

Erratum

Pilla, R. S., Kitska, D. J. and Loader, C. (2005). Statistical analysis of modified complete randomized designs: applications to chemo-orientation studies. *J. Exp. Biol.* **208**, 1267-1276.

The following typographical errors appeared in both the on-line and print versions of this paper.

On p. 1270, second column, second paragraph, the first line is missing a model number and should read:
It is important to emphasize that the model under which the simulated data is generated is different from the model (1).

On p. 1272, second column under 'Results' section, in the first paragraph, the last sentence contains errors with reference to Table 1 and P -value. The sentence should read:
However, the scientific interest is usually in treatment effects and the analysis in Table 2 incorrectly finds a highly significant treatment effect ($P=2.533\times 10^{-5}$), while failing to detect the real...

On p. 1274, first column, first paragraph, the P -value, to match the value in Table 3, should be: $P=0.0307$.

We apologize to authors and readers for any inconvenience these errors may have caused.

Statistical analysis of modified complete randomized designs: applications to chemo-orientation studies

Ramani S. Pilla^{1,*}, David J. Kitska² and Catherine Loader²

¹Department of Statistics and Department of Biology and ²Department of Statistics, Case Western Reserve University, Cleveland, OH 44106, USA

*Author for correspondence (e-mail: pilla@case.edu)

Accepted 26 January 2005

Summary

Often experimental scientists employ a Randomized Complete Block Design (RCBD) to study the effect of treatments on different subjects. Under a 'complete randomization', the order of the apparatus setups within each block, including all replications of each treatment across all subjects, is completely randomized. However, in many experimental settings complete randomization is impractical due to the cost involved in re-adjusting the device to administer a new treatment. One typically resorts to a type of 'restricted randomization', in which multiple subjects are tested under each treatment before the apparatus is re-adjusted. The order of the treatments as well as the assignment of subjects to each block are random. If the data obtained under any type of restricted randomization are treated as if the data were collected under an RCBD with complete randomization within each block, then there is potential to increase the risk of false positives (Type I error). This is of concern to animal orientation studies and other areas such as chemical ecology where it is impractical to reset the experimental

device for each subject tested. The goal of the research presented in this article is twofold: (1) to demonstrate the consequences of constructing an *F*-statistic based on a mean square error for testing the significance of treatment effects under the restricted randomization; (2) to describe an alternative method, based on split-plot analysis of variance, to analyze designed experiments that yield better power under the restricted randomization. The statistical analyses of simulated experiments and data involving virgin male *Periplaneta americana* substantiate the benefits of the alternative approach under the restricted randomization. The methodology and analysis employed for the simulated experiment is equally applicable to any organism or artificial agent tested under a restricted randomization protocol.

Key words: analysis of variance, false positive, false negative, mean square error, olfaction, pheromone, randomized complete block design, restricted randomization, split-plot, type I error.

Introduction

In designed experiments, one measures responses on multiple experimental subjects with the goal of analyzing the effect of changes in controlled experimental conditions, or treatments, on the responses of different subjects. In laboratory experiments, one controls for extraneous conditions, to ensure each experimental run is conducted under similar conditions, so that one is reasonably confident that consistent differences in response are caused by the treatments. However, it is not always possible to ensure that all extraneous conditions are properly controlled. To reduce the effects of extraneous conditions, two techniques are commonly employed: randomization and blocking.

Blocking refers to the division of experimental runs into smaller sub-groups, or blocks. Each treatment is applied randomly to a number of subjects within each block. This design, known as a Randomized Complete Block Design (RCBD), is commonly employed in biological experiments, where, for example, experimental runs on a given day may be

treated as a block (Sokal and Rohlf, 1981). The randomization protocol reduces any bias in favor of particular treatments, while the blocking enables extraneous variation to be absorbed into block effects. Consequently, one obtains better estimates of treatment effects and more powerful tests for treatment differences (Cochran and Cox, 1957).

In the RCBD, the application of treatments to subjects within a block must be completely randomized. If a treatment is applied to five subjects within a block, then the subjects are chosen randomly within the block. However, in many experimental designs, practical constraints prevent this ideal situation from being realistic. The motivation for this research arises from insect pheromone-tracking studies, where a pheromone plume is generated at one end of a wind tunnel. An insect begins at the other end of the tunnel and is challenged to track the plume to the pheromone source. The goal is to detect differences in the response of an insect to different types of plumes or treatments (Willis and Baker, 1984; Linn et al.,

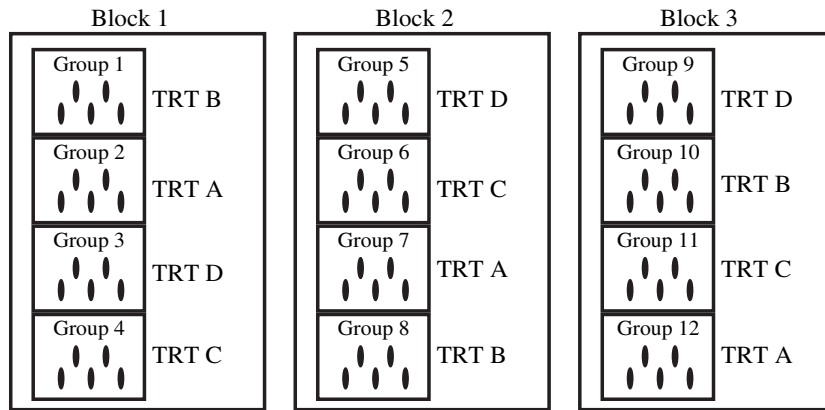


Fig. 1. A design layout under a type of restricted randomization for three blocks, in which multiple subjects were challenged to a single treatment before administering the next treatment. The order of the treatments as well as the assignment of the subjects within each block were randomized.

1988a,b; Mafra-Neto and Cardé, 1998; Zanen and Cardé, 1999; Cardé and Knols, 2000; Dekker et al., 2001; Willis and Avondet, 2005). The effect of various types of formulated synthetic pheromone on different species of walking and flying insects has been studied by various researchers (Linn et al., 1988a,b; Willis and Arbas, 1991; Linn et al., 1996; Cardé et al., 1998; Willis and Avondet, 2005).

Whenever the complete randomization protocol is violated, we refer to the corresponding design framework as 'restricted randomization'. One such type is illustrated in Fig. 1. For instance, in chemo-orientation studies, changing the treatment typically involves dismantling and reconfiguring the experimental device, which is often not practical after each experimental run. A more practical approach often undertaken by experimental scientists is: within each block, multiple subjects are challenged to a single treatment before changing the device to administer the next treatment. Only the order in which the treatments are applied is randomized. By considering subjects in groups, the experiments could be

conducted in a relatively short time span. This randomization protocol is a type of restricted randomization as illustrated in Fig. 1. We demonstrate the effects of a restricted randomization on the analysis and scientific conclusions using a computer-simulated experiment and data involving virgin male *Periplaneta americana*.

