounds. The only reported outcome of the first phase was subsamples of seed from each plot and these subsamples formed the units for the first phase. The second phase was a milling and oil extraction phase, an outcome of which was the extracted oil aliquots that became the units for that phase. The third phase was an oil dilution

and preparation phase resulting in prepared samples. The fourth phase was an analysis phase using High Performance Liquid Chromatography (HPLC); the outcomes were peak areas for a number of analytes from several classes of bioactive compounds — carotenoids, tocopherols, phytosterols and photosterol esters. All stages used partially replicated designs (Cullis, Smith & Coombes 2006), which are unbalanced. The designs for the extraction and preparation phases allowed for the effects of time of processing and multiple instances of the same experimental apparatus, while the HPLC analysis phase allowed for time of processing and positional effects in an autosampler tray. The authors concluded that there is considerable variability arising in the HPLC analysis phase and that the use of multiphase designs is an effective means for analytical chemists to deal with instrument drift or malfunction.

4.3.8. General points

Clearly multiphase experiments often involve two or more experimental designs for assigning allocated objects, and factors, to recipient objects, and factors. That is, single-allocation experimental designs are combined to produce the multiphase design, with orthogonal, nonorthogonal but structure-balanced and unbalanced designs being displayed amongst the designs used. It has been demonstrated that each phase provides outcomes and that the units differ between phases. Sensory, microarray, human and plant breeding experiments potentially include a laboratory phase (Brien et al. 2011). In the past, there has been a reluctance to consider experimental design for the laboratory phase. Some reasons are that (i) researchers do not recognize that trends occur in the laboratory phase, (ii) it is not realized that trends can be allowed for using appropriate experimental designs, and (iii) laboratory analyses are often outsourced and so specification of the order of laboratory analyses is not an option.

5. Accomplishing randomized layouts and analyses

Here some currently available methods and tools for designing and analysing multiphase experiments are described and illustrated with an example. A general point that will become apparent is that, rather than employing tools specialized to multiphase experiments, the appropriate approach consists in adapting tools that are used for designing and analysing standard experiments. Thus, as for standard experiments, mixed models of the following form (Bailey & Brien 2016, section 11.2) are fundamental to both the design and analysis of multiphase experiments:

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\tau} + \mathbf{Z}\mathbf{U} + \mathbf{E},$$

with $\mathrm{E}(\mathbf{Y} \mid \mathbf{U}) = \mathbf{X}\boldsymbol{\tau} + \mathbf{Z}\mathbf{U}$, $\mathrm{cov}(\mathbf{U}) = \mathbf{G}$ and $\mathrm{cov}(\mathbf{Y} \mid \mathbf{U}) = \mathrm{cov}(\mathbf{E}) = \mathbf{R}$, where $\boldsymbol{\tau}$ is the vector of fixed-effects parameters, $\mathbf{X}$ is an indicator-variable matrix for fixed effects with one row for each observation and a column for each fixed effect, $\mathbf{Z}$ is an indicator-variable matrix with a row for each observation and a column for each random effect, $\mathbf{U}$ is the vector of random effects, $\mathbf{E}$ is the vector of random effects for the identity term(s), and $\mathbf{G}$ and $\mathbf{R}$ are symmetric matrix functions of the variance parameters. Identity terms are those whose factors uniquely index the observational units.

This conditional model can be re-expressed in the following marginal form:

$$\mathrm{E}(\mathbf{Y}) = \mathbf{X}\boldsymbol{\tau} \quad \text{and} \quad \mathrm{cov}(\mathbf{Y}) = \mathbf{V} = \mathbf{Z}\mathbf{G}\mathbf{Z}^{\top} + \mathbf{R}, \tag{1}$$

where $\mathbf{V}$ is positive semidefinite. For the set $\mathcal{F}$ of fixed terms and $F \in \mathcal{F}$, both $\mathbf{X}$ and $\boldsymbol{\tau}$ are conformably partitioned into sets $\{\mathbf{X}_F\}$ and $\{\boldsymbol{\tau}_F\}$ so that there is a submatrix and a subvector, respectively, for each fixed term. Likewise, for the set $\mathcal{U}$ of random terms, excluding identity terms, and $U \in \mathcal{U}$, $\mathbf{Z}$, $\mathbf{U}$ and $\mathbf{G}$ are conformably partitioned into the sets $\{\mathbf{Z}_U\}$, $\{\mathbf{U}_U\}$ and $\{\mathbf{G}_U\}$. This partitioning can be specified via symbolic mixed models that list the fixed terms to the left of a vertical line ('|') and random terms to its right. Identity terms are underlined, as in Brien & Bailey (2006). The notation in Brien & Demétrio (2009, table 1) is used.

5.1. Designing

Figure 2 shows a sequence of items recommended for production during the design process. It is based on the approaches of Brien & Demétrio (2009), Brien et al. (2011) and Bailey & Brien (2016). In contradistinction to their approaches, which begin with the design, Figure 2 introduces the anticipated (mixed) model for an experiment as a device to make the model explicit in the choice of a design. Also Figure 2 depicts the derivation of models based on the allocation of factors in the design, not requiring that the allocation be solely by randomization. The focus on the mixed model at the design stage is consistent with the preferred method of analysis being mixed model analysis, as will be argued in Section 5.2. This emphasis conforms to the principle that the analysis should be formulated at the design stage of an experiment.

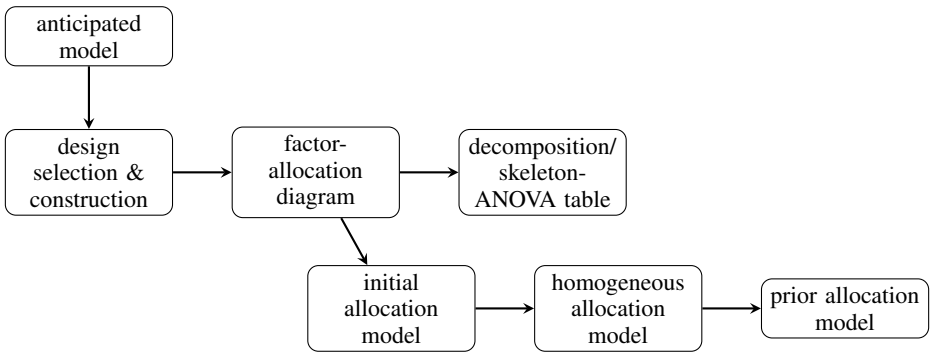

Figure 2. Flowchart of items to produce in designing an experiment.

