Writing up research: a statistical perspective

E. F. Allan, S. Abeyasekera

& R. D. Stern

January 2006

The University of Reading
Statistical Services Centre

Guidance prepared for the DFID Forestry Research Programme

Foreword

The production of this guide and “Writing Research Protocols” has been funded by DFID’s Forestry Research Programme (FRP). This stemmed from a review undertaken by the Statistical Services Centre (SSC), University of Reading, of the use of biometrics by DFID RNRRS research projects (SSC, 2002) which highlighted difficulties faced by scientists in data analysis and reporting, as well as in the planning of investigations, in data management and archiving. One recommendation was that support should be provided on how to write up a data analysis for data-related research work. This guide serves to address this need. It is aimed at discussing good approaches to such reporting procedures. It is focussed on natural resources research, but the general principles are equally relevant in other types of research.

The main audience for this guide is natural resources research scientists. The aim is to provide them with help on how to report results of investigations in a way that clearly demonstrates the evidence base for research recommendations. The guide has been written primarily with the inexperienced user of statistics in mind, but it may also be a resource for an experienced researcher who is writing up research results from studies with multidisciplinary objectives and where multidisciplinary teams are involved.
Our main objective is to support research teams in producing technical reports that are fit for the purpose for which they were intended. Amongst other things, the technical report should serve as a programme manager’s review of the project’s statistical detail with respect to the correct analysis, correct reporting and correct inferences. It is also valuable for the project managers themselves since they can monitor progress to date of research activities within their project and quickly address shortcomings that are identified along the way.

The guidance provided here on principles underpinning the detailed reporting of the data analyses should allow research teams to adopt and adapt these principles in the circumstances of their own individual projects. Further guides on topics such as data management and statistical analysis, provided by the SSC at www.rdg.ac.uk/ssc , and recently updated and incorporated within Stern et al (2004), are available to complement this guide, as is the companion guide on “Writing Research Protocols: a statistical perspective”.

Acknowledgements

We are very grateful to the Forestry Research Programme of the United Kingdom Department for International Development (DFID) for funding this work (Project ZF0183). However, the views expressed are those of the authors and are not necessarily those of DFID.

Index

1. Introduction

1.1 The context

This guide, and its companion guide "Writing Research Protocols", has been developed to assist research scientists with the statistical aspects of reporting their research and in developing appropriate documentation of the research processes.

A research project usually involves a set of activities, each with its own output, where an activity is a clearly identified exercise ranging from a literature review through to a large data collection exercise. To have impact, project results have to be disseminated appropriately, e.g. through research publications, policy briefs, radio broadcasts. It is of course important that in doing so, any messages being disseminated are based on a solid evidence base. Where the message is being drawn from primary data collection exercises, it is all the more important that clear reporting forms the basis of the recommendations being made, so that the user – project leader, programme manager – can assess the value of the work, and decide how best to take it forward.

1.2 Aims of this guide

The SSC's review mentioned above distinguished between "above-deck" and "below-deck" management of the project. It was argued that project leaders, programme managers and donors are already involved in "above-deck" management, i.e. aspects of management that refer to the overall objective-setting and planning, finalisation of outputs and reporting that explicitly relate to scientific objectives. In this context, a policy brief or the text of a radio broadcast is an example of an output that is "above-deck". Such outputs must be underpinned by solid evidence – i.e. the evidence that may be reported in the Final Technical Report, in a detailed activity report, in an annual report, or in a published article.

"Below decks" management refers to planning, actions and reporting during the progress of the research process. The extent to which such reporting is exposed to scrutiny by outsiders varies from project to project. However, the quality of the overall project outputs depends heavily on the quality of this below decks management of the research process. An essential component of this level of management is that it should lead to well-organised, adequately comprehensive and clear reporting. Unfortunately this is rarely done well for a number of reasons:

- Most energy is often put into *conducting* the research, with less time spent on *reporting and justifying* the research results explicitly. There are a variety of reasons for this, not least of which is the project team running out of time and resources. However assimilating and reporting of results is as necessary and productive a part of the research as conducting field data collection.
- There is a lack of awareness regarding the reasons *why* the explicit reporting is important. We argue here that it is important for both the research scientists themselves and their managers.
- Whilst most scientists have knowledge of some statistical methodologies - some of them at quite an advanced level – few are conversant with the whole statistical analysis process, or the general philosophy of statistical modelling that allows them to get the most out of their data.
- Poor recognition of the need for good data management and the need to provide an explanation of data handling steps undertaken as required to ensure the analysis capitalises on and correctly represents the information collected. This is key to successful delivery of outputs. Some pointers in this direction are provided.
- Inadequate awareness of the need to provide a clear demonstration of how the analysis and results relate to the research objectives. This is another area that we address.

- Researchers often lack confidence in how to write up the statistical aspects of their work – possibly because they have never received any training or guidance on how to do so. Most statistics training courses for researchers give little instruction - if any – on the statistical write-up of a data collection exercise.

Ideas necessary for good reporting of statistical results and what constitutes a good report of research results are illustrated in this booklet using examples drawn from the areas of forestry, agro-forestry and agriculture. They have been chosen to illustrate a range of issues and to cover experiments, surveys and participatory work.

2. What constitutes a good report

2.1 The reporting structure

Researchers generally report the findings from primary data collection exercises according to the following structured format, i.e.

i. background and objectives of the research;
ii. materials and methods;
iii. the data and their analysis, presentation of results and interpretation;
iv. conclusions and plans for future work.

However, the statistical aspects of this reporting are often limited. There is usually insufficient detail, or the detail lacks clarity, to demonstrate how the evidence emerging from the research leads to reported conclusions, and how this in turn links to the original objectives of the research. The report should allow a researcher from the same team, or a reviewer, to assimilate all the information about:

- the analysis objectives, and how the particular analysis being reported fits into the wider context of the project;
- the research methodology and the data collection process, together with methods used for the analysis;
- procedures used for managing the data effectively while ensuring data quality;
- the analysis results and what they reveal;
- any limitations with the analysis and uncertainties associated with the findings.

The reason for this is so that a reader (who has an understanding of statistical principles) can judge whether the conclusions drawn are valid and whether the investigation was of sufficient size for reliable conclusions to be drawn in relation to the stated objectives. This means that the write-up must contain sufficient detail to allow the reader to assess whether:

- the analysis as completed was appropriate for the data;
- the analysis results have been appropriately summarised and presented;
- the interpretation of the analysis is correct.

The checklist in Section 2.2 provides an overview of the key components that need to be considered in the write-up process, and why they are essential for good reporting. The themes are then elaborated more fully with illustrative examples. We restrict attention to writing up the analysis for an individual research activity, whether it corresponds to a survey, on-farm or on-station experiment, lab activity, participatory research, or any other type of activity. Our aim is to give guidance on what, and how much, information needs to be provided to ensure effective reporting.

We include as appendices to the guide two examples of analysis write-ups that we regard as well written from a statistical point of view. Appendix A1 is a relatively straightforward piece of work. Appendix A2 involves a more complex data analysis and we have included it to illustrate to the reader that the same "good statistical write-up" principles apply no matter how simple or complicated the piece of work.

2.2 A checklist of components to be reported

Relating the objectives to the intended analysis

Whether the analysis reporting corresponds to the project as a whole or to specific individual activities within the project, the reporting procedure must relate to analyses that are within the context of the research objective(s). These objectives should be stated, together with an outline of the corresponding research activities, so that the relationship between the analysis, the research activities and objectives is clear. Much of this could be drawn from the study protocol (see Wilson & Abeyasekera, 2005).

Summarising "Materials and Methods"

Sufficient detail should be provided of materials and procedures needed to understand the analysis undertaken. Of particular relevance from a statistical point of view is a clear specification of (a) the underlying study population to which the analysis results are targeted; (b) the choice and specification of the study units used in the analysis, and number of units; (c) the data structure and (d) the procedures used in collecting the data. Again, much of this can be extracted from a well-written protocol. See Section 3.2.