Analysis of variance (ANOVA) is the fundamental tool for analyzing data from designed experiments (Cochran and Cox, 1957; Searle, 1971). Chapter 8 of Sokal and Rohlf (1981) emphasizes that ANOVA, while being an effective tool for any modern biologist, may create artificial constructs in the mind of a scientist that could lead to misleading conclusions. In subsequent sections of this article, this is demonstrated in the context of restricted randomization.

In many of the animal chemo-orientation studies published over the past two decades (see for instance, Table 1), an RCBD or a related design was employed to analyze experimental data obtained under the restricted or other modified randomization protocols. One such modified experimental design was employed by Linn et al. (1988a), in which some treatment effects were confounded with block effects. This might be one of the reasons for obtaining non-significance of the treatment effects. In a flight tunnel experiment, Linn et al. (1988b) challenged 5–10 males of each species of *Trichoplusia ni* and *Pseudoplusia includens* to the treatments at one of the two dosages. From their experimental design description it appears that there were multiple levels of blocking and a type of restricted randomization. They analysed the data using a one-way ANOVA, however, ignoring the block effects and restriction in randomization.

Mafra-Neto and Cardé (1998) utilized an RCBD to test the effect of treatments; however, it is not clear from the 'Materials and methods' section whether their experimental design indeed

Table 1. A summary of several experimental studies conducted by various researchers on chemo-orientation studies

Authors (Year)	Journal	Experimental subject or organism (Analysis)
Baker et al. (1984)	<i>Physiol. Entomol.</i>	<i>G. molesta</i> (RCBD)
Cardé and Knols (2000)	<i>Physiol. Entomol.</i>	<i>L. dispar</i> (RCBD)
Linn and Roelofs (1983)	<i>Physiol. Entomol.</i>	<i>G. molesta</i> (ANOVA and regression)
Linn et al. (1988a)	<i>Physiol. Entomol.</i>	<i>G. molesta</i> and <i>P. gossypiella</i> (ANOVA)
Linn et al. (1988b)	<i>J. Chem. Ecol.</i>	<i>T. ni</i> and <i>P. includens</i> (ANOVA)
Mafra-Neto and Cardé (1998)	<i>J. Comp. Physiol. A</i>	<i>Cadra cautella</i> (RCBD)
Vickers (2002)	<i>J. Chem. Ecol.</i>	<i>H. subflexa</i> (RCBD)
Willis and Baker (1984)	<i>Physiol. Entomol.</i>	<i>G. molesta</i> (RCBD)
Willis and Baker (1987)	<i>J. Insect. Physiol.</i>	<i>G. molesta</i> (RCBD)
Willis and Cardé (1990)	<i>J. Comp. Physiol. A</i>	<i>L. dispar</i> (RCBD)
Willis and Arbas (1991)	<i>J. Comp. Physiol. A</i>	<i>Manduca sexta</i> (RCBD)

In these cases, the original experimental design did not have a complete randomization, but the analysis was conducted under such an assumption. In particular, some of these analyses were carried out under the RCBD framework ignoring a certain restriction in the randomization protocol that was inherent in these experimental design structures.

satisfies the complete randomization protocol. From the 'Materials and methods' section of Linn et al. (1996), it appears that they do not have a complete randomization protocol; however, they analysed their experimental data using an ANOVA (although not clear whether it was a one-way or two-way). The effect of light levels and plume structure on the orientation maneuvers of male *Lymantria dispar* (gypsy moths) flying along pheromone plumes has been studied by Cardé and Knols (2000). They used the flight tracks of 20 males per treatment (a total of six treatments), tested in a complete randomized block design. The goal was to study the effects of odor plume structure on the orientation maneuvers of different species of walking and flying insects (Willis and Baker, 1984; Baker, 1990; Cardé and Knols, 2000). Justus et al. (2002) employed an ANOVA; however, they do not state the details of the experimental design and the statistical analysis employed such as a one-way ANOVA or an RCBD. Vickers (2002) considered male *Heliothis subflexa*, which were flown in a wind tunnel to a variety of combinations of synthetic pheromone components admixed on a filter paper disk. Their experimental design violated the complete randomization protocol since groups of 3–5 males were flown under each treatment on any given day; however, they analysed the experimental data using an RCBD ANOVA.

The flight behavior of mosquitoes in host-odor plumes and the effects of the fine-scale structures of such plumes have been studied by Dekker et al. (2001). They considered seven treatments; each treatment had eight replicates and treatments were randomized within each test day. Zanen and Cardé (1999) applied an RCBD to test the treatments on a given day on male *L. dispar*; however, they did not employ a two-way ANOVA to analyse the resulting experimental data. Consequently, their analysis does not match the design. Furthermore, it is not clear whether their design satisfied a complete randomization protocol.

All the above analyses lead to an important question: how does violating the fundamental assumption of complete randomization affect the interpretation of experimental results or scientific conclusions? It is often very difficult (1) to assess the consequences of subtle modifications of the design on the resulting analysis and scientific conclusions and/or (2) to identify an alternative method of analysis corresponding to the randomization protocol scheme at hand by simply referring to the statistical literature (Cochran and Cox, 1957; Searle, 1971; Sokal and Rohlf, 1981).

The goals of this article are to provide insights into the statistical analysis issues embedded within designed experiments when practical constraints impose restrictions on randomization of the treatments. The statistical analyses of simulated experiments and data involving virgin male *P. americana* demonstrate the consequences of overlooking the restricted randomization on the scientific questions being addressed as well as the analysis and interpretation of the results. Our simulated experimental data, presented in the 'Results' section, demonstrate that the RCBD incorrectly finds a highly significant treatment effect that was not present in the

model, while failing to find the real effect present in the model. However, the risk of a false positive (Type I error) indication of the treatment significance is substantially reduced under the alternative analysis. In essence, by employing an RCBD when the underlying assumption is not satisfied, we are more likely to declare an effect exists when it does not. This has implications for the understanding of experimental results as well as the scientific conclusions. It is important to note that the methodology and analysis employed for the simulated experiment is equally applicable to any organism or artificial agent tested under a restricted randomization framework.

Applying an appropriate model to account for the changes in the design is relevant for two reasons. Violation of assumptions in a particular design could result in (1) underestimation of experimental error variance and (2) obtaining false positives (Type I errors). These in turn, may lead to incorrect results or invalidate the analysis employed by a researcher. Therefore, it is important to choose an appropriate model and error structure in considering designed experiments.