Designing an experiment begins by determining the anticipated model for the experiment. Except in optimal design theory for experiments to fit a mathematical function, it has been rare for this model to be explicitly stated, it usually having a subliminal presence. To determine its terms often requires consultation with the researcher as to what effects might occur. From this model, the crossing and nesting relations between the factors in the experiment can be deduced. For example, consider a phase in which a set of plots arranged in a grid are to have one or more treatment factors allocated to them. Indexing the plots with the factors Rows and Columns seems reasonable. There being no further plot factors, one can consider the set of terms that is the complete set of combinations of these factors: $\{\mathrm{Rows}, \mathrm{Columns}, \mathrm{Rows}{\wedge}\mathrm{Columns}\}$. Suppose that the researcher is adamant that column differences will not occur in the area where the experimental plots are located, but that row differences are likely. Then Columns would be omitted from the anticipated model and Rows

retained so that Columns is nested within Rows. Another possibility is that of autocorrelation between plots in the same row or plots in the same column. One will also need to consider the model for the factors to be allocated, the treatment factors. Perhaps two or more crossed treatment factors are to be included.

A design is then decided upon and constructed. What is required is a design for each phase that allocates one or more sets of factors to another set. The usual objective is that this assignment should result in a design that leads to as efficient an experiment as is possible in light of the relative magnitudes of the effects in the anticipated model: that is, a design that is optimal for this mixed model. Brien et al. (2011, Principle 6) recommended that all the designs be constructed prior to commencing the experiment, but, as they, Smith, Lim & Cullis (2006) and Smith et al. (2015) suggested, this is not always possible.

To generate each design, currently available methods for generating single-allocation designs can be used: (i) deploying a straightforward standard design, like a randomized complete-block design; (ii) manually constructing a design, including the use of design keys (Brien et al. 2011, section 7.5) (Bailey 2016); (iii) consulting a catalogue of designs; or (iv) searching for a design using computer algorithms.

If only the systematic plan for one or more of the designs has been obtained using one of these methods and the allocation is to employ randomization, the allocated objects and factors will need to be randomized. The algorithm described in Section 3 has been implemented in the R package (R Development Core Team 2017) dae (Brien 2017b), via the function `designRandomize`, and in GenStat, via the procedure `randomize` (VSN International 2015). Their use for multiphase experiments is exemplified in Supplementary materials A and Supplementary materials B, where scripts for Example 1 and 2 are provided.

In terms of computer searches for designs, one approach is to identify, given the anticipated model, an appropriate class of design and to attempt to determine the optimal design in that class. For example CycDesigN (VSN International 2013) does this for several classes of designs, including incomplete-block, row-column and spatial designs. It maximizes the average efficiency factor which, for equireplicated designs, is equivalent to minimizing the A-optimality criterion (mean of the variances of pairwise differences) (Williams & Piepho 2015) for the model on which the class of designs is predicated. Another approach is to search for a design that is optimal for the anticipated mixed model, as do the R-packages od (‘optimal design’) (Butler 2013) and DiGGer (Coombes 2009); also, there is `PROC OPTEX` in SAS (SAS Institute Inc. 2015). The package od has the capability for producing designs for multiphase experiments (Smith et al. 2015). For a two-phase design, one can obtain the first-phase design, possibly using od, and then use od to obtain the second-phase design, without modifying the first-phase design. A search can be made for the design that minimizes the A-optimality criterion, given a mixed model with specified variance parameter values (Butler 2013; Smith et al. 2015). The function `designRandomize` from dae can be used to provide a randomized starting design.

Given either a systematic or randomized layout for any experiment, its anatomy, in the form of the decomposition table, can be produced using the function `designAnatomy` from dae in R or the procedure `ACANONICAL` (Brien 2017a) in GenStat. Besides a layout, this software requires a model formula for each panel in the factor-allocation diagram. When one or more designs are not structure balanced, the decomposition table is not unique in that it depends on the order in which terms are fitted, with the source for a term being adjusted for all sources that occur above it in the decomposition table. For structure-balanced experiments

and those for which only the first phase is unbalanced and all first-phase, allocated factors are fixed, the decomposition table can be converted manually to a skeleton-ANOVA table by adding the EMS to it; the expressions for the EMS are in Bailey & Brien (2016, section 4) and rules for adding them are in Brien et al. (2011, Web Appendix D).

As indicated in Figure 2, the factor-allocation diagram also provides a basis for deriving the mixed model for the experiment, along the lines of Brien & Demétrio (2009). The three stages involving allocation models in Figure 2 correspond to Stages 1–3 of Brien & Demétrio (2009) respectively, albeit with some renaming of models. The terms in the initial allocation model are obtained from the factor-allocation diagram by forming all possible combinations of the factors, given the crossing and nesting relationships within each panel. The fixed component of the model is then the sum of the terms involving factors allocated in the first allocation and the random component is the sum of the remaining terms. The derivation of the homogeneous allocation model involves swapping terms in the initial allocation model between the fixed and random components of the model as in Stage 2 of Brien & Demétrio (2009). The prior allocation model is obtained by reparameterizing terms in the homogeneous allocation model as in Stage 3 of Brien & Demétrio (2009); for example, autocorrelation or unequal variances might be incorporated.