How the data were managed

Too often little is said in reports about the way research data have been managed, yet good data quality is an essential element of high quality research. Confidence in research results derived from data collection activities will be enhanced by demonstrating that sufficient care has been given to data quality issues and data management, and we encourage more reporting of this area, however brief. See Section 3.3.

Relating the analysis to the objectives

A well-written protocol will have provided an outline of the analysis approaches and how they relate to the objectives (see Wilson & Abeyasekera, 2005). At the data analysis stage of the project these suggested approaches will need to be reviewed in the light of project activities to date to ensure that they are still appropriate and that the data collected addresses the objectives. If not, suitable modifications will need to be made to at least one of objectives and analysis approaches. The analysis report should contain the analysis approaches that were actually used and how they relate to the objectives, with some discussion as to why they were different from the original protocol. Section 3.4 expands on this theme.

Presenting and interpreting the results

Many researchers seem to have difficulty with this part of writing up their analysis. How much to present and how best to present it has to be considered carefully if the readers of the research results are to be convinced about the value of the findings. In Section 3.5 we give hints on the level of detail required for this purpose, while a discussion of the importance of indicating uncertainties associated with the findings is provided in Section 4.

Reporting and presenting conclusions

In reporting and presenting the conclusions, they must be linked not only to the results already presented, but also to the research objectives and the extent to which the results are generalisable. We deal with this area in Section 3.6. When presenting conclusions it is also important to answer the "so what" question for every research finding.

3. The analysis write-up for a research activity

3.1 Objectives

The write-up should have a clear statement of the objective(s) being addressed by the specific research activity under consideration, reasons for undertaking the activity, and a description of the activity. These should be in accordance with objectives and activities set out in the original protocol(s); with any modification to the objectives and/or methodological approaches whilst the study was underway explicitly stated. This means that the write-up should describe what actually happened rather than what was originally proposed in cases where changes occurred. In situations where the activity protocol was followed as planned, a summarised version of the relevant sections from the protocol can be extracted into the report so that it is self-contained, with reference to the protocol for further details.

In a full technical report, the background, objectives and justification for the activity needs to include a demonstration of how it fits within the wider context of the project as a whole. For a shorter write-up, such as a published paper, an introductory section is necessary, but might restrict itself to the primary objectives of the investigation. These same principles apply to conference presentation slides or an in-house slide presentation to colleagues.

3.2 Summarising "Materials and Methods"

Background to the analysis write-up can easily be extracted from relevant components of the materials and methods section of the protocol. This provides the elements necessary to understand and "verify" the analysis. These are (a) the characteristics of the study population and representativeness of the study sample, and (b) a complete description of the data, the data collection process and the data structure, including any data hierarchy.

3.2.1 The target population

The reason it is important to give a clear statement of the target population under study is that it relates to the objectives and hence has a direct bearing on the inferences that can be drawn from the study. We return to this issue in section 3.5. Elements that help to clearly describe the study population are both the biophysical and/or socioeconomic background information as well as more specific details relating to the sampling units themselves, e.g.

- physical location (single location or several named sites, including their agro-ecological and/or other locational characteristics)

For instance, if the reporting makes it clear that trees under study are saplings grown in a hot arid environment, the reader knows that conclusions from the study cannot be extrapolated to other cooler and wetter environments. If a study is aimed to understand mechanisms through which people obtain information relevant to improving their livelihoods, focussing on areas targeted by local NGOs will only allow extrapolation to the NGO coverage rather than to all rural communities in the region.

- characteristics of communities/farmer groupings/households etc. e.g. rural communities, farmers growing bananas as a cash crop, households below a defined poverty line

A study of market potential for improved, disease resistant bananas may be restricted to a study population of only those who grow bananas as a cash crop, rather than all banana growers. If the research focus is to help poor communities move towards greater diversification of their livelihood strategies, and the study population is restricted to "poor" communities then a clear indication of what the researchers define as being a "poor" community is required.

- species of trees/animals under study, age of persons/animals, etc.

For example, for the reader to be informed of the extent to which the results can be generalised, the report of a long-term forestry trial should indicate not just how many, but which, species are being studied; and a livestock experiment should indicate which breeds were used and the age range of the animals.

3.2.2 "Treatments"

In a designed experiment, whether it is an on-farm, on-station or laboratory experiment, the experimental treatments must be stated; and it should be made clear what form, if any, the "control" treatment takes – e.g. is it a "non-active" treatment or is the experimental unit left untreated. The reason for such detail is for the reader to see whether the stated objectives can be addressed by the particular experiment. Most of this information can be taken from the protocol; but any variations need to be mentioned. A clear statement of the treatment (and objectives) also flags to the reader the types of treatment comparisons to expect in the subsequent data analysis – e.g. factorial main effects and interactions, relevant treatment contrasts.

3.2.3 Selection of units and sample size

A basic assumption in most data analysis is that the sample studied is representative of a larger population of interest, and so the description of the sampled units needs to be reported. The sample size is also key to demonstrating the value of the work and the extent to which the results can be generalised to the wider target population. In small studies, describing the sampling procedure and number of units is usually straightforward – e.g. in a farm survey of constraints on cassava production in one single district, a random sample of farms will be appropriate if the agro-ecological conditions within the district are relatively uniform and if farms in the district do not fall into obvious socio-economic groupings.

Larger studies are not so straightforward – the population often has some particular features, e.g. different poverty groups, or clusters of communities with varying levels of ethnic diversity; different types of farming practice. Such diversity – resulting in different strata or clusters - has to be taken into account when selecting the sample, and these need to be described in the reporting, to demonstrate how the sampling features enter into the data analysis.

If appropriate, the multistage or hierarchical nature of the sampling process also needs to be described, e.g. sampling of districts, then villages, then communities and finally farming households. Not only should such sampling stages be specified, the actual method of selection at each stage should also be given e.g. stratified, purposive or other. Such details would normally be included within the sampling protocol, but a summary of this is also needed in the analysis write-up.

Reporting the sampling process is particularly important in those participatory research activities that intend to generalise the findings to a wider target population. A brief outline of the process used to select communities and the process for bringing together stakeholders within the community should be a component of the write-up. This will provide some defence against the common criticism that participatory work is little more than case studies. Similarly, when the sampling is restricted in some way, the reader is able to put some "mental caution" on the reported findings.

In a designed experiment - be it on-farm, on-station or a laboratory experiment - the design layout, a specification of the type of experimental unit (e.g. farm, field, plot, pot, animal), and the way treatments were allocated to the units all needs to be outlined. This information can be extracted from the protocol. For the design layout, it is not sufficient to say that an experiment was "carried out as a randomised block design". The report should state the features used to block the experiment – e.g. sloping ground, different types of farming practice – and whether or not the blocks were complete or incomplete. The layout and treatments both help to determine the subsequent data analysis. If these are clearly detailed in the write-up, the reader is able to assess the appropriateness of the analysis in later sections of the report.

3.2.4 The data and the data structure

A key component of the analysis write-up is an outline of (i) what data were collected; (ii) how they were collected, and (iii) a description of the "*data structure*" resulting from the data collection process. The first two of these would be a summary extracted from the relevant parts of a protocol, but any departure from what was originally planned (and why) needs to be discussed. We say more about the actual data in Section 3.4.

By *"data structure"* we mean the way(s) in which the data can be visualised and tabulated to show the numbers of observations falling within different locations, strata, time points and other sampling features and/or data hierarchies. This allows the reader to form an idea of the size of the study and the characteristics of the data that relate to the statistical analysis.

