**Strengthening Capacity in Environmental Physics, Hydrogeology and Statistics for Conservation Agriculture Research**

CEPHaS
Capacity for Conservation Agriculture Research

# Statistical Checklist for Planning Experiments

This document presents a checklist to be used when planning experiments. It is produced initially for the use of Working Groups of the CEPHaS project, but is also offered as a wider resource for capacity strengthening in conservation agriculture research. In preparing this checklist we are glad to acknowledge the inspiration of J.R.N. Jeffers's *Statistical Checklist* series (e.g. Jeffers, 1978). Jeffers's checklist is available in open-access form (see the references for a link) and we recommend that it is read in conjunction with this checklist.

This checklist is not a substitute for a discussion with a statistical advisor prior to committing to a particular experimental design. Rather it is intended that it should help identify possible issues which a particular experiment might face, and to facilitate discussion with a statistician by ensuring that key issues have been thought about in advance. In the context of the CEPHaS project it is proposed that this checklist be completed and responses recorded in a separate document and sent to the co-leads of Working Group 4 (Murray Lark and Joseph Chimungu) along with any additional commentary as a basis for consultation and advice.

CEPHaS Working Group 4. 27th February 2018.



RESEARCH COUNCILS UK

CEPHaS is funded by Research Councils UK through its Global Challenges Research Fund programme

GCRF
Global Challenges Research Fund

## How to use this checklist.

Please use this checklist in collaboration with *all* staff involved in planning and executing your experiment.

1. Consider each point in turn and write a short response (one or two sentences). Record these in a separate document. A response might be a ***specification*** (e.g. a list of treatments in response to A1), a ***decision*** (e.g. a choice of acceptable statistical power in response to D14) , a ***query*** that must be resolved within the experimental team (e.g. if further discussion is needed to identify all orthogonal contrasts of interest in response to B6), a ***query*** that must be resolved in discussion with other experts (e.g. what is the minimum yield response of practical relevance, in response to A4) or a ***query for clarification*** to the statisticians in Working Group 4 (e.g. if the correct analysis of variance table for a proposed experiment is not clear).

2. If any point is unclear, or not understood, then indicate this in your response. Feel free to contact Working Group 4 for guidance before finishing your checklist document. Do not worry if some of the terms used in the checklist questions are not familiar to you. Deal with them as fully you can and ask for clarification.

3. You should address all sections in the checklist (although one of sections E and F might be irrelevant and can be ignored).

4. Identify any other issues or queries that the checklist raises.

5. In a case where you are modifying an existing experiment, append as much information as you can about that experiment, including a sketch that shows its layout in the field and an example analysis of variance table.

6. Identify any additional queries that you may have about your experimental design. We suggest that you examine the checklist of Jeffers (1978), available online (see the link in the references), to help with this.

7. Send the completed response to the Working Group 4 co-leads (Murray Lark and Joseph Chimungu). A response will be returned as soon as possible.

8. Please note that planning an experiment is part of the scientific process, and requires time and effort. Do not leave it to the last minute. Also, recall Jeffers's words "There is usually little that a statistician can do to help you once you have committed yourself to a particular experimental design". In the similar (if more chilling) words of R.A. Fisher "To consult the statistician after an experiment is finished is often merely to ask him [*sic*] to conduct a post mortem examination. He can perhaps say what the experiment died of".

The co-leads of Working Group 4 are Murray Lark (murray.lark@nottingham.ac.uk) and Joseph Chimungu (josephchimungu@gmail.com)

# Statistical checklist for planning experiments in conservation agriculture research.

## A. Treatments

1. Choose treatments to answer questions within the aims of your research. List them so that you can apply them to distinct plots in your experiment and ensure that they are complete.

2. If you have two or more factors to consider then combine the treatments in a single experiment with a factorial design. For example, if your factors are manure (M) and lime (L), and each factor has two levels (0: none applied, A: a standard dressing applied) then the four combinations, {L0,M0}, {L0,MA}, {LA,M0}, {LA,MA}, are four treatments in a 2 by 2 factorial design.