Materials and methods

The effect of a restricted randomization on the analysis of experimental data is best illustrated through a simulated experiment. In general, if there is a restriction on randomization at a given level in experimentation, there will be a 'split' in the design, leading to a split-plot design nested within an RCBD structure. There are two types of experimental units: the larger units, the groups of subjects (for example, insects), are called the whole-plots and the smaller ones, individual subjects, the sub-plots. A split-plot design creates a nesting within the design structure since the whole-plots are nested within blocks and the sub-plots are in turn nested within the whole-plots. The design structure for the whole-plot experimental units is essentially an RCBD. A split-plot design has two advantages over a simpler ANOVA: (1) if treatment can be applied to the whole-plot at once, rather than separately to sub-plots, this may reduce costs, and (2) because sub-plots are usually more uniform, parameters measuring comparisons among conditions may be estimated more precisely.

Several examples of split-plot designs can be found in the biological literature. Linn and Roelofs (1983) considered a total of 100 treatments in a $5 \times 4 \times 5$ factorial design such that two of the three factors were varied between days and one factor was varied within days. This design has a split-plot structure with days serving as whole-plots. The experiments of Linn et al. (1988a) consider two species, *Grapholita molesta* and *Pectinophora gossypiella*. They challenged 5–10 males to each treatment per day, with a total of 70 males for each treatment–temperature combination. Both the treatments and temperatures were randomized over the experimental period. This experimental design has a split-plot structure with unbalanced data. In both of these examples, the analysis was based on ANOVA and regression techniques, rather than a split-plot analysis.

Model for a hypothetical split-plot design

Every ANOVA is associated with a linear model specifying the effects being considered. The linear model for a split-plot ANOVA includes hierarchies of terms modeling both the block effects and treatment effects. The key concept in constructing models for split-plot designs is recognizing the different sizes of experimental units and consequently identifying the corresponding design structures and treatment structures.

Consider an experiment in which several treatments are administered to different subjects and the experiment is conducted over several blocks. Suppose that a biological experiment consists of subjects of different ages, challenged to various treatments such as pheromones on different blocks (for example, days). The age factor may be included in the model to determine the performance of behavior of an animal as it develops over time. The split plot design originated in agricultural field trials and, in this setting, one factor may be fertilizer and another factor irrigation level.

Consider a hypothetical design such that Y_{ijk} denotes a response measured on a subject at the k th ($k=1, \dots, n$) age when challenged to the j th ($j=1, \dots, t$) treatment in the i th ($i=1, \dots, b$) block. The simplest possible model (1) can be written as

$$Y_{ijk} = \mu + \alpha_i + \beta_j + \gamma_{ij} + \delta_k + \varepsilon_{ijk} \quad \text{for } i=1, \dots, b; j=1, \dots, t; k=1, \dots, n, \quad (1)$$

where μ represents the overall mean, α_i is the block effect, β_j is the effect of the j th treatment applied at the whole-plot level, γ_{ij} is the whole-plot effect, and δ_k is the age effect. The term ε_{ijk} measures random error.

The scientific interest is in the treatment effect β_j and the age effect δ_k . The basic statistical problem is to detect the significance of these effects and estimate the size of any effects that are present. The effects β_j and δ_k are treated as fixed effects or parameters in the model. The block effect α_i and the whole-plot effect γ_{ij} are of no inherent interest; however, these can cause considerable variation from block to block and whole-plot to whole-plot. Therefore, these effects are treated as random. A typical assumption is that these effects follow distributions $N(0, \sigma_\alpha^2)$ and $N(0, \sigma_\gamma^2)$ respectively, where σ_α^2 and σ_γ^2 are unknown variance components corresponding to the block and whole-plot effects, respectively. The random errors ε_{ijk} are assumed to follow a $N(0, \sigma_\varepsilon^2)$ distribution.

Simulated experiment: generation of data

We demonstrate how to specify a model for split-plot design and how to construct appropriate F -test statistics through simulated experimental data. The design is generated as follows:

- (1) Sixteen runs, numbered 1 to 16, were performed for each of five blocks.
- (2) Each block was divided into four whole-plots of four runs. The four treatments, such as pheromone plumes, were randomly assigned to the four whole-plots.
- (3) Four subjects, one of each age, are assigned in a random order to the runs within each whole-plot.

The important feature of this design is that two factors,

treatment and age, are being varied across experimental runs. The treatment is varied randomly at the whole-plot level, while the age is varied randomly at the level of individual runs or sub-plots.

After creating the experimental design, responses were generated according to the following model (2):

$$Y_{ijk} = \mu + \alpha_i + \lambda_i r_{ijk} + \delta_k + \varepsilon_{ijk} \quad \text{for } i=1, \dots, 5; j=1, \dots, 4; k=1, \dots, 4, \quad (2)$$

where subscripts i, j and k denote the block, treatment and age, respectively. (1) The overall mean is $\mu=100$; (2) the block effect α_i has a normal distribution with mean zero and $\sigma_\alpha^2=25$; (3) the δ_k term representing the age effect takes values $1, \dots, 4$ corresponding to subjects of four different age groups (for example, insects of 10 days, 20 days, 30 days and 40 days old, respectively); (4) the component $\lambda_i r_{ijk}$ represents drift of experimental conditions over time, with r_{ijk} taking the run number $1, \dots, 16$ of the (j, k) treatment combination in the i th block; (5) the coefficient λ_i follows a $N(0, 1)$ distribution; and (6) ε_{ijk} has a normal distribution with mean zero and $\sigma_\varepsilon^2=1$. That is, we assumed that different random effects contribute differently to the level of variation in the final measurement of Y_{ijk} . The statistical model used to generate the data includes an age effect, a random block effect and a random drift effect within the block. However, the model does not include any treatment effect β_i and, therefore, the responses are independent of the treatment employed.

It is important to emphasize that the model under which the simulated data is generated is different from the model. In particular, the experimental drift in the model corresponding to the simulation experiment does not match exactly the assumption of the split-plot model. This effect was intentionally introduced into the model, since in practice one does not know the precise form of any uncontrolled variation. It is important that the statistical analysis is robust to misspecification of this term. The data was generated using the statistical language S (Venables and Ripley, 2002) and the model (2) was fitted using the `lm()` function in S. The model fitting and data analysis can also be performed using R, a freely available open source statistical language available from <http://www.r-project.org>.

P. americana experiment

The *P. americana* experiment is an example of a split-plot design, characterized by multiple levels of blocking. The experiment involved 3–18 weeks old virgin males of *P. Americana*, which were challenged to track wind-borne plumes of (–)-periplanone-B 2 h into their scotophase (12 h:12 h L:D cycle). The animals were video recorded as they tracked wind-borne plumes of the female sex-pheromone (–)-periplanone-B in a laboratory wind tunnel. Each videotaped walking path was digitized using a computerized motion analysis system.