It is common practice in surveys to report on any sampling hierarchy. On-farm experiments or participatory studies where data are collected at different levels also need to report this level of detail. For example, in a participatory study on access to markets, the statement that information was collected from 185 respondents is incomplete if it omitted to state that the responses were from 4 focus-group meetings, one in each of four villages. Here some data may be collected at the focus group level, some at a person level; and the hierarchical nature of the data must be recognised and used in the analysis.

Hierarchical, or multilevel, structure is a particularly important feature of data structure, and always needs to be described. We have discussed it already in terms of survey sampling. Other examples of multilevel data also exist; for example when several samples are made on a plot from the same field in an on-farm study. Another common, and important, type of "data structure" is when measurements are made on the study units over time. These data complexities - multilevel structures and repeated measurements – can result in analysis complexity, and failure to properly recognise them can result in misguided statistical analyses and incorrect conclusions. A clear description of the data structure is therefore essential to flag to any reader that the data structure has been properly identified and dealt with.

One difficulty with recording "detailed information" or "clarity" is to know what is the correct amount and what is not. Box 1 shows some examples where there is room for misinterpretation in what has been written.

Many of these elements that constitute "materials and methods" – population, units, sampling strategy, data structure - cannot be easily separated into different sections in the way that we have tried to do here. Often it is best to describe them all together into one or two well-worded paragraphs. Box 2 provides an illustration of this.

Box 1: Describing the data fully (or not!)

Example 1 – inconsistent description

Description: "The study examined the tree growth of two particular species. Each species was grown in twelve 4-tree plots, and the circumference (dbh) was measured on trees in the plots."

The above implies that, if all the trees are measured, a total of 96 measurements would be available for analysis. If the subsequent analysis section indicates, without explanation, only 48 observations then the reader is left wondering what happened to the other 48. Was the original plan to measure only 2 of the 4 trees in each plot, in which case the dataset is complete, or did something unexpected happen in the trial to render so much data missing.

Had the above description said "…circumference (dbh) was measured on two trees, selected at random in each plot" or "…the circumference (dbh) was measured on all trees in the plots" then there would have been no doubt.

Example 2 – plot size not given

Description: "The study examined the tree growth of two particular species. The circumference (dbh) was measured for 48 trees of each species".

Unlike the example above, we have not been given the plot size. Plots of trees often contain several trees, thus the above statement could apply to a situation with twenty-four 2-tree plots per species, sixteen 3-tree plots per species, twelve 4-tree plots or eight 6-tree plots or even 48 single-tree plots. The report should contain either an explicit statement of the actual replication (the twenty-four, sixteen, etc. plots) or sufficient detail of the hierarchy so that the actual replication can be determined.

Example 3 – number of plots not clear

"… 12 plots were harvested on each of the three dates, 14th March, 1st April and 16th April".

From this statement it is not clear whether there were 36 plots in total, with 12 different plots harvested on each date; or whether each of the 12 plots had a section harvested on 14th March, another section harvested on 1st April and a third section harvested on 16th April. The statistical analysis is different for these two scenarios.

Box 2: Description of the sampling and data structure for a participatory study

This example corresponds to a study aimed at establishing an understanding of instruments and mechanisms by which people obtain information from available sources, the perceptions that different client groups have of the quality of information itself and the reasons for choosing different information sources. Studies were undertaken at selected locations in the North-West and North-East of Bangladesh for this purpose. The reporting of the data structure below is for work carried out in the North-West with fieldwork carried out by members of the PRA Promoters' Society of Bangladesh (PPS-BD), with support from the major NGO and collaborating partner institution in the north-west, i.e. the Rangpur-Dinajpur Rural Services (RDRS).

After several rounds of pilot testing, the final sampling procedure was restricted to RDRS's target population, i.e. those with no more than 1.5 acres of cultivable land. For reasons described in the full sampling protocol, it was decided to select male groups as well as female groups from each of five purposively chosen upazilas (sub-districts) in the north-west of Bangladesh. The purposive selection of upazilas ensured crop based agriculture was a primary livelihood, provided a reasonable coverage of the north-west and represented the two major agro-ecological zones in the area. Within the chosen upazilas, unions were selected as randomly as possible, but ensuring they were non-neighbours. At the time of the field work however, some changes were made because of practical difficulties encountered by the field team.

Tables 2.1 and 2.2 show the distribution of the number of groups across the upazilas covered during field work and the total number of people participating in the discussions.

Table 2.1 Distribution of groups by gender and whether in RDRS working area

		Number of participating groups				
Zone	Upazila	RDRS Male	RDRS Female	Control Male	Control Female	Total
Tista-Korotoa Flood Plain	Aditmari	3	1	1	1	6
	Kaliganj	2	2	1	1	6
Old-Himalayan Piedmont zone	Panchagarh		4	1	1	6
	Pirganj	1	3	1	2	7
	Thakurgaon		4	3		7
	Totals	6	14	7	5	32

Table 2.2 Participating numbers by gender and whether in RDRS target area

		Number of participating groups				
Zone	Upazila	RDRS Male	RDRS Female	Control Male	Control Female	Total
Tista-Korotoa Flood Plain	Aditmari	58	17	15	15	105
	Kaliganj	39	42	19	21	121
Old-Himalayan Piedmont zone	Panchagarh		68	15	15	98
	Pirganj	17	46	15	28	106
	Thakurgaon		71	45		116
	Totals	114	244	109	79	546

3.3 Managing the data

A prior condition to effective reporting of the data analysis is the recording, managing, and reporting of project information appropriately. We strongly advocate that project teams discuss and set up procedures for managing, reporting and archiving of the raw data and meta-data, at an early stage of the project. Here, meta-data refers to all other information associated with research activities, e.g. maps, questionnaires, field reports, etc. If all this is done at the start of the project, it will help to overcome issues that might be contentious later such as a partner claiming sole-ownership of data they have been funded to collect.

Systematic procedures for managing the information allow the data to be easily accessible to other project-team members as required. This will assist the research team to assess its own progress and enable a speedier start on the data analysis, immediately following data collection, computerisation and verification. There will sometimes have to be further work on the data to ensure they are suitably anonymised so the confidentiality of individual respondents is respected.

The reporting of data analysis results should include an outline of steps taken to assure data quality throughout the data-chain from collection to computerisation and analysis. There should be a full description in a separate data management report, as recommended in “Writing Research Protocols”. This is not usually seen in many projects but should be regarded as essential to give confidence that research results are based on well-managed and accurate information.

Donors who pay for the collection usually expect the data to be made available eventually for use more widely. This requirement can be a useful spur for the project team to define a suitable system for data management early in the project. Details of how this can be done are described elsewhere, e.g. Murekezi *et al* (2004), Stern *et al* (2004).

This increased interest in data availability means that archiving the data is becoming a more essential element of reporting research. In data archiving, confidentiality can sometimes be an issue, but this should be kept separate from the need for managing the information sensibly and according to procedures initially agreed by the project team. There are usually no issues of confidentiality concerning the meta-data, and hence they can be reported immediately. One way to supply the raw data early is to include them with password protection. A simple way is to zip the file and add a password at the same time.

The above issues would be mentioned in the analysis write-up of the *Final Technical Report*, and reference should be made to further details being available in a well-documented *Data Management Report*. A CD that contains the data archive should also include a copy of this report. This transparency of the steps taken to ensure that data quality and management of the data has been maintained to a high standard will enhance the value readers place on project results.

3.4 The statistical analysis

3.4.1 Analysis variables

The starting point of any analysis generally involves breaking down the overall research objectives into several more specific sub-objectives that can be linked to the data. Each such sub-objective is associated with one or more data variable(s) so that the manner in which the data relates to the objectives is clear. Such analysis variables need to be stated. In a full *Technical Report*, this would include all variables subjected to statistical analyses; plus a mention of those that were collected but not analysed (giving reasons). Shorter reports, such as published papers, need only mention the key response variables that address primary objectives.

The units associated with data variables should be given if the variable is quantitative. If a rating scale is used, both the number of points and the description of the points - e.g. a 3-point scale where 1=bad, 2=moderate, and 3=good – should be given.