If the design is constructed to be consistent with the anticipated model then the anticipated and prior allocation models should be the same and comparing them provides a check on the suitability of the design. When only randomization that respects the crossing and nesting relationships between the factors is used in allocating sets of objects, and their factors, and the terms in the model reflect these relationships then the initial allocation model is designated as the randomization model. When at least one allocated factor has not been randomized, but the factors to which they are allocated have been taken into account in the design construction, the initial allocation model is termed a quasi-randomization model. On the other hand, if not all the factors and their nesting relationships are taken into account in the design construction, the initial allocation model is called an assumed model. The initial allocation model for Example 1 is a quasi-randomization model.

5.2. Analyzing

As noted in Section 3, Bailey & Brien (2016, section 8) characterized some experiments as ANOVA-applicable. However, the availability of software for doing a multiphase ANOVA is limited. The only software known to the author is the procedure `AMTIER` (Brien & Payne 2012) in GenStat (VSN International 2015), but it does not produce predictions.

The use of mixed-modelling software, which is widely available, appears to be a more viable option and has the advantage that the same principles as are used for single-allocation experiments also apply to multiple-allocation experiments. Smith, Lim & Cullis (2006), Brien & Bailey (2006), Brien & Demétrio (2009) and Bailey & Brien (2016) advocated the use of randomization-based mixed models for analysis, with Bailey & Brien (2016, section 11) having a detailed discussion of different classes of mixed models. By a randomization-based mixed model is meant a model that is formulated by beginning with the terms appropriate given the randomization, as in Stage 1 of Brien & Demétrio (2009). However, as has been seen, not all factor allocation is random and then allocation-based mixed models described in Section 5.1 are employed. The prior allocation model provides a starting point.

A problem with mixed models is that, as mentioned in Example 1, often the prior allocation model cannot be fitted because not all variance parameters are estimable. Decomposition/skeleton ANOVA tables can warn of such problems. Sources not present in a table, because they are aliased with sources from the same tier, need to have their terms omitted for a successful fit of the model. A source that has other sources confounded with it and has no Residual source is said to be exhausted by them and is an alert that the model may not fit. It is impossible to fit all terms when the numbers of objects in two or more tiers equals the number of observational units as this produces exhausted sources (see Brien 2017c). In Example 1, all the sources in the locations tier are exhausted and as a result none of ϕ_{BL}, ϕ_{MAT}, ϕ_{B} and ϕ_{M} is estimable. For the canonical component for a source to be inestimable, it is not sufficient that the source be exhausted. However, if the sources confounded with an exhausted source are confounded only with it then its canonical component and one or more of the those for the other sources are not separably estimable. When some canonical components are not estimable, as suggested in Example 1, one or more terms will have to be removed and the resulting model is a 'model of convenience'. Further, convergence of the fitting for a particular data set can be problematic — see Bailey & Brien (2016, section 11.3).

5.3. Illustrative examples

In addition to a reprise of Example 1, a second nonorthogonal example illustrates the the application of the process for designing multiphase experiments described in Section 5.1.

Example 1 (continued) (A nonorthogonal two-phase athlete training experiment). This example involves structure-balanced designs in both phases, the design in the second phase being orthogonal. The first-phase design was obtained by consulting a catalgoue of designs and the second-phase design is a standard design. The need for an incomplete-block design in the first phase is likely to arise for practical reasons, such as the anticipation that some athletes would lose enthusiasm over time and so not complete all nine tests. Hence an incomplete-block design with structure balance in the first phase and an orthogonal design in the second phase results in a multiphase design that is optimal for the anticipated mixed model.

An R script that shows how to use the `designRandomize` function to do the randomization for this multiphase design is given in Supplementary materials A.1 and a GenStat script that uses `randomize` is in Supplementary materials A.2. Also included is the use of the R function `designAnatomy` and the GenStat procedure `ACANONICAL` to produce the decomposition table on which Table 1 is based.

Mixed model software is likely to be used to analyse the data from an experiment based on the design, but ANOVA could be used if an 'intrablock' analysis is all that is required.

Example 2 (A two-phase athlete training experiment with lab-order effects). For this example, as in Example 2 from Brien et al. (2011), it is presumed that primary interest is in surface differences, with intensities included to observe the surfaces over a range of intensities. Otherwise the situation in the first phase is the same as for Example 1. However, the difference means that a split-plot design is now appropriate for the first phase, with months acting as blocks, intensities randomized to athletes within a month and surfaces randomized to tests within each athlete within a month. Such a design will be optimal for allocation of (i) Intensities and (ii) Surfaces as described.

In the second phase of both Example 2 from Brien et al. (2011) and Example 1, Locations is innately crossed with Batches, yet a design in which Locations is regarded

Table 2. Row-column design for assigning 3 Intensities × 3 Surfaces to 4 Batches × 9 Locations in the second phase of Example 2, a two-phase athlete training experiment with lab-order effects.†

	Locations								
Batches	1	2	3	4	5	6	7	8	9
1	1,2	1,3	1,1	2,2	2,3	2,1	3,2	3,3	3,1
2	1,3	1,1	1,2	2,3	2,1	2,2	3,3	3,1	3,2
3	2,1	2,2	2,3	3,1	3,2	3,3	1,1	1,2	1,3
4	3,1	3,2	3,3	1,1	1,2	1,3	2,1	2,2	2,3

†The pair of numbers in a cell of the table is the Intensity level followed by the Surfaces level to be applied to the Batch-Location combination for that cell.

as nested is proposed. This is appropriate when no consistent Locations differences across the Batches are foreseen and so the anticipated model would not include the term Locations. Here it is accepted that laboratory-processing order needs to be taken into account so that the anticipated model would include Locations:

$$\text{Batches} + \text{Months} + \text{Intensities} + \text{Surfaces} + \text{Intensities}\wedge\text{Surfaces} \quad | \\ \text{Locations} + \underline{\text{Batches}\wedge\text{Locations}} + \text{Months}\wedge\text{Athletes} + \underline{\text{Months}\wedge\text{Athletes}\wedge\text{Tests}}.$$

The terms Batches and Months are included in the fixed component of the model because they do not have sufficient degrees of freedom to enable their canonical components to be estimated with any precision.

