NATURAL RESOURCES INSTITUTE

NRI Report No. 2591

Project B0108

Combining Quantitative and Qualitative Survey Work

Methodological Framework, Practical Issues, and Case Studies

March 2001

Report for DFID NRSP (SEM) Project R7033

This report is an output from a research project funded by the United Kingdom Department for International Development (DFID) for the benefit of developing countries.
The views expressed here are not necessarily those of DFID.
Natural Resources Systems Programme – Socio-Economic Methodologies

Natural Resources Institute

DFID Department for International Development

Statistical Services Centre
The University of Reading

<u>**Table of Contents**</u>

Abbreviations

ANOVA	Analysis of Variance
ASD	Actellic Super Dust
CIP	International Potato Centre
CPHP	Crop Post-Harvest Programme
DFID	Department for International Development
FRP	Forestry Research Programme
FSIPM	Farming Systems Integrated Pest Management
ICRISAT	International Centre for Research in the Semi-Arid Tropics
IGA	Income Generating Activities
IIED	International Institute for Environment and Development
IIRR	International Institute of Rural Reconstruction
LGB	Larger Grain Borer (*Prostephanus truncatus* (Horn))
NRI	Natural Resources Institute, University of Greenwich
NCAER	National Council of Applied Economic Research (India)
NFM	Natural Forest Management
NGO	Non-Governmental Organisation
NRSP	Natural Resources Systems Programme
PLA	Participatory Learning and Action
PRA	Participatory Rural Appraisal
RNR	Renewable Natural Resources
RNRRS	Renewable Natural Resources Research Strategy
RRA	Rapid Rural Appraisal
SEM	Socio-Economic Methodologies
SSC	Statistical Services Centre, University of Reading
UPWARD	User's Perspectives with Agricultural Research and Development

INTRODUCTION

This document is an output of the DFID funded research project R7033 "Methodological Framework for Integrating Qualitative and Quantitative Approaches for Socio-Economic Survey Work". The project's main objective was to improve survey methodologies used in the natural resources sector, leading in turn to improved quality of information required for decision-making at the various stages of natural resources programmes and projects. The project, which was jointly implemented by the Natural Resources Institute and the Statistical Services Centre, attempted to bridge the gap between informal / qualitative methods on the one hand and formal / quantitative ones on the other.

The document is presented in the form of a manual, allowing the user to read each of its parts and sections independently of each other. Nevertheless, cross references indicate links to other parts of the document where appropriate. Part I provides a methodological framework, which will also be available as a Best-Practice Guideline published by the Natural Resources Systems Programme of DFID. Part II, which contains six theme papers, looks into a number of practical issues encountered when combining quantitative and qualitative survey techniques. And finally, in Part III case studies are used to demonstrate how the two approaches can be integrated in survey work and experiments related to the natural resources sector. The framework document, theme papers and case studies have been produced between 1998 and 2000.

The authors of the various sections of this manual would like to thank all the different project collaborators who have contributed to this manual in one way or another. Particular thanks are due to the many farming communities where testing of the methodology took place. Last but not least, the project team gratefully acknowledges the valuable advice and encouragement provided by Louise Shaxson and Elizabeth Warham, DFID.

Ulrich Kleih and Ian Wilson
Project Leaders

PART I

A METHODOLOGICAL FRAMEWORK FOR COMBINING QUANTITATIVE AND QUALITATIVE SURVEY METHODS

Marsland N[1], Wilson I[2], Abeyasekera S[2], Kleih U[1]

[1] Natural Resources Institute, University of Greenwich
[2] Statistical Services Centre, The University of Reading

A METHODOLOGICAL FRAMEWORK FOR COMBINING QUANTITATIVE AND QUALITATIVE SURVEY METHODS

Introduction

Qualitative survey methods started to gain prominence in development projects during the 1980s, primarily in response to the drawbacks of questionnaire type surveys, which were considered time-consuming, expensive, and not suitable for providing in-depth understanding of an issue (Chambers, 1983 and 1994; Pretty *et al* 1995). This led to a polarisation in collection and analysis of information with 'traditional', quantitative techniques on the one hand, and qualitative methods, on the other[3].

The result of this polarisation of approaches and the associated shortcomings was that the users of information were often dissatisfied with the quality of data and the resulting analytical conclusions. At the same time, it was recognised that there are areas/interfaces where the two types of approach can benefit from each other, leading in turn to improved quality of information which is required for intelligent decision-making at the various stages of RNR projects and programmes.

During the second half of the 1990s, attempts were made to highlight the complementarity of the two types of approach, e.g. in relation to poverty assessments in Africa (Carvalho and White, 1997; IDS, 1994). Other work e.g. Mukherjee (1995) examined the pros and cons of each type of approach and the potential for synergy in a general development context. In the field of renewable natural resources research it was realised that whilst some research practitioners were combining methods as a matter of course whilst conducting field research, experiences were often not documented. Moreover, several avenues of potential remained untapped. It was in this context that in 1997 the Socio-Economic Methodologies component of DFID's Natural Resources Systems Programme commissioned a three year research project

[3] This paper recognises that the terms "qualitative" and "quantitative" are not without potential problems. In their study of participation and combined methods in African poverty assessment, Booth et. al. (1998) make the distinction between "contextual" and "non-contextual" methods of *data collection* and between qualitative and quantitative *types of data*. Contextual data collection methods are those which "attempt to understand poverty dimensions within the social, cultural, economic and political environment of a locality" (Op. Cit. 54). Examples given include participatory assessments, ethnographic investigation, rapid assessments and longitudinal village studies. Non-contextual types of data collection are those that seek generalisability rather than specificity. Examples of these methods include: epidemiological surveys, household and health surveys and the qualitative module of the UNDP Core Welfare Indicators Questionnaire. The distinction between contextual and non-contextual is a useful one, and the current paper does not make this distinction explicitly. In practice however, this paper's use of the terms "qualitative *method*" and "informal *method*" correspond to Booth et. al's use of the term "contextual", insofar as these terms are applied in the context of the design and data collection stages of the information cycle (see Table 1). Similarly, this paper's use of the term "quantitative *method*" and "formal *method*" corresponds to Booth et. al's use of the term "non-contextual", insofar as these terms are applied in the context of the design and data collection stages of the information cycle (see Table 1). As Booth et. al. note however, contextual and non-contextual and qualitative / quantitative are best viewed as continua. There is no dividing line between what is contextual / qualitative / informal and what is non-contextual / quantitative / formal. This paper goes beyond the scope of Booth et. al. in that it examines *analytical* combinations as well. The meaning of the use of the terms qualitative and quantitative, formal and informal in the *analytical* context become clear on inspection of Table 2 and in the section entitled Type B: Sequencing.

"Methodological framework integrating qualitative and quantitative approaches for socio-economic survey work".

This paper, which is an output of the above project, tries to offer practical guidance for field staff and project managers, allowing them to select the most appropriate data collection and analysis methods when faced with information *objectives* and *constraints* in the data collection and analysis process. The paper aims to address in general terms the basic question: "Given a set of information objectives on the one hand, and constraints such as time, money and expertise on the other, which combinations of qualitative and quantitative approaches will be optimal?" The guidelines are relevant for research involving both socio-economic data (e.g. livelihoods, wealth, gender) and natural scientific information (e.g. entomology, epidemiology). They are relevant for data collected within a "formal" setting as part of an experiment or a survey, and also in the context of participatory activities within a research or development context.

Practical Aspects of the Selection of Survey Techniques

In order to work out the most appropriate combinations of methods for a given task, it is necessary to consider both objectives and constraints.

Objectives: Investigation of a problem or phenomenon. This may be seen as the overall *goal* of data collection. Researchers need to decide:

- *What characteristics* (e.g. precision, scope of extrapolating from findings) the information ought to have.
- *For whom* is the information being collected? (e.g. project managers, policy makers, etc.).
- *Degree of participation:* In most (many) research activities there will be objectives which relate to *how* information is collected and analysed.
- *Training objectives:* There may be *training objectives* attached to the collection and analysis of information guiding the choice of methods.

Constraints. An important point to note in this context is that objectives interact with each other: having one objective will affect the extent to which other objectives can be achieved. In this sense, one objective can become a constraint to the achievement of another. This is because resources of time and money and expertise are limited. These resources will often shape the parameters of a fieldwork just as much as objectives.

Time: One of the reasons why informal methods came into greater use in the 1970s and 1980s was that practitioners and managers were fed up with the excessive time taken to conduct, analyse and disseminate sample surveys. Whilst in practice it is not possible to say *unequivocally* that participatory exercises are quicker than sample surveys - everything depends on the particular circumstances including expertise, logistics, and institutional constraints (see below for more details on these points) - it *does* appear that informal work is quicker than formal *more often than not*. Certainly, this is the - somewhat tentative - conclusion of Mukherjee (1995) who notes that "On balance...by and large...PRA method takes relatively less time".

In most project situations, time is at least as important as cost per day. For many project managers, the quicker turn-around time of informal work is a powerful argument for undertaking such work. It is important to compare like with like in terms of quality and quantity of coverage: a weak sample may be a false economy.

Cost: Received wisdom has it that sample surveys are expensive and PRA/ RRA type exercises are cheap. Gordon (1996), argues however that "there are certain "hidden" costs associated with informal surveys which should not be overlooked". Indeed, as Mukherjee (Op.Cit.) notes: "It is not easy to arrive at a relatively simple comparison of cost for the two methods [sample surveys and PRA]". There are a host of factors to be considered in this regard which can influence both actual cost and imputed cost for undertaking conventional survey or PRA-type studies. As a consequence, it is not possible to say categorically that one type or collection of methods will automatically be more expensive than another type or collection, thus cost per se cannot be reliably used in a blueprint sense to select methods. Each case needs to be taken on its merits.

Expertise: As a general statement, informal survey work requires a greater array of skills per researcher than formal work, and formal work requires a greater number of people to undertake the research process. In addition, the need for a degree of multi-disciplinarity is greater in informal work, which derives much of its internal consistency from "triangulation" - including that achieved by the debate between investigators from different disciplines. For informal work, the interviewer normally will need to be highly skilled in interview techniques, and - often - to be familiar with a range of instruments. He or she will probably also be required to analyse the data at high speed, much of it in the field itself. Characteristically, in formal work a number of different individuals will be involved in the task of research design, training of enumerators, data collection, design of data entry programmes, analysis and write up.

Trustworthiness of information. The value of information depends on its trustworthiness. Here it is argued that the trustworthiness of information will be greater if quantitative and qualitative approaches to data collection and analysis are *combined* rather than being used separately. The following four tests of trustworthiness can be discerned:

- *Internal validity or Credibility*. The key question here is: How confident can we be about the "truth" of the findings?
- *External validity or Transferability*: Can we apply these findings to other contexts or with other groups of people?
- *Reliability or Dependability*: Would the findings be repeated if the inquiry were replicated with the same or similar subjects in the same or similar context?
- *Objectivity or Confirmability*: How can we be certain that the findings have been determined by the subjects and context of the inquiry, rather than the biases, motivations and perspectives of the investigators?

Internal and external validity, reliability and objectivity are the terms used in conventional scientific research. Credibility, transferability, dependability and confirmability are the terms put forward by Pretty (1993), after Lincoln and Guba

(1985) to describe the equivalent criteria implicitly and routinely used in much participatory field research.

Obviously, the size of the target population has a bearing on the importance of these criteria for a particular study. For example, external validity plays less of a role if the target population is small (e.g. a small number of villages in the case of an NGO led development project). On the other hand, research projects covering entire regions or countries depend on results representative of these areas. Overall, formal work has probably most to gain from informal in the area of credibility and objectivity, whereas informal work (if it is to be generalised) can borrow from formal methods to improve external validity.

Types of Combinations. *Merging* is one way of combining qualitative and quantitative approaches. It consists of swapping tools and attitudes from one tradition to the other. In addition to *merging*, there are two other types of combining: *sequencing* and *concurrent* use of tools and attitudes. If they are to lead to integrated conclusions, sequenced and concurrent combinations should be followed by a *synthesis* of the information collected. Box 1 illustrates the differences between the different types of combinations with some examples.

Within a particular RNR research or development project dealing with the sustainability of livelihoods, any mixture of these types of combination can be used. Of them all, *sequencing*, has probably been the most widely practised in the past. Whilst aspects of types A, B and C have undoubtedly been used in the field for some time, it is only relatively recently that examples have been documented and disseminated widely (see e.g. PLA Notes 28, and Carvalho and White, 1997). The latter paper stresses the importance of synthesising of information obtained through combinations of survey techniques.

Box 1: Types of qualitative and quantitative combinations that may be used in sample surveys and experiments

Type A: Swapping tools and attitudes: " Merging"

- Thinking about sampling in designing enquiry based on qualitative methods.
- Coding responses to open-ended questions from qualitative enquiries.
- Using statistical techniques to analyse unbalanced data sets and binary, categorical and ranked data sets, arising from participatory enquiry.
  - creating frequency tables from coded data.
  - modelling binary and categorical data generated from ranking and scoring exercises.
- Using mapping to generate village sampling frames for: questionnaire surveys; type 2 or type 3 on-farm trials.
- Using attitudes from participatory methods, e.g. to reduce the non-sampling error in questionnaire surveys or farmer-researcher misunderstandings in on-farm trials.

Type B: "Sequencing"

- Using participatory techniques in exploratory studies to set up hypotheses, which can then be tested through questionnaire based sample surveys, or via on-farm trials.
- Choosing a random sample and conducting a short questionnaire survey to gain information on key variables which are then investigated in-depth by participatory enquiry.

Type C: Concurrent use of tools and methods from the different traditions: "Mixed Suite"

Concurrent use of:

- Survey of statistically selected sample members, using pre-coded questionnaires to determine target population characteristics of a qualitative (e.g. opinions on a new technology) or quantitative (e.g. crop production) nature.
- Setting up scientific experiments (on-station or type 1 trials) to study the effects of specific interventions in a controlled environment (e.g. on-station or "contract" research).
- Using aerial photographs, GIS.

along with:

- Participatory enquiry for attitudes, beliefs and perceptions of the target population.
- Type 3 trials.

Note:
Type 1 on-farm trials are those designed and managed by researchers. Type 2 trials are designed by researchers but managed by farmers. Type 3 trials are designed and managed by farmers and monitored by researchers. (Coe and Franzel: 1997).
Source: Marsland et al (1998)

Combinations, objectives, trustworthiness and researcher-researched relationship. Figure1 shows how combinations of survey instruments form part of a continuum in relation to the objectives of a given research project. The different types of combinations need to be seen in relation to the different stages of the research process where they can be applied. Although this paper focuses on survey techniques, it is important not to lose sight of the other stages leading to a research output.

Table 1 presents types of formal and informal combinations at the various stages of the research cycle, and their relationship to aspects of trustworthiness. The latter will be enhanced as a result of "*examining, explaining, confirming, refuting, and / or enriching* information from one approach with that from another" (Carvalho and White, 1997). Table 2 demonstrates the link between research objectives and survey techniques in more detail highlighting at the same time the researcher – researched relationship.

Figure 1: Continuum of Objectives and Combinations of Instruments

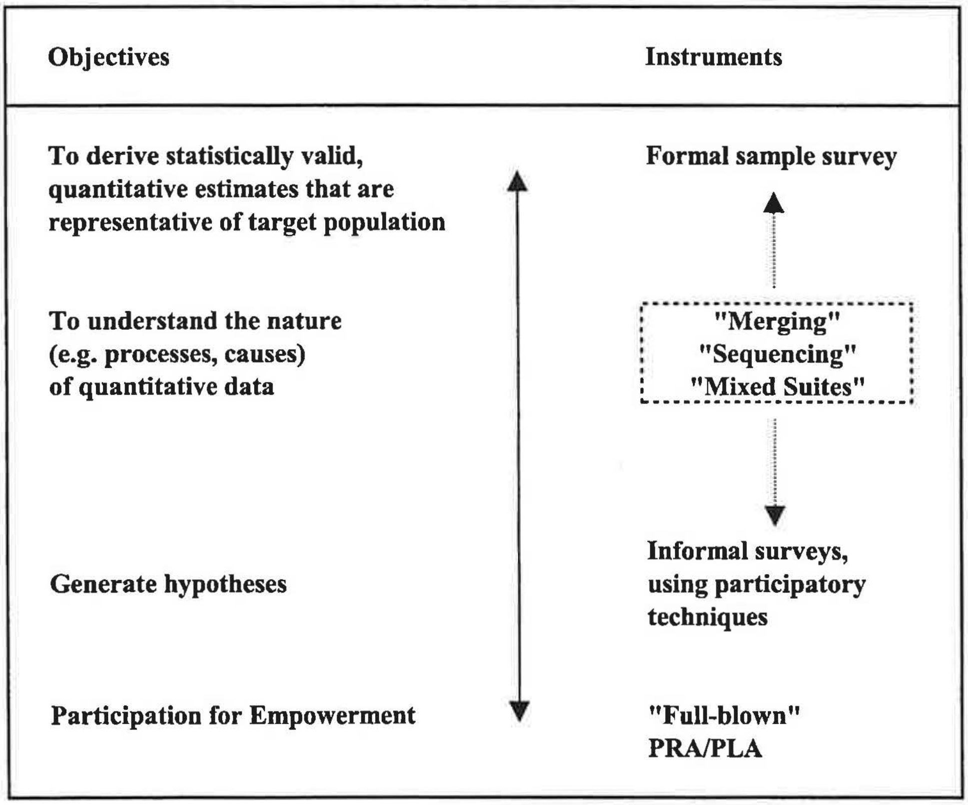


Table 1: Types of informal / formal combinations and their relationship to aspects of trustworthiness

Stage in information cycle	Type of combination	Explanation/Example	Function: Relationship to elements of trustworthiness.
Design	Merging	• Formal sampling procedures for informal work	• Reduced sampling error: better external validity for informal work
		• Informal attitudes for formal work	• Reduced non-sampling error: better internal validity for formal work.
		• Use of social mapping for formal work	• Reduced time and cost for household listing and sampling.
	Concurrent	• Correct use of different instruments for different variables within the same survey/ experiment	• Better internal validity for "qualitative" variables - belief, motivations etc. alongside better external validity for quantitative variables - rates, proportions etc. • "Enriching": The outputs of different informal and formal instruments adding value to each other by explaining different aspects of an issue
Data collection	Sequential	• Analysis of informal outputs feeding into the design of formal instruments i.e. using informal studies to "map out" key issues and approaches to be explored further in formal work e.g. using informal work to generate hypotheses to be tested in formal work.	• "Enriching"
Analysis	Sequential	• Analysis of formal outputs with informal approaches. e.g. testing null hypotheses; investigating unexpected outcomes.	• "Refuting": Where one set of methods disproves a hypothesis generated by another set of methods. • "Confirming": Where one set of methods confirms a hypothesis generated by another set of methods • "Explaining": Where one set of methods sheds light on unexpected findings derived from another set of methods.
	Merging	• Applying statistics to categorical and unbalanced data sets.	• Improved credibility of analytical conclusions from informal work.
		• Coding responses from informal work	• Enhances possibilities for aggregation, thus facilitating generalisation. • Enhances possibilities for stratification of sample for subsequent sample survey
Synthesis	Merging	• Blending the analytical outputs from informal and formal work into one set of policy recommendations.	• Higher quality policy recommendations

Table 2: Information Objectives, Approaches to Data Collection and Analysis and Researcher - Researched Relationship.

Information objectives	Type(s) of instrument(s)	Researcher - researched relationship
1. To derive quantitative estimates (number, rate or proportion) of parameters representative of project, regional or national parameters; data to be replicable and verifiable. When quantitative estimates are needed for "credibility".	Formal surveys • Random sampling • Some use of secondary data	Researchers design, execute analyse, present. Researched are passive.
2. To derive quantitative estimates reflecting the area under consideration, willing to accept lower levels of precision because of resource limits; make maximum use of prior knowledge with purposive sampling.	Formal surveys • Purposive sampling • Greater use of secondary data	As above
3. To obtain quantitative data with an understanding of processes causes (diagnosis); data could be used as benchmark data to assess trends, therefore method repeatable with high degree of confidence.	"Merging" or "mixed suite" • Stratification of sample • Use of ranking and scoring and statistics to analyse data • Use of questionnaires • Use of secondary data and grey literature is important	Researchers interact with researched: there is dialogue; semi-structured formats.
4. To understand the nature (causes, trends, add-ons) of quantitative data already available, either national, regional or project formal surveys.	As above	As above
5. When qualitative data (description and analysis of situations, events, people, interactions and observed behaviours) are appropriate to make a decision; when researching characteristics, cultural patterns, motivations and attitudes.	As above - less emphasis necessary on quantification.	As above, but greater use of visualisation techniques; longer time period per data collection event; more open-ended structure.
6. When very little is known about a project area or topic, or wish to move to the next stage of an investment or other action.	As above	As above
7. When the intention is to introduce a project with a high degree of participation and the local people must be involved at the outset and at all subsequent phases. Quantification still possible.	No necessary requirement for sampling; methodology highly location specific and open-ended.	**From:** Researchers working as equal partners with researched; **To:** researchers acting only to facilitate - translating the wishes of the researched.

Derived from Longhurst (1992)

Types of Combinations

Type A: Swapping tools and attitudes: Merging

Informal contributions to formal approaches

(i) Informalising and contextualising interviews in surveys and experiments

Including semi-structured interviewing in a structured questionnaire format can improve the quality of data generated due to increased flexibility and openness, allowing the questionnaire as a whole to adapt better to particular local environments (Ziche (1990) quoted in Mukherjee (1995)). This adaptation ranges from contextualising of questionnaires through use of appropriate locally specific vocabulary, to being better able to deal with certain types of information within a questionnaire format. To some extent qualitative response is routinely incorporated in many questionnaires, with the inclusion of open-ended questions. The addition of a checklist of points or hints for probing on particular issues takes this process one step further and introduces a greater degree of interaction on the part of the interviewee. Summarising any substantial number of such responses requires a careful coding exercise.

(ii) Using maps to create village sampling frames

Once villages in a region are chosen for a study, based on (say) agro-ecological conditions, social mapping can be used to generate a list of households, together with their physical locations within a village. This can then be used as a sampling frame in sample selection. In a 1993 study, India's National Council of Applied Economic Research (NCAER) found that social mapping compared favourably with standard household listings often employed in sample surveys. Box 2 provides an illustration drawn from Marsland et al (1999).

Box 2 : Use of village mapping to generate sample frame

The sample design for project households in a study on co-management of forest products in Malawi was based on a single-stage cluster sample within each of the stratified substrata, with villages as clusters. Project villages were stratified first by association with particular co-management blocks in each reserve and then by proximity to the reserve (i.e. near and far). Because of time and resource constraints, a systematic sampling method was used to select households within the selected villages. The sample frame was generated through a process of village mapping, with villagers marking out the number and location of each dwelling unit in the village, together with the name and sex of the household head. All the names and numbers were recorded by the RRA field teams and a systematic sample was taken. This process was found to be useful for three main reasons. First, it served as an initial ice-breaker, allowing the RRA team to interact with members of the village. Second, and more importantly perhaps, it provided a very rapid and accurate way of generating a comprehensive sampling frame for selected villages. Characteristically, the whole process would take between 1 and 2 hours for Chimaliro Extension Planning Area (EPA) and 1 to 3 hours for Liwonde EPA. The process was slightly longer in Liwonde than in Chimaliro owing to the larger village sizes in Liwonde. Finally, the existence of an accurate village map helped greatly in planning the actual enumeration and dividing tasks between enumerators.

(iii) Using qualitative understanding to inform classification procedures

Cluster analysis is a technique commonly applied to quantitative data by statisticians. Based on a survey, it entails agglomerating the respondents into groups on the basis of "similarity" with respect to responses to some set of survey questions. The starting point is a choice of "cluster seeds" to which others are then joined in the process of cluster formation. If these seed respondents – core members of groups – have been studied intensively and are well understood through qualitative work, clusters formed on the basis of similarity to the seeds will have an understandable character. Ideally, seed respondents are prototypical of what could become effective strata or recommendation domains.

Formal contributions to informal approaches

In some instances, researchers have found it necessary to incorporate more structure into a previously unstructured exercise. For example, one general conclusion of the IIRR/CIP-funded review of Participatory Monitoring and Evaluation (UPWARD, 1997) was that "with the emphasis on participation and learning processes, much of the PM&E experiences started off with using qualitative and semi-structured methodologies. However, there is an emerging recognition of the need to build into current participatory methodologies some of the quantitative tools to provide for better triangulation of information and greater acceptability of the results when endorsed as inputs to policy. This includes paying greater attention to establishing baseline data to more systematically monitor progress and facilitate ante and post evaluation procedures."

(i) Sampling and Stratification

Pretty (1993) argues for the trustworthiness of participatory inquiry, citing the four characteristics of credibility, transferability, dependability and confirmability. It is interesting and important to note, however, that the case for transferability (equivalent to external validity in structured research) appears to be considerably weaker than the one he makes for the other characteristics (Op. Cit., 27-28). It is perhaps in the question of transferability that the most obvious "Achilles heel" of informal research lies, at least insofar as its practitioners try to generalise their findings in much the same way as sample surveys. Effective and statistically based methods of sampling are needed if the domain of validity of research conclusions is to be extended.

Many issues have to be considered in the sample selection process if results are to be generalised to a wider population. Some important issues are (a) a clear identification of the recommendation domain; (b) the use of secondary data and relevant grey literature in assessing the availability of a suitable sampling frame; (c) where a sampling frame is unavailable, evaluating the feasibility of adopting a hierarchical sampling procedure so that sampling frames can be built up for just selected units in the hierarchy; (d) clearly defining the sampling units most appropriate for study objectives; (e) methods to be used in sample selection, in particular, including an element of randomness in the procedure; (f) being open to the possibility of post-stratification at the data analysis stage; (g) sample size considerations. Wilson (2000) gives more detailed consideration to these elements.

(ii) Applying statistical analysis to unbalanced, binary, categorical and ranked data sets

During the 1990s, practitioners of informal surveys and PRA type work in developing countries have started to recognised the potential for applying modern statistical methods to unconventional data sets. Martin and Sherington (1996) and Abeyasekera (2000) amongst others have outlined some of the ways in which statistical techniques can play a useful role for such data.

One starting point is *coding open-ended questions from informal work.* This is common in questionnaire work. What is less common is coding of information collected informally. Certain types of information collected during informal work can be coded readily, and others with rather more careful thought.