Other essential information when discussing the analysis variable(s) are whether:

(i) the raw data were analysed;

(ii) a derived variable was analysed;

(iii) a transformation was applied, and if so which one and why?

A derived variable is one that is calculated from the raw data. It could be a simple average or total or something more complicated. An example could be "growth rate" for each tree in a study – where growth is determined from the raw *dbh*[1] collected over a period of time. Details of how the variable was derived should be given.

A transformation can be applied to the raw data or to a derived variable.

Sometimes there are "problematic" values associated with the use of a transformation. It may be useful, in a detailed report, to comment on how many of these there were, and how they were dealt with. A common example is the addition of a small number to zero values when the transformation used is the logarithm. The requirement to comment on such issues is sometimes a spur to consider an alternative approach. For example, zero values are often better approached by splitting the data into those units corresponding to the zeros and those corresponding to the non-zero values. The zero values are then discussed separately, followed by analysis of just the units with the non-zero values.

3.4.2 Describing the methods of analysis

All primary data collection exercises require some form of synthesis, and this usually means starting with an exploratory data analysis (EDA) before moving to more formal statistical methods.

There are two instances where one might want to report the results of an EDA. The first is to give confidence to the reader that the data has been scrutinised for oddities and for curious patterns that may be missed in a formal analysis. The second is to explain why a particular approach was taken in the subsequent analysis, e.g. to provide a justification for fitting a statistical model.

It is essential that the formal analysis, aimed at addressing project objectives is reported without ambiguity. The amount of detail depends on the type of report – a full report should contain the information for all the variables that were analysed, whereas published papers or presentations need only concentrate on the relevant subset. Our experience indicates that many reports include some statement of the methods of analysis but it is often not detailed enough to be properly informative.

The key analysis variables, together with the objectives, point towards the type of analysis required. Hence any reader, who has an understanding of statistics, would be looking to see if each primary analysis variable "matches up" with the analysis method. In experiments or surveys with a comparative element and a primary response variable that is quantitative - such as yield, *dbh*, etc. - statistical methods like t-tests, analysis of variance, linear regression might be expected. Categorical data, especially yes/no type data – might be analysed using chi-square tests or logistic regression. But more than this is needed in the reporting; the commonly used statement that an "analysis of variance (ANOVA) was performed" is too vague. Full details of the statistical model need to be given, i.e. of features in the analysis that:

- relate to the design structure, e.g. blocking variables or stratification variables;

[1] *dbh* - diameter at breast height

- treatment factors, in the case of designed experiments;
- confounding factors, e.g. distance to nearest road if the study was on marketable produce; and
- any other relevant factors, such as socio-economic characteristics of households if the study was (say) about the effects of spread of a plant disease on peoples' livelihoods

Thus for instance "analysis of variance with effects for block and varieties" might be a correct and complete description of the two-way analysis of variance used to analyse data from a randomised complete block design (RCBD) conducted on a research station. But data from an on-farm variety evaluation trial would most likely have been subjected to something more complicated, so that fuller (more complicated!) descriptions are required. The phrase "… analysis of variance with effects for farms ("blocks"), varieties, and different management practices, as well as the variety-by-management interaction" might be appropriate in such a situation.

Complex statistical modelling of data is possible with current software, and with this comes a more modern approach to describing the data analysis. Whilst we have no quibble with "analysis of variance with effects for block and treatment", the analysis of variance methodology, like linear regression, is part of the general linear model family of methods, and many writers nowadays, particularly those with messy data structures and more complex statistical models, are using phrases such as "… a general linear model with effects for farms, varieties and management practices …" to describe their analysis of variance.

Box 3 gives an example of a statistical model fitted to data from an on-farm experiment. We are not advocating that everyone's analysis should be as complicated as this - but if it needs to be, then the details need to be given. All the elements of the statistical model are given, with a description of which terms in the model are grouping factors (species, manure etc.) and which are co-variates – i.e. treated as being linearly related to the response - as in the case of the amount of additional fertiliser. To have used only the words "analysis of variance was performed" would not do justice to the analyst's effort.

Box 3: Describing the statistical analysis

This example corresponds to an on-farm experiment looking at six different agro-forestry practices, i.e. the planting of 3 different species implemented with and without manure. These 6 practices were replicated once on each of 6 different farms.

The following words allow readers to visualise the analysis of variance table below. They can check (i) whether any other effects are omitted from the analysis and (ii) whether the residual degrees of freedom are sufficient to reliably estimate residual variation on which statistical tests are based.

"A general linear model with main effects for farm, species and presence/absence of manure, and the species×manure interaction were fitted to the data. Farm, species and the presence/absence of manure were grouping factors in the model; and the amount of additional fertiliser was included as a covariate".

Analysis of Variance

Source	**d.f.**
Farm	5
Fertiliser	1
Species	2
Manure	1
Species×Manure	2
Residual	24
Total	**35**

The above is an example of a clear statement of the model for the analysis. This is not however being promoted as a particularly good design. [Many on-farm studies result in a farm by treatment interaction - for example, because of differences in their initial soil status different farms may respond differently to the different species. The full statement of the model, as given above, shows that there has been no allowance for a farm by treatment interaction in the model.]

3.4.3 Method of analysis – getting it right

Sometimes those who have issued the blanket statement "the data were analysed using ANOVA", merely report mean values – often only those for the main effects, i.e. with no discussion or reporting of interactions. They are likely to have ignored many potential sources of variation that should have entered the ANOVA, and also the standard errors in their reporting. In this case the simpler statement that "tables of means were produced" better describes the approach that was taken.

Often the justification put forward for limiting the analysis to the production of simple tables is that "the readers for whom this research is intended would not appreciate, nor be able to understand the concepts of a statistical model." This view has wrongly combined two different stages in the research - the first is the appropriate analysis of the data, as dictated by the objectives of the study, and the second is the dissemination of the results and recommendations.

For example, a simple analysis that makes one recommendation for a whole region can only be justified if the technologies being evaluated in the research showed no evidence of interaction with one or more characteristics of different sites within the region. Often, different recommendations may be needed for different sub-groups of the study area because, for example, the analysis showed different needs for farmers on land prone to flooding, compared to those who farm on high ground. In each case the results can be disseminated as simple tables, or graphs, or stories involving "typical farmers", but the story has to be underpinned and defended by an appropriate analysis with standard errors and confidence intervals.

There are instances when simple tables alone are the appropriate analyses. They may occur when the investigation is a pilot one or some other equally small study whose only objective is to identify some general pattern of response, or where the data are largely qualitative and simple descriptions in the form of tables and graphs are sufficient to fulfil the study objectives.

3.4.4 Method of analysis – other elements

If the project required extensive statistical modelling, then the writer should consider including more details on aspects surrounding the model fitting process rather than just the final model selected. Examples would include (a) more detail on the model selection process and (b) some information about model checking with a brief discussion of the findings. Some comments regarding limitations of the model fitting process is also valuable to alert the reader of assumptions made and the omission of potential factors that were impossible to measure and were therefore excluded. Box 4 illustrates an example from a participatory study.

A "danger sign" for a statistical reviewer is when the only details of the model-fitting process are a statement like "the model was produced using stepwise regression". These automatic selection methods are only intended to assist the analyst, and should not be allowed to take control. Choosing the most appropriate model needs the researcher's perceptions and experience, together with suitable modelling software and an understanding of the modelling process, not a "black-box" technique.

It is unlikely nowadays that anyone performs statistical analyses by hand, and so a statement of the statistical software that was used should be given when reporting the analysis. This is particularly important when some specialised methodology, only available in specialist software, has been used. When the analyses are more "main-stream" and were carried out in more readily-available software some statement of the actual software routines or procedures that were used should be given.