3. In the case of a factorial experiment (2 above) consider whether you expect factors to *interact* (i.e. the response to one factor depends on the level of another factor) or to be additive (i.e. the response to one factor is fixed, regardless of the level of any other).

4. Decide whether any one of the treatments can be regarded as a control. A control might be a "conventional treatment" or a "check plot". In the case of an on-farm experiment a control might be "farmer practice". If there is not such a treatment then consider adding one to the experiment.

5. Specify minimum treatment effects (e.g. yield increases) of practical relevance (e.g. a minimum grain yield increase of 100 kg $ha^{-1}$) if you can. This effect size may be based on practical criteria. For example, what grain yield increase is needed for the mean yield in a region to meet annual household calorie requirements for a standard household of six people according to FAO guidelines? What increase in grain yield is thought by economists to be necessary for farmers to consider changing practice? What increase in soil organic carbon content is needed so as to meet critical thresholds for maintaining soil productivity? When thinking about effect size one should also consider factors such as the duration of the experiment and local environmental conditions

## B. Hypotheses

6. State as clearly as you can the hypothesis or hypotheses that you wish the experiment to test.

7. Express the hypothesis or hypotheses in terms of contrasts between your experimental treatments. Ideally identify a set of orthogonal contrasts between treatments; see Webster and Lark (2018) for an example.

8. Decide which possible contrasts between treatments, and interactions between factors, are most important. Think whether any are so unimportant that you could ignore them and thereby design a more efficient experiment (e.g. you might be willing to ignore the three-way interaction in the case of a factorial design with

three factors).

## C. Experimental Units.

9. Decide what the basic experimental units are to be, or, in the case of an existing experiment, identify what the basic experimental units are. In field experiments these will usually be plots. Consider what size or shape of plots is required given the response variable you wish to measure and other considerations such as the need for a buffer between plots with contrasting treatments. If the experiment is already established consider whether the current plot size and layout is satisfactory

10. Plan each experiment with plots of uniform in size and shape wherever possible. If this is not possible, or an existing experiment does not have uniform plots, consider the possible consequences and discuss these with a statistical advisor

11. Through discussion with researchers and technicians who have worked on the same site or neighbouring sites, and from reports or published papers on previous experiments there, summarize what you know in advance about the variation across the experimental site, i.e. among the plots. You should be aware of factors such as slope, susceptibility to flooding or localized pest or disease risk which might create major spatial trends across the site in the measured response. Such variation might be managed by blocking (see 19 below). Also make sure you have information about previous activities on the site. You should obtain, wherever possible, information on land use and management — crops grown, tillage practices, fertilizer management, water management (rainfed or irrigated), crop residue management etc. — for at least the two more recent cropping seasons.

12. Obtain quantitative information on the variability of the measured response in previous experiments on the site or neighbouring sites (noting whether the experiments included blocking). Ideally this information will be expressed as the residual mean square from analyses of variance on past experiments.

13. Assemble relevant information on other variables measured on the plots (e.g. data from previous soil surveys, remote sensor surveys, geophysical surveys such as ground penetrating radar or electromagnetic inductance).

## D. Replication

*Consider this section even when the number of replicates is already fixed, in an existing experiment, or constrained by costs and logistics.*

14. Do you have positive information on the effect size of interest (see 5 above) and on the expected residual variability of the measured response (see 11 above)? If not, reconsider possible secondary sources of information on both (e.g. observed residual mean squares from comparable experiments on similar sites).

15. Decide what *P*-value you would regard as acceptable evidence to reject a null hypothesis of no difference between treatments in your experiment. Recall that the *P*-value is the probability of seeing evidence and strong or stronger than the

evidence your experiment provides if there is no underlying difference between the treatments, so a small value is evidence to reject the null hypothesis. A value $P \leq 0.05$ is conventionally regarded as adequate evidence to reject the null hypothesis, but you may choose another value. We advise you not to automate the interpretation of *P* values, but for purposes of experimental planning a target value is helpful.