Plume structure

For this experiment, four different plume structures were constructed by varying the size, shape and orientation of the

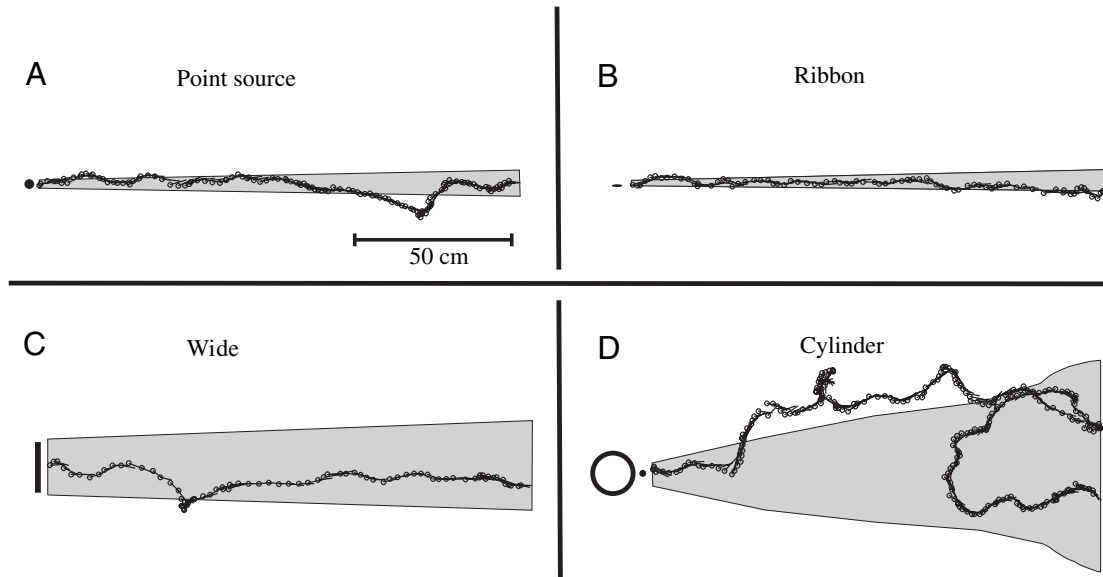


Fig. 2. An illustration of *P. americana* males tracking female pheromone upwind (right to left) in a laboratory wind tunnel, containing the time-averaged plume boundaries of titanium tetrachloride smoke plume in 25 cm s^{-1} wind. Each circle represents the body position at every 0.083 s. (A) Point source plume, 2.4 cm wide at the source spreads to 7.7 cm wide at the downwind end. (B) Ribbon plume, 1.5 cm wide at the source spreads to 6.1 cm wide at the downwind end. (C) Wide plume, 17 cm wide at the source spreads to 26.5 cm at the downwind end. (D) Cylinder plume, 7.6 cm wide at the source spreads to 68.1 cm at the downwind end.

pheromone source (Fig. 2). The point source plume was constructed using a 0.7 cm diameter circular filter paper disk (Whatman No. 1, Eastbourne, East Sussex, UK) held perpendicular to the airflow with an insect pin (Fig. 2A). A ribbon plume with a chemical source 0.05 cm wide was constructed by rotating the 0.7 cm filter paper disk 90° , so that the disk shape was parallel to air flow in the wind tunnel, resulting in a very narrow plume (Fig. 2B). The third plume was created by increasing the surface area of the source by ca. 25 times while also proportionally increasing the dosage of pheromone solution applied to the source. The wide plume treatment source was 14.3 cm wide \times 0.7 cm tall (Fig. 2C). The cylinder plume structure was generated by placing a Plexiglas[®] cylinder (81.28 cm tall \times 7.62 cm diameter) 5 cm upwind of the 0.7 cm diameter circular filter paper disk held perpendicular to airflow (Fig. 2D). The reader is referred to Willis and Avondet (2005) for further details on the materials employed in conducting this experiment.

The aim of this experiment was to test the hypothesis that male cockroaches steer their walking while tracking female pheromone using a chemotactic strategy characterized by counter turning (turning-back) when they experience a sharp pheromone-clean air edge (Willis and Avondet, 2005).

Measurements

Response variables measured from the digitized insect movement tracks included: track angle (degrees), track width (cm), ground speed (cm s^{-1}), body axis (degrees), net velocity (cm s^{-1}), inter-turn duration (s), the number of times each animal stopped, and the duration of each stop (s). For the purposes of the analysis, Willis and Avondet (2005) considered

(1) a turn as the location at which the head reached a local maximum or minimum value with respect to the lateral frame of reference of the wind tunnel, and (2) an animal to be in stopping position if there was no movement between two sequential positions of the head point. Measurement of these response variables from one animal was considered as one trial. The response variable is the average of the measurements for an entire walking track of a single animal.

The animals are expected to have peak response during a specific period in each scotophase, and only a limited number of experimental runs can be performed each day. The experiment was therefore carried out over 5 days. The design can be summarized as follows:

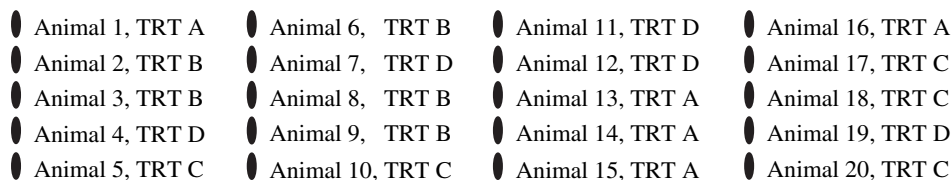
- (1) The experiment was run over 5 days.
- (2) Each day was divided into four whole-plots. Each of the four pheromone treatments were randomly assigned to one of the whole-plots.
- (3) Within each whole-plot, five animals were tracked. Each animal or sub-plot representing the smallest experimental unit.

This gives a total of 100, corresponding to $5 \times 4 \times 5$ experimental runs. Three animals did not respond when challenged with the experimental conditions so that 24 observations in total were completed for each treatment, except for the second treatment which yielded 25. The analysis therefore includes a total of 97 observations.

Split-plot model for the *P. americana* experiment

The key feature of the *P. americana* design is that the treatments (the pheromone plumes) were varied at the level of whole-plots, and not at the level of individual experimental runs. In the simulated experimental data described above, a

Fig. 3. A completely randomized design layout for a single block in which every subject and every treatment were individually randomized.



second factor of age was varied at the level of an experimental unit; however, in the *P. americana* experiment, this second factor is absent.