Abeyasekera and Lawson-McDowall (2000) describe how qualitative information from farmer activity diaries collected as part of the Farming Systems Integrated Pest Management Project in Malawi (FSIPM) was computerised using the spreadsheet programme Excel and analysed using a statistics package, SPSS. Studies that are relatively large may justify the use of specialist software packages (e.g. NUD-IST) for computerising this type of qualitative data, although these may be time-consuming and difficult to use.

ANOVA: The principal method for the statistical analysis of data from on-farm participatory trials is the analysis of variance (ANOVA). The power of the method lies in its ability to "disentangle", "correct", or in a loose sense, "explain" the effects of one or more factors (e.g. new technologies) on response variables such as results from participatory scoring exercises (Abeyasekera, Op.Cit.). When the data structure is "balanced" (equivalent to equal numbers of observations in cells of 2-way tables concerning factors of interest), the ANOVA is relatively straightforward and is quite well-known. Although "balance" is rare in participatory on-farm trials, the ANOVA technique can allow the simultaneous study of several factors (qualitative, as well as quantitative), and the study of interactions between them. The procedures are easily available in many statistics packages, but their use is generally less well known and they appear not to have been widely applied to on-farm trials. Simple treatment means, which suffice for balanced data, can be misleading in the analysis of unbalanced designs. Martin and Sherington (1996) illustrate this with data from the project "Management of *Imperata cylindrica* in Smallholder Farming Systems". They compare (i) simple means of % *imperata* cover for different groups, and (ii) adjusted means from an unbalanced ANOVA. The authors were able to separate the effects of the farming system on *imperata* cover from those of herbicide use, which simple means could not do.

Generalised linear models for binary data: Martin and Sherington (1996) also show how categorising farmers' preference rankings of tree species as "good" or "poor", allows the resulting binary data to be analysed via a generalised linear modelling approach to determine factors which affect their preference. In particular, the dependence of preference ranking on ethnic groups is demonstrated.

Multi-level Models: A recent set of statistical developments extends the idea of general linear models to multi-level models which explicitly acknowledge and model hierarchical information, as found for instance where some data are at community level, some at household and some at individual level. The power of the multi-level modelling method lies in "separating out, "accounting for", or loosely "explaining" the effects of several factors at different hierarchy levels. These up-to-date models do not as yet appear to have been applied to data collected using informal methods in developing countries, but there is clear scope to improve the quality of data analysis by doing so. Pending further development, the above modelling can be quite technical and is likely to require the use of a professional statistician. With time and funds, however, it should be possible to make modelling more user friendly to the NR research practitioner.

Qualitative Residuals: A general idea which runs through regression and ANOVA modelling as well as generalised and multi-level modelling is that of the "residual", the difference between the observed result and that suggested by the model fitted. There is a residual for every observation after a model has been fitted, and the set of residuals corresponds to what is "left over" or "unexplained" after "correcting for" known influencing factors. If a large body of qualitative data is collected, say from a substantial number of separate informants, it is time-consuming and labour-intensive to summarise it. The analogue of quantitative residual analysis is first to account for common features in the qualitative data in a systematic way such as the above, so as then to focus attention specifically on explaining the more individual characteristics.

Ranking and Scoring: Ranking and scoring data arise from activities where precise numerical measurement is inappropriate, including a range of qualitative work, some of it participatory. Ranking entails an ordering e.g. between a set of crop varieties in terms of cooking characteristics. For the same task, scoring would entail assessing each variety separately on a fixed scale, say a four-point scale with values 1, 2, 3, and 4. Simple scoring and ranking data can be analysed very straightforwardly (see Box 3 and Box 4), but where the study has more structure, statistical methods can be used to correct for respondent grouping factors, e.g. respondent's ethnic group and gender. In a substantial number of cases, scoring data can be treated by relatively standard statistical methods, so the results can be modelled and simultaneously corrected for a range of "explanatory factors", even when these occur in an unbalanced fashion (Abeyasekera, Op.Cit.).

Bayesian Statistics: Bayesian statistics is based on the notion of subjective probability or degree of belief. Briefly, the Bayesian paradigm consists of modelling beliefs *before* observing data, by *prior* probabilities, and using Bayes' theorem to combine information from observations with the prior distribution to obtain a *posterior* distribution. Thus, an inference about an unknown is a blend of observed data and subjective degrees of belief. There has been much recent research on the so-called *elicitation* process; this is the process of obtaining the prior probabilities. One area where Bayesian ideas show some promise is in the analysis of causal diagrams (Burn, 2000). These are a popular tool in qualitative enquiries, and recent work by Galpin, Dorward, and Shepherd (2000) has generated "scored" causal diagrams, where participants generate scores for the importance of cause-effect pathways within the diagrams. One set of such data constitutes a descriptive profile of a problem analysis. The question has arisen of combining or comparing several such diagrams, independently elicited. The Bayesian approach to statistical modelling involves a similar type of elicitation, and recent developments in Bayesian networking methods show promise as a toolkit for comparing and combining structured sets of uncertain information. Burn (Op.Cit.) provides more details.

(iii) Procedural aspects of applying statistical analysis to qualitative data sets

A compromise needs to be struck so that informal data can be analysed by using statistical techniques. Some of the flexibility inherent to RRA/PRA exercises needs to be given up in favour of a minimum of rigor, making the data suitable for cross-site analysis. Nevertheless, if well blended into the exercise, this can be done without seriously restricting participation.

The following are a number of aspects to respect during survey design and data collection when considering the application of statistical analysis to qualitative data sets, in particular if the research is to lead to generalisable results:

- The study group needs to be adequately large and representative of the target population
- There has to be an element of randomness in the selection of the study units
- The format of the data collection tool should remain the same throughout the survey (e.g. use of the same format of matrix throughout the exercise; use of a uniform scoring system)
- Well-defined consistent recording of information so that e.g. results from individual PRA practitioners can be coded in a coherent way and put together for analysis
- Clear and complete recording of meta-data, i.e. details of where and how the information was collected, so that information summaries can be based on a clear-cut rationale, and have proper support for any claim to generalisability.

Box 3: Example of a first stage analysis of scored data

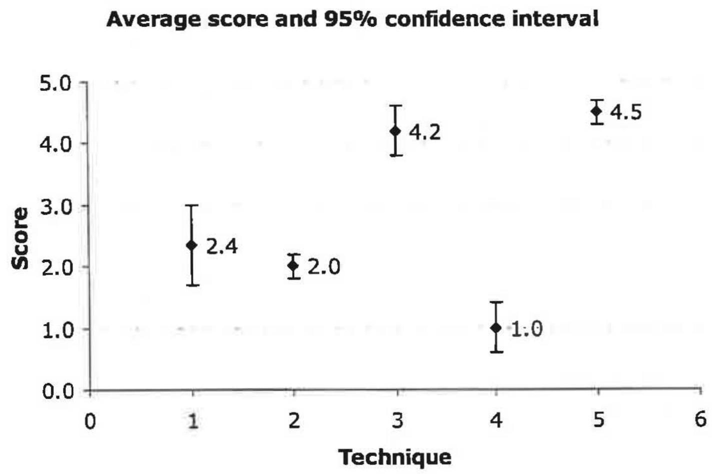
Data: Five techniques were compared by 20 farmer groups as part of a matrix scoring exercise, giving results below.

Group	*Tech 1*	*Tech 2*	*Tech 3*	*Tech 4*	*Tech 5*
1	2	1	5	0	4
2	5	2	3	2	5
3	1	3	5	0	5
4	1	2	5	1	5
5	1	2	5	1	5
6	2	2	3	0	5
7	5	2	3	1	5
8	4	3	4	0	5
9	3	2	3	0	4
10	2	1	3	1	4
11	1	2	4	2	4
12	3	2	4	2	4
13	1	2	5	2	5
14	2	2	5	0	5
15	1	2	5	1	5
16	4	2	4	2	4
17	4	2	5	1	4
18	1	2	3	1	4
19	2	2	5	1	4
20	2	2	5	2	4

Statistical analysis of the scores for each technique:

Technique	N	Mean	95% confidence interval for the mean
1	20	2.4	(1.7, 3.0)
2	20	2.0	(1.8, 2.2)
3	20	4.2	(3.8, 4.6)
4	20	1.0	(0.6, 1.4)
5	20	4.5	(4.3, 4.7)

Graphical display of results:



Source: Barahona (2000)

Box 4: Example of analysing ranked data from a study in Tanzania

The Larger Grain Borer (LGB), *Prostephanus truncatus* (Horn), was first reported in Africa in 1981. The beetle, a severe pest of farm-stored maize and dried cassava was initially a major problem to farmers in western Tanzania.

The two principal objectives of the study were to:

- To assess the role played by *P. truncatus* in determining changes in production, storage, and marketing of the maize and cassava crop during the period between the time of the establishment of the beetle and today.

- To assess the factors determining the role played by *P. truncatus* in these stages of the maize and cassava commodity system, in particular the impact of the insecticide treatment.

In order to achieve these objectives, a combination of sample survey and rapid rural appraisal (RRA) techniques was required.

In pursuing one component of the above objectives, attempts were made to apply statistics to the ranking data derived from the RRA exercises. Chi square tests and variants thereof were used to test for changes in rankings of the importance of *P.Truncatus* when farmers compared the situation at the time of establishment of the pest with the situation at the time of the survey (i.e. 1998). As an example, in one exercise farmer groups were asked to rank the importance of the pest in comparison to all other storage problems (a) at the time of establishment and (b) for the present day. The ranks were then compared and analysed using McNemar's test. The following table illustrates how ranking data for the past and present can be summarised.

		Present	
		Rank = 1	Rank > 1
Past	Rank = 1	24	13
	Rank > 1	2	4

The cells representing no change give no information about how the ranking of LGB has changed over the years. Only the bottom left and top right cells give information about change. McNemar's test (sign test in this case) can be used to test the null hypothesis of no change in attitude. The test gives a p-value of 0.0045, which indicates strong evidence for rejecting the null hypothesis. It is clear from the table that there was a significant increase in the ranking, giving significant evidence for a reduction in the role of LGB as a storage problem.

Source: Marsland, Golob and Abeyasekera (1999)

Type B: Sequencing

Informal before Formal in different stages of the research / development process

(i) Formulating and testing hypotheses

Survey work: The use of informal tools before structured questionnaires is an accepted and common practice. The reasons for conducting an open-ended enquiry before a more closed but geographically broader one are well known. Open-ended diagnostic studies help in the process of formulation of hypotheses, which can then be tested rigorously by structured tools such as a questionnaire administered to individuals selected through an unbiased sampling procedure. As noted by McCracken, Pretty and Conway (1987), the primary role of the informal study is to "define and refine hypotheses that are then tested, either formally or informally".

Interestingly, the practice of undertaking informal studies before formal ones has been standard practice in mainstream market research for at least 30 years. The reasons given for this by the Association of British Market Research Companies (ABMRC) are very relevant to renewable natural resources research and development:
"Prior to any large-scale quantitative study particularly in a relatively unknown market, it is strongly recommended that a qualitative phase of research is initially conducted, the main purpose being to understand the vocabulary and language used by customers as well as understanding their motivations and attitudes towards given services, products and usage occasions. The findings of the qualitative research provide invaluable input to the quantitative stage in terms of the line and tone of questioning, and of course the overall structure and content of the quantitative phase" (ABMRC,1989:26)

Experiments: Before formal scientific experiments are designed and implemented, the use of informal studies performs very much the same function in experimental work as it does in survey work. Prior to a programme of on-farm experimentation, it is necessary to get an understanding of local farmers' knowledge, perceptions, beliefs, and practices, and to scope the range of circumstances which may fall in the recommendation domains of conclusions from formal studies. Conroy and Rangnekar (in press) describe the use of participatory techniques (e.g. 'herd history', problem tree analysis), as part of the identification and research issues prior to undertaking on-farm goat feeding trials in Semi-arid India.

(ii) Rejecting null-hypotheses

Casley and Kumar (1987) and Casley and Lury (1982) have commented on the use of informal surveys as diagnostic studies (i.e. to build up hypotheses) and also as case studies to reject null hypotheses in survey work by producing counter-examples. Thus Casley and Kumar note that informal surveys "can be used to disprove a null hypothesis (for example, that a certain constraint does not exist) or to indicate that an assumption of the project plan is not holding true in the cases studied". Casley and Lury point out that "one advantage of the case study method [is that] one may not be able to generalise from it, but one may be able to reject existing generalisations".

(iii) Building up rapport

Formal work, such as on-farm experimentation, requires the development of farmer-researcher understanding and a degree of consensus on the programme of work. This preparatory phase is then likely to provide a pool of potential collaborators who can be "sampled". Participatory activities conducted prior to formal work, can, irrespective of any other benefits, generate rapport and a degree of confidence between farmers and researchers.

It is inevitable that the selection of participants for a long-term activity will involve compromises. For example, the selection of farmer participants in an on-farm study will depend on the willingness and capability of the candidate farmers. A note of caution is needed, because this may affect the "population" to which conclusions can be claimed to generalise; if the non-compliant are likely to be more resistant to adopting new processes, the effects of a research intervention may be over-estimated. It may be valuable to carry forward informal estimates of the participation rate and of the type and importance of differences between those willing and those not willing to be involved.

Formal before Informal in different stages of the research / development process.

Survey work: Whilst the use of informal studies before formal work is the most common form of sequencing, in some cases, researchers and practitioners may conduct a questionnaire survey before a more in-depth informal study. In such cases the questionnaire survey acts as a kind of baseline, the results indicating areas requiring further probing and analysis through informal methods. This type of sequencing will work best in situations where most of the key issues are known or strongly suspected, but further information is needed on causes e.g. in the context of a project or programme that has been going for some time and for which a lot of information has already been collected through an M&E system.

The information from the formal questionnaire both poses the issues which should be addressed in greater depth in follow-up, and provides a basis for selecting individuals whose further participation is solicited. Respondents may be post-stratified or clustered into groups on the basis of information from the questionnaire. This may be deliberately done:-

- so that a particular grouping comprises those targeted for follow-up,
- so that the group followed up are broadly representative of all the clusters found in the population, and the follow-up study is made "representative".
- so that differences amongst the clusters can be explored - particularly relevant if the cluster definitions lay the foundations of recommendation domains.

Formal and informal methods used in sequence throughout the research and development project cycle

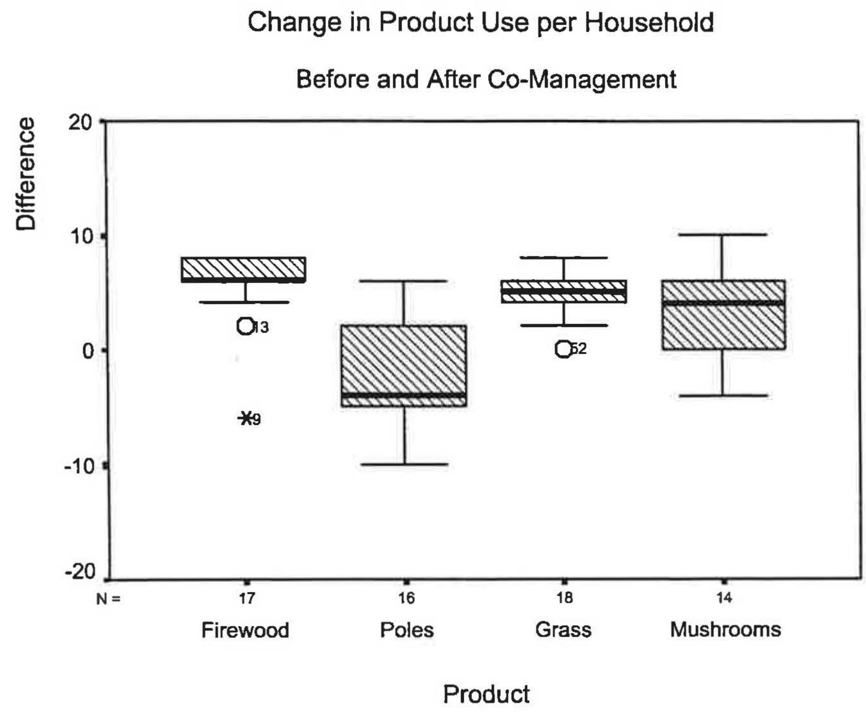
Through defining and refining hypotheses, correcting misapprehensions, providing depth and causal linkages, the informal survey is used in series with formal methods throughout the project cycle from needs assessment to ex-post evaluation. There are several examples of formal and informal methods being used in concert in both research and development contexts.

For example, in relation to research, Hagmann et al (1995) have commented on the benefits of a symbiotic relationship between participatory on farm research (type 2) and formal on-station research in the context of the Conservation Tillage Project in Zimbabwe. In the project, the qualitative results from the on -farm research fed into the on-station work, were quantified, modified and then fed back into the on-farm research and so on. The authors report that the process of integration of formal research into participatory technology development enabled "..both farmers and researchers to develop technologies and had the benefits in terms of data (researchers and policy makers) and a deeper understanding of processes (farmers and researchers)" (Op. Cit., 13).

Commenting mainly in relation to development projects, McCracken et. al. note that "The advent of RRA has .... greatly enriched the availability of methods of analysis for rural development. Techniques can be chosen on the basis of the problem, the local situation and the resources to hand. In particular, different techniques, both formal and informal, can be blended to produce a project cycle..." (Op. Cit., 76).

Box 5: Change in consumption of forest products

As part of the DFID Forestry Research Programme project *"Sustainable Management of Miombo Woodland by Local Communities in Malawi"*, an RRA exercise involved asking farmer groups (on a 1-10 scale) to estimate the magnitude of the change in forest product use before and after co-management. An initial description of the data (see box-plots below) indicated that the majorities of groups felt that access to all products apart from poles had improved under co-management. The box-plots provide a useful summary since (a) the middle line indicates the median; (b) end-points give the 1st and 3rd quartiles; (c) lines on either side of the box extend to maximum and minimum values; and (d) outliers are highlighted. Further analysis via analysis of variance confirmed that the differences were significant and that this was largely due to the mean for poles differing from all the other means.



Source: Marsland, Henderson, and Burn (1999)

Type C: Concurrent uses of tools

Survey work: NCAER (1993) found several benefits in using informal and formal techniques together in its evaluation of the "India's National Programme on Improved Chullah". The NCAER experience concerned a geographically broadly spread sample in which a questionnaire was used to collect quantitative or quantifiable information on a limited number of variables. Other mainly qualitative data was collected through RRA / PRA methods from a smaller sample, spread across fewer villages picked from all regions. The questionnaire results provided "representativeness", whilst the RRA / PRA work provided "contextual linkages for explaining behavioural patterns,...[and]....additional in-depth qualitative data which could be helpful during analysis and report writing stages" (NCAER, Op. cit.) Overall, "The blending of the two approaches can lead to a more reliable data base"; in other words there was a definite "trustworthiness payoff". (See Box 6).

As reported by Abbott and Guijt (1997), Schoonmaker-Freudenberger (1996) makes precisely this point, arguing that we should not attempt to extrapolate from PRAs, but instead use the findings to stimulate, "a more accurate debate about a policy issue by identifying the diversity of local conditions. By combining PRA with questionnaires or remote sensing techniques which capture broader *spatial* information, one can derive 'an attractive combination of range and depth of information'". Abbott and Guijt (Op.Cit.). Martin and Quan (2000) demonstrate how Geographical Information Systems (GIS) and PRA can draw from each other.

Table 3 shows the concurrent use of both PRA exercises and formal household questionnaires, while Box 7 shows a similar exercise used for purposes of triangulation.

Box 6: Combinations of broad, formal survey and narrow, in-depth study.

It often makes sense to think of a combination of a broad shallow study which provides good "representativeness" and one or more deep narrow studies which provide the depth referred to above. This combination may be thought of as providing a table or platform supporting the research conclusions. When such a combination of studies is planned, it is of course desirable that the sampling structure be planned so that effective merging of conclusions can follow. This implies that the in-depth studies are planned with special attention to how their selection relates to the broad shallow study. For more information on this, refer to Wilson (2000).

Experimental work: A further type of concurrent combination is that which involves detailed scientific measurements on the one side and informal investigations of perceptions, beliefs and attitudes on the other. An example of this is the qualitative and quantitative sorghum loss work conducted in India by NRI. This seeks to compare detailed laboratory-based analysis of mycotoxins, pest damage of stored sorghum with farmers' perceptions of the importance of losses (Hodges, NRI , *pers. comm.*).

Table 3: Concurrent use of research tools: LGB study

Thematic area	Research approach
1. Changes in role of crop production in household food security strategies comparing 1985 with 1998.	RRA (Groups of men and women – some single gender groups - ranking strategies for 1985 and 1998)
2. Changes in farmers' perceptions of the importance of maize and cassava, comparing 1985 with 1998.	RRA (Groups of men and women – some single gender groups - ranking both crops against all other crops for 1985 and 1998)
3. Influence of P.truncatus on production, storage and marketing outcomes • Production levels • Role of P.truncatus in maize and cassava harvests • Role of P.truncatus in the choice of maize and cassava varieties • Role of P.truncatus in the duration of storage and volume of sales at farm level	Household sample questionnaire
4. Is P.truncatus still regarded as a problem? • P.truncatus in the context of major agricultural problems • P.truncatus in the context of other storage problems	RRA (Groups of men and women – some single gender groups - ranking strategies for 1985 and 1998)
5. Coping strategies for P.truncatus • Actellic Super Dust perceptions • Storage operations and structures	Household sample questionnaire

Source: Marsland, Golob, and Abeyasekera (1999).

Box 7: Concurrent use of tools for triangulation

Both formal questionnaire surveys and informal RRA exercises were carried out concurrently in 1998/99 as part of the DFID Forestry Research Programme project "*Sustainable Management of Miombo Woodland by Local Communities in Malawi*". Regarding the importance of the forest products, Table below shows how the results of the RRA confirmed the results of the questionnaire survey.

Comparison of responses from questionnaire survey with RRA exercise: Importance of different forest products

Product	Questionnaire survey		RRA exercise
	%	Rank	Rank
Firewood	94	1	1
Grass/thatch	84	2	2
Mushroom	70	3	3
Poles/timber	58	4	4
Rope fibres	28	5	5
Medicine/herbals	24	6	6
Fruits	22	7	7

Source: Marsland, Henderson and Burn (1999)

Conclusions

There are a variety of ways in which qualitative and quantitative methods may be combined to improve the trustworthiness of survey and experiment findings. Several combinations are already known to practitioners in the field, whilst others have not yet found practical expression. It is clear that the choice of particular instruments and combinations will be conditioned not only by the extent to which they improve trustworthiness, but also by time, money, expertise and other factors which can act as constraints to the process of data collection and analysis. Clearly, all information objectives need to be resourced, and, in many cases, the types of instruments used will be as much - or more - a reflection of resource constraints as they are of objectives.

Both, objectives and resource constraints have implications for the selection of survey teams. Aside from the typical multidisciplinary combination of social and natural science inputs, there is a need to consider inputs from statisticians, especially in the more complex cases.

Case study exercises have shown that it is important that survey teams are sufficiently trained and familiar with approaches and have been provided with sufficient resources to achieve their targets. Supervision can be a problem, in particular if exercised over long distances without direct contact. Unforeseen circumstances can push a relatively inexperienced survey team to the limits of its capabilities. If in doubt about the experience of the team and the tasks expected, it may be more appropriate to choose a less demanding survey design.

Well synthesised survey results are required so that decisions can be taken by project leaders or policy decision makers. A unified set of recommendations should reflect a balanced use of tools, which ultimately led to more trustworthy information. Aside from swapping tools for the collection and analysis of data (i.e. *merging* of techniques), findings obtained through the use of one approach can be confirmed, enriched, or refuted by research results obtained from the concurrent or sequenced use of the other approach.

This paper identifies a range of possible combinations of qualitative and quantitative survey techniques, some of which were tested as part of DFID research project R7033. Copies of the various case studies and theme papers written as part of this project can be found at the following website addresses: http://www.reading.ac.uk/ssc/ and http://www.nri.org. The fact that some approaches are relatively untried requires a certain degree of flexibility during design and implementation of research and development projects aiming to improve natural resource use and livelihoods.

References and Suggested Reading

Abbot, J. and Guijt, I. (1997) *Methodological Complementarity: Creativity and Compromise in PLA Notes* 28 February 1997.

Abeyasekera, S. (2000) *Analysis approaches in participatory work involving ranks and scores.*

Abeyasekera, S. and Lawson-McDowall, J. (2000) *Computerising and Analysing Qualitative Information from a Case Study Concerning Activity Diaries.* Case study produced as part of DFID research project R7033.

ABMRC (1989) *Buying Market Research (*Key Note Guides)

Barahona, C. (2000) *Generalising Results from Matrices of Scores.* Case study produced as part of DFID research project R7033.

Burn, R.W. (2000) *Causal diagrams.* Theme paper produced as part of DFID research project R7033.

Carvalho, S. and White, H. (1997) *Combining the Quantitative and Qualitative Approaches to Poverty Measurement and Analysis: The Practice and the Potential.* World Bank Technical Paper No 366.

Casley, D.J. and Kumar, K. (1988) *The Collection, Analysis and Use of Monitoring and Evaluation Data* (Oxford)

Casley, D.J. and Lury, D.A. (1982) *Data Collection in Developing Countries* (Oxford).

Chambers, R (1983) *Rural Development: Putting the Last First.* London: Longman.

Coe, R. and Franzel, S. (1997) *Designing On-farm Experiments: Lecture Notes* (Unpublished). ICRAF, Nairobi.

Davis, R. (1997) *Combining rapid appraisal with quantitative methods: an example from Mauritania.* In PLA Notes 28.

Floyd, S. (1998) When is quantitative data collection appropriate in farmer participatory research and development? Who should analyse the data, and how?, Chatham: Natural Resources Institute.

Galpin, M., Dorward, P., and Shepherd, D. (2000) *Participatory Farm Management Methods for Agricultural Research and Extension Needs Assessment*: A Manual. Departments of Agriculture and Agricultural Extension and Rural Development, The University of Reading.

Hagmann et. al. (1995) *Integrating Formal Research Into A Participatory Process in:* ILEA Newsletter Vol. 11 No 2.

Leach, M. and Kamanga, J. (1997) *Shotgun wedding or happy marriage? Integrating PRA And sample surveys in Malawi.* In PLA Notes 28.

Lincoln, Y.S. and Guba, E.G(1985) *Naturalistic Enquiry* (Sage).