3.5 Results, their presentation and interpretation

The meat of the report is the section that presents and discusses the results. It is essential therefore that the wording of this section is accompanied by good quality tables and graphs so that the reader can follow the writer's train of thought easily. (See SSC, 2001a).

Equally important is putting the results into context - not just with regard to this particular data collection exercise (i.e. its objectives and underlying population), but also in the larger research area. Thus once the results have been correctly interpreted they need to be presented in the context of other similar studies that may have been done, and how these results influence the direction of future work.

3.5.1 The wider population

The results of a statistical analysis can only be generalised to the population that the data represent. For instance, if an investigation into the effects of a new approach to fisheries co-management on people's livelihoods has been restricted to sampling from regions where a certain NGO is active because they will assist in the data collection, then strictly speaking the results from the study can only be generalised to the areas where the NGO is active.

It is important that the reporting distinguishes between the evidence from the research and the opinions of the researcher. It is reasonable to include both, but the balance must always be in favour of the reported research. For example, the reporting may say "these results concerning constraints faced by rice growing farmers strictly apply to flood-plain areas in the north-west region of Bangladesh. However, the author's view is that the results are also likely to be relevant in other flood-plain areas of Bangladesh where rice growing is a major livelihood activity."

Box 4: Reporting results of statistical modelling using data from a participatory study

This example comes from a study undertaken to evaluate the effectiveness of Participatory Action Plan Development (PAPD) as a new systematic approach to building consensus among local stakeholders on problems and solutions for fisheries co-management. One key variable in this evaluation was an indicator of the change in social cohesion since the start of co-management activities. Following the use of an arrow diagram in the field, a numeric form of the indicator was derived on a –5 to +5 scale and obtained from results of each person participating in 18 focus group discussions at sites where PAPD was used and 18 where PAPD was not used. The key response for analysis was the average across group participants, thus 36 observations in total.

To compare results from PAPD sites with those from control sites, a general linear modelling procedure was used, allowing for six other potential variables (confounders) expected to affect the change in social cohesion, e.g. type of waterbody, proportion of fishers in the community. To determine which of the confounders contributed to explaining the variation in the indicator of social cohesion, a backward elimination procedure was used, whereby non-significant terms were dropped one by one and the model re-assessed to determine the significance of each of the remaining variables at every stage. Table 4.1 shows the final set of variables in the model. The results concerning the PAPD effect are summarised in Table 4.2.

Table 4.1 Model results showing factors affecting mean change in social cohesion

Variables in the model	d.f.	Significance probability
Waterbody type	3	0.017
No. of development activities in area	1	0.017
Percentage of better-offs in community	1	0.030
PAPD effect	1	0.007

Table 4.2 Mean social cohesion score for sites with / without PAPD

Site type	Mean	Std. Error	Lower Bound	Upper Bound
PAPD	4.69	0.372	3.92	5.45
Control	3.09	0.392	2.29	3.89

Mean values have been evaluated for covariate: percent of better-offs in community at 21.9%.

3.5.2 What statistics do I present?

Tables and graphs contain summaries of the data, and so their contents count as "statistics" that should be presented. Most researchers are already comfortable with presenting tables of means or proportions - even if they don't think of them as being "statistics". They seem less comfortable with presenting estimates of precision, i.e. standard errors, alongside their means or proportions. However these are essential for proper interpretation of a study's findings, and need to be reported.

Where the key variable for analysis is a quantitative response, it is important to consider whether to present a standard error (se) of a mean, or proportion, or a standard error of a difference (sed). Standard errors of differences should usually be presented when the study is a comparative one, i.e. where data are summarised and compared for different subsets – perhaps different tree species, different levels of manuring, different agro-ecological zones, or different wealth groups. The investigation is focussing on a difference, and the extent of the difference, and so the measure of precision presented should relate to that difference. If the objective of the study centres on a particular estimate of a population, e.g. the amount of forest wood collected annually by women in an area where a forest co-management scheme was in operation, then the standard error of the estimate, or a confidence interval (SSC, 2001b) should be reported.

Where formal significance testing or statistical model fitting has been carried out, the degrees of freedom[2] associated with the statistical test, and wherever possible, the exact p-value should be given, either in the text of the report or in a table, or even as an annotation on a graph. Statistics software packages now give exact p-values and so there should be very few instances where the writer can only use notation such as $p<0.05$ or $p<0.001$ or *, **, or *** to indicate different levels of significance.

We have no strong objection to including a complete analysis of variance (ANOVA) table in a detailed data analysis report; but this is likely to be convenient only in simple studies with one or two primary response variables. Where there are many different response variables, the ANOVA tables could be relegated to an appendix if their inclusion was felt to be essential.

Another area that needs attention is the role of percentages in technical reports. Researchers sometimes express the results in terms of percentage change. These are often difficult to use and to interpret. The main problem is that they may hide the importance of the change, because the changes cannot be meaningfully interpreted without knowing exactly what corresponds to 100%.

For example, suppose the usable grain weight of maize in upland areas of southern Malawi was 2900 kg/ha without *Gaucho* seed dressing and 3500 kg/ha with seed dressing for the control of whitegrubs. A report states that the results showed an increase of 20.7% with seed dressing. There are two problems here if this percentage is the only result that is given. First is the possible hiding of the scale. The second is that this is a ratio of the increase compared to the control treatment, expressed as a percentage, and it can be difficult (but not impossible) to attach a measure of precision to the estimate. It is often better to present the two actual means, and the standard error of the difference between them, in a table and give a percentage in the text, to demonstrate the value of the increase in relation to the control treatment.

3.5.3 Interpreting the analysis

Understanding what a p-value means in the context of hypothesis testing is essential for a correct interpretation of the results, and we strongly advise against the old-fashioned practice of rejecting or accepting a hypothesis based on whether the significance is $p<0.05$ or not. With statistical software now providing exact p-values, today's analyst should be more comfortable avoiding such a hard-and-fast rule and accepting the idea that a low p-value is providing some indication that the null hypothesis is not true, and the smaller it becomes, the greater is the evidence against the null hypothesis. The tricky part is to determine when a p-value is or is not "low enough". This is where statistical significance needs to be interpreted alongside estimates of means or other summary statistics, and the sample size.

This still leaves the writer with the problem of the words to write. It is certainly insufficient to stop at describing something as "not significant" or "statistically significant". Any write-up is poor if it describes a p-value of 0.061 as "not significant" without any further clarification. It is equally inappropriate to describe a result as "significant ($p<0.05$)" without further discussion. The result has to be seen in the context of (i) whether the finding matches expectations and (ii) the magnitude of the finding (e.g. a weight increase of 10 kg with an improved diet). The report should

[2] Degrees of freedom for a statistical test can be thought of as the number of independent pieces of information available to measure the background variability against which the test is made.

contain some statement of the magnitude of the quantity of interest, be it a mean, a proportion or a difference, together with an associated confidence interval.

The actual written words such as "of borderline significance", "approaching significance", "statistically significant" or even "highly significant" become less important provided that the exact p-value is given, an estimate of the quantity of interest is provided and there is some discussion about whether or not the results suggest the presence of a real effect.

We are not suggesting that a full-blown report should be littered with lots of discussion about non-significant results; only that a non-significant result may be of interest when the response variable is one of primary importance. Some discussion about why the statistical test did not achieve significance where it was expected is useful. Significance testing should be regarded as an aid in helping to assess whether the finding is a real one rather than a chance occurrence, but not to be used on its own without attention to common sense judgement or reference to scientific expectations.

Box 5 gives an example to illustrate some of the above issues.

Box 5: Statistical significance versus meaningful significance

Results of four hypothetical experiments to compare the yields of a new and standard wheat variety are presented in the table below. Units are t/ha.