Since the treatments were applied to groups of animals within each apparatus setup, the treatments must be associated with the whole-plot part of the design. Therefore, in order to make an appropriate inference regarding treatments, the *F*-statistic denominator must include the random variation between the whole-plots.

In the context of the *P. americana* experiment, the response Y_{ijk} in model (1) represents the ground speed (cm s^{-1}) averaged over different time points; α_i and γ_{ij} are the day and whole-plots effects, respectively; β_j is the effect of the *j*th pheromone applied at the whole-plot level and δ_k is the effect of the treatment applied at the subject level. However, since no factor was varied at the subject level in this experiment, the δ_k term is absent in the model.

Appropriate *F*-test statistics

The analysis involves computing the sums of squares (SS) and mean squares (MS) due to each of the terms in the model; see Sokal and Rohlf (1981) and Searle et al. (1992) for the partitioning of the SS. In order to test for the significance of treatment effects, one forms an *F*-ratio using the treatment MS (MS_{trt}) and the error MS as:

$$F = \text{MS}_{\text{trt}} / \text{MSE}, \quad (3)$$

where the mean square error (MSE) is an unbiased estimator of the error variance σ_e^2 , the variation between subjects within groups. The above *F*-statistic is an appropriate one to employ when the only source of random variation in the estimated treatment effects are the random errors. This is the case for the age effect δ_k term in the model (1). Since each age occurs once in each whole-plot, any block effects must influence all treatments equally. Consequently, the presence of block effects does not inflate the treatment MS.

The *F*-statistic in Equation (3) is no longer valid when one

Table 2. ANOVA under the RCBD for the simulated experimental data assuming the complete randomization within each block, such as day, while treating the block effect as fixed

Source of variation	d.f.	MS	<i>F</i> -statistic	<i>P</i> -value
Block	4	1553.9	60.8653	<0.00001
Treatment	3	241.9	9.4756	0.00002
Age	3	32.4	1.2688	0.29190
Residual	69	25.5		
Total	80			

is testing for the treatment effects applied at the whole-plot level, the β_j term in model (1). In the *P. americana* experiment, there may be additional whole-plot variation due to differences in responsiveness of animals during the scotophase, or due to any random variation in resetting the experimental device. The MSE estimates only the subject-to-subject variation while ignoring these other potential sources of random variation. Therefore, the *F*-statistic in Equation (3) corresponding to the treatment effects is biased upwards, leading to false indications of significant treatment effects (Type I error).

The split-plot analysis overcomes this difficulty by modeling whole-plot variation by the random γ_{ij} term. The appropriate denominator for the *F*-ratio is the MS attributed to γ_{ij} ; i.e. MS_{intr} . Consequently, the *F*-ratio becomes:

$$F = \text{MS}_{\text{trt}} / \text{MS}_{\text{intr}}.$$

Results

We demonstrate the consequences of the randomization protocol on the analysis of experiments and scientific conclusions using the simulated and real data.

Comparison of the RCBD and split-plot analyses: simulation experiment

In terms of the practical significance, the main findings from our simulation experiment are summarized in Tables 2 and 3. Table 2 presents the results of an RCBD ANOVA, which assumes complete randomization as illustrated in Fig. 3. The analysis shows a highly significant block effect ($P < 0.00001$). However, scientific interest is usually in treatment effects and the analysis in Table 1 incorrectly finds a highly significant treatment effect ($P = 2.533 \times 10^{-5}$), while failing to detect the real age effect ($P = 0.2919$). The restricted randomization in this

Table 3. ANOVA under a split-plot design for the simulated experimental data assuming the restricted randomization

Source of variation	d.f.	MS	<i>F</i> -statistic	<i>P</i> -value
Whole-plots (groups of subjects)				
Block for groups (Day)	4	1553.9	11.9941	0.00040
Treatment	3	241.9	1.8673	0.18890
Residual (Treatment-by-block)	12	129.6		
Sub-total	19			
Sub-plots (Individual subjects)				
Age	3	32.4	8.9233	0.00006
Residual (Subjects within groups)	57	3.6		
Total	80			

design has led to both false positive (Type I error) and false negative (Type II error) results.

The restricted randomization means that the treatments are submitted in whole-plots of four runs, and for the purpose of analyzing treatment effects, the correct analysis is to treat these four runs as a single experimental unit. This leads to the split-plot ANOVA. The results of applying the split-plot analysis to the simulated experimental data are shown in Table 3. This analysis correctly identifies that there is no significant treatment effect ($P=0.1889$) and that the age effect is highly significant ($P=6.126 \times 10^{-5}$). Therefore, the risk of a false positive indication of the treatment significance is substantially reduced under the split-plot design. In essence, by employing an RCBD when the underlying assumption is not satisfied, we are more likely to reject a true null hypothesis. This has implications to the understanding of experimental results as lack of treatment effect would be expected to be of relevance.

Comparison of the RCBD and split-plot analyses: the P. americana experiment

We present the analysis of the data from the *P. americana* experiment using RCBD and split-plot models to demonstrate when and how likely false positives can occur and their consequences on the biological questions. This illustrates how violations of the underlying assumption for the RCBD leads to underestimation of the error variability and inflating the statistical significance of the treatment effects.

The response variable, ground speed (cm s^{-1}), is shown in Fig. 4 sorted first by day and next by pheromone, and in Fig. 5 sorted first by pheromone and then by day. It is clear that the pheromone D (cylinder source) is behaving differently. It also appears that treatments behave differently on different days. For example, a comparison of pheromone A (point source) vs pheromone B (ribbon source) shows that animals are responding more rapidly to the point source on days 2 and 4 and more rapidly to the ribbon source on days 1 and 5.