Marsland, N. and Golob, P. (1997) *Rapid Rural Appraisal of Post-Maturity Issues in the Central Region of Malawi.* (NRI - Unpublished).

Marsland, N., Wilson, I., Abeyasekera, S., and Kleih,U. (1998) *A Methodological Framework for Combining Quantitative and Qualitative Survey Methods -* Background Paper: Types of Combinations. Report written for DFID Research project R7033.

Marsland, N., Golob, P., and Abeyasekera, S. (1999) *Larger Grain Borer Coping Strategies,* Case study produced as part of DFID research project R7033.

Marsland, N., Henderson, S., and Burn, R.W. (1999) *On-going Evaluation of FRP project: Sustainable Management of Miombo Woodland by Local Communities in Malawi,* Case study produced as part of DFID research project R7033.

Martin, A. and Sherington, J. (1996) *Participatory Research Methods - Implementation, Effectiveness and Institutional Linkages.* Paper prepared for SEM Workshop, ODA, 28-30 April 1996.

McCracken, Pretty J.N. and Conway (1987) *An Introduction to Rapid Rural Appraisal for Agricultural Development.* (IIED).

Mukherjee (1995) *Participatory Rural Appraisal and Questionnaire Survey.* (Concept, New Delhi).

NCAER (1993) *Comparative Study of Sample Survey and Participatory Rural Appraisal Methodologies with Special Reference to Evaluation of National Programme on Improved Chullah.*

Pretty J.N. (1993) *Participatory Enquiry for Sustainable Agriculture.* (IIED - Unpublished).

Pretty, J.N., Guijt, I., Thompson, J., and Scoones, I. (1995) *Participatory Learning and Action: a Trainer's Guide.* London: International Institute for Environment and Development.

Shah, P. (1993) *Questionnaires: Participatory. Reliable and Interesting?* A note prepared for IIED/IDS Workshop on Alternatives to Questionnaire Surveys.

Schoonmaker-Freudenberger (1996) *The use of RRA to inform policy: Observations from Madagascar and Guinea.*

Shanley, P. Galvao, J., and Leda, L. (1997) *Limits and strengths of local participation:* A case study in Eastern Amazonia. In PLA notes 28.

UPWARD (1997) *Participatory Monitoring and Evaluation: The Asian Experience.*

Wilson I. M. (2000) *Sampling and Qualitative Research.* Theme paper produced as part of DFID research. Project R7033.

Ziche (1990) *Some Critical Aspects in Gathering Socio-Economic Data in Rural Areas of Non-Western Societies, in: Hoeper, Qualitative versus Quantitative Approaches in Applied Empirical Research in Rural Development.* (Zed)

PART II

PRACTICAL ISSUES RELATED TO COMBINING QUANTITATIVE AND QUALITATIVE SURVEY METHODS

Theme Paper 1

PARTICIPATION AND THE QUALITATIVE – QUANTITATIVE SPECTRUM

Neil Marsland, Natural Resources Institute, University of Greenwich

PARTICIPATION AND THE QUALITATIVE – QUANTITATIVE SPECTRUM

"I remembered one morning when I discovered a cocoon in the bark of a tree, just as the butterfly was making a hole in its case and preparing to come out. I waited a while, but it was too long appearing and I was impatient. I bent over and breathed on it to warm it. I warmed it as quickly as I could and the miracle began to happen before my eyes, faster than life. The case opened, the butterfly started slowly crawling out and I shall never forget my horror when I saw how its wings were folded back and crumpled; the wretched butterfly tried with its whole trembling body to unfold them. Bending over it, I tried to help it with my breath. In vain. It needed to be hatched out patiently and the unfolding of the wings should be a gradual process in the sun. Now it was too late. My breath had forced the butterfly to appear, all crumpled, before its time. It struggled desperately and, a few seconds later, died in the palm of my hand".

NIKOS KAZANTZAKIS

Zorba the Greek

Introduction

"Participation" is normally a word that is applied to those for whom a particular research or development project is intended. Such people are normally referred to as the primary stakeholders, a term which has now superseded the former term "beneficiaries". In the context of this project, the primary stakeholders are people living in rural areas in developing countries. The degree of participation in design of data collection instruments, actual data collection itself and analysis of results depends on the objectives of the enquiry in question, the constraints in achievement of the objectives and should also depend on the types of variables being investigated.

If we wish to encourage primary stakeholders fully to participate in the information system of a project then we need to accept the fact that the process may take longer than we would have liked. If we attempt to rush the process unduly then participation may fail entirely – this is the message of the experience of Zorba the Greek.

What is "Participation"?

The range of meaning of participation is similar whether one is referring to research or development projects or whether one is referring to surveys or experiments. DFID makes use of a simple model of stakeholder participation[1], which can be applied to the research, development, survey and experiment context. For DFID, participation is seen as a spectrum with a range of possibilities:

[1] See – originally - ODA (1995) Guidance note on how to do a stakeholder analysis of aid projects and programmes. This has now been incorporated into DFID (1995).

- being in *control* and only consulting, informing or manipulating other stakeholders
- *partnership* (equal powers of decision-making) with one or more of the other stakeholders
- being *consulted* by other stakeholders who have more control
- being *informed* by other stakeholders who have more control
- being *manipulated* by other stakeholders

Manipulation can in some ways be regarded as a failure of participation. As noted in DFID (1995), "No one likes being manipulated...but the other parts of the spectrum are all equally valid roles, selected on the basis of the specific circumstances".

As Floyd (1998) reports, in the specific context of on-farm trials (i.e. experiments), Biggs (1987) has framed the spectrum of participation in terms of researchers' and farmers' *relative degree of control over the research agenda*:
a) Contract - researchers only set the agenda; farmers' only involvement is that researchers carry out trials on their land;
b) Consultative - researchers consult farmers in order to diagnose problems and modify research plans, but retain control over decision-making;
c) Collaborative - researchers and farmers work as equal partners, and decisions over what research should be done, and how, are made jointly;
d) Collegiate - the research agenda is farmer-driven, with farmers having the final say in all decisions.
All except type a) are participatory in the sense that the research process takes some account of farmers' opinions and priorities, and it is this typology which is often cited in literature on FPR. (also see Case Study 5 on this distinction).

Coe (1997), meanwhile, distinguishes three categories of on-farm trial, based on three *broad groupings of objectives*:
a) Type 1 - researcher-designed and managed;
b) Type 2 - researcher-designed and farmer-managed;
c) Type 3 - farmer-designed and managed.
Types 2 and 3 are clearly participatory in the sense that farmers are involved in implementing the research. Type 1 may also be seen as 'participatory' if farmers and researchers have decided together that such a trial would be useful (research mode is collaborative or collegiate), or if farmers are involved in assessing the outcome of the trial[2].

[2] Note: Readers are referred to the paper "When is quantitative data collection appropriate in farmer participatory research and development? Who should analyse the data, and how?" by Floyd (NRI 1998) for more details on quantitative methods in the on-farm trial context.

"Participation does not equal Qualitative".

Participation and the qualitative – quantitative spectrum

Participation in data collection and analysis should not be confused with qualitative data collection. Data collection and analysis may be highly participatory and highly quantitative, equally it may be highly non-participatory and highly qualitative. Table 1 illustrates the importance of distinguishing between "participation" and "qualitative".

Table 1: Participation and the qualitative – quantitative spectrum.

Data collection and analysis	Participation	
	Low	High
Highly quantitative	1	2
Highly qualitative	3	4

Key:

1. Low participation by primary stakeholders, highly quantitative data collection and analysis. This could characterise a sample survey with closed questions investigating quantified variables (e.g. educational achievement, number of pit latrines, family structure and ages), and carrying out analysis using a statistical package.

2. High participation by primary stakeholders, highly quantitative data collection and analysis. This could characterise an inventory of some sort undertaken by primary stakeholders and analysed by them using simple arithmetic. e.g. counting the number of species and trees in a communal woodlot.

3. Low participation by primary stakeholders, highly qualitative. This could characterise a situation in which the opinions of secondary stakeholders such as field staff form the central part of a study and primary stakeholders are consulted briefly.

4. High participation by primary stakeholders, highly qualitative. This could characterise an in-depth PRA type of study that produces a large quantity of narrative and results of diagrams such as Venn diagrams, maps and transects.

Objectives, constraints and participation outcomes

The way in which participation engages with the qualitative – quantitative spectrum depends on the configuration of information objectives and constraints that will include the boundaries set by motivation to and interest in participating, money to fund participation, time to be participatory and expertise to facilitate participation[3].

Building on Table 1, Table 2 illustrates how information objectives and constraints can shape the degree of participation of primary stakeholders in data collection and analysis.

Table 2: Participation outcomes in eight different scenarios.

Data collection and analysis	Low resource capacity to enable participation		High resource capacity to enable participation	
	Need for generalisation of results		Need for generalisation of results	
	Low	High	Low	High
Highly quantitative	1	2	5	6
Highly qualitative	3	4	7	8

Key:

1. Low resource capacity to enable participation, low need for generalisation of results, highly quantitative data collection and analysis: **Example:** locationally specific exercise undertaken by external stakeholders e.g. social anthropologist observing the number of times women in one village collect water during one day and analysing the results using simple arithmetic. **Outcome:** primary stakeholders are subsequently informed.

2. Low resource capacity to enable participation, high need for generalisation of results, highly quantitative data collection and analysis: **Example:** a sample survey with closed questions investigating quantified variables (e.g. educational achievement, number of pit latrines, family structure and ages), and carrying out analysis using a statistical package. **Outcome:** primary stakeholders are consulted.

3. Low resource capacity to enable participation, low need for generalisation of results, highly qualitative data collection and analysis. **Example:** Study concerning key informants with geographically limited local knowledge using highly qualitative tools

[3] Readers are referred to Table 2 and Pages 3-4 of the Methodological Framework in Part I for a more general treatment of the role of objectives and constraints in determining the qualitative – quantitative mix.

e.g. maps and other diagrams and open ended discussion. **Outcome:** primary stakeholders may be consulted, informed or manipulated.

4. Low resource capacity to enable participation, high need for generalisation of results, highly qualitative data collection and analysis. **Example:** key informants representing different points of view or with special knowledge of issues which have a wide applicability over space are consulted for their opinions. **Outcome:** primary stakeholders are informed.

5. High resource capacity to enable participation, low need for generalisation of results, highly quantitative data collection and analysis. **Example:** intensive location specific quantitative exercise undertaken by primary stakeholders e.g. an inventory of some sort undertaken by primary stakeholders and analysed by them using simple arithmetic. e.g. counting the number of species and trees in a communal woodlot. **Outcome**: primary stakeholders are in partnership with external facilitators.

6. High resource capacity to enable participation, high need for generalisation of results, highly quantitative data collection and analysis. **Example:** Extensive exercise involving generation of ranking and scoring matrices from different village that are then compared and analysed. **Outcome:** primary stakeholders are consulted

7. High resource capacity to enable participation, low need for generalisation, highly qualitative data collection and analysis. **Example:** highly inclusive process, perhaps extending over several years using locationally specific tools and approaches. **Outcome:** primary stakeholders are ultimately in control.

8. High resource capacity to enable participation, high need for generalisation, highly qualitative data collection and analysis. **Example:** Engagement in a process which generates narrative and visual material such as maps, transects in a number of different sites which need to be compared with general conclusions derived. **Outcome:** primary stakeholders are consulted / in partnership.

Participation in Project Information Systems: the Example of Monitoring and Evaluation

Over recent years, it has become more common for primary stakeholders to be involved in defining, gathering and analysing monitoring and evaluation information. This type of involvement is commonly referred to as Participatory Monitoring and Evaluation (PM&E). In line with the points made in the introduction, PM&E does not equate with the use of qualitative methods instead, it is, as Guijt (1999) notes,

"...about negotiating what needs to be assessed and measured, and then finding appropriate methods. Combinations of quantitative and qualitative and more and less participatory or natural science oriented methods are likely to emerge from discussions".

More generally, when the spectrum of participation is translated into a PM&E context, it determines the nature of the following tasks: indicator selection, selection of method(s) to measure indicators; collection of data; collation of data; analysis of data; disseminating data and taking action on the basis of the analysed information.

There are many different examples of PM&E, covering a wide spectrum of participation (from informing to controlling). The degree of participation is closely related to objectives of the M&E, whether these be organisational self-assessment; citizen monitoring of government programmes; villagers monitoring externally driven projects, or resource users monitoring the state of their own environment. As well as objectives, there are constraints. Unwillingness to be involved in PM&E is a critical potential constraint. Factors conditioning people's willingness to get involved in M&E have been summarised by Gujit (1999) see Box 1.

BOX 1: WHAT INFLUENCES PEOPLE'S PARTICIPATION IN MONITORING AND EVALUATION?

- perceived benefits (and partial or short term costs) of PM&E
- relevance of M&E to the priorities of participating groups
- quick and relevant feedback of findings
- flexibility of the PM&E process to deal with diverse and changing information needs
- meeting expectations that arise from PM&E, such as acting on any recommendations that are made
- degree of maturity, capabilities, leadership and identity of the groups involved, including their openness to sharing power
- local political history, as this influences society's openness to stakeholders initiatives
- whether short term needs of participants are dealt with, while considering the longer term information needs of PM&E (especially in natural resources management)
- incentives to make the PM&E possible (e.g. pens, books etc)

Source: Guijt (1999: 11)

Another constraint for PM&E can be the need to get consensus form different stakeholder groups on the types of indicators used and the frequency of reporting. A donor or government need for some degree of uniformity in indicator type over a particular geographical area, may limit the degree of participation by primary stakeholders in defining indicators[4].

Conclusion

The relationship between participation and the qualitative – quantitative spectrum in data collection and analysis is not self-evident. It will be influenced by the configuration of objectives and constraints. In the context of a project cycle, the configuration may change

[4] For a discussion of the procedural aspects of applying statistical analysis to qualitative data sets, see the Methodological Framework in Part I.

implying different points along the qualitative – quantitative spectrum as far as information is concerned, and different levels of participation.

References

Biggs S. (1987) Proposed methodology for analysing farmer participation in the ISNAR OFCOR study. *Agricultural Administration (Research and Extension) Network Newsletter*, No. 17 Overseas Development Institute, London.

Coe. R (1997) Aspects of on-farm experimentation: lecture notes (unpublished).

DFID (1995) Stakeholder Participation and Analysis (*Social Development Division*)

Floyd, S. (1998) When is quantitative data collection appropriate in farmer participatory research and development? Who should analyse the data, and how?, Chatham: Natural Resources Institute.

Guijt, I. (1999) Participatory Monitoring and Evaluation for Natural Resource Management and Research. *Socio-Economic Methodologies for Natural Resources Research – Best Practice Guidelines*, Chatham, UK: Natural Resources Institute.

Marsland et.al. (in press) A Methodological Framework for Combining Quantitative and Qualitative Survey Methods, *Socio-Economic Methodologies for Natural Resources Research – Best Practice Guidelines*, Chatham, UK: Natural Resources Institute.

Theme Paper 2

SAMPLING AND QUALITATIVE RESEARCH

I. M. Wilson, Statistical Services Centre, University of Reading

SAMPLING AND QUALITATIVE RESEARCH

Summary

This paper addresses one particular part of the search for knowledge and understanding - the *principles* of sampling. It does not set out to discuss what should be done with an entity which has been sampled. Different disciplines approach issues differently. Maybe a quantitatively minded scientist would measure something which she thought characterised an entity, say a household. Maybe a qualitative practitioner would conduct a wide-ranging discussion with members of a household he worked with. Quite aside from that, either should have some answer to questions as to how they came to sample any particular entity or the overall set which they selected. Many other issues of research design will also arise for either investigator, but these are not developed herein.

This paper concentrates on issues of particular concern to those who are working near the qualitative/quantitative "interface". An SSC guideline booklet produced for DFID (SSC, 2000) and available on the SSC website at http://www.reading.ac.uk/ssc covers other general topics not developed here. The main message is that the underlying ideas work equally effectively for qualitative or quantitative approaches, when these are concerned with collecting information about a population as the basis for coming to some broadly applicable conclusions - generalisation. The very wide applicability of the ideas means that a brief description as here is rather abstract.

The issue of section 2 is described as "site" selection. We assume here that a multi-stage or hierarchical sample involves the choice of some large units such as villages, then selection of smaller units within these, e.g. households, and maybe then of individuals within those. "Sites" are the large primary units. Hierarchical sampling is neither universal nor essential, but it is the focus here because it is extremely common and poses a number of problems often given scant attention in qualitative settings. One of these is that sample size determination has to be thought out in relation to the structure, so that not only the total number needs to be decided but the spread of effort across the stages. It may be useful to think of hierarchical sampling in sequencing terms, so that a relatively large initial sample of the primary units is studied, maybe cursorily in a first phase, and the selection from them is based on observations taken. Some phasing ideas are developed.

Not infrequently, reviewing the information on a large sample of sites is laborious and a method which allows comparison of smaller subsamples is useful. A relatively little known method, of ranked set sampling, is advocated.

Section 3 concerns the following idea. The way samples are chosen may be much easier to think through if it is clear whether the main objective is (a) separate description of each one, (b) a synthesis intended in some sense to "represent" the whole population sampled, or (c) a comparison with respect to one or several characteristics, e.g. is access important? Are the results (qualitative or quantitative) very different between accessible and remote sites, say? Where comparison is the key objective, comparability is usually much more important than overall representation of the population as SSC (op. cit.) explains.

Section four addresses the question of whether studies can and should be broken down into modules, so that due consideration is given to how much information is needed on each issue, and to how the modules fit together. On the other hand larger programmes of work are often first considered as free-standing studies on what may be unconnected samples from the same population. We draw out the point that linking the studies allows a more effective integration of livelihood information.

Another form of structure concerns the subdivision of the sampled population into separate sections - strata - which may allow more economical sampling if each stratum is internally homogeneous. Several variations on this theme are briefly reviewed.

The concluding section offers some practical starting points for situations where the researchers want to target special segments of the population. The message of this section is that there are interesting ways forward, but some "adaptive research" challenges awaiting when we take concepts from relatively "easy" quantitative settings, and bring them into the more complex settings of qualitative work in developing countries.

Introduction

Broad applicability of geographically small-scale research

It is in the tradition of qualitative work that it aims to build up an accurate, in-depth interpretation of what is being studied through triangulation of different descriptive sources, e.g. according to DFID SLSG (2000) [1]. Of necessity this often means limited breadth" e.g. geographical spread.

Broad applicability is not always a target. Generalisation may be irrelevant to many participatory exercises which make very limited claims beyond local impact of the work done, e.g. empowerment of a particular community. However, for many development projects which use qualitative techniques or the frequent mixture of qualitative and quantitative methods, the issue *is* important: how can it be demonstrated that results or findings will work beyond the immediate setting in which a project has been based? This theme paper is intended to inform people who face this question.

One major purpose of considering sampling issues carefully is to ensure that results can be claimed to be representative - of a population or a clearly-defined part of it. That the results are representative is the basis of generalisation from a small sample to a statement about the whole. The themes developed in this paper all have this purpose in common.

"The plural of anecdote is not data"

Evidence-based generalisation is a characteristic purpose of research. If the claim is made to or by a funding agency, such as DFID, that work is worthwhile because the results will be widely applicable, the issue of generalisability should be addressed seriously. Often research or development project proposals are vulnerable to the criticism that their results will "just" be case studies and that they will do very little to provide knowledge rather than hearsay about the wider population supposed to be the beneficiaries of development budgets. The first aim below is to describe some procedures whose application can help to support the claim to representativeness.

One common difficulty is that the *target* population is *not* a readily ascertainable, maybe national, population, but a subset whose size and boundaries are rather ill defined. This is addressed in section 6 below.

Value of "statistical" sampling concepts
Sampling methods are most usually formalised in the context of quantitative "statistical" research, especially if the primary aim is to infer population characteristics with some assurance of representativeness. These ideas are often couched in inaccessible language, and in such idealised terms that practical application seems like far too much trouble. Yet many of the concepts of sampling are very general and can be applied, in a qualitative form, to qualitative information-gathering studies. The purpose of doing so is to bring some of the generalisability arguments to bear in support of study conclusions. The purpose of this paper is thus to show how the swapping of tools and attitudes mentioned in Marsland et al. (2000) [2] relates to the application of "statistical" sampling ideas in more qualitative enquiries.

High-quality modern examples of the "statistical" sampling literature include Thompson (1992) [3] and Levy and Lemeshow (1999) [4]. These are wide-ranging and technically deep. For the purposes of this paper, it is adequate to have an acquaintance with the brief pamphlet by SSC (2000) [5] or Lindsey (1999) [6].

Site Selection

Aim
In this section we discuss ways in which sampling concepts can help the in-depth, probably qualitative, maybe participatory, study to be justified as being more than "just a case study". By borrowing conceptual tools from "statistical" sampling the study carried out in rather few sites can be underpinned a little more effectively. We feel that establishing the credibility of a claim to representativeness is a very important issue for much qualitative research. Even with careful, and well-documented procedures this is difficult with small samples and there is no magic answer to the problem. In the following we set out some of the concepts which we feel help to support the claim as effectively as possible with given resource limits.

A prototype example
Rather than cite or criticise any actual project, we use an artificial study which illustrates a number of the points we make. This synthesises a subset of the real features of a considerable number of research proposals made to DFID.

The researchers are concerned with "livelihoods" aspects of a set of potential innovations which might be taken up by poor rural communities, households or individuals in a swathe of Africa. Given the holistic nature of the approach, a few communities have to be studied quite intensively. The innovations are not all equally suitable for everyone, because of variations in resources. In some cases adoption is primarily a household decision, in others effective utilisation depends to a greater degree, perhaps crucially, on community acceptance or involvement, while at a larger scale topographical factors may limit the geographical range of applicability of any part of the set of innovations.

Hierarchical Sampling

Taking it as an unarguable given fact that "a few communities have to be studied quite intensively", researchers often focus on the question "how?" and delve immediately into the methodologies to be used within communities, qualitative, quantitative or a combination. The first sampling point is that the selection of communities, then of households within communities, and maybe individual members within households is a hierarchy of *connected* decisions. The choice of the number of primary sampling units, frequently communities, and their selection, is all too often described and justified very poorly indeed in research proposals

The Problem of Study Design Compromises

If we select twenty communities instead of two - for equally intense study in each one - the depth attainable in any one community becomes much less and the study loses some richness of detail about human variety, social structure and decision-making *within* communities.

If we select two communities instead of twenty, though, our sample of communities is a sample of size two! If our two communities turn out to be very different from one another, or have individual features that we suspect are not repeated in most other places, the basis of any generalisation is extremely weak and obviously suspect. Even if the two produce rather similar results, generalisation is still extremely weak, though less obviously so, insofar as we have almost no primary information about *between*-community variability (* see below). To quote a DFID research manager, we may have "another meaningless case study".

It is perhaps useful to elaborate the above argument. The point made here is about the intrinsic nature of variability, so it is easiest to demonstrate by considering one community-level measurement, X, and doing so without too much distraction because of the context. Simple quantitative examples could be something like the water-related outgoings of households with piped water in their compounds, or the off-farm income of heads of household with three or more years of schooling.

If we have two values of X from two communities A and B, say $X_A = 220$ and $X_B = 260$, there are many unanswerable questions, e.g. is either value "typical" of the "population" of communities from which A and B are sampled, do both fall in a "usual range" for the "population", which of many differences in variables other than X are closely related to the difference here?

The informal interpretation of variability is much more assured if we have five communities, A - E, rather than two which provide X-values. Three possible samples, (i), (ii), and (iii) of five communities each, are hypothesised below, and each of (i) - (iii) points towards a clearer supposition if not a firm conclusion.

	A	B	C	D	E
(i)	*220,*	*260,*	215,	220,	215
(ii)	*220,*	*260,*	285,	190,	240
(iii)	*220,*	*260,*	215,	265,	270

- In hypothetical case (i), the value 220 looks like a representative part of a grouping, while 260 may be anomalous or possibly representative of something less common - it looks like an "outlier" relative to the others.
- In (ii), both 220 and 260 fit into a range of variability - there is considerable variability but no apparent pattern within it.
- In (iii) there is a bit of a suggestion that there may be two groups - around 220, and around 265.

As these examples show, five is still an extremely small sample from which to claim that we have an overall picture. If the sample entailed 20 values rather than five, any one of the above data patterns, and the corresponding conclusions, might achieve much greater plausibility.

The above arguments are not to do with what exactly is being measured or how. At a conceptual level they apply equally well to qualitative or quantitative research instruments. Even the argument immediately above, presented in terms of one number per community, can be thought through in terms of a profile comprising several numbers, or in qualitative terms.

Where is effort expended?
Crudely, one can think of [total effort] = [effort per community] x [number of communities]. If one factor on the right hand side goes up, the other must come down for a total resource fixed in terms of effort, or project budget. Including an extra community often entails a substantial overhead in terms of travel, introductions and organisation, so that [effort per community] = [productive effort + overhead]. If this means [total overhead] = [overhead per community] x [number of communities], there is a natural incentive to keep down the number of communities, because this overhead then eats up less of the budget. This seems efficient, but there needs to be a deliberate, well argued case that an appropriate balance has been struck between this argument about overhead and the counter-argument at * above about having too few units from the top level of the hierarchy.

Phased sampling of hierarchical levels
One commonplace concept which can help if used effectively is that of phased sampling. This is only a formalisation of what people do naturally, and it is an example of what Marsland et al. describe as "sequencing". As a first phase a relatively large number of communities - tens or hundreds - may be looked at quickly, say by reviewing existing information about them. Alternatively it could be done by primary data collection of some easily accessible, but arguably relevant, baseline information on tens, probably not hundreds, of them. Either route can lead to quantitative data such as community sizes or to qualitative classifications or rankings.

Sampling of communities for more intensive study can then be informed by this first-phase data if the two phases are thoughtfully linked up, e.g. if phase one gives us an idea of the pattern of variability of X above, we may be able to argue somewhat plausibly that a small sample of communities can be chosen for intensive investigation in the second phase and yet be "representative" of a larger population - as far as X is concerned. Of course this argument has been simplified here in referring only to one variable X; in reality a selection of key items of first-phase data would be used to profile communities, and the argument would be made that the communities chosen for the second phase are reasonably representative *with respect to this profile.*

The real purpose of random sampling as the basis of statistical generalisation is objectivity in the choice of units - here communities. However, a very small sample may be obviously unrepresentative if chosen at random. Given that the sample size is unavoidably small, the claim to "objectivity" is more or less achieved if the target communities for the second phase are *selected* on the basis of reasonable, clearly stated criteria using first phase information. This is elaborated in SSC (2000, op.cit.)