16. Decide on an acceptable power for your experiment. The power of an experiment is the probability that an effect of specified size can be detected as significant with the *P* value specified in response to (15). A power of 0.8 is commonly specified, but again you may have different views. If you are not familiar with the idea of statistical power then make sure you highlight this concern in your responses to this checklist and obtain guidance from Working Group 4.

17. Calculate the number of replicates of your experiment required to achieve your target power, given answers to (12, 14, 15 & 16). Again, request assistance from Working Group 4 if this is a procedure with which you are not familiar.

18. Take into account any constraints in the design and layout. These are likely to be a limit on the number of replicates constrained by the budget and the sizes of the plots constrained by the area that can be managed. If a feasible number of replicates does not give your target power, calculate the power you can achieve and review plans with the whole project team, requesting input from Working Group 4 as necessary.

## E. Establishing a new experiment.

19. Consider how a randomized block design might improve the efficiency of your experiment. In the simplest case a block will contain one complete set of all treatments, and differences between blocks will correspond to sources of variation across the site identified in response to (11), such as trends associated with slope, proximity to trees, etc.

20. If recording data on the experiment takes a long time (as harvesting a crop might) then make the measurements one block at a time so that any effect of time will be encompassed in the block effect. The same applies if several technicians are needed: assign each technician to one block and only one block so that differences between them are encompassed in the block effect.

21. If the number of treatments is large, it might be difficult to accommodate one replicate of each per block without making the blocks too large. Consider the following options.

i. *Incomplete blocks*, with some treatments missing from any one block. If some contrasts or interactions are less important than others then further efficiency may be gained by using unbalanced blocks.

ii. *Confounded designs*. Efficiency gains might be possible if some contrasts or interactions can be ignored altogether.

Request help from Working Group 4 if you decide to consider the option of incomplete blocks or confounded designs.

22. It might be impracticable to replicate all factors at the same scale (i.e. on plots of the same size), as for example where one factor is a cultivation method and the other the application of manures. In this case, consider a split plot design ( 28 below).

23. Having considered all the above points in this checklist, write down a proposed experimental design, ensuring that all treatments are replicated and randomized and blocking as appropriate. At this stage you need not come up with your final randomized allocation of treatments to plots. **However, you must write down the outline analysis of variance table for your experiment, showing the different sources of variation (blocks, treatments, error) and their degrees of freedom.** Doing this is always an acid test of whether you have fully thought through your experimental design.

## F. Working with an established experiment

*In this section we consider a case where you are using an existing experiment. This might or might not be modified for your purposes.*

24. As highlighted in (11) above, make sure that you have complete information on previous practices on the site, in addition to the design of the experiment.

25. Make sure that you can specify the design of the experiment fully, including any blocking (and the basis on which blocking was done, e.g. perpendicular to a slope), subdivision of plots or other changes made to the design after the start of the experiment and any other constraints, such as the use of spatially balanced designs. Careful scrutiny of reports and papers, and discussion with other workers may be necessary to ensure that this is done adequately. Again, the acid test is that you can write down an outline ANOVA table for the experiment.

26. Consider whether the established experiment is sufficiently replicated for your purposes. If not, do the benefits outweigh the possibly inadequate power? This may require an appropriate power analysis (see section D, 14–17).

27. Consider whether you wish to modify the design during the course of the experiment, for example, by subdividing the plots to incorporate contrasting crop rotations. Do so with care ensuring that the new design is sound and that data from it can be analysed statistically. Again, draft an outline ANOVA table for the new design, if necessary in consultation with Working Group 4.

28. Consider whether you need to impose one or more new factors on the experiment (e.g. by adding a manuring treatment). If you do then you must decide how to accommodate the new factor(s). Split-plot designs will often enable you to do that. If a split-plot design is to be used, again draft an outline ANOVA table, and ensure that you know how to account correctly for the original and new randomization. Ask for assistance from Working Group 4 if necessary.

## G. Measurements

29. Have established and approved protocols for measuring experimental responses. Consider WG 4's Checklist: *Protocols for Plot-scale Sampling*.