This anticipated model dictates the use of a row-column design with Batches corresponding to the rows and Locations to the columns. In constructing the new design, prime consideration needs to be given to Intensities and Surfaces, rather than ignoring them and considering only Months, Athletes and Tests as in Example 1. R. A. Bailey (pers. comm.) proposed the systematic version of the balanced factorial design (Hinkelmann & Kempthorne 2005, section 12.5) given in Table 2 to assign two three-level factors to 4 rows × 9 columns, such that the split-plot nature of the first-phase design is retained. It is constructed by dicyclic development of an initial column (John & Williams 1995, section 3.5) that consists of four pairs of numbers, the first number in a pair being a level of Intensities and the second being a level of Surfaces. The initial column has (i) all three levels and a repeated level of Intensities and of Surfaces; and (ii) the repeated level of Intensities occurs in two pairs and that of Surfaces in the other two pairs. The remaining eight columns are obtained by cyclic development of the levels of the two factors in turn; that is, adding one modulo three to the previous column of levels for Surfaces and then adding one modulo three to the previous set of three columns of levels for Surfaces.

To complete the design for the experiment the Athletes and Tests in each levels of Months have to be associated with the Intensities and Surfaces combinations. For example, the athlete for the first month that was assigned level 1 of Intensities is associated with the first three combinations in the first row of Table 2; then the levels of Tests for this athlete are associated with the Surfaces levels according to which surface was assigned to which test in the first phase. This is repeated for all combinations in the table. This allocation of Athletes will result in one of 3 groups of 4 Athletes being allocated to each of Locations 1–3, 4–6

and 7–9. Considering the combinations of Intensities and Surfaces as nine Conditions, the allocation of training conditions to locations is optimal because Conditions is orthogonal to Batches and is structure balanced with respect to the Locations, it not being possible to obtain a design in which Conditions is orthogonal to Locations.

The factor-allocation diagram for this experiment is in Figure 3. Again it shows two randomizations in a chain. However, now the randomization of tests to locations cannot be done in ignorance of the outcome of the randomization of training conditions to tests. The two randomizations are randomized-inclusive randomizations (Brien & Bailey 2006, 2009). The initial allocation model for it, which is a quasi-randomization model, is the same as the anticipated model, except that Batches and Months are included in the random component of the model.

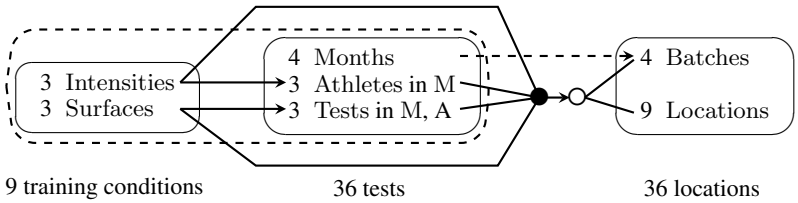

Figure 3. Factor-allocation diagram for Example 2, a two-phase athlete training experiment with lab-order effects: training conditions are randomized to tests, then training conditions and tests are randomized to locations; the '●' indicates that the observed combinations of the levels of Intensities, Surfaces, Athletes and Tests are randomized to locations; the '○' indicates that a nonorthogonal design was used in this randomization to the combinations of the levels of Batches and Locations; the dashed arrow indicates that Months were systematically allocated to Batches; the dashed oval indicates that all factors from the first phase form a pseudotier and all are actively involved in determining the allocation to locations; M = Months and A = Athletes.

The anatomy of this design, as summarized in its decomposition table, can be obtained using `designAnatomy`, which requires the systematic or randomized layout and three model formulae derived from the panels in Figure 3: `Batches*Locations`, `Months/Athletes/Tests` and `Intensities*Surfaces`. An R script is in Supplementary materials B.1. It can also be obtained using `ACANONICAL` and a GenStat script is available in Supplementary materials B.2. The skeleton-ANOVA table is derived from the decomposition table by including the manually formulated EMS; see Table 3. It shows that most of the information about Intensities, Surfaces and their interaction is orthogonal to Batches and Locations. However, on noting that the two sources for Athletes [Months] have efficiency one, it is clear that the eight degrees of freedom for Athletes [Months] has been split into two orthogonal subspaces: two degrees of freedom confounded with Locations and the remaining six with Batches # Locations; the two degrees of freedom correspond to differences between the three groups of Athletes confounded with Locations. Consequently the degrees of freedom for the Residual associated with Intensities have been reduced from six to four, which obeys Multiphase law 1 (DF never increase) from Brien et al. (2011) and Brien (2017c). Also, the 24 degrees of freedom for Tests [Months∧Athletes] have been split into two orthogonal subspaces with six and 18 degrees of freedom.

In this example, the first-phase design is orthogonal and, with the introduction of pseudosources for Athletes [Months] and Tests [Months∧Athletes], the tests sources remain orthogonal to locations sources in the second phase. Nevertheless there is linear

Table 3. Skeleton ANOVA for Example 2, a two-phase athlete training experiment with lab-order effects: B = Batches; L = Locations; M = Months; A = Athletes; I = Intensities; S = Surfaces.

locations tier		tests tier			training-conditions tier		EMS‡						
Source	df	Source	df	eff†	Source	df	ϕ_{BL}	ϕ_{L}	ϕ_{B}	ϕ_{MAT}	ϕ_{MA}	ϕ_{M}	$q(\cdot)$
Mean	1	Mean	1	1	Mean	1	1	4	9	1	3	9	$q(\mu)$
Batches	3	Months	3				1		9	1	3	9	
Locations	8	Athletes [M]	2	$\frac{1}{16}$	Intensities	2	1	4		1	3		$\frac{1}{16}q(\mathrm{I})$
		Tests [M∧A]	6	$\frac{1}{16}$	Surfaces	2	1	4		1			$\frac{1}{16}q(\mathrm{S})$
				$\frac{1}{4}$	I # S	4	1	4		1			$\frac{1}{4}q(\mathrm{IS})$
B#L	24	Athletes [M]	6	$\frac{15}{16}$	Intensities	2	1			1	3		$\frac{15}{16}q(\mathrm{I})$
					Residual	4	1			1	3		
		Tests [M∧A]	18	$\frac{15}{16}$	Surfaces	2	1			1			$\frac{15}{16}q(\mathrm{S})$
				$\frac{3}{4}$	I # S	4	1			1			$\frac{3}{4}q(\mathrm{IS})$
					Residual	12	1			1			