Some phase one information simply demonstrates that selected communities "qualify" e.g. maize growing is the primary form of staple food production, or e.g. they fit with DFID's poverty focus; other items that they are "within normal limits" e.g. not too untypical in any of a set of characteristics. A few key items of information show that the selected communities represent the spread in the set e.g. different quality of access to urban markets. Representation can be of differing sorts. For example we might choose two with easy access and two with difficult access. A different strategy could select four taken 1/5, 2/5, 3/5, and 4/5 of the way through a ranked list, omitting the very untypical extremes. The case studies can then be seen or shown to be "built" on a more solid evidential foundation in terms of what they represent.

The above process can work if the first phase and subsequent work are qualitative, quantitative or mixed, but it is particularly apt to use simple quantitative data followed by qualitative work, since the first phase described above is intended to be cheap and cheerful rather than rich and deep. Note that the above is phrased so as not to limit or direct in any way the phase 2 methodological choices made by discipline specialists working in development settings. What is suggested is just a systematic approach to selection of settings, which offers some strengthening of the foundations on which the work is based.

How does this relate to "statistical" approaches?

In many people's minds, *a priori* "statistical" sample size argument is all about formulae and tables derived from them e.g. as in Lemeshow et al. (1990) [7], used to decide in advance how big a data collection exercise needs to be. Almost all formulaic argument is based on considering one variable at a time so is primarily relevant to single-issue studies, e.g. if we are looking at the coverage of a vaccination campaign our one primary variable of interest is the proportion of qualifying subjects vaccinated.

Even when a quantitative study gets one or two steps more complex relevant formulae become scarce. When the aim is to observe several characteristics and cross-tabulate results, statistical work is concerned with checking that the patterns of relationships are adequately ascertained. This involves situation-specific and detailed consideration of what pattern and adequacy mean. In such cases, what statisticians actually do is usually to elicit from the researcher the overall pattern of data (s)he expects to find and the analyses (s)he is likely to need to do. It *may* then be possible to pin down critical parts of the data collection and set up size targets or sampling strategies to ensure those are adequately covered.

It is a logical fallacy to think there is any easy formulaic *a priori* answer to the sample size question for a complicated study design issue where the analysis is to incorporate integrative or holistic approaches. There is no easy or universally applicable formula which can be used as a substitute for thinking things through. What is almost always required is to think carefully and in detail about circumstances, expectations and the essential core of the analysis which is foreseen. The present context - phasing - serves to emphasise another point which very few formulae are able to assimilate: most sampling decisions quite rightly incorporate prior knowledge about the research setting. Of necessity this will usually be a mixture of relatively explicit and somewhat vague information, of local detail and analogy with other better-known settings. The most valuable role of the experienced sampling statistician is often to help the qualitative or quantitative researcher to focus this partial understanding into structured best guesses about what data collection can be expected to yield, and to ensure a data collection procedure will yield the most rewarding material possible.

A subsequent phase
A second sequencing idea can be added on to strengthen the claims of in-depth research carried out in a few communities. The knowledge gained, thought to be applicable to the wider population, should lead to recommendations which can be tested in communities other than those impacted by the main research work. These can be sampled from the set considered in the first phase, possibly as a "replicate" of the set chosen for intensive phase 2 work. Once again this applies in the context of our prototype example, whether the work is qualitative or quantitative in the various phases.

Ranked Set Sampling
A rather different concept from sampling theory also operates at a conceptual level where it is equally applicable to qualitative, including participatory work, and brings with it some claims to objectivity of selection, lack of systematic bias, and generalisability. The ranked set sampling approach is illustrated here in a simple case, where there is a single key measure or characterisation used to determine that the sample chosen is reasonable.

One frequent problem where ranked set sampling can help is the following. In the context of our prototype example, the discussion above assumed that a baseline study or existing sample frame is available, so that all the site selections could be made from a reasonable if not comprehensive list of communities. If there is no such list, how might we proceed, using more localised knowledge to help choose a few communities?

Example
A participatory problem diagnosis study is to be carried out in four food-insecure communities in one of the eight Agricultural Development Divisions (ADDs) of Malawi. Four* Extension Planning Areas (EPAs; sub-units) are selected at random from those which have featured recently in the Famine Early Warning System

(FEWS) as having food-insecure communities. A set of "qualifying" criteria are set up which exclude unusual or untypical communities e.g. trading centres adjacent to metalled roads. Four* village communities per EPA are selected and it is verified that they "qualify", e.g. they have *not* been the setting for any village-based development project in the recent past. Knowledgeable extension staff from each EPA are asked to think about the last five years and to rank the set of four villages in terms of the proportion of their population who suffered three or more months of hunger in the year with the worst rains out of the last five.

The 1, 2, 3, 4 rankings from the four EPAs are brought together. Taking the sets of ranks in an arbitrary order the community ranked 1 in the first EPA is selected, that ranked 2 in the next EPA is selected, that ranked 3 is taken from the EPA that happens to be third in the review, and that ranked 4 from the fourth.

This set of four selected villages now has one per EPA, but also some claim to span the range of levels of food insecurity in the target area, and not to represent unconscious selection biases of the researchers, insofar as it has some elements of objectivity in its selection. The four villages selected are a set chosen in a "random" way to be representative of a larger sample of 16. This sampling process has in no way affected the research methodology decisions which can now be made by the qualitative researchers working in each of the four villages. Of course the status in the entire population of the four villages ranked 1, say, will not be identical, and *a fortiori*, the differences between those ranked 1 and those ranked 4 will not be the same. This does not matter to the argument that we have an "objective" subset of a first sample of 16, and an enhanced claim to representativeness.

Of course the above is a conceptual description and practical safeguards need to be in place. The initial selection of the set of villages within each EPA is assumed here to be made in an objective way e.g. unless it is properly incorporated in the analysis we prevent any influence due to extension officers' perceptions of transport difficulties.

Extra ranked sets
As a by-product of this selection process we also have some "spare" similar ranked sets. For example the set comprising that ranked 3 in the first community, that ranked 4 in the second, that ranked 1 in the third, and that ranked 2 in the fourth should provide a "comparable" ranked set, because any single ranked set of 4 is supposed to be representative of the original 16. One use of this second rank set could be for a later phase of recommendation testing, as described above. See also section 3 below.

The process of ranking considered in this section is, in some cases at least, very quick and easy compared with "the real work" which follows. The sample of size n (4 above) was selected from n^2 villages (16 above), but in some circumstances a larger "starting" or "comparison" set can be managed. Each set of EPA officials might be asked to rank eight villages as four pairs, and the researchers could then choose one at random from each pair. They would end up with a sample of the same size n grounded in a comparison set of 2 n^2, i.e. 32 in the example. Many variations are possible on this theme.

* The two numbers should be the same - as the method description indicates - but of course four is just an example.

Site Comparison and Impact Assessment

Aim

In this section we discuss ways in which sampling concepts can contribute to comparisons, for example those useful to impact assessment. This section talks mainly in terms of a project, or even a project activity, where explicit examples of the statistical principles can easily be briefly stated. As explained in SSC (2000, op.cit.), it is at the lower levels in a hierarchy, where the reader will not have detailed information about the individuals involved, that well-documented objective sampling procedures are most important. At programme level there is greater richness of information about the units sampled, and greater emphasis on qualitative judgment in sample selection by the assessor. The importance, and implications, of hierarchical, or multi-stage, sampling are discussed in SSC (2000, op.cit.) and this discussion is not repeated here.

A specimen study design

It is well-known e.g. Cook & Campbell (1979) [8] that the skeletal form of research design which provides a truly effective demonstration that an intervention had an impact is to have "before", or "baseline", and "after" observations, both in a sample of sites where "the intervention" occurred and in a comparable "control" set of sites where there was believed to be no effect due to the intervention.

Per pair of sites	Before	During	After
Intervention = X	O	X	O
Observation (O) only	O		O

Once again this component of the research design for a set of studies does not dictate the use, let alone sole use, of quantitative measurement nor quantitative comparison of before/after information. The reader is referred to Cook and Campbell, and other sources in the very large literature on research design, for study design considerations wider than sampling: this paper makes no claim to cover that wider field.

Say the research design involved comparing two strategies, each tried in four villages. Two ranked sets as described in the last paragraphs of section 2 above are, as far as we can tell, "comparable" with respect to the criterion used for rank setting. It is plausible to use one set for project work (the intervention) and the other set as the control*.

Matching

A different approach to choosing controls, rather more expensive than using a second ranked set, but relevant however the intervention set was determined, is matching. Pairwise matching requires that for each unit of the intervention set, each village in

* Arguments are made against control sites on the grounds that expectations are raised unfairly, and that cooperation is poor if no potential benefit is offered. Also costs are raised both by time spent "unproductively" on controls and by any compensation to those used in this extractive way. Research managers must trade these difficulties off against longer-term problems. For example when DFID's ten-year RNRKS strategy is being appraised around 2005, weak ability to demonstrate impact could have harsh consequences. For the purposes of this paper, we take the view that there is a logical place for using controls, and address ourselves to situations where others decide that this is feasible.

the above, we find a matched control that shares appropriate characteristics. The difficulties in doing this include both practical and conceptual ones.

To illustrate the conceptual difficulties consider a case where the units are individual heads of household and the aim is to help ensure beneficiaries enjoy sustainable levels of social and financial capital. If the matching is established at the outset, comparability needs to be established in respect of factors which *may* turn out to be important, and that set could be large and burdensome. If the matching is retrospective, as when the comparison is an afterthought, the aim must still be to compare individuals who would have been comparable at the start of the study. The results may be masked or distorted if the possible effects of the intervention are allowed to interfere with this. Of course a targeted approach of this sort often amounts to rather subjective accessibility sampling. Its effect depends upon the perceptions of the person seeking the matches, and comparability may be compromised by overlooking matching factors or distorting effects.
Less demanding matching procedures look to ensure that groups compared are similar overall, e.g. with respect to the mean of a quantitative variable such as village population, or the presence of a qualitative attribute e.g. access to a depot supplying agricultural inputs.

A compromise approach
A sampling-based scheme for selecting comparable controls can be used to set up several potential controls. For example after selecting one ranked set sample of size four from sixteen communities, there remain three which were ranked 1 but not chosen, and similarly for ranks 2, 3, and 4. It is then quite rational to select a comparison set comprising the "best matching" one out of the three for each rank. This may or may not be constrained to represent all four EPAs.

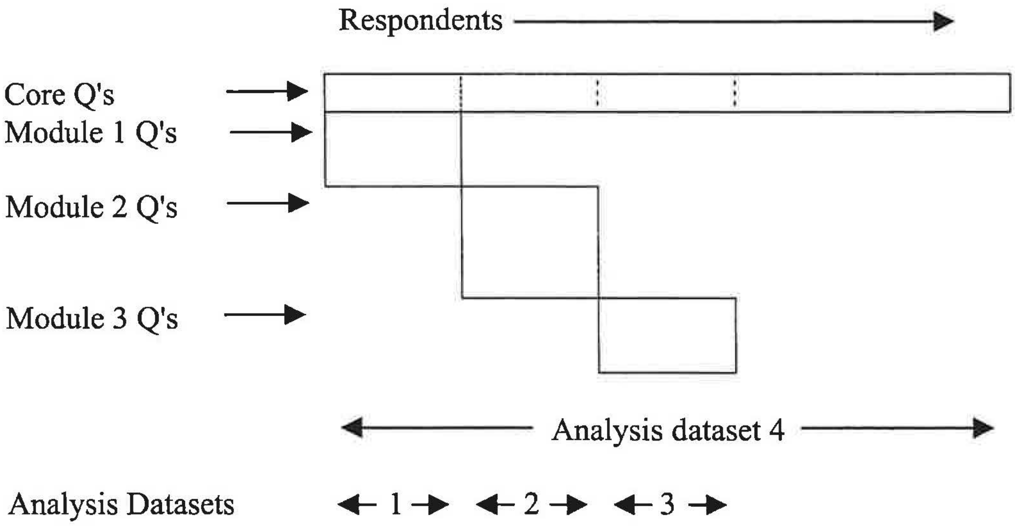
This approach makes the scope of the matching exercise manageably small, the process of selecting the best match being based on the criteria most relevant in the very small set used. Of course the original sampling limits the quality of matching achieved, but the match-finding workload is much reduced. The claims to objectivity, thanks to starting with a sampling approach, are much easier to sustain than with most matching approaches.

Combinations of Studies

We first illustrate the ideas here by reference to a conventional survey - formal or semi-structured. The wider application of the approach is discussed later.

Segments of one survey
Very often, a single survey is effectively a combination of segments each comprising some questions or themes for semi-structured investigation. Having conceived of one round of data collection, there is then a tendency to treat the survey instrument as a unitary entity. Every respondent answers every question, and especially where the survey has been allowed to become large, the exercise gets far too cumbersome. Often this is blamed on the survey method, on the basis that a bad workman blames his tools.

Segmenting a survey into carefully thought out sections can save unnecessary effort. A core set of questions is usually crucial e.g. to establish a baseline for evaluation or impact assessment purposes. Often this set should be small. Other questions which one or more of the research team would like to pursue may not need or justify such a large sample of respondents. Such themes can be set up as modules to be answered by only a subset of the respondents, maybe in one community, or by a structured sub-sample of the respondents within each community. If there are (say) three such modules of lower-priority information the survey can comprise (a) a relatively large sample responding to the core questionnaire, (b) three sub-samples also responding to one of the modules. Diagrammatically, this can be presented as below, where all respondents contribute to analyses of the core questionnaire - analysis dataset 4 - while analyses of module 1 questions and their inter-relationships with core questions are restricted to the relevant subsets of respondents, i.e. analysis dataset 1. The modules are deliberately shown as different sizes. There is no reason they have to be of equal length. There is no time dimension implied here.



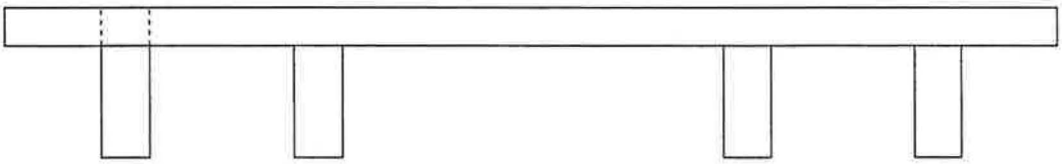
There is a rather common misconception about "balance", perhaps resulting from limited understanding of the analogy with designed experiments, where things are much easier if the study is balanced. In that case the balance is over the measurements of a single quantity, whereas above we are looking at different audiences for *different* questions, and the study described above need not be unbalanced. Warning bells should ring if one starts to take responses to module 1 from one set of people and to "compare" them with responses to module 2 from different people. Modules should be so defined that this is *not* required.

Phasing as modularisation

Of course the idea described immediately above divides a portmanteau survey into modules, but can equally well be thought of as linking a series of what could be separate studies together. In the previous context of phased research, the idea sketched out was of a broad but shallow baseline study, perhaps quantitative, which led on to a smaller sample of locations for more in-depth study in a second phase these having been demonstrated to spread over, or perhaps represent, the wider set.

The diagram equivalent to that above could be as follows, the "table-top" representing the wide but shallow phase 1 study, the "legs" representing narrower, but deeper follow-up work in a few communities. This applies regardless of whether the phases involve collecting qualitative or quantitative data or a mixture.

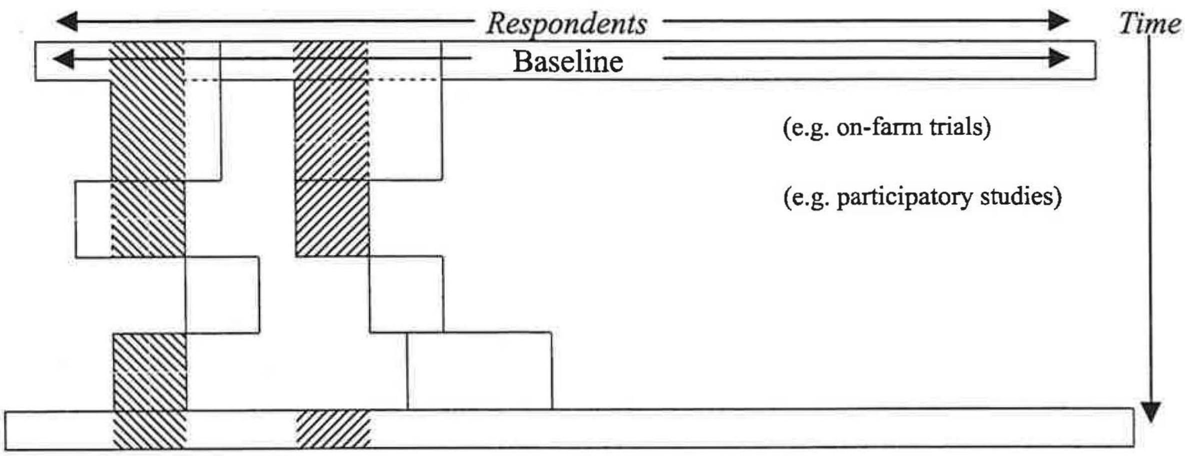
If the data collected in phase 1 can be connected directly to some of that from phase 2, then the analysis after phase 2 can incorporate some of the earlier information, and this is shown diagrammatically in the shaded "leftmost leg" below. Such "read-through" of data becomes more powerful when it adds a time dimension or historical perspective to the later-phase data, as is required in impact assessments. It also serves the purpose of setting the small case study and its phase 2 information



(in the "table leg") in context. The shaded part of the tabletop can be considered relative to the unshaded to show how typical or otherwise the case study setting is.

Project activities as modules

The same style of thinking applies if we consider a development project where samples of farmers are involved in a series of studies. These studies will very likely be conceived at different points of time by different members of a multi-disciplinary team, but may still benefit from a planned read-through of information which allows triangulation of results, time trends, cause and effect, or impact to be traced as the recorded history gets longer. This is not inconsistent with spreading the research burden over several subsets of collaborating rural households, for example. Once again the horizontal range in the following diagram corresponds to a listing of respondents involved, the vertical showing a succession of studies.



Livelihoods

An important point to make in this context is that the separate interventions, processes or studies within a project sequence are of greatest interest to the individual researchers who thought them up. These correspond to a "horizontal" view of the relevant sections of the above diagram. Looking at the project information from the householders' point of view the information which relates to them and their livelihoods is a "vertical" view. Two groups of respondents who have been involved in a particular set of project exercises or interventions are shown diagrammatically by two shaded areas in the above diagram. If a project indeed sets out to focus on the livelihoods of the collaborating communities, it ought to be of interest to look "down" a set of studies at the overall information profile of groups of respondents. As a corollary of this, the programme of studies should be designed, and the data organised, so that such "vertical" analysis can be done sensibly for those livelihood aspects that benefit from triangulation or repeated study, rather than a snapshot view.

Sample Stratification

Set in the context of communities, households and individuals, the above discusses the sampling issues that arise even if the units of study are all treated as interchangeable e.g. one community is the same as any other when sampling at primary unit level, within villages households are treated as being undifferentiated, and so on. Of course this is usually not the case.

Effective stratification
The statistical concept of stratification is widely cited, but not always relevant. Its essential meaning is not technical, and can be expressed clearly by considering a wildly extreme case: suppose a population comprises subsets of individuals where every member is identical within each subset in terms of the response we observe, even though the subsets differ from each other. We then need only a very small sample (of one) from each subset to typify it. In combination with information about how big the subsets are we can typify the whole. In reality stratification is effective if the members who form a subgroup are a *relatively* homogeneous subset of the population i.e. have a greater degree of similarity to one another in response terms than would a completely random subset. The sex of head of household, the land tenure status, or other such factor used for stratification, brings together subsets of people who have something in common. Relatively small numbers of people can be supposed to typify each group, so the method is economical in fieldwork terms. Also it is common that a report of a study will produce some results at stratum level, so it is sensible to control how many representatives of each stratum are sampled, so the information base is fit for this purpose, as well as to represent the whole population.

Ineffective stratification
Populations are often divided into subgroups for administrative reasons, and results may be needed for separate subdivisions e.g. provinces. Unless the administrative grouping happens to coincide with categories of participants who are homogeneous in response, it is not an effective stratification in the above sense. As an over-simplified example, if every village contains farmers, traders and artisans in vaguely similar proportions, villages will be of little relevance as a stratification factor if the main differences in livelihood situation are between farmers, traders and artisans.

The above suggests that the subsets by occupation correspond to clearly distinguished, identifiable groups, internally similar to each other but very different from group to group. In this clear situation, stratification - by occupation - is an obvious sampling tactic. In many cases, however, the groups are by no means so distinct, and the subdivisions may be as arbitrary as the colonial borders of some African states. Usually this makes for ineffectual and delusory stratification.

Pre- and post-stratification
Where stratification is meaningful, it is sensible to pre-stratify where the groupings can be detected before study work commences. In some cases the information for stratifying only becomes apparent during the study, and the cases are sorted into strata after the event - post-stratification.

Participatory stratification
It is sometimes suggested that useful subdivisions of community members within communities can be achieved by getting them to divide into their own groups using their own criteria. This provides useful functional subdivisions for participatory work at local level. If results are to be integrated across communities, it is important that the subgroups in different villages correspond to one another from village to village. Thus a more formal stratification may require (i) a preliminary phase where stratification criteria are evolved with farmer participation, (ii) a reconciliation process between villages, and then (iii) the use of compromise "one size fits all" stratification procedures in the stratified study. If so the set of strata should probably be the set of all subsets needed anywhere, including strata that may be null in many cases, e.g. fisher folk, who may only be found in coastal villages.

Quantile subdivision
Stratification is not natural where there is a continuous range rather than an effective classificatory factor. If there is just one clear-cut observable piece of information which is selected as the best basis to be used, a pseudo-stratification can be imposed. For example a wealth ranking exercise may put households into a clear ordering, and this can be divided into quantiles, e.g. the bottom, middle and top thirds, or four quartiles, or five quintiles. This permits comparisons between groups derived from the same ranking e.g. the top and bottom thirds of the same village. Since the rankings are relative, they may be rather difficult to use across a set of widely differing communities, some of which are overall more prosperous than others.

Sub-sampling
The last paragraph hints at one way of choosing sub-samples for later phase, more detailed work, the "table legs" of the metaphor used in section 4. A result of the broad, shallow "table top" study - maybe a baseline study - could be a ranking or ordering of primary study units such as communities, and it would then be plausible to select a purposive sample to represent quantiles along the range of variation found.

Stratification for Comparing Groups
Another approach to site selection would arise if the baseline study had classified rather than ranked the primary units. For example, villages might be classified as Near/Remote from a metalled road, their land as mostly Flat/Steeply Sloping, their access to irrigation water as Good/Poor - three stratification factors each at two crudely defined levels. The 8 possible combination characterisations such as [Near, Flat, Good] suggest we might have 8 sub-samples if possible.

If that is too many to handle a suitable selection of 4 permits each factor to appear twice at each level e.g. [Near, Flat, Good], [Near, Sloping, Poor], [Remote, Flat, Poor] and [Remote, Sloping, Good]. For purposes of comparison across levels this provides some "representativeness" especially if the chosen villages are reasonably selected with respect to other characteristics. This idea is developed further in SSC (2000, op.cit.).

Targeted Sampling

The processes described above are mainly concerned with ensuring that the sample selected can be justified on the basis of being representative. In some cases the aim is to target, exclusively or mainly, special segments of the general population e.g. members of a geographically dispersed socio-economic or livelihood subgroup. The problem is that there is not a sampling frame for the target group and we are never going to enumerate them all, so methods are based on *finding* target population members. There are several approaches to doing this.

General population screening
If the target population is a reasonably big fraction of the overall population, and if it is not contentious or difficult to ascertain membership, it may be possible to run a relatively quick screening check that respondents qualify as target population members e.g. "Are there any children under 16 living in the household now?" As well as finding a sample from the target population, this method will provide an estimate of the proportion which the target population comprises of the general population, so long as careful records are kept of the numbers screened. If the target population is a small proportion of the whole, this method is likely to be uneconomical.

Snowball sampling
The least formal method of those we discuss is "snowball" sampling. The basis of this is that certain hard-to-reach subgroups of the population will be aware of others who belong to their own subgroup. An initial contact may then introduce the researcher to a network of further informants. The method is asserted to be suitable in tracking down drug addicts, active political dissidents and the like. The procedure used is serendipitous, and it is seldom possible to organise replicate sampling sweeps. Thus the results are usually somewhat anecdotal and convey little sense of how completely the subgroup was covered. One interesting account is in Faugier and Sargeant (1997) [9].

Adaptive Sampling
This relatively new method is discussed in Thompson (1992, op. cit.) and allows the sampling intensity to be increased when one happens upon a relatively high local concentration of the target group during a geographical sweep such as a transect sample. It provides some estimation procedures which take account of the differing levels of sampling effort invested, and is efficient in targeting the effort. Until now this method has been developed primarily for estimating the abundance of sessile species and it is not yet in form suitable for general use with human populations. It does not carry any suggestion of networking through a succession of connected informants and is not a straightforward route to formalising snowball sampling.

Protocol-derived Replicated Sampling

In conclusion we offer a possible solution to the targeted sampling problem. The combination of ideas, and the suggestion to use it in the development setting make this solution novel in the sense of being untried. It clearly needs further development through practical application. The notion of replicated sampling is discussed by Kalton (1983) [10] and is highly adaptable as a basis of valid statistical inference about a wider population[1].

We need to combine that idea with two other notions introduced here before using that of replication. The first is the idea of developing a prescriptive sampling protocol to be used in the field as a means of systematic targeting, say of particular households.

The protocol prescribes in detailed terms how to reach qualifying households in practice. As an example, suppose our target comprises "*vulnerable, female-headed rural households*" in a particular region. This involves sorting out all necessary procedural details. One element thereof might concern interviewing key informants at primary unit level, e.g. NGO regional officers - maybe presenting them with a list of twelve areas within the region and getting them to agree on two areas where they are sure there is a high number of target households. There would be numerous procedural steps at several hierarchical levels. In the preceding example, the use of key informants is just an example; it is not an intrinsic part of every such protocol.