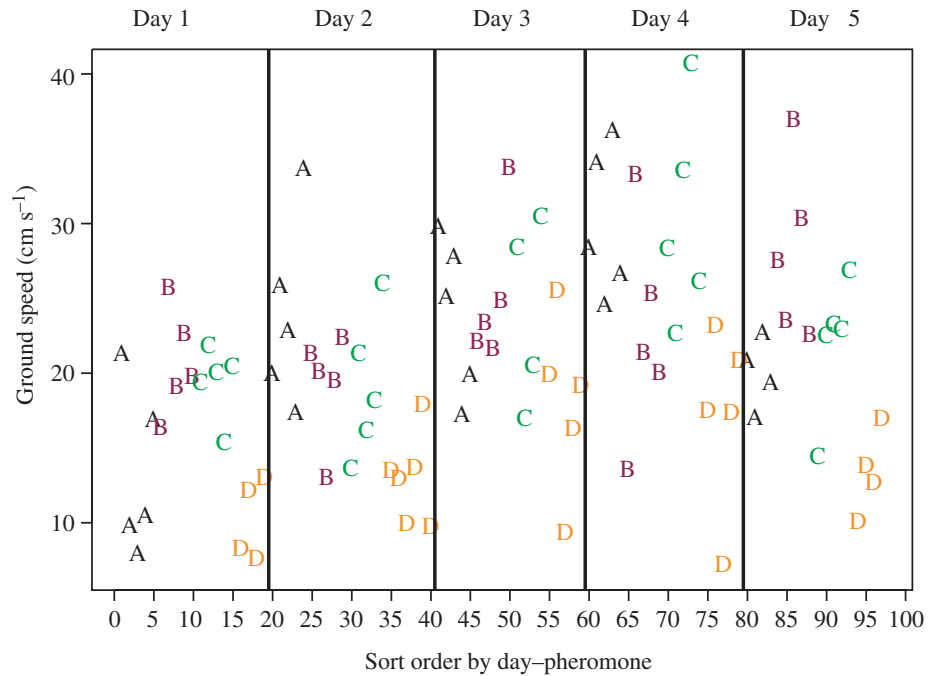


Fig. 4. The response variable ground speed (cm s^{-1}) for the *P. americana* experiment sorted first by day and next by pheromone.

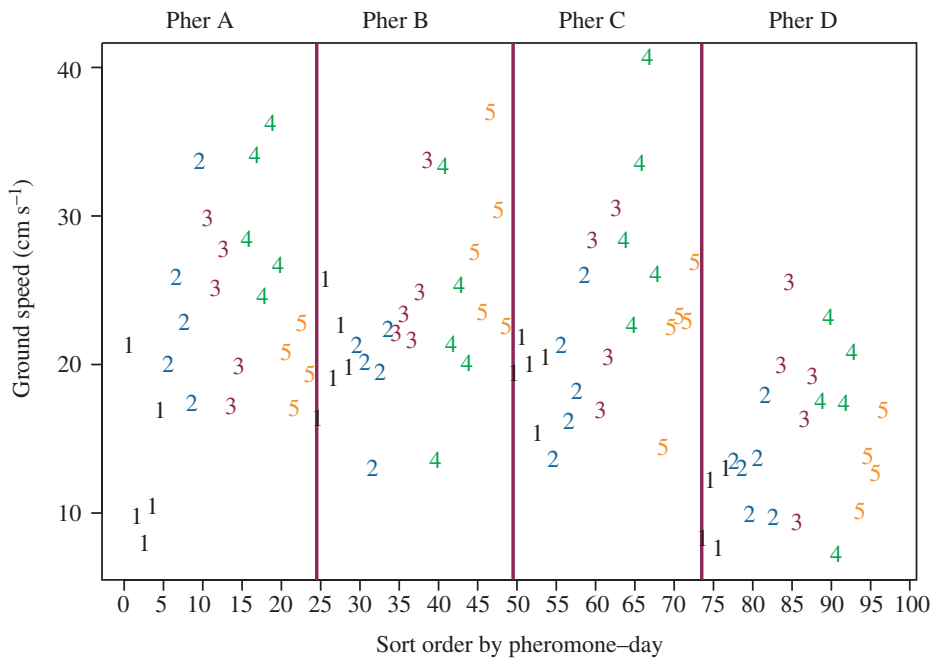


Fig. 5. The response variable ground speed (cm s^{-1}) for the *P. americana* experiment sorted first by pheromone (Pher) and then by day.

Analysis of the data was conducted using the SAS PROC MIXED program (Littell et al., 1996). Table 4 presents a two-way ANOVA for the response variable ground speed (cm s^{-1}). While the usual analysis of an RCBD includes only main effects for treatments and blocks, in the present experiment there are multiple replications of the pheromones on each day.

Table 4. ANOVA under the RCBD for the *P. americana* experiment assuming the complete randomization within each day while treating day effect as fixed for the response variable ground speed (cm s^{-1})

Source of variation	d.f.	MS	F-statistic	P-value
Pheromone	3	424.3216	16.63	<0.0001
Day	4	241.7842	9.48	<0.0001
Pheromone-by-day	12	52.2499	2.05	0.0307
Residual	77	25.5112		
Total	96			

Therefore, we are able to include an interaction term between the treatments and blocks. The table shows a highly significant day and pheromone effects ($P < 0.0001$) and a significant pheromone-by-day interaction effect ($P < 0.0307$). These results are consistent with those obtained by Willis and Avondet (2005).

How should the pheromone-by-day interaction effect be interpreted? What are the consequences in terms of the questions of biological interest such as responsiveness of animals to different pheromone plumes? The pheromone-by-day interaction effect means that the animals have responded more to some plumes than others on different days. However, the biological interest lies in the overall response to the individual pheromones and there is no inherent interest in the individual days. It is of little use to state that the animals respond more rapidly to the ribbon source on day 1, and to the point source on day 2, as the data here indicate.

The resolution to this paradox is to model the interaction effect, the γ_{ij} term in model (1), as a random effect, introduced by the restricted randomization of the design. However, these random effects influence estimates of the treatment means. Therefore, the SS attributed to the pheromone effects in Table 4 is inflated by these interactions. As a result, comparing the pheromone MS (MS_{tr}) with the MSE is inappropriate.

Table 5 presents the ANOVA using a split-plot model for the *P. americana* experiment under the restricted randomization while treating the block effects as random. The point estimates of the variance components and the F -statistic for the treatment effect in the table, were generated using the SAS PROC MIXED program (Littell et al., 1996). The F -statistic for the pheromone effect is now obtained as the MS for pheromone ($MS_{\text{tr}}=424.3216$) divided by the MS for pheromone-by-day interaction ($MS_{\text{intr}}=52.2499$).

The analysis in Table 5 is further complicated by the three deleted observations, causing imbalance in the design (i.e. $n_{ij} \neq n$ for all $i=1, \dots, b$ and $j=1, \dots, t$, where n_{ij} is the number of observations in the i th block for the j th treatment). The consequence is that a Satterthwaite approximation (p. 24 in Milliken and Johnson, 1984) must be used for the degrees of freedom (d.f.) of the F -statistic, namely the df_2 . Note that the df_1 is simply $(t-1)$. The F -statistic is much smaller (8.18 instead of 16.63) under the split-plot analysis; however, the treatment effect is still significant ($P=0.003$). The 95% confidence intervals for the variance components were

Table 5. ANOVA under the split-plot design for the *P. americana* experiment assuming the restricted randomization while day effect is treated as random for the response variable ground speed (cm s^{-1})

Fixed effect	df_1	df_2	F-statistic	P-value
Pheromone	3	12.2	8.18	0.0030
Variance components	Value	95% C.I.		
σ_α^2	9.8372	(1.6579, 58.367)		
σ_γ^2	5.5723	(1.1605, 26.757)		
σ_ϵ^2	25.4590	(18.575, 34.895)		

The df_1 is the numerator degrees of freedom that equals $(t-1)$ and df_2 is the denominator degrees of freedom obtained via the Satterthwaite approximation. The point estimates and the 95% confidence interval estimates for the variance components are provided.

generated by the lme() function in S-Plus (Pinheiro and Bates, 2000). The results presented in Table 5 confirm that both σ_α^2 and σ_γ^2 , representing the day and whole-plot effects respectively, are significant.