Expt.	New	Standard	Difference (New - Standard)	s.e.(diff)	95% Confidence Interval	Results of Hypothesis Test
1	5.01	4.99	0.02	0.003	0.014 to 0.026	Significant
2	6.00	4.00	2.00	3.00	-4.00 to 8.00	Non-significant
3	5.01	4.99	0.02	0.03	-0.04 to 0.08	Non-significant
4	6.00	4.00	2.00	0.03	1.94 to 2.06	Significant

In both experiment 1 and 4, the null hypothesis of no difference in expected yields between new and standard varieties gives a "significant" result, yet the situations are very different. In experiment 4, the confidence interval shows that the true difference is likely to be of practical importance. In experiment 1 the difference is likely to be less than 0.026t/ha, much too small to be of interest. If the variability was the same in all experiments then this one was wasteful on resources as it gives much higher precision than was needed.

In experiments 2 and 3 the hypothesis test is not significant, but again the situations are quite different. In experiment 3 the confidence interval indicates that if there is a difference in yields it is likely to be less than 0.08t/ha – too small to bother about. On the other hand in experiment 2 the confidence interval is so wide that it is uninformative about the extent of the treatment difference. The experiment has been imprecise and therefore wasteful of the few resources that it used.

3.5.4 Reporting uncertainties

The analysis write-up should also include a discussion about uncertainties associated with the results as reported. Such uncertainties can arise in a number of ways, and recognising them gives assurance to the reader that the limitations of the analysis have been considered and the extent to which results are acceptable.

The types of uncertainties that can arise are:

- those associated with the sampling structure as measured by sampling errors;
- uncertainty caused by the inability to measure and account for all potential sources of variation;
- uncertainty due to possible un-identified errors during data collection, coding and analysis
- uncertainty that relationships and other findings established with the data remain as they were at the time when the research was done;
- uncertainty in the viability and validity of the model used in the analysis.

It is not necessary for the write-up to include a sentence, or paragraph, on each of these bullet points. We merely advocate that attention is paid to these aspects (where they occur) and that they – or any others that arise - be included as appropriate.

3.6 Reporting and presenting conclusions

The full reporting of the analysis as advocated in this guide is primarily to provide the detailed evidence that will subsequently be given more simply and concisely in papers, presentations, policy briefs and similar documents. The preparation of these simplified reports should draw on the evidence that is available in the detailed reports.

It is often appropriate for the detailed reports to include suggestions on the types of simplification that can be made. These simplifications would be such that they can be defended with the evidence from the research.

In some studies the methods of dissemination are themselves part of the research. In this case the way the results are disseminated would be fully described in the final technical report. But often the simplified reporting is a part of the wider dissemination process and discussions concerning the implications of the research. We discuss in this section the type of simplifications that may be used in reporting and presenting conclusions in summary form.

Simplifications may be of the following types:

- Results for sub-groups of the population are replaced by an overall population summary

For example, in a study where the primary objective is to investigate factors that affect poverty levels in a country, the findings may be largely similar across the country's districts. Hence results corresponding to the primary objective may be given for the country overall, with a footnote to indicate instances where there are important differences across districts with respect to some individual features.

- The tables in the detailed reports are largely replaced by simple or multiple bar charts.

These we accept. We do have some concerns at the following (over-)simplifications:

- Omission of all measures of precision.

Ignoring the measures of precision is only acceptable if the research has established the results, "beyond all reasonable doubt". If not, then the simplified reporting can later be accused of "over-selling" the results. A danger is that all results from the research may be ignored, if one or two elements have been over-sold.

- Insufficient attention to the extent to which results can be generalised.

Summary presentations often ignore or pay scant attention to the actual population to which results can be generalised. In some cases it can be difficult to separate what can be defended from the evidence of the research results from the researcher's own views about the extent to which the results are believed applicable more widely than the population to which the results actually relate. We do not wish to prevent researchers from adding their own views, but the distinction between them and the results from the research must be made clear.

- Elements of risk are sometimes ignored.

Results that are "true in general" can give rise to generalised statements that may be wrong or dangerous if users are allowed by the results presentation to overlook the underlying variability. For example, after some study it might be concluded that a country or region is food-secure in that there is enough staple food available that the population could in principle be sustained – yet in reality household food security may be seriously inadequate for large numbers of at-risk households within the population, because of inadequate distribution systems, poverty, or any number of other inequality reasons.

The detailed reporting of the project activity should either link to the simplified reports, or include sections that describe the simplified and defendable ways the results can be reported. Researchers should not assume that a report that presents the key results simply is simple to prepare!

4. Developing a system for reporting

Small projects usually involve just a few activities, so it is relatively simple for the project leader to monitor the progress of the work. Larger projects, often multi-disciplinary and multi-centre or even multi-national, can produce a multitude of reports and it is then difficult for a team-leader (or a donor, or a reviewer) to monitor the progress. Then procedures are needed to monitor progress and maintain reports in a systematic way.

For monitoring the statistical aspects, we ask only for information that is usually needed by project leaders in their managerial capacity. They require detailed information in just the same way as is necessary for a statistical reviewer, or anyone else who needs to fully understand the details of the work. A relatively small, short-term and discipline-based project probably has a simple management structure and relatively straightforward lines of communication for its results. We do not recommend over-complicating these smaller exercises.

The comments below are focussed on more complex enterprises, and apply particularly where substantial, long-term or longitudinal research is set to generate datasets that may be important to successive generations of users. It is then incumbent on those establishing such projects to ensure that potentially long-lived datasets are computerised, cleaned, checked, documented with full metadata, linked in time to further rounds of data, and put into public archives on a reasonable timescale.

As well as work at the data and datafile level, organisation should extend to the other documents without which the value of the data is limited, and tends to atrophy quickly as project staff move on, or forget details. A project could well include protocols, research instruments such as questionnaires, a wide range of data files, trip reports, many interim and final analyses and reports, workshop proceedings and journal articles. There is considerable advantage in establishing a consistent computerised system for reporting, and in including keywords. These can be used to establish links between documents, and to search documents for particular topics. This must be made easy to do, or the team members will ignore the requirements.

For example, a report on the analysis could include a link to the corresponding protocol document. This would save time in constructing a précis of the protocol information, unless the summary is needed for other purposes, such as an article for publication. These links and keywords would make the files searchable, and also possibly allow a tree-structure to be constructed, to structure the report files. The larger the project, the more this step is necessary.

This guide has concentrated primarily on the production of the detailed technical reports that underpin articles, and other information. When articles or results briefs are published, or submitted, readers and reviewers will increasingly expect a web site to contain supporting information. Publications in medical and epidemiological research set a standard that other areas, such as agriculture are likely to emulate in the future. Lim et al (2005) provide a simple, light-hearted example that shows a high standard of reporting, together with the provision for easy access to further information. This is a worthwhile read for those who may find the agricultural examples in the appendices outside their own areas of specialism.

Another area where medical research currently leads the way is in the assessment of research quality, rightly an increasing concern of donors, specifically looking at the technical fitness of reported findings to provide support to evidence-based decision-making. A developing concern is to ensure that research processes include effective strategies to deliver messages to policy-makers that can have a substantial influence. The moves that we advocate, towards greater organisation of research materials and better reporting of research results, are steps in the direction of achieving these goals.

References

Lim, M. S. C., Hellard, E. M., and Aitken, C. K. The case of the disappearing teaspoons: longitudinal cohort study of the displacement of teaspoons in an Australian research institute.

Murekezi, C., Abeyasekera, S., Mulumba, Y., Rwakatungu, A., Kubiriba, J. and Tushemereirwe, W.K. (2004). *Guidelines and Procedures for Effective Data Management (with emphasis on banana research)*. National Banana Research Programme, Kawanda Agricultural Research Institute, Kampala, Uganda, 35 pp.