Samples are often derived in some such manner: they get at qualifying respondents cost-effectively, but the method usually carries overtones of subjectivity, and of inexplicit individual preference on the part of the selector. The protocol is supposed to address these difficulties. Naturally its development is a substantial process involving consultation, some triangulation, and pilot-testing of its practicability. It is thus a specially developed field guide which fits regional circumstances and study objectives, incorporating e.g. anthropological findings, local knowledge, and safeguards against fraud and other dangers. The protocol is a fully-defined set of procedures such that any one of a class of competent, trained fieldworkers could deliver a targeted sample with essentially interchangeable characteristics.

The second added notion is that if the protocol development involves appropriate consultation, brainstorming and consensus building, then the protocol can be used to define the *de facto* target population being reached. Developers of the protocol can effectively sign up to (i) accepting a term such as "*vulnerable, female-headed rural households*" as the title of the population who are likely to be sampled during repeated, conscientious application of the protocol, and to (ii) accepting that the population so sampled is a valid object of study, and a valid target for the development innovation(s) under consideration in the particular locale for which the protocol is valid.

[1] The original idea concerned a standard quantitative survey, probably with a complication such as multi-stage structure. If this could be organised as a set of replicates - miniature surveys, each with identical structure - then an estimate of some key measure could be derived from each one and that set of estimates treated just as basic statistics treats a simple random sample of data. The replicate-to-replicate standard error would incorporate the whole set of complexities within the stages of each miniature survey and we would get an easy measure of precision of the final answer.

Repeated application of the procedure would produce equivalent "replicate" samples. These carry some "statistical" properties, provided that (i) the sampling is regulated as described above, and (ii) the information collection exercise within any given replicate is standardised. When the procedure is replicated, it is necessary that at least a common core of results should collected in the same form, and recorded using the same conventions, in each replicate and it is for these results that we can make statistical claims.

For example, suppose we record the proportion (x) of respondents within a replicate sample who felt their households were excluded from benefits generated by a Farmers' Research Committee in their community. The set of x-values from a set of replicate samples from different places now have the properties of a statistical sample from the protocol-defined population. Even though the protocol itself encompassed various possibly complicated selection processes, we can, for example, produce a simple confidence interval for the general proportion who felt excluded.

The important general principle which follows from this is that if we can summarise more complicated conclusions (qualitative or quantitative) instead of a single number x, from each replicate, then we can treat the set as representing, or generalising to, the protocol-defined population. There are interesting ways forward, but the practical development and uptake of such a notion poses "adaptive research" challenges if the concept is put to use in the more complex settings of qualitative work in developing countries.

References

1. DFID SLSG (2000) *Sustainable Livelihoods Guidance Sheets*. DFID, London.
2. Marsland et.al. (in press) A Methodological Framework for Combining Quantitative and Qualitative Survey Methods, *Socio-Economic Methodologies for Natural Resources Research – Best Practice Guidelines*, Chatham, UK: Natural Resources Institute.
3. Thompson, S.K. (1992) *Sampling*. Wiley-Interscience, New York
4. Levy, P.S. and Lemeshow, S. (1999, 3rd ed.) *Sampling of Populations: Methods and Applications*. Wiley-Interscience, New York
5. SSC (2000) *Some Basic Ideas of Sampling*. Booklet in series written for DFID.
6. Lindsey, J.K. (1999) *Revealing Statistical Principles*. Arnold, London.
7. Lemeshow, S. et al. (1990) *Adequacy of Sample Size in Health Studies*. WHO/Wiley.
8. Cook, T.D. & Campbell, D.T. (1979) *Quasi-experimentation: Design and Analysis Issues for Field Settings*. Houghton-Mifflin, Boston..
9. Faugier, J. and Sargeant, M. (1997) Sampling hard to reach populations. Journal of Advanced Nursing, vol. 26, pp. 790-797.
10. Kalton, G. (1983) *Introduction to Survey Sampling*. Sage: Quantitative Applications in the Social Sciences.

Theme Paper 3

ANALYSIS APPROACHES IN PARTICIPATORY WORK INVOLVING RANKS OR SCORES

Savitri Abeyasekera, Statistical Services Centre, University of Reading

ANALYSIS APPROACHES IN PARTICIPATORY WORK INVOLVING RANKS OR SCORES

Introduction

In participatory work, ranking and scoring exercises are often conducted to elicit farmers' views on issues specific to study objectives. Some typical examples are as follows.

(a) Identifying the most important constraint to crop production
(b) Determining which pests are mainly responsible in bean yield losses
(c) Prioritising uses of farmer-managed water bodies within a given community
(d) The main problem(s) faced by farmers during storage of maize and cassava
(e) Criteria of greatest importance to the farmer when evaluating pigeonpea varieties
(f) Assessing which elements of policies or institutional arrangements cause most problems for small traders.

In each of the above examples, the objective is to determine or identify one, or a few items, as being the most important. This list may be an open list (e.g. different farmers using different criteria in variety evaluations), or a closed list (e.g. a specified list of water bodies as in (c) above). Issues of interest are often investigated separately for men and women farmers with the aim of determining whether there are gender differences in their opinions. If the study is conducted in several regions, or within different socio-economic groups, it is often also of interest to evaluate whether differences exist between regions and/or groups.

In studies of this nature, the sampling of (say) farmers in an appropriate fashion is crucial so that results from the **sample** of farmers can be generalised to a wider **population** of farmers. The latter target population has to be identified clearly at the outset. Issues concerning sampling are elaborated in an accompanying parallel document entitled "Sampling and qualitative research". Here, we will assume that sampling aspects have been correctly dealt with and are in accordance with study objectives.

In this document, we discuss advantages and limitations in the use of ranking and scoring exercises in participatory rural appraisal (PRA) work. We describe ways in which information gathered from such exercises can be analysed, keeping in mind the study objectives. These vary from simple descriptive procedures to more advanced methods of statistical analysis. Examples provide brief illustrations of the methodological approaches and show where assumptions are needed.

Ranks or Scores?

Preference ranking has been a popular tool in PRA activities for over a decade (Bayer, 1988; Chambers, 1988). The general argument for using ranks rather than scores is that farmers are better able to judge, from a given list of items labelled as (say) A, B, C,..., whether one item is better or worse, more or less important, than another item. The main difficulty in using ranks is the following. Say for example that one farmer gives a higher rank for item E than item B, and another farmer does exactly the same.

However, the reality may be that the first farmer thought E was only slightly better than B while the second thought E was a lot better than B. In other words, it is not possible to attribute a "distance" measure to differences between numerical values given to ranks.

Ranks only represent an ordering of a list of items according to their importance, as viewed by the respondent, for the particular issue under consideration. In interpreting ranks, it is important to keep this in mind. Thus ties should be allowed, i.e. permitting two or more items to occupy "equal" positions in the ordered list. This is because the researcher is only eliciting an ordering. It is unrealistic to oblige the farmer to make a forced choice between two items if she/he has no real preference for one over the other. Each item involved in a tie should be given the average value of ranks that would have been allocated to these items had they not been tied.

Thus for example, suppose six items A, B, C, D, E and F are to be ranked. Suppose item B is said to be the best; item C the worst; item F second poorest but not as bad as C; and the remaining items are about the same. Then the ranks for the six items A, B... F should be 3, 1, 6, 3, 3, 5. This set of ranks is obtained by using 3 as the average of ranks 2, 3 and 4, i.e. the ranks that items A, D and E would have got if the respondent had perceived some difference in these items. One reason for using the full range from 1 to 6 rather than ranking the items as 2, 1, 4, 2, 2, 3 is that mis-leading results will otherwise be obtained in any further data summaries which combine information across farmers. Fielding *et al* (1998) give a fuller discussion on the advantages of using ties sensibly.

Scores are typically used in situations where criteria of importance in evaluating a series of items (e.g. pest-tolerant bean varieties) are first identified. The items are then scored with respect to each criterion in turn. This is what is essentially referred to as matrix scoring (Pretty *et al*, 1995). It is common to use scores from 1 to 5 although a wider range should be attempted where possible since the data are then more amenable to standard statistical methods of analysis.

It is sometimes argued that scores have the disadvantage in being appropriate only when dealing with numerate respondents. However, if a fixed number of seeds or beans are given to the respondent with a request to allocate more seeds for the preferred items and fewer for the less favoured items, then reasonably accurate results should result even with non-numerate respondents.

The major advantage of scores over ranks is that they are numerically meaningful. Differences between scores given to different items show the strength of the preference for one item over the other. Scores provide a ranking of the items but also something extra - a usable distance measure between preferences for different items. The availability of a meaningful measurement scale makes scores more suitable for use in statistical analysis procedures, so scores are a good deal easier to summarise across respondents.

Various forms of scoring methods exist (Maxwell and Bart, 1995). Fully open scoring, where each item to be scored can be given any value within a particular range (say 1-5) is the most flexible since it leads to observations that are "independent" of each other – a requirement for most simple statistical analysis procedures. However, restricted forms of scoring, where a fixed number of "units" (e.g. bean seeds, pebbles) are distributed between a set of items, has to be used with some care.

Say, for example, that a farmer is asked to allocate 10 seeds among four bean varieties, giving more seeds to the variety she/he most prefers. If the farmer allocates 5 seeds to variety C and 3 seeds to variety A because C and A are the varieties she/he most prefers, then the farmer is left with just 2 seeds to allocate to the remaining varieties B and D. This is a problem if she/he has only a marginal preference for A over B and D. Such a forced choice suffers the same difficulties faced by ranks in not being able to give a meaningful interpretation to the "distance" between the scores. A larger number of seeds, say 100, partially gets over this problem because now the farmer has more flexibility in expressing his/her strength of preference for one variety over another. One hundred seeds may however still be too few if the number of cells into which the seeds must be distributed is large (e.g. in matrix scoring exercises).

Do Ranks and Scores Provide Consistent Responses?

If an ordering of a list of items is all that is required, there will be little need for using scores instead of ranks. Researchers may however be interested in using both forms of enquiry for purposes of triangulation. Consistent results across different forms of enquiry can provide greater reliance on the findings, while contradictory evidence can give useful insights concerning the issue under consideration (Moris and Copestake, 1993). Using both forms of enquiry can also be useful in a pilot study so that the more suitable one can be used in a subsequent larger study.

We provide an example below to illustrate some insights gained in a study involving both ranks and scores. The example is drawn from the Farming Systems Integrated Pest Management (FSIPM) Project carried out in Malawi. It concerns a study conducted in four villages within the FSIPM Project area to determine the status of existing formal and informal networks of communication and their potential utilisation in the dissemination of IPM strategies (Lawson-McDowall and Kapulula, 1999). In each village, focus group discussions were held, separately for men and women, limiting each group to a maximum of 10 members in order to encourage all members to take part. One component of the study involved asking participants both to rank and to score sources of information about agriculture from a list of seven such sources elicited from participants in a previous round of discussions (see Table 1). The ranking exercise took place first using pictures of the sources to help participants identify different sources of communications. Subsequently they were then asked to score, on a 1-5 scale (5 being the most popular source), each of these sources of information so that the results could be quantified.

Comparisons are straightforward if the methods give exactly consistent results. However, there is no exact method that can be used to determine whether ranking and scoring leads to *approximately* the same results. An ad-hoc method can be adopted to put the ranks on a similar footing to the scores. We allocated a value of 7 to items ranked first (since 7 was the maximum number of items ranked), a value of 6 to items ranked second and so on, with a value of 0 to any items omitted during the ranking. This automatically overcomes the question of what should be done with tied ranks. Table 1 summarises the average scores (means) across the four villages, and the averages (again over villages) of values *allocated* to the farmers' ranks. Note that the means based on farmers' scores are given in increasing order in the first column of numerical values in Table 1. In the next column, the ordering of the mean values based on farmers' ranks is indicated by a superscript.

Table 1. Average values based on scores and ranks given by groups of farmers

Women	Mean based on farmers' **scores**	Mean based on farmers' **rank**	**Men**	Mean based on farmers' **scores**	Mean based on farmers' **rank**
FSIPM Project	5.0	4.8^{3}	Radio	4.3	6.3^{1}
Radio	3.7	6.0^{1}	Friends	4.3	4.8^{3}
Friends	3.0	5.5^{2}	Extension worker	4.0	4.5^{4}
Written material	2.7	2.8^{7}	Other places	4.0	3.7^{5}
Extension worker	1.8	$4.3^{4.5}$	Written material	3.3	6.0^{2}
Own experimentation	1.8	$4.3^{4.5}$	FSIPM Project	3.0	0.3^{7}
Other places	1.5	3.5^{6}	Own experimentation	1.7	2.3^{6}

The findings from the scoring and ranking results are not entirely consistent, particularly with the male groups. This example illustrates that farmers may be basing their responses on different criteria. There was some indication that they may have been using ranking as a tool to indicate the importance that different sources of information should have (for example in disseminating FSIPM findings), and scoring to indicate experience. This was plausable in Kambuwa with the men's discussion group where the extension worker was ranked second but given a score of 1/5 because there was at the time no Field Assistant for this section. Similarly, in Magomero, the women's group gave radio second place in the ranking but only 2/5 as a score since not enough people have radios for this to be a practical source of information.

In general practice, it would not be a good idea to carry out both ranking and scoring exercises with the same group at the same time for the same purpose. Firstly, it takes twice as long and secondly, there is the danger that farmers may be confused by the two concepts, in view of 1 being the best rank but the worst possible score.

The example above suggests that there may have been a communication difficulty which would not have been noted without methodological triangulation. For example, is the question being asked "Which is the most *important* method through which you learn about agricultural innovations?" or "What do you think is the most *appropriate* method that we can use to disseminate to you information about findings from the FSIPM Project?" Whichever method is used, it is important to have a very clear cut definition beforehand of what is being requested of the farmers during the ranking or scoring exercise. Translation of English into the locally spoken Chichewa may also have been an issue here, combined with enumerator misunderstandings concerning the actual question of interest.

Some Statistical Analysis Procedures

There is no attempt here to give a full coverage of analysis methods. Some situation specific examples are given and statistical analyses appropriate for these situations are described. In general, choosing the most appropriate form of analysis depends on the objectives of the study. For example, the main objective of a wealth ranking exercise may be to divide households into different strata. Grandin (1994) describes in detail how ranks given to households by four key informants can be converted to scores and then summarised and interpreted so that the strata are identified.

Whatever the objective, it is usually advisable to begin by thinking carefully about the data structure and then to produce a few simple graphs or summary statistics so that the essential features of the data are clear. Often this form of summary may be all that is needed. In section 4.1 below, we discuss data structures arising from ranking and scoring exercises and simple methods of presentation. This is followed in sections 4.2 and 4.3 by a brief discussion of more advanced methods of analysis applicable to ranking and scoring data. Poole (1997) provides fuller coverage.

Simple methods of summary

A typical example that arises in farmer participatory work concerns studies where the aim is to identify farmers' assessment of the "best" or "most important" item from a list of items. Some researchers address this aim by asking farmers to allocate a number of counters (e.g. pebbles, seeds), say out of a total of five counters per item, to indicate their views on the importance of that item. The number allocated then provides a score, on a 1-5 scale with 1 being regarded as being "worst" or "least important". Other researchers request only that each farmer place the items in rank order. In either case, the number of items presented to the farmer (or farmer group) may be a fixed number (e.g. new varieties tested on farmers' fields), or it may vary from farmer to farmer. For example, each farmer may be asked to list, and then rank, each known pest problem affecting his/her bean crop.

The data structures that result from ranks/scores given to a fixed number of items, are shown using fictitious data in Tables 2 and 3. Although the numerical values shown look similar, a number of essential differences, as highlighted in section 3, must be recognised in interpreting this information. For example, mean values for columns in Table 2 are meaningful summaries and give some indication of the most serious pest problem. However, producing column totals for the information given in Table 3 assumes a common "distance" between any two consecutive ranks. This is particularly problematical if there are missing cells in the table, because some respondents have not ranked some items.

Table 2. An example data set showing scores given by farmers to four items, i.e. an identified set of pests attacking beans

Farmer	Ootheca	Pod borers	Bean stem maggot	Aphids
1	4	2	1	2
2	5	4	1	3
3	4	1	2	1
4	4	5	1	4
5	1	2	1	1
6	1	4	1	2
7	5	1	1	5
8	2	5	5	3
Mean	3.3	3.0	1.6	2.6

Table 3. An example data set showing ranks allocated (without allowing ties) by five groups of farmers to four pigeonpea varieties

Farmer Group	ICEAP 00040	ICEAP 00053	ICP 9145	Local
1. Mixed	2	3	4	1
2. Women	1	4	2	3
3. Men	1	2	4	3
4. Mixed	3	4	1	2
5. Mixed	2	3	1	4

If an overall ranking of the varieties in Table 3 is required across all farmer groups, a simple procedure is to give varieties receiving a rank 1, 2, 3 and 4, corresponding scores of 4, 3, 2, 1, and then to average the resulting set of scores. This gives mean scores of 3.2, 1.8, 2.6 and 2.4 respectively for varieties ICEAP 00040, ICEAP 00053, ICP 9145 and Local. Of course, some would argue that this will lead to exactly the same conclusions as would be obtained if the numerical values given to each rank were averaged across farmer groups. This is true, but the conversion to scores makes the assumptions involved in the averaging process clearer, i.e. the assumption that the degree of preference for one item over the next within an ordered list is the same irrespective of which two neighbouring ranks (or corresponding scores) are being considered, and the assumption that a missing value corresponds to a numerical zero score.

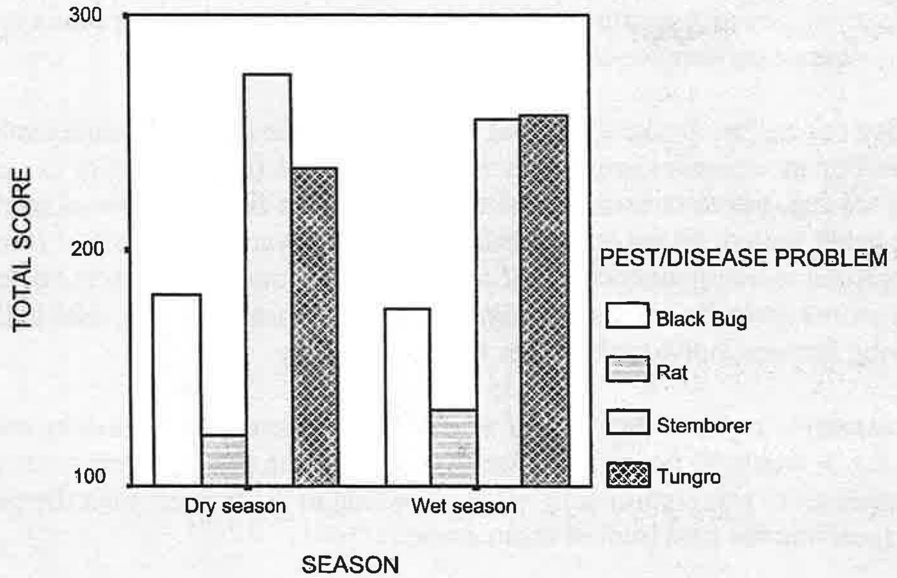
Alternative transformation scales could also be considered on the basis of comments made by those who rank the items. For example, ranks 1, 2, 3, 4 might be converted to scores 9, 5, 2, 1 respectively, if it were apparent during discussions with respondents that there is a much clearer preference for items at the top of the priority scale than those lower down.

Consider now a situation where the specific items being ranked vary from farmer to farmer. Again, a transformation from ranks to scores can be made with a zero score for items omitted during the ranking. See Abeyasekera *et al* (2000) for details. Box 1 provides a further example aimed at investigating whether a disease, known to cause high yield losses in a crop, is regarded by farmers as being the most damaging to the crop compared to damage caused by other pests and diseases. The results are shown in Figure 1. The main point to note in this example is the difficulty of summing the ranks given to pest/disease problems since many problems had been mentioned by the 226 farmers. If one problem was mentioned by only a few farmers, but ranked highest by them, its average rank, based on just those farmers, would be unrepresentative of the farming population as a whole. Scores on the other hand, include a meaningful zero, so averages work out correctly.

Box 1. An example drawn from Warburton (1998) illustrating a descriptive summary of ranks given to the three most important pests and diseases

In a study concerning farmers' practices, experiences and knowledge of rice tungro disease, 226 farmers in the Philippines were asked to name and rank the three most damaging pests or diseases affecting their rice crop. Each pest and disease named by a farmer was given a score of 3, 2 or 1 according to whether they ranked the pest or disease as being the first, second or third most damaging in their attack on the crop. A score of zero was given to any pest/disease not mentioned by the farmer. These scores were *totalled* over all farmers and are shown in Figure 1 below, for each of two seasons, for pests/diseases receiving the four highest overall scores. Tungro is clearly a recognised problem but in the dry season, stemborer is identified as being the greater problem.

Figure 1. Farmer perceptions about four pests/diseases



Statistical procedures for analysing farmer evaluation scores

Consider the researcher who wishes to determine whether observed differences, in mean values of scores given by farmers to different items, signal real differences in preferences for these items among farmers in the target population. This is a situation where generalisability of study results is of primary concern. Such a question can be answered by applying appropriate statistical techniques to a data structure of the form shown in Tables 2. But first, the overall data structure must be recognised and the objectives stated more explicitly. For example, consider a situation where many more than 8 farmers provided importance scores for the four pests listed in Table 2. Suppose farmers are grouped by gender and by village. Possible questions that may be posed are:

1. Do the data demonstrate evidence that farmers regard one pest as being more problematical than another?
2. Is there evidence to demonstrate that male and female farmers differ in their overall perceptions of pest problems?
3. Are pest problems more severe in one village compared to another as perceived by sampled farmers?

Analysis of variance (*anova*) procedures, involving statistical modelling of the data, enable the above questions to be answered. The procedure allows a comparison of pest mean scores, after making due allowance for possible farmer to farmer differences. (See Box 2 for an example).

Assumptions associated with the *anova* technique must also be kept in mind. The method assumes that (a) the farmer evaluations are independent of each other; (b) scores given to each pest problem come from populations of scores that have a common variance; and (c) that the data follow a normal distribution. Of these, (c) is the least problematical and can be mitigated if there are a sufficient number of farmers or farmer groups involved (see an accompanying document "Sampling and qualitative research" for some guidance on sample sizes).

Assumption (b) can be checked via what statisticians refer to as a residual analysis. Data collection procedures should ensure that assumption (a) is met. For example, restricted scoring, where farmers are asked to distribute a fixed number of seeds among the items being scored, do not lead to independent observations. Farmers' responses can be regarded as being independent if scores given by one farmer do not directly influence scores given by another farmer. This can be ensured during data collection by interviewing farmers individually rather than collectively.

If *anova* assumptions hold, then further analysis subsequent to the two-way *anova* is possible, e.g. it would be possible to ascertain whether the pest problem receiving the highest mean score was significantly more important to the farmers than the pest problem receiving the next highest mean score.

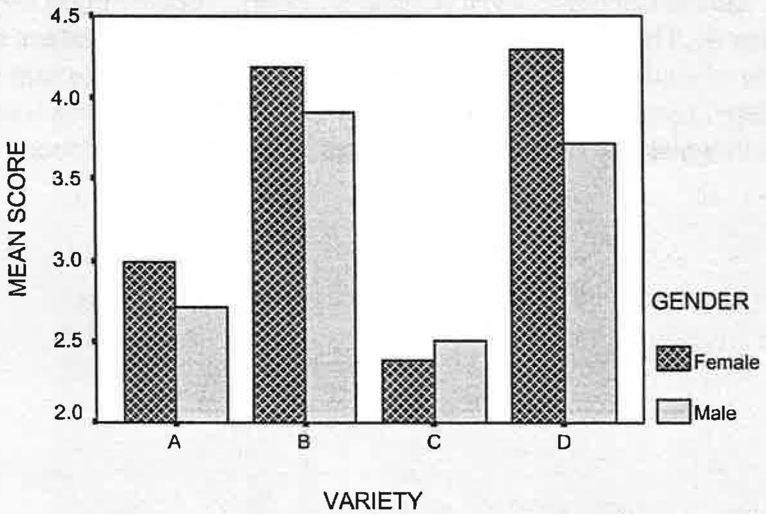
Box 2. An example illustrating results from a statistical modelling exercise to compare four sweet potato varieties for their tolerance to *cylas puncticollis*

As part of an integrated pest management project in Malawi, four sweet potato varieties were studied in an on-farm trial for their tolerance to sweet potato weevils (*Ophiomyia* spp). One component of the variety evaluation involved asking twenty farmers to score separately, on a 1-5 scale (1=least tolerant; 5 = most tolerant) their assessment of each variety with respect to its tolerance to weevils. All farmers scored all four varieties resulting in 80 observations from 20 farmers. However, farmers were not divided equally by village or by gender, as is seen in the frequency table below. Modelling the data is still possible.

Village	Pindani	Chiwinja	Lidala	Maulana	Total
Female	5	1	2	2	12
Male	2	4	3	1	8

The modelling analysis demonstrated that there were no significant gender differences or differences between villages with respect to farmers' views on pest tolerance by the different varieties. There was however strong evidence ($p<0.001$) to indicate that two varieties, A and C, were regarded as being more susceptible than varieties B and D. Figure 2 shows some results. The modelling enables a valid comparison between male and female farmers, between villages and between varieties, despite the unequal replication of male and female farmers across the four villages. The modelling also allowed varieties to be compared "free" from possible effects due to gender and village differences, i.e. we excluded the alternative explanations that these effects might account for, or mess up, the picture of varietal differences.

Figure 2. Farmer scores given to four varieties for their tolerance to sweet potato weevils



The general technique used here is quite powerful and is based on an underlying general theory that can be applied even if some data are missing or if the number of items being evaluated differ across farmers. Appropriate statistics software (e.g. SPSS, Genstat, Minitab) is needed to deal with such data structures and other complexities. More advanced methods of analysis do exist when large numbers of farmers are involved. For example, the frequency of farmers giving different ranks to each item can be modelled via use of proportional odds models (Agresti, 1996). The interpretation of results is then based on the odds of farmers preferring one item compared to another item (see Box 3 for an example).

4.3 Statistical procedures for analysing farmers' ranking of several items

We consider here the type of data shown in Table 3, but restricted to situations where such data are available for a larger number of farmers or farmer groups. Sometimes, a simple summary may be all that is needed. For example, in a study in India, aimed at investigating the potential for integrating aquaculture into small-scale irrigation systems managed by resource-poor farmers, 46 farmers were asked to rank four different uses of water bodies according to their importance (Felsing *et al* 2000). The data summary shown in Table 4 clearly indicates irrigation as the primary use of water bodies. Thirty-four farmers (74%) rank this as the most important use of water.