Pairwise comparisons

To further understand the treatment differences, we performed pairwise comparisons of treatment means. The analyses under RCBD and split-plot design in Table 6 correspond to the results in Tables 4 and 5, respectively. In comparison with the RCBD, the split-plot analysis returns larger standard errors and in turn, yields smaller T -statistics and larger P -values. In fact, if we choose the level of significance to be 0.01, then the pairwise difference of A vs D (point source vs cylinder) is significant under the RCBD while not significant under the split-plot analysis. Moreover, the P -values corresponding to the pairwise differences of B vs D (ribbon vs cylinder) and C vs D (wide vs cylinder) are closer to 0.01 than to 0.0001. The RCBD analysis has substantially overstated the statistical significance; see Curran-Everett and Benos (2004) for a discussion of why choosing 0.01 for a significance level is appropriate for certain situations.

The expected mean squares (Sokal and Rohlf, 1981; Searle et al., 1992) may be used to determine how much the F -statistic for treatment affects will be inflated, on average, when blocks such as days are treated as fixed and the RCBD analysis is employed for analysis as if the design is carried out under the complete randomization protocol. In this scenario, the F -statistic for treatment effects would use the MSE as its denominator, which has an expected value of σ_ϵ^2 . However, under the split-plot analysis with block effect treated as random, the denominator would be MS_{intr} , which has an expected value of $(t\sigma_\alpha^2 + n\sigma_\gamma^2 + \sigma_\epsilon^2)$. In the balanced case (i.e. $n_{ij}=n$ for all $i=1, \dots, b$ and $j=1, \dots, t$), an approximate indication of the inflation in the F -statistic is provided by the ratio:

$$(n\hat{\sigma}_\gamma^2 + \hat{\sigma}_\epsilon^2) / \hat{\sigma}_\epsilon^2,$$

where $\hat{\sigma}_\gamma^2$ and $\hat{\sigma}_\epsilon^2$ are the estimators of σ_γ^2 and σ_ϵ^2 ,

Table 6. Pairwise comparisons under the RCBD and split-plot analyses for the *P. americana* experiment

Treatment comparison	Analysis under RCBD			Analysis under Split-plot design		
	<i>T</i> -statistic	Estimated standard error	Tukey–Cramer <i>P</i> -value	<i>T</i> -statistic	Estimated standard error	Tukey–Cramer <i>P</i> -value
A vs B	−0.67	1.4463	0.9076	−0.44	2.0768	0.9700
A vs C	−0.49	1.4639	0.9607	−0.32	2.0882	0.9876
A vs D	5.35	1.4697	<0.0001	3.77	2.0903	0.0120
B vs C	0.17	1.4463	0.9981	0.12	2.0767	0.9994
B vs D	6.08	1.4522	<0.0001	4.23	2.0793	0.0054
C vs D	5.84	1.4697	<0.0001	4.09	2.0908	0.0068

respectively. Recall that γ_{ij} is confounded with the setup variation of the experimental device as well as any potential treatment-by-block interaction under the restricted randomization protocol and this combined variation is given by σ_{γ}^2 . The *F*-statistic given in Equation (3) under an RCBD does not include σ_{γ}^2 in its denominator and therefore ignores the setup variation entirely. In other words, this *F*-statistic ignores the ‘loss in efficiency’ when one cannot randomize between individual subjects due to practical difficulties involved in conducting an experiment.

The simulated experimental studies presented earlier demonstrate the generality of these conclusions and are consistent with our above treatment of the real data. Recall that the simulation results show a change from highly significant to non-significant treatment effects.

Discussion

A randomized complete block design is one of the most widely used designs by experimental scientist in studying the effects of treatments on subjects. Often, treatments are replicated within each block to obtain separate estimates of the error variance and any potential treatment-by-block interaction. Experimental constraints can prevent complete randomization within each block. Previous work in identifying the statistical significance of treatment effects on the behavior of response has used either a one-way ANOVA or an RCBD. In this article, we have demonstrated that a split-plot model can be applied to analyse data under a restricted randomization protocol. Furthermore, we have demonstrated that overlooking the effect of restricted randomization on inferences from RCBD analyses can lead to various spurious interaction effects as well as potentially serious Type I or Type II errors. In particular, if the restricted randomization is ignored and an RCBD analysis performed, then there is a risk of overstating the significance of treatment effects. In contrast, the split-plot analysis provides a powerful alternative to the analysis of data collected under the restricted randomization protocol. The proposed methods are illustrated using a real data from chemo-orientation studies; however, they extend directly to other studies where it is impractical to completely randomize the treatments given to individual experimental subjects. The techniques presented in this article can be implemented using widely available statistical software.

Our findings clearly substantiate the consequences of ignoring the restricted randomization and have the following implications. (1) Under the restricted randomization (Fig. 1), one has two sets of experimental subjects: (i) the subjects nested within groups, which in turn serve as blocks for the subjects, and (ii) groups nested within blocks. The appropriate analysis is to employ a split-plot ANOVA by considering the groups of subjects as the whole-plots and the individual subjects as the sub-plots. (2) The expected mean squares usually assume complete randomization. Under the restricted randomization, one must be cautious in calculating the expected mean squares. In particular, one cannot test the treatment-by-block interaction through an *F*-statistic given by $F = MS_{\text{intr}}/MSE$, and it is possible only if experimental setups are completely randomized between animals.

An important conclusion of our work is to demonstrate the significance of describing completely the design employed and statistical analysis performed on any experimental data. As we have shown, small changes to the design protocol can have a major effect on the validity of a statistical analysis. There are many different types of ANOVA, and employing an inappropriate analysis to a dataset, can result in incorrect conclusions. The main function of the ‘Materials and methods’ section of a scientific article is to provide sufficient details and information so that a knowledgeable reader with access to the original data can verify and reproduce the reported results (Curran-Everett and Benos, 2004). Our applications, as well as the citation of the literature, have been limited due to incomplete description of the experimental designs and statistical methods employed in many of the articles we reviewed, thereby making it difficult if not impossible to replicate the experiments or the statistical analysis. Statistical methods and analysis are inherent to many allied fields and underpin the scientific discovery process. As stated eloquently by Curran-Everett and Benos (2004), misunderstanding and misuse of the statistical techniques as well as misinterpretation of the analysis jeopardizes the scientific discovery process as well as accumulation of scientific knowledge. We hope that this article will serve to improve the caliber of statistical information as well as the reporting and presentation of the statistical techniques in allied scientific publications.