SSC (2001a*). Informative Presentation of Tables, Graphs and Statistics*. Statistical Guidelines Series supporting DFID Natural Resources Projects, Statistical Services Centre, The University of Reading, UK. www.reading.ac.uk/ssc/publications/guides.html

SSC (2001b). *Confidence and Significance: Key Concepts of Inferential Statistics*. Statistical Guidelines Series supporting DFID Natural Resources Projects, Statistical Services Centre, The University of Reading, UK. www.reading.ac.uk/ssc/publications/guides.html

Stern, R.D., Coe, R., Allan, E.F., and Dale, I.C. (eds) (2004). *Good Statistical Practice for Natural Resources Research*. CAB International, Wallingford, UK. 388 pp.

Wilson, I.M. and Abeyasekera, S. (2006). Writing Research Protocols: a statistical perspective. Available at www.reading.ac.uk/ssc/publications/guides.html

Appendix A1
Reporting a standard analysis

This example is from the same study as the protocol in Appendix 2 of the "Writing Research Protocols" booklet. It is drawn from a DFID bilateral farming systems integrated pest management project in Malawi. It addresses the key components needed in the analysis write up from a statistical viewpoint.

Details of the study objectives etc. can be found in "Writing Research Protocols". The results of the study have been published in the paper *On-farm trials of technologies for the management of Striga asiatica in Blantyre / Shire Highlands* by C.S.M. Chanika, et. al. in *Integrated Crop Management Research in Malawi: Developing Technologies with Farmers* (ISBN: 0 85954 519 9).

Objective: To explore technologies for management of *Striga asiatica* in Blantyre/Shire Highlands.

Study design: The basic structure of the experiment was a split-plot design with four main plots per block. There were a total of 10 blocks from 6 farms; with three farms having a single block each, two farms having two blocks each and the remaining farm having three blocks.

The main plot treatments were:

- *Tephrosia*
- Cowpea
- No legume (i.e. no *Tephrosia* or cowpea), replicated twice within each block.

Each main plot was subdivided into two, thus making a total of 80 subplots in the experiment.

The subplot treatments were:

- No fertiliser
- A dollop of 50kg N/ha fertiliser.

Only a part of the analysis write-up is given below.

Data collected:

- Measurements made with respect to *Striga* counts included the number of *Striga* plants that emerged, had flowered or were dead (before flowering). These data were collected at 2-week intervals from a quadrat of 0.9 x 0.9 m set up within each plot.
- Maize yield data were collected as total grain weight and usable grain.
- All grain yields were adjusted for moisture content and converted to kg/ha prior to analysis.

All statistical analyses were carried out using the software package Genstat Release 4.2, Fifth edition.

← Genstat: powerful statistical software, good for analysing data from designed experiments.

Response variables and description of data analysis:

(a) Maize yields

Two yield responses were considered for analysis, i.e. total grain weight and usable grain weight.

These responses were subjected to analysis of variance (ANOVA) procedures to investigate the effect of fertiliser, legume treatment effects and their interaction, allowing for possible variation between farmers, and interaction between farmer effects and the two treatment factors.

> ← Analysis method is an appropriate one for a continuous variable.
>
> ← Sensible general linear model analysis that takes into account other known or potential sources of variation before looking at the effects of interest.

(b) Striga counts

Three responses related to *Striga* incidence were analysed, namely the maximum over six sampling occasions of (i) the total number of live plants of *Striga* emerging from 3 quadrats, (ii) the number of *Striga* plants that flowered and (iii) the number of *Striga* plants dead without flowering.

ANOVA procedures similar to those carried out for the maize yields were performed on each of these *Striga* count responses scaled up to a per square metre basis. Subsequent residual analysis gave a clear indication that the variance homogeneity assumption of the ANOVA procedure was violated. The data, therefore, were transformed to logarithms and the analysis repeated.

> ←Two points here.
>
> (a) The mention of residual analysis shows the analyst used model checking techniques to verify the appropriateness of the analysis. When it turned out not to be, the data were transformed to address the departure from model assumptions.
>
> (b) Log transformation is commonly used when data are counts.

Results:

(a) Maize yields (1997/98)

Data on maize yields were available for eight of the ten blocks. The missing blocks were from Farmer 5 who had harvested her maize crop early in the season because she was worried about theft. Only a total of 64 plots were therefore initially available for the analysis. However, a subsequent residual analysis revealed six outliers in the data. These appeared to be data recording errors, and were omitted from the analysis. The final analysis was therefore based on 58 observations.

> ← Makes clear that attention has been paid to checking data quality and taking appropriate action.

The resulting analysis showed clear evidence that the application of fertiliser has an effect on total grain weight and on usable grain weight ($p<0.001$). There was little evidence of a legume effect, but strong evidence of a legume by fertiliser interaction ($p<0.01$).

> ← Exact p-values are preferred; but when the p-value is low using $p<0.001$ or $p<0.01$ is acceptable.
>
> Note the results "show evidence". They do not yield definite conclusions. With hypothesis testing there is still an element of uncertainty.

The presence of interaction effects between legume and fertiliser for actual grain weight can best be understood by examining the corresponding two-way table of mean values in Table A1.1. These are means adjusted to take account of unequal replication of treatment factors due to missing values.

The results show the absence of any adverse effects when *Tephrosia* is undersown in a maize field. This could be a result of slow growth of *Tephrosia* in the earlier part of the season which increases after maize is harvested.

Table A1.1 Adjusted means for maize yields

	Grain weight (kg/ha)		Usable weight (kg/ha)	
Legume	No Fertiliser	Fertiliser applied	No Fertiliser	Fertiliser applied
No legume	675	1972	432	1342
Tephrosia	622	1868	424	1093
Cowpea	734	1509	433	734
SED (19 df)	136	237	130	240
P	0.718	0.100	0.997	0.024

← Large enough df for good estimate of variation; hence reliable estimate of precision.

↑ A couple of points about presenting standard errors of differences (SEDs).

1. Presenting SEDs allows the reader to do some informal significance testing. For instance a difference must be more than about twice the SED to be significant at the 5% level. Hence the author states that the table shows the absence of Tephrosia effects, because the difference is not greater than 2 x SED.
2. In studies with a comparative element present SEDs, rather than standard errors of means, because one is more interested in the difference than in estimating the means.

(b) Striga counts (1997/98)

↓ Exact p-values quoted.

Results showed little evidence of a legume effect on *Striga* emergence, flowering or deaths. Fertiliser effects were evident for emerged *Striga* plants (p=0.004) and for the number that flowered (p=0.022). There were more *Striga* plants when fertiliser was applied than when fertiliser was not. This is contrary to previous findings of some researchers who have reported fewer plants emerging with fertilised plots than unfertilised ones.

Possible explanation for these conflicting results could include the amount and the way fertiliser was applied to the maize crop. Earlier reported work had fertiliser applied either as banding or deliberately mixed with the growing medium so that the *Striga* seed was on contact with the fertiliser, in which case *Striga* seed germination would have been affected.

[The original write-up went on to discuss other reasons that might explain the difference; but we have omitted that here.]

The analysis showed no evidence of any interaction effects between legume and fertiliser treatments. Adjusted mean counts on a log-scale appear in Table A1.2, while the means retransformed to the original scale of measurement (on a per square metre basis) are shown in Table A1.3. In general plots with *Tephrosia* appeared to have lower *Striga* counts than plots with no legume or with cowpea. The exception was for *Striga* emergence without fertiliser where the mean for the plots with no legume had lowest incidence.

Table A1.2 Mean log counts of emerged live and flowered *Striga* (1997/98)

	log *Striga* emerged		**log *Striga* flowered**	
Legume	No Fertiliser	Fertiliser applied	No Fertiliser	Fertiliser applied
No legume	1.83	2.81	1.64	2.09
Tephrosia	2.05	2.18	1.15	1.49
Cowpea	2.51	2.57	1.74	1.86
SED	0.425	0.432	0.290	0.433
(df)	15	17	21	16
P	0.300	0.339	0.154	0.741

Table A1.3 Mean actual counts of emerged live and flowered *Striga* (1997/98)

	Striga emerged		**Striga flowered**	
Legume	No Fertiliser	Fertiliser applied	No Fertiliser	Fertiliser applied
No legume	6.2	16.7	5.1	8.1
Tephrosia	7.8	8.9	3.2	4.4
Cowpea	12.3	13.0	5.7	6.4

↑ Good to present the summaries of the data that were analysed first – i.e. the log transformed values - and then to convert back to units that make more sense i.e. actual counts.