Table 4. Number of farmers giving particular ranks to different uses of water bodies

Rank	Irrigation	Livestock consumption	Household use	Clothes washing
1	34	6	5	1
2	8	16	14	8
3	1	16	14	15
4	3	8	13	22
N	46	46	46	46
R_i	65	118	127	150

A more formal statistical test is available to demonstrate that farmers' water uses differ significantly. This is Friedman's test (Conover, 1999). A description of this test is provided in Box 4. The test demonstrates that the water uses ranked are not all perceived to be of equal importance. Further tests showed evidence that irrigation was indeed considered more important than other uses, and that there was insufficient evidence to distinguish the remaining three uses in terms of importance.

Box 3. An example drawn from Poole (1997) to illustrate the application of proportional odds modelling of frequency data summarising farmers' evaluation scores

In a farmer participatory study in Kenya concerning six tree species, a local board game, bao, was used to enable farmers to allocate 1-5 beans (1=poor; 5=good) to each species on the basis of each of six criteria, i.e. browsing resistance, drought resistance, termite resistance, growth, survival and biomass (Franzel *et al*, 1998). The mean scores across the six criteria was calculated for each farmer and each species and rounded to the nearest integer – call this a species performance score. A summary appears in the table below in terms of the number of farmers who give performance scores 1-2, 3, 4, 5 for each species. (Note: Scores of 1 and 2 were combined since very few farmers gave a score of 1).

	Performance Score				
Tree species	1 - 2	3	4	5	No. of farmers
Grevillea R.	3	8	22	11	44
Casuarina	24	13	3	2	42
Calliandra C.	0	4	23	18	45
Leucaena L.	2	14	22	6	44
Leucaena D.	2	12	27	3	44
Eucalyptus	6	21	16	2	45

The aim is to determine whether there is a preference for one tree species over another. Application of a chi-squared test is not strictly appropriate because (a) the scores given to each species are not independent (all farmers score all six species); and (b) the scores have an ordinal structure. The latter concern is overcome by subjecting the data to a proportional odds model. The modelling also allows other socio-economic factors concerning the farmers (e.g. wealth status, gender) to be taken into account, although this was not done in this study.

The results of the analysis are presented below in terms of the odds of a farmer scoring a species the same as Eucalyptus or worse. Comparisons with *Eucalyptus* were relevant since this was a well-known tree in the area and provided a "control" with which other species could be compared. The results demonstrate that *Casuarina* performs significantly worse than *Eucalyptus* in respect of farmers' evaluations. The odds (chance) of *Casuarina* scoring low is about 8 times higher than for *Eucalyptus*.

	Odds ratio*	Sig. Prob.
Grevillea R. versus *Eucalyptus*	0.20	< 0.001
Casuarina versus *Eucalyptus*	8.19	< 0.001
Calliandra C. versus *Eucalyptus*	0.08	< 0.001
Leucaena L.. versus *Eucalyptus*	0.38	0.0156
Leucaena D.. versus *Eucalyptus*	0.39	0.0184

* The odds ratio provides a comparison of the odds of a score lower than or equal to j (j=1-2,3,4) with the corresponding odds for Eucalyptus.

If ties occur in the data, an adjustment to the Friedman's test statistic is available. However, two other associated problems need consideration. The first is that the test is based on an approximation to the chi-squared distribution. If *n* is the number of farmers and *k* is the number of items being ranked, then *nk* must be reasonably large, say ≥ 30, for the approximation to be reasonable. Secondly, the test makes no allowance for missing data. If missing data do occur, because (say) the number of items ranked varies from farmer to farmer, then ranks need to be converted to scores in some fashion (Abeyasekera et al, 2000) and then analysed using procedures described in 4.2 above.

Box 4. An example drawn from Felsing *et al* (2000) to illustrate Friedman's test procedure for analysing ranks.

Friedman's test can be applied to ranks given by a number of farmers (say n) to each of a specified number of items (say k). For the data shown in Table 4, n=46 and k=4.

First calculate the sum of ranks (Rj) for item j (j=1 for irrigation, j=2 for livestock consumption, etc. Thus for data in Table 4,

$$R_1 = (1 \times 34) + (2 \times 8) + (3 \times 1) + (4 \times 3) = 65$$

Next calculate the Friedman's test statistic X^2 given by

$$X^2 = \frac{12}{nk(k+1)} \sum_{j=1}^{k} \left[R_j - \frac{n(k+1)}{2} \right]^2$$

$$= \frac{12}{(46)(4)(5)} \left[\left(65 - \frac{(46)(5)}{2} \right)^2 + \ldots + \left(150 - \frac{(46)(5)}{2} \right)^2 \right]$$

$$= \frac{12}{(46)(4)(5)} \times 3878 = 50.58$$

Compare $X^2 = 50.58$ with tabulated values of a chi-squared distribution with (k-1) degrees of freedom. In this example, the test statistic is highly significant ($p < 0.001$), and we can conclude that farmers' views with regard to the importance of the uses of the four water bodies differ significantly.

Some Final Remarks

It is important to recognise that proceeding beyond straightforward data summaries and graphical presentations to formal statistical procedures and tests of significance has little value in helping research conclusions if the sample size is an inadequate representation of the farming community. Here the sample size refers to the number of independent assessments obtained, either by interviewing farmers individually or by having separate discussions with a number of farmer groups. How large a sample is needed will depend on the specific objectives behind eliciting farmer opinions via ranks or scores. The parallel theme paper entitled "Sampling issues in participatory research" provides some useful guidelines.

In situations where the sample size is adequate and the sample has been appropriately chosen to represent the target population of interest, the application of statistical methods will provide greater validity to research conclusions. This theme paper provides some guidance about suitable statistical analysis approaches for use in participatory work that gives rise to ranks and/or scores. However, the exact approach for a particular study will be closely associated with the study objectives and other data collection activities (e.g. those related to on-farm trials). It would therefore be desirable if possible to make decisions concerning sampling and analysis aspects in consultation with a statistician experienced in survey applications.

References

Abeyasekera, S., Lawson-McDowall, J., and Wilson, I.M. (2000). Converting ranks to scores for an ad-hoc procedure to identify methods of communication available to farmers. Case-Study paper for project on "Integrating qualitative and quantitative approaches in socio-economic survey work"

Agresti, A. (1996). *An introduction to Categorical Data Analysis*. Wiley, New York.

Conover, W.J. (1999) *Practical non-parametric statistics*. 3rd Edition. Wiley, New York.

Felsing, M., Haylor, G.S., Lawrence A. and Abeyasekera S. (2000). Evaluating some statistical methods for preference testing in participatory research. Submitted for publication.

Fielding, W.J., Riley, J., and Oyejola, B.A (1998). Ranks are statistics: some advice for their interpretation. PLA Notes 32. IIED, London.

Franzel, S., Ndufa, J.K., Obonyo, I.C. and Coe, R. (1998). Farmer-Agroforestry Trials: Farmers' experiences in Western Kenya. Working Paper, ICRAF, Nairobi.

Grandin, B.E. (1994). Wealth Ranking. Chapter 2 in *Tools for the Field: Methodologies Handbook for Gender Analysis in Agriculture*. Edited by Feldstein, H.S. and Higgins, J. IT Publications Ltd.

Lawson McDowall, J and Kapulula, P.K. (1999). Networks of Communication: How farmers learn about innovations in agriculture. FSIPM Project Mimeo.

Maxwell, S., and Bart, C. (1995). Beyond Ranking: Exploring Relative Preferences in P/RRA. PLA Notes 22, pp. 28-34. IIED, London.

Moris, J. and Copestake, J. (1993). *Qualitative Enquiry for Rural Development.* IT Publications Ltd.

Poole, J. (1997). Dealing with ranking/rating data in farmer participatory trials. MSc thesis. University of Reading.

Pretty J.N., Gujit I., Scoones I. and Thompson J. (1995). IIED Participatory Methodology Series. Int. Institute for Environment and Development, London.

Warburton H., Villareal, S. and Subramaniam, P. (1998). Farmers' rice tungro management practices in India and Philippines. Proceedings of a workshop on Rice Tungro Management. IRRI, Philippines, 9-11 November 1998.

Acknowledgements

Valuable comments by Ulrich Kleih, Ian Wilson and Roger Stern on initial drafts of this document are very much appreciated.

Theme Paper 4

CONVERTING RANKS TO SCORES FOR AN AD-HOC ASSESSMENT OF METHODS OF COMMUNICATION AVAILABLE TO FARMERS

S. Abeyasekera[1], Julie Lawson-McDowall[2], and Ian Wilson[3]

[1] Statistical Services Centre, University of Reading
[2] Farming Systems Integrated Pest Management Project, Malawi
[3] Statistical Services Centre, University of Reading

CONVERTING RANKS TO SCORES FOR AN AD-HOC ASSESSMENT OF METHODS OF COMMUNICATION AVAILABLE TO FARMERS

Introduction

The Social Anthropology Team of the FSIPM project conducted an investigation of existing formal and informal networks of communication among farmers and the potential utilisation of these methods in dissemination of recommended IPM strategies (Lawson-McDowall *et al*, 1997). One component of this study involved focus group discussions (one for men, one for women), in each of the four FSIPM Project villages where farmers views were sought concerning existing methods of communication. The groups were limited to a maximum of ten members in order to encourage all members to take part. Table 1 shows the distribution of the numbers involved in the discussion from each of the villages. Of the 66 farmers involved, 55 were farmers who were participating in on-farm experimental trials.

Table 1. Farmers participating in farms group discussions

Village	Men	Women	Total
Lidala	8	10	18
Chiwinja	6	9	15
Magomero	7	7	14
Kambuwa	9	10	19
Total	30	36	66

Ranking Data Collected During Focus Group Discussions

During an initial round of meetings with farmer groups, existing methods of communication amongst farmers and methods by which farmers acquired sources of information about agricultural innovations, inputs, etc were ascertained. In the second round of meetings, these items were presented to each group and they were asked to rank these items, giving a score of 1 to the item which they viewed as being the most effective method for learning about new aspects of agriculture. The ranks as elicited are shown in Tables 2(a) and 2(b).

Table 2(a) Ranks for sources of information regarding agriculture (Women groups)

Information source	Village			
	Lidala	Chiwinja	Magomero	Kambuwa
FSIPM Project	2	1	-	2
Radio	3	2	2	1
Friends	3	3	1	3
Extension Worker	1	7	3	4
Own Experimentation	5	6	1	3
Written Material	-	4	5	4
Other Sources	4	5	4	5

Table 2(b) Ranks for sources of information regarding agriculture (Men groups)

Information source	Village			
	Lidala	Chiwinja	Magomero	Kambuwa
FSIPM Project	-	7	-	-
Radio	4	1	1	1
Friends	2	6	2	3
Extension Worker	1	3	-	2
Own Experimentation	5	4	-	6
Written Material	1	2	1	4
Other Sources	3	5	-	5

A First-stage Summary and its Limitations

Any data summary and possible follow-on analysis must pay close attention to the specific objectives underlying the study. Here the objective may be stated as:

Identifying the most important information sources through which farmers learn about new agricultural innovations.

This objective needs further clarification. Who are the "farmers" to whom this applies? What precisely does "new agricultural innovations" mean? What is the set of "sources" from which the most important are to be selected? We will therefore define the starting position more precisely as follows:

The population of interest are resource-poor farmers growing sustainable food crops (primarily maize) in each of four selected villages in the extension planning areas of Matapwata and Chiradzulu, located within the Blantyre Shire Highlands Rural Development Programme area in Malawi. A pilot study involving focus group discussions with a few farmers in each of these villages, have identified seven sources of information by which they learn about new aspects of agricultural production. Conditional on the assumption that these are the only likely means which enable farmers to learn about methods of enhancing their agricultural produce, the objective of this study is to determine which of these seven sources are regarded by farmers in the area as being the most effective means of learning about agriculture.

With the above objective in mind, the first stage in the summary process involves the conversion of tied ranks to an "average rank". This process is justified and explained in the paper by Fielding et al (1998). The corresponding data appear in Tables 3(a) and 3(b). Women in Lidala for example, give a rank of 3 to both "Radio" and "Friends". These two items jointly occupy 3rd and 4th positions in the ordered list of seven items. They are therefore given an average rank of 3.5. This results in "other sources" being allocated a rank of 5 and "Own experimentation" a rank of 6.

The final columns in Tables below give the mean rank and a ranking of these means. This is the procedure typically followed by those involved in PRA work. But there are limitations in this method as discussed below.

Table 3(a) Ranks for sources of information with ties replaced by average ranks (Women groups)

Information Source	Village				Mean rank	Ranked Mean rank
	Lidala	Chiwinja	Magomero	Kambuwa		
FSIPM Project	2	1	-	2	1.67	1
Radio	3.5	2	3	1	2.38	2
Friends	3.5	3	1.5	3.5	2.88	3
Extension Worker	1	7	4	5.5	4.38	5
Own Experimentation	6	6	1.5	3.5	4.25	4
Written Material	-	4	6	5.5	5.17	6
Other Sources	5	5	5	7	5.50	7

Table 3(b) Ranks for sources of information with ties replaced by average ranks (Men groups)

Information Source	Village				Mean rank	Ranked mean rank
	Lidala	Chiwinja	Magomero	Kambuwa		
FSIPM Project	-	7	-	-	7.00	7
Radio	5	1	1.5	1	2.13	1
Friends	3	6	3	3	3.75	4
Extension Worker	1.5	3	-	2	2.17	2
Own Experimentation	6	4	-	6	5.33	6
Written Material	1.5	2	1.5	4	2.25	3
Other Sources	4	5	-	5	4.67	5

When summaries such as those in the final columns of Tables 3(a) and 3(b) are produced, the main limitation is the assumption that ranks represent a "distance" measure, i.e. a rank of 2 is assumed to be 3 units ahead of a rank of 5; a rank of 4 is also 3 units ahead of a rank of 7. Thus, they are assumed to reflect an interval scale of measurement. However it must be recognised that the degree of preference for an item ranked 1 over an item ranked 2 may not necessarily be the same as the degree of preference for the item ranked 4 over the item ranked 5.

As an example, consider men in Chiwinja village. The item "radio", given rank 1, may be a much more important source of information for the farmers than the item "written material" ranked 2. On the other hand, the difference in importance between ranks 4 and 5 for information sources "own experimentation" and "other sources" may be much smaller. It is also quite likely that in a focus group discussion involving several people, items ranked highest (highest = best = rank 1) are those which are quickly agreed by all as being the most important. Those mentioned by only a few farmers in the group are likely to have a lower rank. If this happens, then a greater weight must be placed on items that have higher ranks.

The second difficulty in the use of rank values 1, 2, 3, ..., for items listed from "best" to "worst", is that the lowest number (rank) is given to the item regarded as "best". In some cases, a small numerical value is intuitively taken as "bad" and a large value taken as "good". Therefore any numerical summary like the average rank can be easily misinterpreted.

The third difficulty lies in the interpretation of the overall ranks for men and women. The non-response by men in 3 of the villages to the FSIPM project as an information source indicates they did not have much association with project activities. This, and other non-responses possibly imply that some farmers were basing their opinions only with respect to sources of information of which they had had some experience. Women on the other hand are usually more keen to be cooperative and generally ranked all items. They may possibly have been giving ranks according to what they thought would be the most effective means of communication if all these sources were available to them.

Note also that all groups did provide a rank for "radio". It is unlikely that all 66 farmers involved in the study had access to a radio. Were they therefore referring to the fact that their neighbour has a radio and thereby they received information directly or indirectly from their neighbour's radio? All men groups ranked "written material". Were all the 30 male farmers who participated literate? These questions are unimportant if we assume that the given ranks were conditional on literacy, availability of a radio, etc.

We now proceed assuming that farmers had a clear understanding that the ranking was to be done in such a way that the highest rank was given to the source that they collectively thought was the most effective means by which they could learn new approaches to agricultural production.

The simple summaries given above indicate that there are gender differences. But are the summary ranks sufficiently inconsistent to declare that men and women have differing opinions? Can this be checked in some quantitative fashion? Even if gender differences are recognised and results therefore presented separately for men and women, how can the "best" methods of learning about agriculture be judged on the basis of the ranks given to the seven items?

In the next section we explain an ad-hoc method that converts ranks to scores which can then be subjected to statistical analysis procedures. We will assume that a second aim is to assess gender differences as well as village differences.

An Ad-hoc Method for Converting Ranks to Scores

Scores are much more informative than ranks and should be used where possible in PRA work (Maxwell & Bart, 1994). However, where ranks are used, some attempt should be made to elicit from the farmer, how much better an item ranked 1 is than an item ranked 2, how much better the item ranked 2 is from the item ranked 3, and so on. In other words, establish some "distance" between successive ranks. This will enable the numerical ranks to be converted to meaningful scores. In the absence of this additional information, some broad assumptions must be made to convert ranks to scores. A general approach, when only ranks are available is described below using the example data set given in Table 2.

The ranks in Tables 2(a) and 2(b) are given for either 3, 6 or 7 items. Let us arbitrarily decide that we will allocate 20 units to each of the ranked items, making the "distance" between the top ranks (say 1 and 2) larger than the "distance" between the lower ranks. There is no clear rationale for this decision except on the grounds that lower ranks are likely to have a less clearer distinction in the farmers' minds than items regarded as having greater importance.

One example of the distribution of 20 units to each of n ranked items (n=3, 4, ..., 7) appears in Table 4 below. This can be regarded as a conversion scale that allows each rank to be converted to a score. Where all 7 items have not been ranked, we give a score of zero to the omitted items to give them a negligible contribution towards subsequent summaries of the scored data.

Table 4. A possible scale for converting ranks to scores

No of items ranked	Conversion scale							
7	Rank	1	2	3	4	5	6	7
	Allocated score	8	5	3	2	1	1	0
6	Rank	1	2	3	4	5	6	-
	Allocated score	8	5	3	2	1	1	0
5	Rank	1	2	3	4	5	-	-
	Allocated score	8	5	3	2	2	0	0
4	Rank	1	2	3	4	-	-	-
	Allocated score	8	6	4	2	0	0	0
3	Rank	1	2	3	-	-	-	-
	Allocated score	8	7	5	0	0	0	0

The conversion scale above was applied to the ranks in Tables 3(a) and 3(b). Where there were ties, we have allocated the average score. Thus a rank of 3.5 occurring with 7 ranked items, was given the average score of 3 and 2, i.e. 2.5. Tables 5(a) and 5(b) show the resulting set of scores for female and male groups.

The scores in Table 4 above have some useful characteristics when producing statistical summaries.

(i) They all add to the same total of 20 so each group is given the same weight however many sources they cited – not obviously possible with ranks.

(ii) If a group omits one or more sources, these have default scores of zero, which fit into the same numerical scale as the other scores for analysis purposes.

(iii) The most important source is given the same weighting for each group (8), thus limiting the extent to which one group can influence the overall mean score, as in Tables 5(a) and 5(b).

Table 5(a) Scores allocated to ranks for women groups

Information Source	Village Lidala	Chiwinja	Magomero	Kambuwa	Mean scores
FSIPM Project	5	8	0	5	4.50
Radio	2.5	5	3	8	4.63
Friends	2.5	3	6.5	2.5	3.63
Extension Worker	8	0	2	1	2.75
Own Experimentation	1	1	6.5	2.5	2.75
Written Material	0	2	1	1	1.00
Other Sources	1	1	1	0	0.75

Table 5(b) Scores allocated to ranks for men groups

Information source	Village Lidala	Chiwinja	Magomero	Kambuwa	Mean scores
FSIPM Project	0	0	0	0	0
Radio	1	8	7.5	8	6.13
Friends	3	1	5	3	3.00
Extension Worker	6.5	3	0	5	3.63
Own Experimentation	1	2	0	1	1.00
Written Material	6.5	5	7.5	2	5.25
Other Sources	2	1	0	1	1.00

The scores from Table 4 used above in Table 5(a) and 5(b) are arbitrary. The top limit value of 8 could have been another number. The distribution of scores could have been different so that a rank of 7 would be scored as 1, not zero.

There may be some misgivings about using such scores. We have found that if two or more variant allocations of scores to ranks are produced and used as above, the relativities of the means are little affected. If a concensus is reached amongst those using, and communicating in terms of, a scoring system, the element of arbitrariness soon becomes unimportant compared to the benefit of conveying summary information in an efficient standard way.

Participatory Use of Scores

Ranking is often the preferred way to elicit people's views, because scoring is a less natural and perhaps less reliable approach for individuals and groups. Scoring is sometimes attempted by asking respondents to share out a number of beans among items to be ranked. In our example, this would be equivalent to sharing 20 beans between seven information sources. A possible two-stage approach is to elicit rankings with a particular group, then to convert these to scores. These can then be counted out as sets of beans and the group asked to correct the bean distribution if necessary, starting from a fairly reasonable representation. After any adjustment by the group, the final numbers should represent true participatory scores.

In situations where the farmers or farmer groups involved in the participatory exercises have been appropriately selected from a larger well-defined target population, it is also possible to carry out formal statistical procedures (e.g. analysis of variance techniques) to determine whether the observed differences in mean scores represent a real difference or just a chance effect. This leads to a significance test of the hypothesis that there is no real difference. Rejection of the hypothesis, with a small chance of error (e.g. less than 5%), is indicative of a true difference in mean scores in the wider target population of interest.

Further analysis is also possible to investigate whether there are real gender differences in farmers' views regarding the importance of information sources. These analysis procedures are possible using standard statistics software (e.g. SPSS, MINITAB).

It must be noted that statistical tests do exist for analysing ranked data without conversion to scores. However, the methods do not allow a full exploration of the data to disentangle the different sources of variation (e.g. socio-economic characteristics) that may influence the ranks given by the respondents. Moreover, such methods are generally limited to situations where there are no missing values and where the data structure is relatively simple.

Therefore, in farmer participatory evaluations where a major aim involves selecting the "best" from a range of items (e.g. selecting a "best" variety for distribution to farmers or selecting a particular information source for dissemination of agricultural information), conversion of ranks to scores is beneficial. It provides the means to make formal assessments to address study objectives with a greater degree of objectivity and statistical validity.

References

Fielding, W.J., Riley, J. and Oyejola, B.A. (1998) Ranks are statistics: some advice for their interpretation. PLA Notes 32. IIED. London.

Lawson McDowall, J., Kapulula, P.K. and Chiumia-Kaunda C. (1999) Networks of Communication: How farmers learn about innovations in agriculture. FSIPM Project Mimeo.

Maxwell, S. and Bart, C. (1995) Beyond Ranking: Exploring Relative Preferences in P/RRA. PLA Notes 22, pp. 28-34. IIED, London.

Theme Paper 5

COMPARING CHANGES IN FARMER PERCEPTIONS OVER TIME

Savitri Abeyasekera[1] and Neil Marsland[2]

[1] Statistical Services Centre, University of Reading
[2] Natural Resources Institute, University of Greenwich

COMPARING CHANGES IN FARMER PERCEPTIONS OVER TIME

Introduction

The larger grain borer (*P. truncatus*) is a major pest of stored maize. The beetle is able to develop and reproduce in maize and dried cassava, and because of its boring activities, it is capable of damaging a large variety of commodities including other food commodities, wooden objects and drying timber, and leather.

As part of the project "Coping strategies adopted by small-scale farmers in Tanzania and Kenya to counteract problems caused by storage pests, particularly the Large Grain Borer", funded by DFID's Crop Post-harvest Research Programme, a survey was conducted in Tanzania to determine ways farmers have coped with the larger Grain Borer (LGB) problem in the past two decades. The survey covered seven districts spread over six regions, with 2 to 5 villages being visited in each district (Table 1). The villages were chosen on the basis of known LGB infestation.

Table 1. Study villages showing district and region locations.

Region	District	Villages
Iringa	Iringa	Kiwere; Mgera; Nzihi; Chamdindi
Rukwa	Mpanda	Songambele; Mnyaki; Ikologo
Morogoro	Kilosa	Rubeho; Msingisi; Ukwmani; Ihenje
Arusha	Babati	Mamire; Chasimba; Singe; Dareda; Riroda
Kilimanjaro	Mwanga	Lembeni; Mwembe
	Hai	Rundugai; Magadani
Tabora	Tabora Rural	Isikizya; Magiri; Inala; Itonjanda

One major objective of the survey and associated RRA approaches, was to assess the role played by LGB in determining changes in production, storage and marketing of the maize and cassava crop during the period between the time of establishment of the beetle (mid to late 1980's) and today.

Many hypotheses related to the objective above and others were to be investigated in the study (Marsland, et al 1999). Here, just one hypothesis is considered to illustrate a simple approach to the data analysis. The hypothesis is the following:

> *The role of crop production in household food security strategies has reduced during the period between the establishment of LGB and today.*

This document demonstrates a method of analysis to test the above hypothesis and discusses the limitations associated with the method.

Data Collection Methodology

RRA techniques were used to understand farmers' perceptions of post-harvest problems, within a more general context of livelihood and food security strategies. The method involved interviews with 47 groups of farmers, with information being recorded on a data recording sheet by an interview team of 4 enumerators (See Appendix 1 for an example of the first of several sheets used).

In each village visited enumerators first explained the meaning of household food security (HFS) strategies, i.e. that it did not just refer to production of food for own consumption, but also referred to food produced for cash since cash sales make a very important contribution to food security. Farmers were then asked to rank HFS strategies in order of importance, starting with a rank of 1 for the most important strategy. This was done in order to place the importance of LGB damage to maize and cassava in context. The specific aim was to determine whether crop production was much less important than 15 years ago. The question may be stated more specifically as:

Does the ranking of food security strategies when
- (a) LGB was first sighted: and
- (b) in recent years (since 1995);

demonstrate a change in the importance farmers place on agricultural production.

Data Analysis

During farmer group discussions, many food security strategies were mentioned. A full list appears in Appendix 2. The number of strategies used by any one community group varied in general from about 5 to 11 in regions Morogoro, Rukwa and Iringa, and from 10 to 17 in regions Arusha, Kilimanjaro and Tabora. The higher numbers in the latter three regions arose because farmer groups in these regions gave ranks to specific crops they grew, while in regions Morogoro, Rukwa and Iringa, all crops were considered under a single heading as "Agriculture". This was the result of different understandings of the term "food security strategy" by the different enumerators visiting the two groups of regions.

Thus two difficulties were encountered in summarising the data.