R.S.P.’s research was supported in part by the National Science Foundation (NSF) DMS 02-39053 and Office of

Naval Research (ONR) grants N00014-02-1-0316 and N00014-04-1-0481. C.L.'s research was supported in part by the NSF grant DMS 03-06202 and ONR grant N00014-04-1-0481. The authors thank the Editor and the reviewers for constructive comments and suggestions that led to significant improvements of the manuscript. We especially thank Mark Willis for providing the *P. americana* experimental data, for important discussions with regard to the relevant biological literature and for several comments on the manuscript. The authors are grateful to Joseph Koonce and Christopher Cullis for stimulating discussions and for their comments on the earlier version of the manuscript.

References

- Baker, T. C.** (1990). Upwind flight and casting flight: Complementary phasic and tonic systems used for location of sex pheromone sources by male moths. In *Proc. 10th Int. Symp. Olfaction and Taste* (ed. K. B. Døving), pp. 18-25. Oslo, Norway: Graphic Communications Systems.
- Baker, T. C., Willis, M. A. and Phelan, P. L.** (1984). Optomotor anemotaxis polarizes self-steered zigzagging in flying moths. *Physiol. Entomol.* **9**, 365-376.
- Cardé, R. T. and Knols, B. G. J.** (2000). Effects of light levels and plume structure on the orientation manoeuvres of male gypsy moths flying along pheromone plumes. *Physiol. Entomol.* **25**, 141-150.
- Cardé, R. T., Staten, R. T. and Mafra-Neto, A.** (1998). Behavior of pink bollworm males near high-dose, point sources of pheromone in field wind tunnels: Insights into mechanisms of mating disruption. *Ent. Exp. Appl.* **89**, 35-46.
- Cochran, W. G. and Cox, G. M.** (1957). *Experimental Designs*. New York: Wiley.
- Curran-Everett, D. and Benos, D. J.** (2004). Guidelines for reporting statistics in journals published by the American Physiological Society. *Am. J. Physiol. Regul. Integr. Comp. Physiol.* **287**, 247-249.
- Dekker, T., Takken, W. and Cardé, R. T.** (2001). Structure of host-odour plumes influences catch of *Anopheles gambiae* s. s. and *Aedes aegypti* in a dual-choice olfactometer. *Physiol. Entomol.* **26**, 124-134.
- Justus, K. A., Schofield, S. W., Murlis, J. and Cardé, R. T.** (2002). Flight behavior of *Cadra cautella* males in rapidly pulsed pheromone plumes. *Physiol. Entomol.* **27**, 58-66.
- Linn, C. E., Campbell, M. G., Poole, K. R., Wu, W.-Q. and Roelofs, W. L.** (1996). Effect of photoperiod on the circadian timing of pheromone response in male *Trichoplusia ni*: relationship to the modulatory action of octopamine. *J. Insect. Physiol.* **42**, 881-891.
- Linn, C. E., Campbell, M. G. and Roelofs, W. L.** (1988a). Temperature modulation of behavioral thresholds controlling male moth sex pheromone response specificity. *Physiol. Entomol.* **13**, 59-67.
- Linn, C. E., Hammond, A., Du, J. and Roelofs, W. L.** (1988b). Specificity of male responses to multicomponent pheromones in noctuid moths *Trichoplusia ni* and *Pseudoplusia includens*. *J. Chem. Ecol.* **14**, 47-57.
- Linn, C. E. and Roelofs, W. L.** (1983). Effect of varying proportions of the alcohol component on sex pheromone blend discrimination in male oriental fruit moths. *Physiol. Entomol.* **8**, 291-306.
- Littell, R. C., Milliken, G. A., Stroup, W. W. and Wolfinger, R. D.** (1996). *SAS System for Mixed Models*. SAS Institute Inc., Cary, NC.
- Mafra-Neto, A. and Cardé, R. T.** (1998). Rate of realized interception of pheromone pulses in different wind speeds modulates almond moth orientation. *J. Comp. Physiol. A* **182**, 563-572.
- Milliken, G. A. and Johnson, D. E.** (1984). *Analysis of Messy Data*. Volume 1: *Designed Experiments*. New York: Van Nostrand Reinhold.
- Pinheiro, J. C. and Bates, D. M.** (2000). *Mixed-Effects Models in S and S-Plus*. New York: Springer.
- Searle, S. R.** (1971). *Linear Models*. Wiley, New York.
- Searle, S. R., Casella, G. and McCulloch, C.** (1992). *Variance Components*. New York: Wiley.
- Sokal, R. R. and Rohlf, F. J.** (1981). *Biometry: The Principles and Practice of Statistics in Biological Research*. Second edition. New York: W. H. Freeman and Company.
- Venables, W. N. and Ripley, B. D.** (2002). *Modern Applied Statistics with S*. Fourth edition. New York: Springer.
- Vickers, N. J.** (2002). Defining a synthetic pheromone blend attractive to male *Heliothis subflexa* under wind tunnel conditions. *J. Chem. Ecol.* **28**, 1255-1267.
- Willis, M. A. and Arbas, E. A.** (1991). Odor-modulated upwind flight of the sphinx moth, *Manduca sexta* L. *J. Comp. Physiol. A* **169**, 427-440.
- Willis, M. A. and Avondet, J. L.** (2005). Odor-modulated orientation in walking male cockroaches *Periplaneta americana* L., and the effects of odor plumes of different structure. *J. Exp. Biol.* **208**, 721-735.
- Willis, M. A. and Baker, T. C.** (1984). Effects of intermittent and continuous pheromone stimulation on the flight behavior of the oriental fruit moth, *Grapholita molesta*. *Physiol. Entomol.* **9**, 341-358.
- Willis, M. A. and Baker, T. C.** (1987). Comparison of manoeuvres used by walking versus flying *Grapholita molesta* males during pheromone-mediated upwind movement. *J. Insect Physiol.* **33**, 875-883.
- Willis, M. A. and Cardé, R. T.** (1990). Pheromone-modulated optomotor response in male gypsy moths, *Lymantria dispar* L.: upwind flight in pheromone plume in different wind velocities. *J. Comp. Physiol. A* **167**, 699-706.
- Zanen, P. O. and Cardé, R. T.** (1999). Directional control by male gypsy moths of upwind flight along a pheromone plume in three wind speeds. *J. Comp. Physiol. A* **184**, 21-35.