†Each Eff is the value of the nonzero canonical efficiency factors for a training-conditions source confounded with a tests source.
‡Each ϕ is a canonical component that, except for ϕ_{BL} and ϕ_{MAT}, is allowed to be negative. Their subscripts are comprised of the first letter of each factor in the corresponding term and the numbers in the table are the coefficients of the canonical components in the EMS. The q-functions are the same quadratic functions of the expectation as the corresponding mean squares are of the data.

dependence amongst the canonical components and so the variance structure for the experiment is a linearly-dependent, commutative variance structure (LDCVS) (Bailey & Brien 2016). Furthermore the second-phase design has induced nonorthogonality between training-conditions sources and those sources derived from locations and tests. Nonetheless the design is structure balanced and an intrablock analysis, in which only the training-conditions information associated with Batches # Locations is used, could be achieved using ANOVA. On the other hand, a full analysis of data from an experiment employing the design requires the use of mixed-modelling software. The initial allocation model needs modifying as per Stages 2 and 3 of Brien & Demétrio (2009). In particular, as in the anticipated model, Batches and Months should be designated as fixed rather than random. In practice this change will have little impact because no training-conditions information is associated with them. It will result in the prior allocation model being the same as the anticipated model. Another point is that the same canonical components as in Example 1 are not separately estimated and, as outlined for that Example, two terms will need to be omitted and a 'model of convenience' fitted.

6. Discussion

Multiphase experiments involve two or more, possibly overlapping, stages of experimentation and the units involved in each phase are different; each phase produces one or more outcomes (Brien et al. 2011, section 4). In designing multiphase experiments, an experimental or sampling design is required for each phase. In some ways, the selection of a design for a phase has much in common with standard or single-allocation studies so that many of the classical experimental and sampling design principles and techniques apply. To account for early phases when designing later phases, all factors from the earlier phases are allocated to the recipient factors in the later phases using the design chosen. In cases like Example 1, the allocation for previous phases plays no explicit role in the design for the later phases; only the recipient factors from the first phase need be involved in the design for the second phase, all other factors from previous phases follow, according to their allocation, these recipient factors. In other cases, more than just the recipient factors from previous phases are involved in the design. In Example 2 both the allocated and recipient factors from the first phase had to be taken into account in the design for the second-phase. This is the difference between composed and randomized-inclusive randomizations (Brien & Bailey 2006).

Our focus has been comparative multiphase experiments, comparative experiments being those that investigate treatment differences or contrasts. The emphasis in the paradigm presented in Section 5 is a little different from the approach in Brien & Demétrio (2009), Brien et al. (2011) and Bailey & Brien (2016), where the starting point was a design to be used as an example and the objective was to formulate a (prior) mixed model for an experiment based on that design. In Section 5 it was asserted that factor-allocation diagrams have their origin in an anticipated model and that the objective is to find a design that is optimal for this model. Making this explicit emphasizes the commonality of the design of comparative experiments and the design of experiments whose purpose is to fit a mathematical function to an experiment (Goos 2002; Atkinson, Donev & Tobias 2007), such as response surface experiments. A particular feature of comparative experiments is that, usually, the model for which an optimal design is sought is a mixed model and, for this, A-optimality and related criteria are favoured. An advantage of using optimal design software like the R package od (Butler 2013) and `PROC OPTEX` in SAS (SAS Institute Inc. 2015) to do this is that it makes the specification of the anticipated model explicit for comparative experiments. For multiphase designs, the designs are constructed sequentially, starting with the first phase and proceeding to the last. For each phase, the designs for the previous phases are fixed and the model for which an optimal design is sought contains the terms for all phases up to and including the current phase. Thus, the resulting design takes into account the designs for the previous phases, in a manner analogous to randomized-inclusive randomization.

The tools for designing experiments described in detail herein have yet to be applied to experiments that are not structure balanced. Further, Brien et al. (2011) gave a set of principles for designing multiphase experiments using orthogonal designs, but the same has not been provided for nonorthogonal designs. These matters are dealt with in Brien (2017c).

Supporting information

Additional supporting information may be found in the online version of this article at `http://wileyonlinelibrary.com/journal/anzs`:

Appendix A. Example 1, a nonorthogonal two-phase athlete training experiment.
Appendix B. Example 2, a two-phase athlete training experiment with lab-order effects.
Supplementary bibliography for multiphase experiments.

References

ATKINSON, A.C., DONEV, A.N. & TOBIAS, R.D. (2007). *Optimum Experimental Design, with SAS*. Oxford Statistical Science Series, Oxford: Oxford University Press.

BAILEY, R.A. (2016). Design keys for multiphase experiments. In *mOda 11—Advances in Model-Oriented Design and Analysis*, eds. J. Kunert, C. Müller & A.C. Atkinson. Contributions to Statistics, Cham: Springer, pp. 27–35.

BAILEY, R.A. & BRIEN, C.J. (2016). Randomization-based models for multitiered experiments. I. A chain of randomizations. *Annals of Statistics* **44**, 1131–1164.

BATE, S.T. & CHATFIELD, M.J. (2016). Using the structure of the experimental design and the randomization to construct a mixed model. *Journal of Quality Technology* **48**, 365–387.

BOXRIKER, M., BOEHM, R., KREZDORN, N., ROTTER, B. & PIEPHO, H.P. (2017a). Comparative transcriptome analysis of vase life and carnation type in *Dianthus caryophyllus* l. *Scientia Horticulturae* **217**, 61–72.