- (a) The varying numbers of strategies mentioned by different community groups;
- (b) The use of "Agriculture" as a single HFS strategy by some community groups while others listed a number of crops as being important HFS strategies.

To overcome these problems, the following approaches were adopted.

Firstly, the strategies were divided into five main HFS categories, namely (i) skilled income generating activities (IGAs), (ii) Trading activities; (iii) Unskilled IGAs; (iv) Agriculture and (v) Livestock. This meant that if specific crops were mentioned as important HFS strategies, they were included under strategy category (iv), i.e. Agriculture.

Secondly, each HFS strategy was allocated the rank of the most important strategy listed within that category. The reason for selecting the highest rank rather than (say) the average rank is clear if the example below in Table 2 is considered. This corresponds to information from one community group.

Table 2. HFS strategies mentioned by one community group

Strategy	Rank When LGB first sighted	Main respon sibility	Rank Recent years	Main respon sibility
1.AGRICULTURE	1	M&W	1	M&W
2.LOCAL BREWING AND SELLING	2	W	3	W
3.PETTY BUSINESS	5	W	4	W
4.LIVESTOCK KEEPING	3	M&W	5	M&W
5.FIREWOOD COLLECTION AND SELLING	6	M&W	6	M&W
6.POTTERY	7	W	7	W
7.MAKING AND SELLING BASKETS	9	W	8	W
8.NEEDLEWORK AND EMBROIDERY	10	W	10	W
9.MASONRY	8	M&W	9	M&W
10.CASUAL LABOURING	4	M&W	2	M&W

Strategies listed in Table 2 fall into the five HFS strategy categories as follows:
Items 2, 6, 7, 8 and 9 fall within HFS category (i), i.e. Skilled IGAs.
Item 3 falls within HFS category (ii), i.e. Trading activities.
Items 5 and 10 falls within HFS category (iii), i.e. Unskilled IGAs;
Item 1 falls within HFS category (iv), i.e. Agriculture.
Item 4 falls within HFS category (v), i.e. Livestock.

The use of the highest rank, corresponding to the time when LGB was first sighted, leads to categories (i), (ii), (iii), (iv) and (v) receiving ranks 2, 5, 4, 1, 3 respectively.

The use of the average rank leads to categories (i), (ii), (iii), (iv) and (v) receiving ranks 7.2, 5, 5, 1, 3 respectively. This is less appropriate than the use of the highest rank since category (i) now receives a lower rank than categories (ii), (iii) and (v) although it included the second most important HFS strategy.

Hence for analysis purposes, the highest rank of items mentioned within each HFS category was used. The corresponding ranks are given below in the first two numerical columns of Table 3 for the period when LGB was first sighted and in recent years.

Table 3. Conversion of ranks to scores

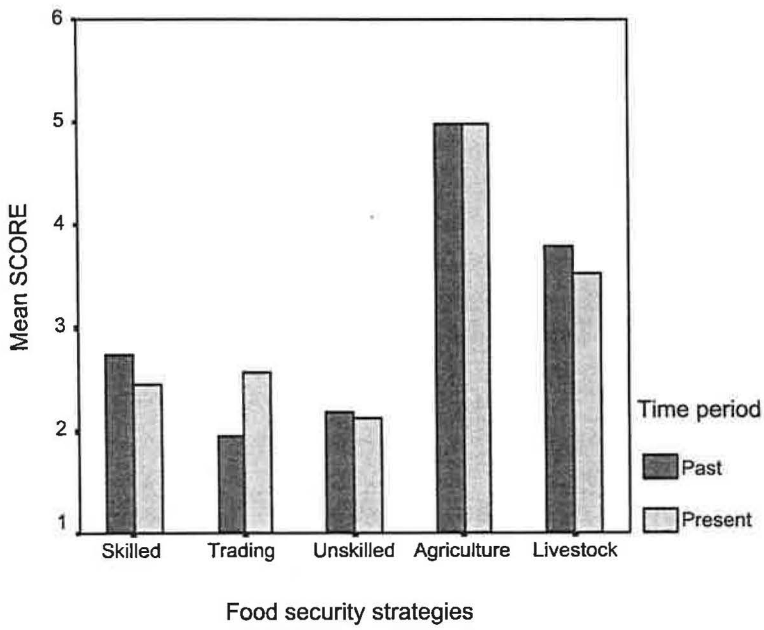
HFS strategy category	Rank when LGB was first sighted	Rank in recent years	Score for rank when LGB first sighted	Score for rank in recent years
1 = Skilled IGAs	2	3	4	3
2 = Trading	5	4	1	2
3 = Unskilled IGAs	4	2	2	4
4 = Agriculture	1	1	5	5
5 = Livestock	3	5	3	1

The ranks were then converted to scores so that the most important item, i.e. the one having the smallest rank, received a score of 5, the second most important item received a score of 4, and so on. The resulting scores are shown in the final two columns of Table 3. The conversion of ranks to scores was necessary to justify the subsequent use of numerical summaries from results of the data collected on ranks of HFS strategies. (For further discussion concerning ranks and scores, see the parallel document entitled "Analysis approaches in participatory work involving ranks or scores").

Similar scores for each of the 47 farmer groups were then averaged over groups to give Figure 1. In this figure, the higher the "bar", the greater is the importance of the specific category it represents. This gives an overall comparison of changes in farmers' perceptions towards livelihood strategies since the first sighting of LGB. There is little evidence of change, but Agriculture is clearly the most important food security strategy, followed by Livestock related activities.

A more formal comparison of past perceptions with current perceptions can be made for each of the food security strategies by looking at the change in scores from past to now across the 47 groups of farmers. For each of the HFS strategy categories, Table 4 shows the total number of farmer groups involved with each HFS strategy and the number who show a change in their opinion about the importance of a particular HFS strategy. An increase in the strategy score indicates that the strategy is more important now than it was in the past. A decrease shows that the strategy is less important now than it was in the past.

Figure 1. Mean Importance Scores when LGB was first sighted and in recent years#



It is interesting to note that even though trading was not the most important strategy, compared to any other strategy type, its importance was felt by farmers to have increased since the mid 1980's.

Table 4. Numbers of farmer groups showing a change in their perception of the importance of different HFS strategies.

HFS Strategy	Past to Present: No Change		Increase in importance score		Decrease in importance score		Total	
Agriculture	45	(96%)	1	(2%)	1	(2%)	47	(100%)
Livestock	31	(66%)	5	(11%)	11	(23%)	47	(100%)
Skilled IGAs	17	(41%)	6	(15%)	18	(44%)	41	(100%)
Unskilled IGAs	13	(54%)	4	(17%)	7	(29%)	24	(100%)
Trading	12	(38%)	17	(53%)	3	(9%)	32	(100%)

Results in Table 4 demonstrate that several farmer groups showed no change in their perceptions of the importance of difference HFS strategies. For those that do, a simple statistical test, i.e. the sign test (Siegel and and Castellan, 1988) can be used to determine whether the increase or decrease in the importance score for a particular HFS strategy indicated a real difference in farmers' perceptions or whether it was merely a chance effect. The relevant null hypothesis for this test (say for Livestock) is the following.

Null Hypothesis (H_0) : Farmers' perception of the importance of livestock as a household security base has not changed from past (when LGB was first sighted) to recent years.

If this hypothesis holds, then, of the 16 farmers (Table 4 for Livestock) who showed a change in their perception, we would expect exactly half to show an increase in their importance score and half to show a decrease in their importance score. The following table then results giving observed frequencies and frequencies expected under H_o.

Table 5. Table of frequencies to test whether farmers' perceptions of livestock as a HFS strategy has changed over a 15-year period.

	Past to Present: Increase in importance score	Decrease in importance score
Observed frequecies (O_i)	5	11
Expected frequenceis (E_i) under H_0	8	8

A chi-square test can now be carried out on these frequencies, i.e. calculate X^2 as:

$$X^2 = \sum(O_i - E_i)/E_i$$

$$= \frac{(5-8)^2}{8} + \frac{(11-8)^2}{8} = 2.25$$

This result is non-significant (p = 0.134, based on 1 degree of freedom). We may therefore conclude that there is insufficient evidence in the data to demonstrate that farmers' perception of the importance of livestock as a HFS strategy has changed.

Note that the ch-square test is an approximate test. The approximation improves as the sample size increases. Here, sample size refers to the number of persons who give different ranks for a particular HFS strategy at the different time points. The sample size determines the expected frequencies for the two categories shown in Table 5. The approximation is poor when the expected frequency in any one category is less than 5.

The idea of comparing observed frequencies with expected frequencies is quite general and extends to more than two categories. If there are k categories, the degrees of freedom associated with the test is k-1, and a rough guideline to sample size requirements is to ensure that no more than 20% of the expected frequencies are less than 5 and none are less than 1. If these conditions are not met, then specialised software, which provide exact tests for small sample sizes, have to be used.

Lessons Learned

Two key issues are illustrated in this case study.

(a) How to over come the thorny problem of enumerator misunderstanding with respect to ranking. This was achieved by grouping household food security strategies into five major categories and using the rank of the most important strategy listed within each category. The problem arose largely because of limited supervision – inevitable when the project leader, is not available on a full-time basis in-country. Despite solid training given to enumerators during this study, mistakes do happen in the field and this paper has demonstrated how remedial action can be taken at the data analysis stage.

(b) Some statistical procedures for analysing ranked data are described in a parallel document (Abeyasekera, 2000). Here the ranks have been taken at two different periods of time and interest is focused on assessing the change over time. It is important in this case to recognise that a lack of a change in rankings from past to present provides no information about time changes. Therefore corresponding statistical tests for change use only information from community groups that do identify a change over time. This paper illustrates how an appropriate statistical test can be carried out to determine whether there is a significant change in participants' views regarding the role of different household food security strategies during the period between the establishment of LGB and today.

References

1. Marsland, N., Golob, P. and Abeyasekera S. (2000) Case Study on LGB coping strategies project. (Output to project R7033).

2. Siegel, S. and Castellan N.J. (1988) Nonparametric Statistics. 2nd Edition. McGraw-Hill Book Company.

APPENDIX 1

DATA RECORD SHEET

DATE:	20.02.98	GROUP TYPE (m/f / mixed):	FEMALE
REGION:	MOROGORO	ENTERED BY:	JOE LUSEYA
DISTRICT:	KILOSA	CHECKED BY:	ETC LUGALIKE
VILLAGE:	UKWAMANI		TEAM LEADER

1(a) Household Food security base:

Strategy	Rank When LGRS [illegible]	Main responsibility	Rank in Recent years	Main responsibility
AGRICULTURE	1	M&W	1	M&W
PETTY BUSINESS	—	—	2	M&W
LIVESTOCK KEEPING	2	M&W	3	M&W
EMBROIDERY	—	—	11	W
POTTERY.	3	W	10	W
BRICK MAKING	—	—	4	M&W
LOCAL BREWING	4	W	7	W
GRASS COLLECTION AND SELLING	6	M&W	8	W
FIRE WOOD COLLECTION	5	W	9	M&W
WATER FETCHING AND SELLING	—	—	6	M
TRADING	7	M.	5	M.

1. GRASS COLLECTION AND SELLING IS NOWADAYS DONE BY WOMEN ALONE WHILE MEN HAVE SWITCHED TO WATER FETCHING AND SELLING
2. THE GROWING POPULATION HAS CAUSED A HIGH DEMAND FOR WATER

Key points (CONTINUE OVERLEAF IF NECESSARY):

SINCE THE SOURCE IS STILL AT MINIMUM. THIS HAS LED TO SELLING OF WATER WHICH WAS NOT THE CASE IN THE PAST.

3. POTTERY IS RANKING LOW TODAY BECAUSE POTS HAVE BEEN REPLACED BY INDUSTRIAL COOKING POTS.
4. VILLAGERS ARE NOWADAYS MAKING BRICKS FOR BUILDING THEIRS HOUSES (IN THE PAST HOUSES WERE BUILT USING TREES BECAUSE THERE ARE NO ENOUGH TREES AS BUILDING MATERIALS DUE TO DEFORESTATION.

APPENDIX 2

List of all food security strategies mentioned by farmers

Group 1 – Skilled income generating activities

- baskets/mats
- blacksmith
- brewing/wine
- brick making
- carpentry
- charcoal
- cult. with tractor
- embroidery
- fish rearing
- fishing
- handcraft
- herbs & medicine
- hiring plough
- honeybee
- horticulture
- hunting
- mango/fruits
- masonary
- mechanics bicycle
- mining
- needlework
- onion
- photo
- pottery
- technical work
- tomato
- weaving

Group 2 - Trading activities

- large scale trading
- Petty business

Group 3 - Unskilled income generating activities

- casual labour
- firewood
- gardening
- grass collection and selling
- lumbering
- sand digging
- water fetching and selling

Group 4 - Agriculture

- general agriculture
- bambara
- banana
- beans
- bullrush millet
- cassava
- castor seed
- coffee
- cotton
- cowpea
- dolcon bean
- dolichos beans
- finger millet
- green grams
- groundnut
- maize
- pigeon pea
- rice
- round potato
- sorghum
- sugar cane
- sunflower
- sweet potato
- tobacco
- wheat

Group 5 - Livestock

- cultivation with ox
- general livestock
- oxen

Theme Paper 6

QUANTIFYING AND COMBINING CAUSAL DIAGRAMS

R.W. Burn, Statistical Services Centre, University of Reading

QUANTIFYING AND COMBINING CAUSAL DIAGRAMS

Introduction

Causal diagrams have been found by practitioners of PRA to be a useful tool in empowering farmers both to identify problems and help understand causal relations between them. A recent development, *scored* causal diagrams, has been introduced and tested in the field by Galpin *et al* (2000). They describe a simple method for assigning scores to the relations in a causal diagram. The idea results from recognising that the causal diagram itself does little or nothing to assist in assessing the *relative* importance of various causal links. The method is to assign a number of counters to the end outcome and work back through the diagram, at each stage sharing out the counters in proportion to the relative importance of the immediate antecedent causes. At the end of this process, each of the "root" causes in the diagram (those without any antecedent causes) finish up with a number of counters which should represent its relative importance. This scoring method is examined in more detail below.

In this paper we focus on exploring the value to the *researcher* of the information gathered by means of causal diagrams, without in any way attempting to belittle the value to the farmer (or other participant). Besides the benefits of the exercise to the participant, there are potentially many insights for the researcher to be gained from an understanding of causal diagrams resulting from PRA activities. For this purpose, to overcome the limitation of attempting to make inferences from a case-study based on a single group or village, we are inevitably drawn to the problem of combining several causal diagrams. PRA studies normally involve repeating the same activity with several groups of participants, resulting in a causal diagram for each group. How should they be combined to get an "average" causal diagram? We shall see that a natural way of combining several diagrams is available provided the links or relationships are scored or quantified in some other way.

Some limitations of quantified causal diagrams are identified and an alternative approach is suggested which can overcome these difficulties. This method is based on Bayesian belief networks, which have recently been applied in many fields.

Once we have constructed a pooled causal diagram for several groups or villages, we explore the idea of a measure of difference or "distance" between causal diagrams. One possible use for this could be to identify particular groups or villages which are very different from the "norm", as represented by the pooled diagram. Closer examination of these groups could lead to a valuable understanding of their different perceptions. An analogy can be made with the conventional statistical practice of fitting a regression model to a set of data points and examining outliers, or data points which are markedly different from the expected values as predicted by the model.

To summarise, the aims of this paper are to

- describe a method for combining several quantified causal diagrams;
- indicate some difficulties with quantified causal diagrams;
- suggest a new formulation for quantitatively modelling a causal diagram;
- explore some possible measures of difference between causal diagrams.

Some of the ideas presented here are tentative, and further development and testing are desirable before advocating their use as routine best practice.

Scoring a Causal Diagram

We present here the method proposed by Galpin *et al* (2000), using one of their simple (hypothetical) examples. The diagram is constructed by first identifying the "end" problem, in this case *low income from maize*, immediate causes of the end problem - *low grade* and *low yields*, causes of those - *many pests* and *poor emergence* , and so on until arriving at "root" causes, which have no identifiable causes themselves. Causal links between these various events are also identified, leading to the diagram shown in Figure 1. (It is assumed that there is a single end problem, which is often the case. The scoring method could probably be extended to causal diagrams with more than one end problem.)

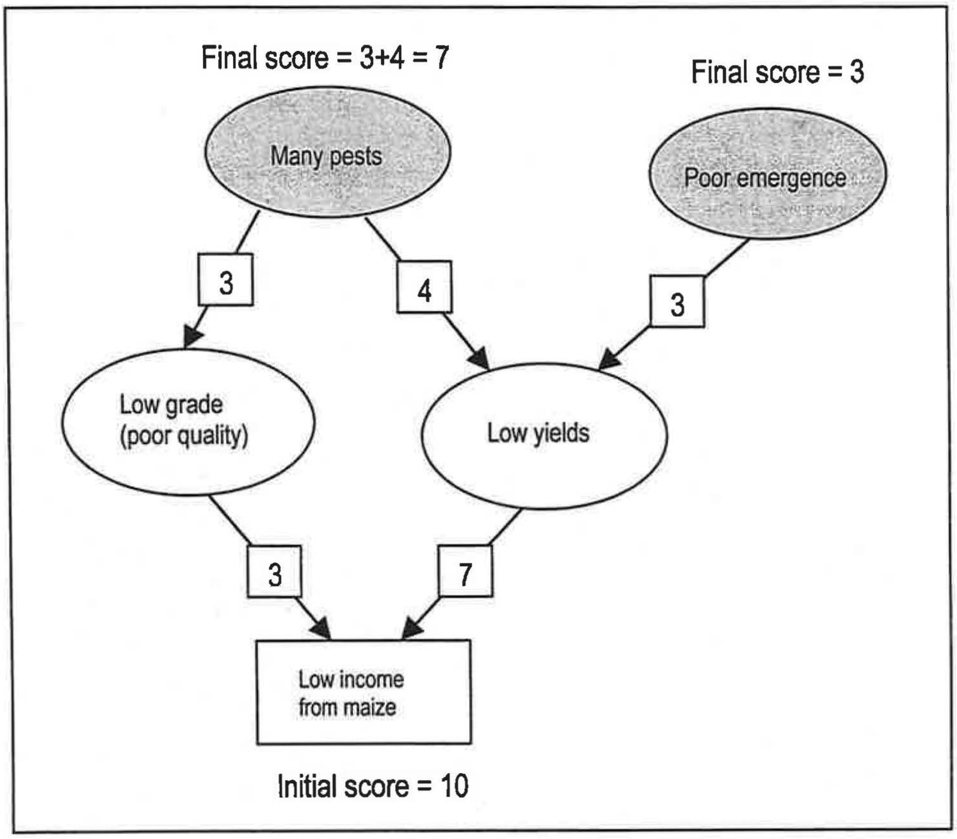


Figure 1

The scoring procedure begins by assigning an even number of counters, 10 in this case, to the end problem. These are then divided to reflect the perceived relative importance of the immediate causes: 3 for *low grade* and 7 for *low yields*. The score of 7 for *low yields* is in turn shared out between its own immediate causes: 4 for *many pests* and 3 for *poor emergence*. There is only one antecedent immediate cause of *low grade*, so its score of 3 is simply transferred to that cause. Once this process has been completed, the total final scores of the root causes are calculated. These scores indicate the relative importance of the root causes with respect to the end problem.

Combining Several Scored Causal Diagrams

Suppose now that we have a number of separate groups, villages for example, each of which has constructed a scored causal diagram. The problem is how to combine these causal diagrams to obtain, in some sense, an "average" or pooled result. Before embarking on the details, we are forced to make some assumptions. The first is that each group is addressing the same basic problem, low income from maize in the above example. Second, we assume that each group is working in the same general environment, so that similar types of causes are likely to emerge from each group. It is not necessary that these causes will be identical, but grossly divergent perceptions of causes could lead to difficulties in combining.

To simplify the description, it is sufficient to combine two causal diagrams, the extension to more than two following in an obvious way by repeated combinations. It is convenient to adopt a different terminology: what were called "causes" or "problems" in the above example will be called *nodes* and the relations between them *links*. This terminology draws partly on graph theory in mathematics (where links are usually called *edges*). In graph theoretic jargon, a causal diagram is a *directed graph* (directed because causal links have a direction from one node to another). This connection with graph theory could provide a natural framework for further theoretical development of causal diagrams, and researchers who wish to pursue these developments should be aware of the rich theory that already exists (Wilson, 1996) to avoid "re-inventing the wheel".

If there is a link from node A to node B, then A is called a *parent node* of B and B is called a *child node* of A. We assume that there are no *cycles* in a causal diagram. That is, there is no node with a sequence of links which start from it and finish on it. This assumption should not be restrictive, because a cycle is a logical impossibility in most situations.

Case 1: Same Nodes and Links

We first consider the case in which the two causal diagrams to be combined have identical nodes and links, but possibly different scores. In practice it may be necessary to resolve apparent differences resulting from non-uniform use of language. What is essentially the same concept may be expressed in different, but synonymous, ways by different groups. Figure 2 illustrates a simple example with two groups, one with 8 and the other with 4 participants. The initial scores do not have to be the same; they were 10 and 16, respectively, for the two groups in this example.

The combined diagram will have the same set of nodes and links. An obvious choice for the combined scores would be the weighted means of the separate scores for each group,

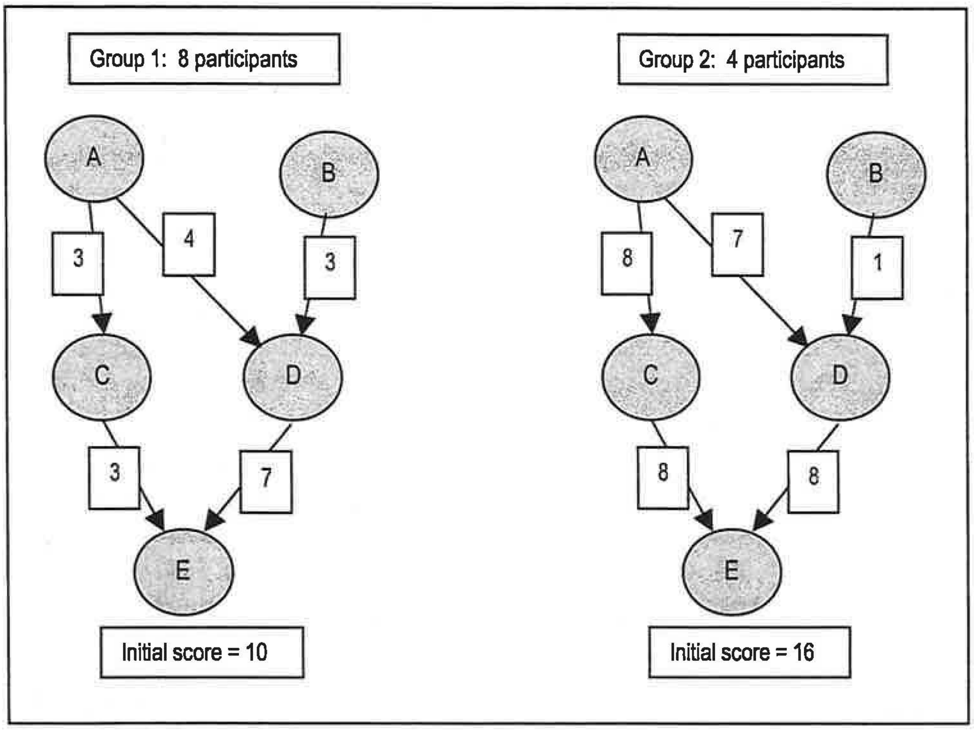


Figure 2

using the numbers of participants as weights. It could happen that some other weighting scheme may be more appropriate, depending on the context, but in the absence of other prior knowledge, group size is a sensible choice (see the discussion at the end of this section).

The combined initial score for node *E* is the weighted mean of the two initial scores 10 and 16, weighted by group size:

$$\overline{N} = \frac{1}{12}(8\times 10 + 4\times 16) = \frac{144}{12} = 12$$

The score for each link is calculated in the same way. For example, for the $C \rightarrow E$ link, the combined score is

$$n_{.1} = \frac{1}{12}(8\times 3 + 4\times 8) = \frac{56}{12} = 4.7$$

The completed combined diagram is shown in Figure 3.

Note that in any scored causal diagram, there is a constraint on the scores imposed by the way in which they are assigned: the sum of all scores on links emanating *from* the node must be equal to the sum of scores on links leading *to* it.

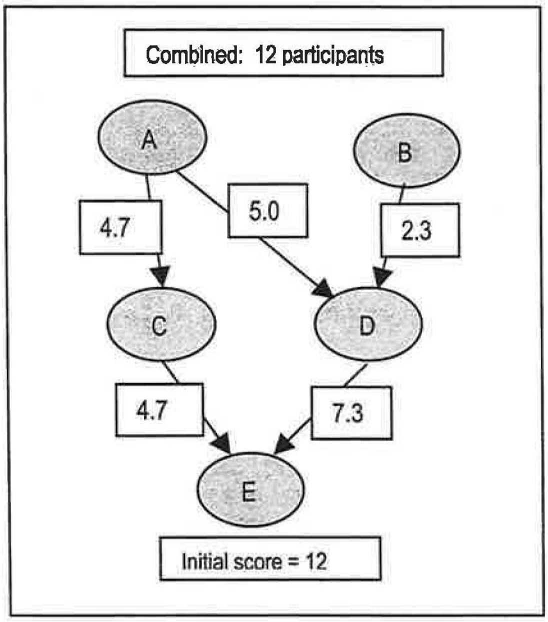


Figure 3

This property is automatically preserved by the weighted average scores. Thus for node D, for example, the score of 7.3 on the $D \to E$ link is equal to the sum of the scores on the $A \to D$ and $B \to D$ links. Furthermore, the total final scores on the "root" nodes A and B are 9.7 and 2.3, respectively. Note that the sum of these is 12, the initial score on the combined diagram.

The general formulae for combining scored causal diagrams from p groups of m_i participants are as follows.

Notation: N_i denotes the initial score for the i^{th} group, n_{ij} denotes the score on the j^{th} link for the i^{th} group and M denotes the total number of participants $M = \sum_{i=1}^{p} m_i$. Then the initial score for the combined diagram is given by $\overline{N} = \frac{1}{M} \sum m_i N_i$, and the score on the j^{th} link is given by $n_{.j} = \frac{1}{M} \sum_{i=1}^{p} m_i n_{ij}$.