Appendix A2
Reporting a non-standard method of analysis

This example is from a study published in the paper *Pod pests and yield losses in smallholder pigeonpea in Blantyre/Shire Highlands*, by J.M. Ritchie et al in *Integrated Crop Management Research in Malawi: Developing Technologies with Farmers* (ISBN: 0 85954 519 9).

The analysis is less straightforward here than in the one reported in Appendix A1 in that there is a particular type of non-normal data – a binomial proportion, i.e. a count of the number of "successes" out of a total number of occurrences of some event. The summaries are therefore different, although the analysis principles are the same as with other analyses such as ANOVA. Again a statistical model is fitted to the data, the model including all relevant terms expected to affect the key response (here a proportion) being modelled.

Objective: Within the overall goal of investigating varietal tolerance to *Fusarium udum* in pigeonpea under a maize/pigeonpea/beans intercropping system, the specific activity of concern here has the aim of exploring the incidence and severity of damage caused by key pests of pigeonpea.

Study design: Data on pigeonpea pests were collected from each of 61 farms included in a larger study of integrated pest management under a maize/piegonpea/beans intercropping system. The study covered two Extension Planning Areas (EPAs) of Blantyre/Shire Highlands, namely Chirazulu and Matapwata EPAs. Two villages were selected purposively from each EPA as areas where intercropping was practiced by predominantly resource poor farmers. Sixteen farmers were selected at random from each village from those willing and able to participate in the on-farm trials. The trials began in the 1996/97 crop season, but 3 farmers dropped out later. Hence only 61 farms were surveyed for pigeonpea pod pests in the 1997/98 season.

Each farmer had four plots to which four pigeonpea varieties were allocated at random, i.e. the local control Chilinga, the wilt resistant variety ICP 9145 and two wilt resistant Kenyan landraces ICEAP 00040 and ICEAP 00053. The procedure for collecting information on pod pests involved sampling five plants at random from each plot and sampling five pods at random from each plant, i.e. 25 pods per plot.

Data collected:

- Presence/absence of external damage on each of the 25 pods per plot
- Number of seeds (in selected pods) damaged by pod borers
- Number of seeds (in selected pods) damaged by pod-sucking bugs
- Number of seeds (in selected pods) damaged by pod flies.

Response variables and description of data analysis:

← Analysis method is appropriate for proportions.

← Model checking procedures used to check the validity of the model assumptions.

Variables analysed were the proportion of pods (out of 25) with visible external damage, and the proportions of seeds damaged by pod-borers, pod-sucking bugs and pod flies.

Since these variables are all counts of a particular "success" out of a known total denominator, they were analysed using logistic regression procedures, i.e. fitting a generalised linear model. One assumption in the analysis here is that the chance of damage remains the same (within a particular plot) for each item examined for damage. Initial analyses and statistical diagnostics used for checking the goodness of fit of the model indicated that this assumption was questionable. All analyses were therefore modified to take account of varying binomial probabilities across seeds and pods.

All analysis procedures were carried out taking account of the data structure as specified by the trial design. Farmers were recognised as having either *dambo* or upland fields and belonging to one of two EPAs. The effect of these two factors and their interaction were studied relative to the farmer to farmer variation, while pest damage across the pigeonpea varieties were studies relative to the plot-to-plot variation.

← The description makes it clear that potential sources of variation have been accounted for in the analysis.

← Different levels of variation have been recognised in the analysis.

Results:

For each of the responses analysed, EPA effects were clearly evident (P<0.001) with a higher incidence of pod/seed damage in Matapwata compared to Chiradzulu. Table A2.1 summarizes the results in terms of predicted percentages of damage to seeds or pods, as well as the odds of damage in Matapwata relative to Chiradzulu.

← Shows summary results are given in terms of percentages because the model errors follow a binomial distribution, in contrast to use of means to summarise models with normally distributed errors.

← Further, comparisons between treatment groups are made in terms of odds ratios, in contrast to the use of differences in means for normally distributed errors.

The odds of damage in Matapwata, relative to Chiradzulu are worst for pod flies, being about five times greater. Odds ratios are found to be approximately three times higher in Matapwata, compared to Chiradzulu, for pod borers and for pods showing external damage.

Table A2.1 Predicted percentages of seeds/pods showing damage across EPAs.

EPA	Pods with external damage	Damaged by pod borers	Damaged by pod-sucking bugs	Damaged by pod flies	Overall damage
Chiradzulu	10.3	3.1	14.7	0.5	18.4
Matapwata	22.7	9.2	20.5	2.5	32.2
P for EPA difference	<0.001	<0.001	<0.001	<0.001	<0.001
Odds of damage in Matapwata relative to Chiradzulu	2.54	3.16	1.50	5.15	2.11

In comparing the severity of pod pests among different varieties in the trial, clear effects were found (P<0.001) for both the overall percentage of damaged seeds and the percentage of seeds damaged by pod-sucking bugs. There was some evidence (P=0.046) of a variety effect also with respect to damage caused by pod flies, but insufficient evidence of a varietal difference for the proportion of pods with visible external damage and for pod borers. The local variety in general gave the lowest damage percentages (see Table A2.2).

← The p-values here are based on a chi-square distribution corresponding to the change in "deviance" due to inclusion of the variety effect in the model after allowing for other sources of variation.

Table A2.3 shows that the odds of damage to variety ICEAP 00040 is about twice the odds of damage to the local variety with respect to overall seed damage, seed damage due to pod-sucking buds and seed damage due to pod flies. The relative increases in odds of damage for ICEAP 00040 compared to the local variety were significant for all these three responses (P<0.001, P<0.001, P=0.007, respectively). However, the odds of damage were only about 1.5 times higher for ICEAP 00053 relative to the local variety for these three responses and about the same for pods showing external damage and damage due to pod borers. Similar results were obtained for odds of damage for ICP 9145 compared to the odds of damage for the local variety.

Table A2.2 Predicted percentages of seeds/pods showing damage across pigeonpea varieties.

Variety	Pods with external damage	Damaged by pod borers	Damaged by pod-sucking bugs	Damaged by pod flies	Overall damage
Local	15.0	5.1	11.6	0.8	17.3
ICEAP 00053	16.9	6.2	16.0	1.1	23.4
ICEAP 00040	13.9	4.8	23.8	1.5	30.4
ICP 9145	12.3	4.6	14.7	1.2	20.7
P for EPA difference	0.317	0.340	<0.001	0.046	<0.001

Table A2.3 Odds ratios for wilt resistant varieties compared to the local variety

Variety	Pods with external damage	Damaged by pod borers	Damaged by pod-sucking bugs	Damaged by pod flies	Overall damage
ICEAP 00053	1.15	1.23	1.45	1.45	1.46
ICEAP 00040	0.92	0.94	2.38	1.98	2.09
ICP 9145	0.80	0.90	1.31	1.54	1.25

Comments:

It is most appropriate for results from a logistic regression to be presented in terms of odds ratios – here the odds of damage for one variety relative to the odds of damage for another variety. The odds are defined as the Probability of damage/Probability of no damage, i.e.

Probability of damage / (1 – Probability of damage).

Two varieties can then be compared by looking at the ratio of their odds. So for example, the odds of damage to ICP 9145 relative to the local variety would be given in terms of the odds ratio as:

$[P1/ (1 - P1)] / [P2/ (1 – P2)]$

where P1 = probability of damage for variety ICP 9145 and P2 = probability of damage for the local variety.