BOXRIKER, M., BOEHM, R., MÖHRING, J. & PIEPHO, H.P. (2017b). Efficient statistical design in two-phase experiments on vase life in carnations (*Dianthus caryophyllus* l.). *Postharvest Biology and Technology* **128**, 161–168.

BRIEN, C.J. (1979a). Analysis of variance tables based on experimental structure. I. A method and some insights. Unpublished manuscript.

BRIEN, C.J. (1979b). Analysis of variance tables based on experimental structure. II. Complex experiments. Unpublished manuscript.

BRIEN, C.J. (1979c). Orthogonal decomposition and analysis sequence for multitiered experiments. Unpublished manuscript.

BRIEN, C.J. (1983). Analysis of variance tables based on experimental structure. *Biometrics* **39**, 53–59.

BRIEN, C.J. (1992). Factorial linear model analysis. Ph.D. thesis, Department of Plant Science, University of Adelaide, Adelaide, South Australia. URL http://hdl.handle.net/2440/37701.

BRIEN, C.J. (2008). Tiers in gene expression microarrray experiments. URL http://www.newton.ac.uk/seminar/20080812140014301. Paper presented at Design of Experiments: Recent Advances in Methods and Applications (DEMA2008), Cambridge, U.K.

BRIEN, C.J. (2011a). Factor allocation in gene-expression microarray experiments. URL http://chris.brien.name/multitier/MTTalks.html. Paper presented at International Conference on Design of Experiments (ICODOE 2011), Memphis, U.S.A.

BRIEN, C.J. (2011b). Robust microarray experiments by design: a multiphase framework. URL http://www.newton.ac.uk/seminar/20110830163017001. Paper presented at Design of Experiments: Recent Advances in Methods and Applications (DEMA2011), Cambridge, U.K.

BRIEN, C.J. (2012). Comments on 'Therapist variation within randomized trials of psychotherapy: Implications for precision, internal and external validity'. *Statistical Methods in Medical Research* **21**, 215–216.

BRIEN, C.J. (2017a). ACANONICAL Procedure. In *GenStat Reference Manual Release 19, Part 3 Procedures*, ed. VSN International. Hemel Hempstead, U.K.: VSN International, pp. 23–25.

BRIEN, C.J. (2017b). *dae: functions useful in the design and ANOVA of experiments*. URL http://CRAN.R-project.org/package=dae, (R package version 3.0-04, accessed July 5, 2017).

BRIEN, C.J. (2017c). Multiphase experiments with at least one later laboratory phase. II. Nonorthogonal designs. Manuscript submitted to the Australian & New Zealand Journal of Statistics.

BRIEN, C.J. & BAILEY, R.A. (2006). Multiple randomizations (with discussion). *Journal of the Royal Statistical Society Series B-Statistical Methodology* **68**, 571–599.

BRIEN, C.J. & BAILEY, R.A. (2009). Decomposition tables for multitiered experiments. I. A chain of randomizations. *Annals of Statistics* **37**, 4184–4213.

BRIEN, C.J. & BAILEY, R.A. (2010). Decomposition tables for multitiered experiments. II. Two-one randomizations. *Annals of Statistics* **38**, 3164–3190.

BRIEN, C.J., BAILEY, R.A., CORRELL, R.L., HARCH, B.D., PAYNE, R.W. & DEMÉTRIO, C.G.B. (2001). Multitiered experiments web site. URL http://chris.brien.name/multitier/. (accessed November 8, 2017).

BRIEN, C.J., BERGER, B., RABIE, H. & TESTER, M. (2013). Accounting for variation in designing greenhouse experiments with special reference to greenhouses containing plants on conveyor systems.

Plant Methods **9**. URL http://www.plantmethods.com/content/9/1/5.

BRIEN, C.J. & DEMÉTRIO, C.G.B. (2009). Formulating mixed models for experiments, including longitudinal experiments. *Journal of Agricultural, Biological, and Environmental Statistics* **14**, 253–280.

BRIEN, C.J., HARCH, B.D. & CORRELL, R.L. (1998). Design and ANOVA for experiments involving a field trial and laboratory analyses. Paper presented at The Ninth International Conference on Quantitative Methods for the Environmental Sciences, Gold Coast, Queensland, Australia.

BRIEN, C.J., HARCH, B.D., CORRELL, R.L. & BAILEY, R.A. (2011). Multiphase experiments with at least one later laboratory phase. I. Orthogonal designs. *Journal of Agricultural, Biological, and Environmental Statistics* **16**, 422–450.

BRIEN, C.J., MAY, P. & MAYO, O. (1987). Analysis of judge performance in wine-quality evaluations. *Journal of Food Science* **52**, 1273–1279.

BRIEN, C.J. & PAYNE, R.W. (1999). Tiers, structure formulae and the analysis of complicated experiments. *The Statistician* **48**, 41–52.

BRIEN, C.J. & PAYNE, R.W. (2012). AMTIER Procedure. In *GenStat Reference Manual Release 15, Part 3 Procedures*. Hemel Hempstead, U.K.: VSN International, pp. 106–108. URL http://www.vsni.co.uk/resources/documentation/genstat-reference-procedure-library/, (accessed June 19, 2016).

BUTLER, D.G. (2013). On the optimal design of experiments under the linear mixed model. Ph.D. thesis, School of Mathematics and Physics, Brisbane, Australia. URL http://espace.library.uq.edu.au/view/UQ:314932.

CAVANAGH, C.R., TAYLOR, J., LARROQUE, O., COOMBES, N., VERBYLA, A.P., NATH, Z., KUTTY, I., RAMPLING, L., BUTOW, B., RAL, J.P., TOMOSKOZI, S., BALAZS, G., BEKES, F., MANN, G., QUAIL, K.J., SOUTHAN, M., MORELL, M.K. & NEWBERRY, M. (2010). Sponge and dough bread making: genetic and phenotypic relationships with wheat quality traits. *Theoretical and Applied Genetics* **121**, 815–828.