30. Is a single measurement to be made on each plot (e.g. crop yield, or soil organic carbon content)?

31. If multiple measurements are to be made, (e.g. more than one soil core is to be analysed separately) then ensure that an appropriate analytical protocol is used reflecting the correlation between observations within the same plot. See Webster and Lark (2018), section starting *"Sampling within experimental plots"* on page 133.

32. Many experiments nowadays require repeated measurements to be made over time. If you repeat measurements in this way then ensure that you analyse the data appropriately in a way that reflects correlation between repeated measurements on the same units. This applies to the analysis of measurements repeated with low frequency (e.g. annual crop yields) or high frequency (e.g. hourly measurements from a sensor).

33. Consider the possible benefits of archiving sampled material for future use. Consult an organization with experience of archiving the relevant material.

34. Have procedures for secure back-up of data from your experiment.

35. Specify a format for the files in which to record the data and one that will be usable on most computer platforms. Bear in mind that some proprietary file formats (such as Excel spreadsheets) might not be readable in future. ASCII .txt or .csv files are good open formats.

36. Store metadata that will allow future users of the data to identify the experimental design and the position of each unit in that design. Ensure that the metadata are retained and backed up with the original data files. If complete metadata can be stored as a header in an open-format data file then the risk for confusion is minimized.

## H. Data Analysis

37. Our goal is reproducible research, with transparent analyses of data which outside observers can repeat for themselves. Provide evidence that randomization was done as claimed in the experimental design. A good way to do this is to use an appropriate R script, with the seed for the random number generator specified and recorded in the script. R scripts for these tasks can be provided or checked by Working Group 4.

38. Specify the analysis to be used, and the hypothesis tests to be undertaken, in advance of completing the experiment. State where you have registered or recorded your analytical plans (ideally in the form of an R script) so that they can be demonstrated subsequently. One approach is to e-mail scripts prior to completing the experiment to Working Group 4 for checking.

39. Demonstrate that you are testing pre-specified hypotheses, making maximum use of your experiment's power. The best way to do this is to record a set of planned orthogonal contrasts in your experimental plan, and to write the basic analytical scripts in advance, see (7). Methods for *post hoc* testing of differences between treatments (e.g. Tukey's Honest Significant Difference, Scheffé's Critical Difference

and the associated use of letters to designate “significantly different treatment means”) are not appropriate as the primary means of inference in an experiment. They are designed for checking additional differences identified in the “wash up” phase of analysis.

40. *Ensure that your analysis fits the design* (Webster and Lark, 2018). All constraints on randomization (e.g. blocking, split plots) must be reflected in the analysis of variance table, which should be reported in any write up of the experiment. You should send your R scripts to Working Group 4 co-leads for advice and comment from the wider group. Note that Working Group 4 aims to assist you in developing scripts for analysis of data using R, but it requires you to have completed this check list first.

**References**

Jeffers, J.N.R. 1978. *Design of Experiments*. Statistical Checklist 1. Institute of Terrestrial Ecology, Natural Environment Research Council, Swindon U.K. Available online, open access at: http://nora.nerc.ac.uk/id/eprint/5271/

Peng, R.D. 2009. Reproducible research and Biostatistics. *Biostatistics*, **10**, 405–408 Available online, open access at: https://academic.oup.com/biostatistics/article/10/3/405/293660

Webster, R., Lark, R.M. 2018. Analysis of variance in soil research: let the analysis fit the design. *European Journal of Soil Science*, **69**, 126–139. A preprint will be available open access from http://nora.nerc.ac.uk/id/eprint/519233/ after 18th January 2019.

**Acknowledgements.**

This checklist was prepared as part of the CEPHaS project which is funded by Research Councils UK through their Global Challenges Research Fund.

This checklist was prepared by staff of Working Group 4 of the CEPHaS project from University of Nottingham, University of Zimbabwe, Lilongwe University of Agriculture and Natural Resources, University of Zambia and Zambia Agriculture Research Institute. We are grateful to Professor Richard Webster, Department of Computational and Analytical Sciences at Rothamsted Research, for helpful comments on an initial draft of this checklist.

This checklist is published with the permission of the Director of the British Geological Survey (NERC).