COCHRAN, W.G. & COX, G.M. (1957). *Experimental Designs*. New York: John Wiley & Sons, 2nd edn.

COOMBES, N.E. (2009). *DiGGer: design search tool in R*. URL: http://nswdpibiom.org/austatgen/software/, (accessed July 16, 2017).

COWLEY, R.B., ASH, G.J., HARPER, J.D.I., SMITH, A.B., CULLIS, B.R. & LUCKETT, D.J. (2012). Application of multi-phase experiments in plant pathology to identify genetic resistance to *Diaporthe toxica* in *Lupinus albus*. *Euphytica* **186**, 655–669.

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

CULLIS, B.R., SMITH, A.B. & COOMBES, N.E. (2006). On the design of early generation variety trials with correlated data. *Journal of Agricultural, Biological and Environmental Statistics* **11**, 381–393.

CULLIS, B.R., SMITH, A.B., PANOZZO, J.F. & LIM, P. (2003). Barley malting quality: are we selecting the best? *Australian Journal of Agricultural Research* **54**, 1261–1275.

CURNOW, R.N. (1959). The analysis of a two phase experiment. *Biometrics* **15**, 60–73.

DIEPEVEEN, D., CLARKE, G.P.Y., RYAN, K., TARR, A., MA, W. & APPELS, R. (2012). Molecular genetic mapping of NIR spectra variation. *Journal of Cereal Science* **55**, 6–14.

ECCLESTON, J. & RUSSELL, K. (1975). Connectedness and orthogonality in multi-factor designs. *Biometrika* **62**, 341–345.

ECKERMANN, P.J., VERBYLA, A.P., CULLIS, B.R. & THOMPSON, R. (2001). The analysis of quantitative traits in wheat mapping populations. *Australian Journal of Agricultural Research* **52**, 1195–1206.

ELIAS, A.A., ROBBINS, K.R., DOERGE, R. & TUINSTRA, M.R. (2016). Half a century of studying genotype environment interactions in plant breeding experiments. *Crop Science* **56**, 2090–2105.

FLAKELAR, C.L., HOWITT, J.A., PRENZLER, P.D., DORAN, G., COOMBES, N. & LUCKETT, D.J. (2017). A multiphase experiment for the analysis of bioactive compounds in canola oil: sources of error from field and laboratory. *Chemometrics and Intelligent Laboratory Systems* **162**, 55–64.

FOITO, A., HACKETT, C.A., BYRNE, S.L., STEWART, D. & BARTH, S. (2015). Quantitative trait loci analysis to study the genetic regulation of non-polar metabolites in perennial ryegrass. *Metabolomics* **11**, 412–424.

FOX, G.P., MARTIN, A., KELLY, A.M., SUTHERLAND, M.W., MARTIN, D., BANKS, P.M. & SHEPPARD, J. (2013). QTLs for water absorption and flour yield identified in the doubled haploid wheat population Lang/QT8766. *Euphytica* **192**, 453–462.

GOOS, P. (2002). *The Optimal Design of Blocked and Split-Plot Experiments*, *Lecture Notes in Statistics*, vol. 164. New York: Springer-Verlag.

HARDNER, C.M., EVANS, K., BRIEN, C., BLISS, F. & PEACE, C. (2016). Genetic architecture of apple fruit quality traits following storage and implications for genetic improvement. *Tree Genetics & Genomes* **12**, 20.

HINKELMANN, K. & KEMPTHORNE, O. (2005). *Design and Analysis of Experiments, Wiley Series in Probability and Statistics*, vol. 2. Hoboken, N.J.: Wiley-Interscience.

HOUTMAN, A.M. & SPEED, T.P. (1983). Balance in designed experiments with orthogonal block structure. *Annals of Statistics* **11**, 1069–1085.

JARRETT, R.G. & RUGGIERO, K. (2008). Design and analysis of two-phase experiments for gene expression microarrays – Part I. *Biometrics* **64**, 208–216.

JOHN, J.A. & WILLIAMS, E.R. (1995). *Cyclic and Computer Generated Designs.* London: Chapman and Hall, 2nd edn.

KEMPTHORNE, O. (1952). *The Design and Analysis of Experiments.* New York: John Wiley & Sons.

KERR, M.K. (2003). Design considerations for efficient and effective microarray studies. *Biometrics* **59**, 822–828.

LEVER, T., KELLY, A., DE FAVERI, J., MARTIN, D., SHEPPARD, J., QUAIL, K. & MISKELLY, D. (2005). Australian wheat for the sponge and dough bread making process. *Australian Journal of Agricultural Research* **56**, 1049–1057.

MCINTYRE, G.A. (1955). Design and analysis of two phase experiments. *Biometrics* **11**, 324–334.

MCINTYRE, G.A. (1956). Query 123. *Biometrics* **12**, 527–532.

MEAD, R., GILMOUR, S.G. & MEAD, A. (2012). *Statistical principles for the design of experiments.* Cambridge: Cambridge University Press.

MILLIKEN, G.A., GARRETT, K.A. & TRAVERS, S.E. (2007). Experimental design for two-color microarrays applied in a pre-existing split-plot experiment. *Statistical Applications in Genetics and Molecular Biology* **6**, Article 20.

MOLENAAR, H., GLAWE, M., BOEHM, R. & PIEPHO, H.P. (2017). Selection for production-related traits in *Pelargonium zonale*: improved design and analysis make all the difference. *Horticulture Research* **4**, 17004.

NELDER, J.A. (1965). The analysis of randomized experiments with orthogonal block structure. I. Block structure and the null analysis of variance. *Proc. Roy. Soc. A Math. Phys. Eng. Sci.* **283**, 147–162.

NETTLETON, D. (2012). Design of gene expression microarray experiments. In *Design and Analysis of Experiments*, ed. K. Hinkelmann, *Wiley Series in Probability and Statistics*, vol. III. Special Designs and Applications. Hoboken, N.J.: Wiley-Interscience, 1st edn., pp. 73–108.

NETTLETON, D. (2014). Design of RNA sequencing experiments. In *Statistical Analysis of Next Generation Sequencing Data*, eds. S. Datta & D. Nettleton, chap. 5. New York: Springer, pp. 93–113.

OAKEY, H., SHAFIEI, R., COMADRAN, J., UZREK, N., CULLIS, B., GOMEZ, L.D., WHITEHEAD, C., MCQUEEN-MASON, S.J., WAUGH, R. & HALPIN, C. (2013). Identification of crop cultivars with consistently high lignocellulosic sugar release requires the use of appropriate statistical design and modelling. *Biotechnology for Biofuels* **6**.

OSBORNE, B.G., FOX, G.P., KELLY, A.M. & HENRY, R.J. (2007). Measurement of barley grain rheology for the quality selection of breeding material. *Journal of the Institute of Brewing* **113**, 135–141.

PAYNE, R.W. & TOBIAS, R.D. (1992). General balance, combination of information and the analysis of covariance. *Scandinavian Journal of Statistics* **19**, 3–23.

PEACOCK, L., HUNTER, T., YAP, M.L. & ARNOLD, G. (2003). Indirect interactions between rust (*Melampsora epitea*) and leaf beetle (*Phratora vulgatissima*) damage on *Salix*. *Phytoparasitica* **31**, 226–235.

PIEPHO, H.P., BÜCHSE, A. & EMRICH, K. (2003). A hitchhiker's guide to mixed models for randomized experiments. *Journal of Agronomy and Crop Science* **189**, 310–322.

PREECE, D.A. (1982). Balance and designs: another terminological tangle. *Utilitas Mathematica* **21C**, 85–186.

PREECE, D.A. (1991). Contribution to the Discussion on Strata for randomized experiments by R. A. Bailey. *Journal of the Royal Statistical Society, Series B (Methodological)* **53**, 67–68.

R DEVELOPMENT CORE TEAM (2017). *R: A Language and Environment for Statistical Computing.* Vienna, Austria: R Foundation for Statistical Computing. URL http://www.r-project.org/, (accessed July 5, 2017).

RAMAN, R., ALLEN, H., DIFFEY, S., RAMAN, H., MARTIN, P. & MCKELVIE, K. (2009). Localisation of quantitative trait loci for quality attributes in a doubled haploid population of wheat (*Triticum aestivum L.*). *Genome* **52**, 701–715.

SAS INSTITUTE INC. (2015). *SAS/QC® 14.1 User's Guide.* SAS Institute Inc., Cary, NC. URL http://support.sas.com/documentation/cdl/en/qcug/68161/HTML/default/viewer.htm#qcug_optex_toc.htm, (accessed June 19, 2016).

SMITH, A.B., BUTLER, D.G., CAVANAGH, C.R. & CULLIS, B.R. (2015). Multi-phase variety trials using both composite and individual replicate samples: a model-based design approach. *Journal of Agricultural Science* **153**, 1017–1029.

SMITH, A.B., CULLIS, B.R., APPELS, R., CAMPBELL, A.W., CORNISH, G.B., MARTIN, D. & 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., CULLIS, B.R., BROCKHOFF, P. & THOMPSON, R. (2003). Multiplicative mixed models for the analysis of sensory evaluation data. *Food Quality and Preference* **14**, 387–395.

SMITH, A.B., CULLIS, B.R., COOMBES, N.E., ALLEN, H. & MARTIN, P. (1999). The use of statistical techniques to improve the estimation of flour extraction. Paper presented at Genstat 99: Applications, Advancement and Enhancements, Lorne, Victoria, Australia.

SMITH, A.B., LIM, P. & CULLIS, B.R. (2006). The design and analysis of multi-phase plant breeding experiments. *Journal of Agricultural Science* **144**, 393–409.

SMITH, A.B., THOMPSON, R., BUTLER, D.G. & CULLIS, B.R. (2011). The design and analysis of variety trials using mixtures of composite and individual plot samples. *Journal of the Royal Statistical Society, Series C* **60**, 437–455.

SNELL, P., REINKE, R., BRAYSHER, B., FITZGERALD, M. & SMITH, A.B. (2002). Selecting for grain quality in a rice breeding program coming to terms with a two-phased environment. In *Plant Breeding for the 11th Millennium. Proceedings of the 12th Australasian Plant Breeding Conference, Perth, WA*, ed. J.W. McComb. Australian Plant Breeding Association Inc., pp. 636–641.

VSN INTERNATIONAL (2013). *CycDesigN for Windows ver. 5.1.* Hemel Hempstead, UK: VSN International. URL http://www.vsni.co.uk/software/cycdesign/, (accessed July 17, 2017).

VSN INTERNATIONAL (2015). *GenStat for Windows 18th ed.* Hemel Hempstead, UK: VSN International. URL http://www.vsni.co.uk/software/genstat/, (accessed July 17, 2017).

WALWYN, R. & ROBERTS, C. (2010). Therapist variation within randomised trials of psychotherapy: implications for precision, internal and external validity. *Statistical Methods in Medical Research* **19**, 291–315.

WILKINSON, G.N. (1970). A general recursive procedure for analysis of variance. *Biometrika* **57**, 19–46.

WILKINSON, R.I., FRISINA, C., PARTINGTON, D.L., FRANZ, P.R., BRIEN, C.J., THOMSON, F., TOMKINS, R.B. & FARAGHER, J.D. (2008). Effects of 1-methylcyclopropene on firmness and flesh browning in pink lady (TM) apples. *Journal of Horticultural Science and Biotechnology* **83**, 165–170.

WILLIAMS, E.R. & PIEPHO, H.P. (2015). Optimality and contrasts in block designs with unequal treatment replication. *Australian & New Zealand Journal of Statistics* **57**, 203–209.

WOOD, J.T., WILLIAMS, E.R. & SPEED, T.P. (1988). Non-orthogonal block structure in two-phase designs. *Australian & New Zealand Journal of Statistics* **30A**, 225–237.