Case 2: Same Nodes, Different Links

It is relatively easy to deal with this case by first forming the combined diagram with the common set of nodes. The set of links of the combined diagram is defined as follows: for each pair of nodes, if there is a link between them in any one diagram then the link is put into the combined diagram. In other words, the set of links is the "set-theoretic"

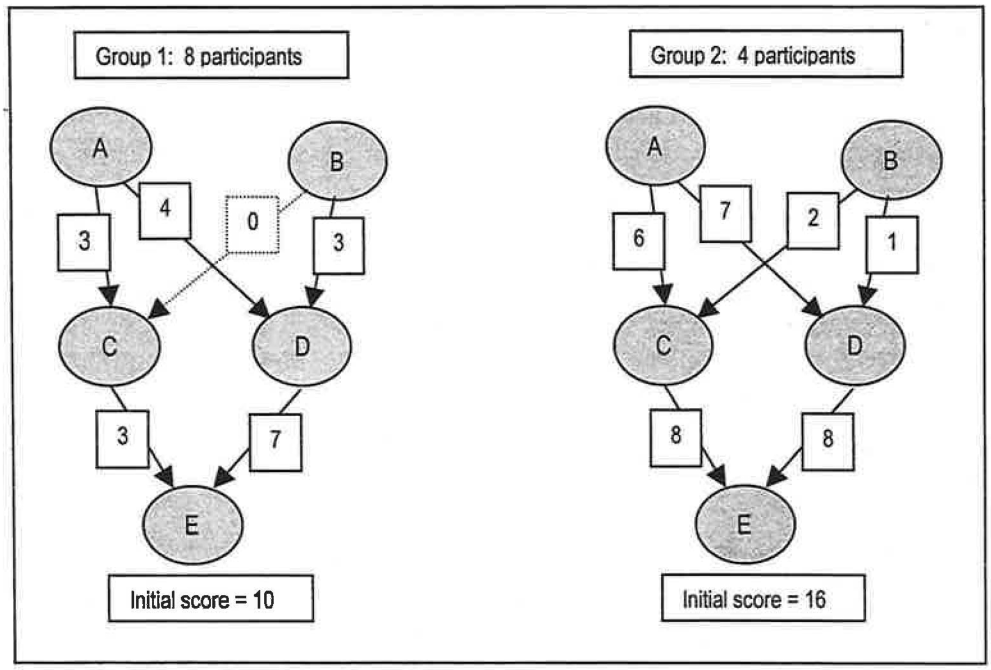


Figure 4

union of all links on the individual diagrams.

We now replace the set of links in each individual diagram with this complete set. If a particular link was absent from the original diagram, it is given a score of zero. The existing links, of course, maintain their original scores. The problem has now been reduced to a special case of combining diagrams with the same nodes and links (albeit with zero scores on some of the individual diagrams), and the above method of calculation applies.

To illustrate, suppose in the previous example that Group 2 had a causal link from B to C with a score of 2 (and correspondingly the score of the $A \rightarrow C$ link is reduced to 6). We insert an imaginary $B \rightarrow C$ link in the Group 1 diagram and assign it a score of zero.

The score for the $B \rightarrow C$ link in the combined diagram will be

$$n_{.6} = \frac{1}{12}\left(8 \times 0 + 4 \times 2\right) = \frac{8}{12} = 0.7$$

and the score for the $A \rightarrow C$ link now becomes $n_{.3} = \frac{1}{12}\left(8 \times 3 + 4 \times 6\right) = \frac{48}{12} = 4.0$. Apart from these, the other scores will be the same as in Figure 3.

Case 3: Different Nodes, Different Links

This case follows similarly, by forming the union of all nodes that appear in all individual diagrams. Difficulties could arise if there is great disparity between the diagrams. A substantial amount of overlap in the node sets would appear to be a pre-requisite for making sense of pooled causal diagrams. We assume that this exists in what follows.

The idea is similar to the case of common nodes with different links. First form the union of all nodes, so that if any one diagram has a node it is included in the combined diagram. Then do the same with links. On each individual diagram, add an imaginary node for each node that appears in the union, and add links with zero scores. In this way, we again arrive at a set of individual diagrams with a common set of nodes and links, and the general method of calculation again applies.

As illustration, suppose that in the example in Figure 2 Group 2 had identified another cause, F, for C, and assigned a score of 3 to the $F \rightarrow C$ link (so that the score for the $A \rightarrow C$ link is reduced from 8 to 5). Group 1 did not have node F. The pooled diagram will have all nodes A to F. Introduce the imaginary node F into Group 1's diagram and give the $F \rightarrow C$ link a score of zero. The situation is shown in Figure 5.

The calculations now proceed as before. The $F \rightarrow C$ link in the pooled diagram has a score of 1.0, and the new score for the $A \rightarrow C$ link is 3.7. The "root" causes have total final scores as follows: 1.0 for F, 8.7 for A and 2.3 for B, so that the sum is still 12, the initial score for the combined diagram.

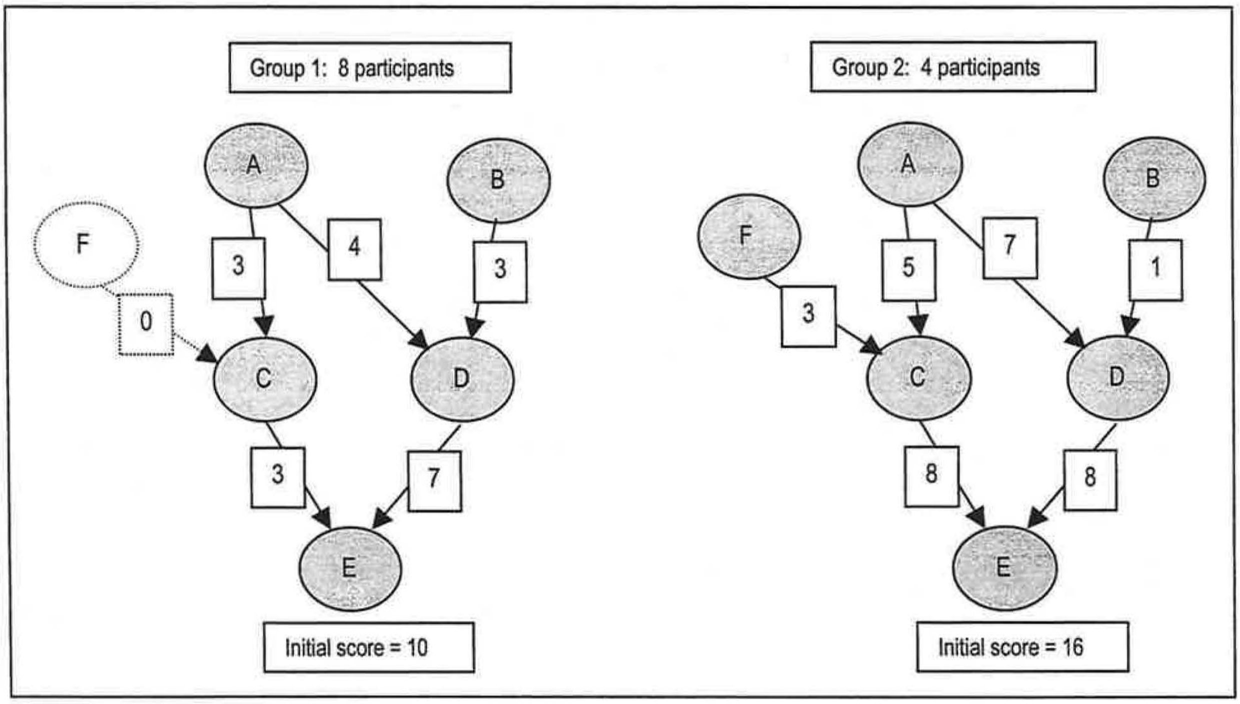


Figure 5

A different kind of complication may arise when some groups interpose an intermediate cause between two nodes while others do not. For example, one group may have a direct link $A \rightarrow E$, while another group may have $A \rightarrow C \rightarrow E$. This situation can be dealt with by again constructing the pooled diagram to be the maximal diagram with all nodes and links that occur in the individual diagrams, and inserting imaginary nodes and links with zero scores in the individual diagrams. Again the general method of calculation will apply.

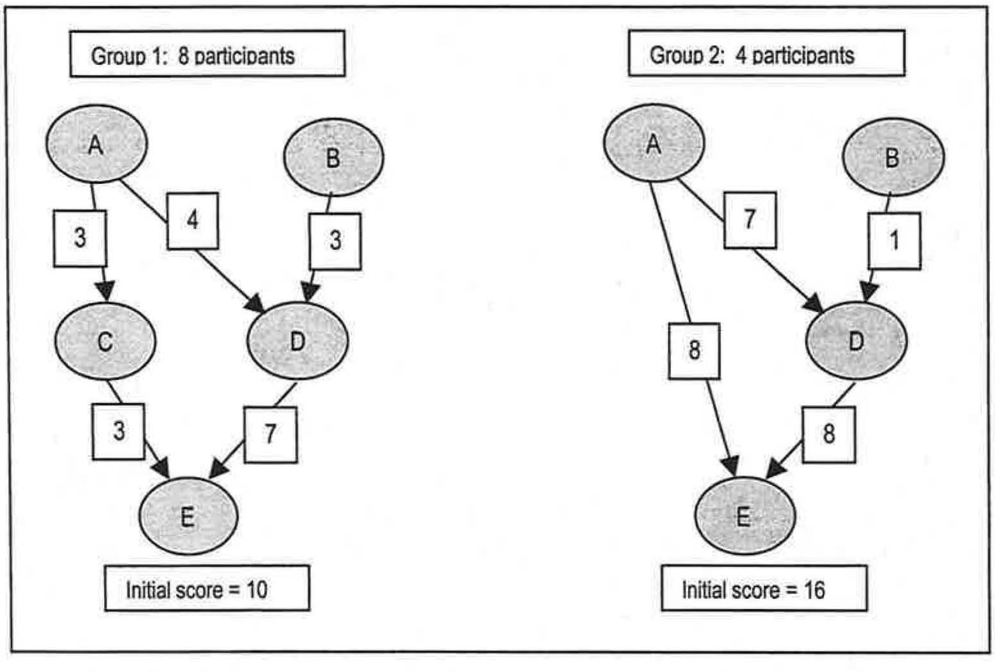


Figure 6

As an example, suppose the situation in Figure 2 is now as shown in Figure 6. Here, Group 1 have an intermediate node *C* between nodes *A* and *E*, while Group 2 do not. The pooled diagram will have all five nodes and the same links as in Group 1, together with an additional link $A \rightarrow E$. Again, the diagram for each group is augmented by imaginary nodes and links, with zero scores where appropriate, so that it resembles the pooled diagram, as in Figure 7.

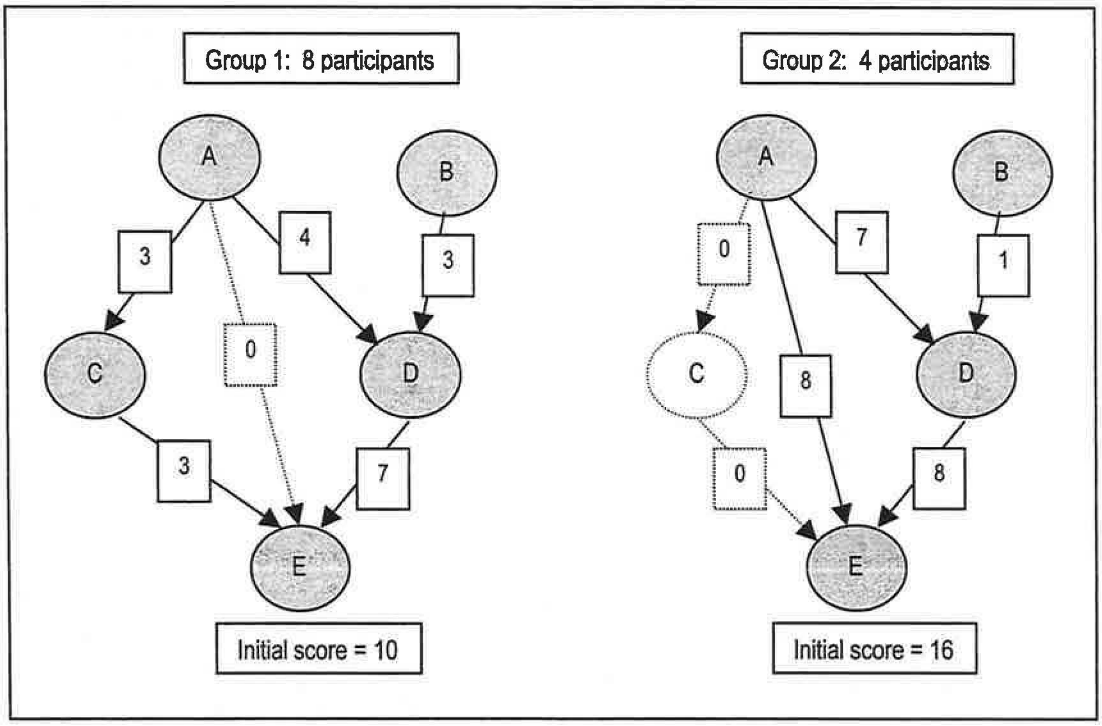


Figure 7

The combined diagram is shown in Figure 8. The calculations for the links $A \rightarrow E$ and $A \rightarrow C \rightarrow E$ are as follows:

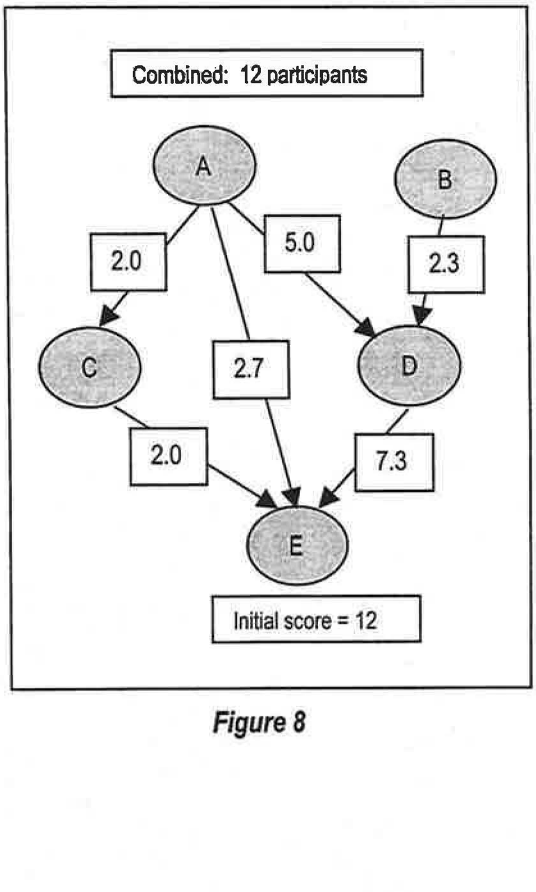


Figure 8

$A \rightarrow E$:

$$n_{.6} = \frac{1}{12}(8 \times 0 + 4 \times 8) = \frac{32}{12} = 2.7$$

$C \rightarrow E$:

$$n_{.1} = \frac{1}{12}(8 \times 3 + 4 \times 0) = \frac{24}{12} = 2.0$$

$A \rightarrow C$:

$$n_{.3} = \frac{1}{12}(8 \times 3 + 4 \times 0) = \frac{24}{12} = 2.0$$

The pooled scores for the other links are calculated as before.

Alternative weighting schemes
Whenever information is combined, there are alternative ways to represent the relative importance of each group.

- Equal weighting for each person in the sample: this is what we have done above.
- Equal weighting, because each group represents a village: this gives equal weight to each village.
- Equal weighting for each person in the village: each group is weighted by the number of people they represent, rather than by the group size.
- Equal weighting for each hectare of land: groups are weighted proportional to the land they represent.

We emphasise the word "equal" that is in all the alternatives above. They are all "fair" weighting schemes, but each by a different definition of "equality". Sometimes it may be sensible to use more than one weighting scheme.

Some Limitations of Scored Causal Diagrams

Adding scores to causal diagrams is a clear advance on the purely qualitative causal diagram of the more traditional PRA approach, both from the point of view of the participant and that of the researcher. Scoring enables the participant to gain insights into the relative importance of root causes of problems. Although it requires more effort to produce, it is a relatively straightforward exercise. The researcher also can learn much from scored causal diagrams, especially with the method outlined above for pooling results from several groups.

However, there are two difficulties with scored causal diagrams. The first concerns the *interaction* of multiple causes. The scoring procedure introduced by Galpin *et al* does not appear to allow for the possibility that when a problem has two or more causes, these causes may not act independently of each other. Interactions are quite common in practical situations. Consider, for example, the hypothetical causal diagram in Figure 1. Causes of *low yields* have been identified as *many pests* and *poor emergence*. It is quite probable that the effect of pests is less important when emergence is poor than when it is good. This is an example of an interaction effect: the two causes do not act independently of each other. The scoring method of Galpin *et al* appears to offer no way of representing this interaction, and therefore makes the implicit assumption that the set of causes for each node are independent. This is a limitation of non-scored causal diagrams also.

The second difficulty with the scoring method is that there appears to be some ambiguity in the meaning of the scores. The scoring method requires participants to assign a score to a cause in proportion to its "importance" in causing the outcome under consideration. What is the precise meaning of "importance" here? We attempt to dissect this idea. Suppose we have an outcome (node) A for which two possible causes B and C have been identified. What precisely is meant by stating that the $B \rightarrow A$ causal link is "more important" or "stronger" than the $C \rightarrow A$ link? One interpretation is that when B occurs then A is more likely to occur than when C occurs. Another meaning could be that the consequences of *A* occurring are more serious when *B* occurs than when *C* occurs. Yet another meaning could be a combination of these two: perhaps *A* is more likely to occur when *C* occurs, but the consequences would be more serious when *B* occurs, and this outweighs the greater likelihood of *A* when *C* occurs. There are two distinct underlying concepts in these interpretations. The first is likelihood, or probability, and the second, using the language of statistical decision theory (DeGroot, 1970), is a loss function or negative utility. A loss function is a quantitative measure of the consequences of an occurrence. The examples presented by Galpin *et al* suggest that both of these underlying ideas have played some part in arriving at their scores.

It could be argued that these two problems fall entirely within the use of the method to the researcher, and do not diminish the value of the exercise to the participant. Identifying and scoring causes of problems could be a valuable exercise in itself for the participants. There is some justification for this point of view, but the practical implications of both of the problems mentioned above should have been rehearsed, at least by the facilitator, before starting the scoring exercise. The problem of interacting causes has possibly more immediate importance for the participant, simply because interacting causes *do* often occur, and the scoring method cannot accommodate them. The other problem, concerning the interpretation of the strength of a causal link may be slightly less worrying, at least for the participant, provided whatever interpretation is understood is used consistently.

The next section presents a reformulation of scored causal diagrams which automatically overcomes the first of the above two difficulties by allowing for interacting causes. This approach, based on Bayesian belief networks, also deals with the second problem by defining "importance" or "strength" of a causal link in terms of probability or likelihood. This could be regarded as a tentative step towards a rational quantitative model for causal diagrams, but further development is needed to explore the role of utility or loss functions.

Causal Diagrams as Bayesian Belief Networks

A Bayesian belief network (BBN) is usually defined, using terminology introduced above, as a directed acyclic graph (Jensen, 1996). A BBN is thus a collection of nodes with a set of directed links between them without any cycles or loops. This, of course, is just the same as what we understand by a causal diagram. In practice, BBNs also have tables of conditional probabilities which measure the strengths of the links in terms of their likelihoods. In recent years, BBNs have been found to be useful tools in a wide range of application areas, and some very powerful computational methods have been developed and implemented through easily available software.

To see how causal links are quantified in a BBN, consider first a node *A* with two causes *B* and *C* (Figure 9). We construct a table of *conditional probabilities*, denoted $P(A \mid B, C)$ (the probability that *A* occurs given that *B* and *C* occur). In this paper, we usually assume, for simplicity, that nodes are *binary*, that is, they have just two possible "states", formally labelled T (true) and F (false), which can be taken to mean "occur" or "not occur". In general BBNs, nodes can have more than two possible states.

The table of conditional probabilities could be as shown in Table 1.

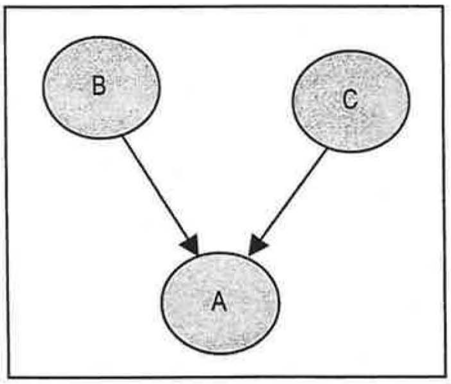


Figure 9

P(A | B, C)			
		B:	
		T	F
C:	T	0.9	0.8
	F	0.4	0.1

Table 1

Thus, there is a 90% chance that *A* occurs if both *B* and *C* occur; a 40% chance if *B* occurs but not *C*, and so on. Clearly this table captures any interaction in the effects of *B* and *C* on *A*. Nodes with no parents ("root causes") have *prior* probabilities assigned to their possible states. Thus for node *B* for example, we might have the probability that *B* occurs = 0.6. (In the absence of any prior information on the state of *B*, we would assume that either state *T* or *F* is equally likely, and assign a probability of 0.5 to each.)

		A	
C	B	T	F
T	T	0.9	0.1
T	F	0.8	0.2
F	T	0.4	0.6
F	F	0.1	0.9

Table 2

An alternative, more commonly used, layout for the table of conditional probabilities is shown in Table 2. Note that the row sums are all equal to 1.

If node *A* had three states (low, medium, high for example), there would be three columns under *A*.

The left hand side of the table for a node *A* with three binary parent nodes, *B, C* and *D* would list all eight possible combinations of their states.

Elicitation of the Probabilities

Obtaining these probabilities from a group of participants is a more demanding task than the scoring method of Galpin *et al*, and would consequently take more time. In the first place, most people do not think explicitly in terms of probabilities, although often their expression of strength of a causal link can be interpreted that way (see the remarks in the preceding section). Furthermore, there are more quantities required: four probabilities for two binary causes, and in general 2^p probabilities for p binary causes, which compares with p quantities for the scoring method.

There is scope here for future work on developing practicable methods for the elicitation of probabilities in participatory activities. The method of phrasing questions which seek quantified assessments of probabilities needs to be thought through, especially with regard to synonymous or proxy concepts of probability which may be more familiar to participants. There is also the need for methods of checking the robustness of probability assessments. Facilitators or researchers trying to build BBNs should ensure that the reasoning leading to probability estimates is properly recorded and that whenever possible, there is triangulation with different groups of stakeholders (e.g. farmers and extension workers). Some recent research (O'Hagan, 1998, for example) has explored systematic methods for eliciting probabilities in the context of working with panels of experts, and it is possible that some of these ideas could be adapted to the present context.

We take the example causal diagram introduced earlier and rebuild it as a BBN. There is no direct relationship between the scores that were allocated in the previous treatment with the probabilities assigned below. To give more scope for possible outcomes, we have made the end-problem node (low income from maize) a node with three states: *low, medium, high.* After some deliberation (and/or elicitation), we produce the conditional probability table shown here for this node.

		Income from Maize		
LG	***LY***	Low	Medium	High
True	True	0.85	0.15	0.00
True	False	0.35	0.55	0.10
False	True	0.65	0.30	0.05
False	False	0.05	0.10	0.85

Abbreviations for the nodes are:

IM – income from maize
LG – low grade
LY – low yield
PE – poor emergence
MP – many pests.

		Low Yield	
MP	***PE***	True	False
True	True	0.80	0.20
True	False	0.60	0.40
False	True	0.50	0.50
False	False	0.10	0.90

This is the conditional probability table for the node *low yield.*

The conditional probabilities for the node *low grade* are simpler because there is only one parent node. And there are only prior probabilities for the nodes *Many pests* and *Poor emergence* because they have no parent nodes:

	Low Grade	
MP	True	False
True	0.80	0.20
False	0.10	0.90

Many Pests	
True	False
0.60	0.40

Poor Emergence	
True	False
0.25	0.75

Along with Figure 1, this completes the specification of the BBN. It is useful at this point to use one of the several software packages available for designing, manipulating and analysing BBNs. In principle, many of the computations could be done manually, but if we wish to make use of the powerful algorithms which have been developed for analysing BBNs, then the software becomes essential. The examples in this paper have been analysed using the *Netica* software (Norsys, 1998).

Once the conditional probabilities have been specified, the probabilities are automatically *propagated* through the diagram. This means that, although the original conditional probabilities were given locally for each node, they all affect each other as we move through the network. Thus we see from the completed network (Figure 10) that a consequence of the conditional probabilities that have been entered is that, for instance, there is an overall 48% chance of low income from maize.

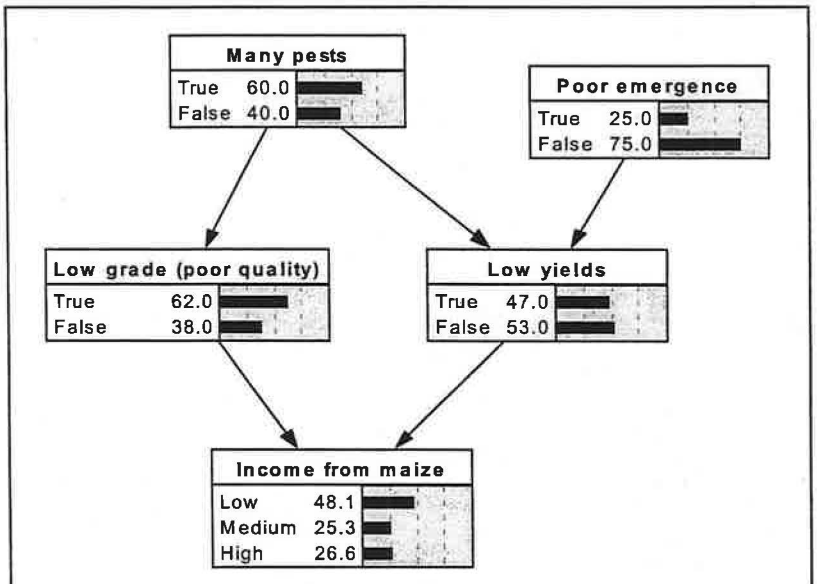


Figure 10

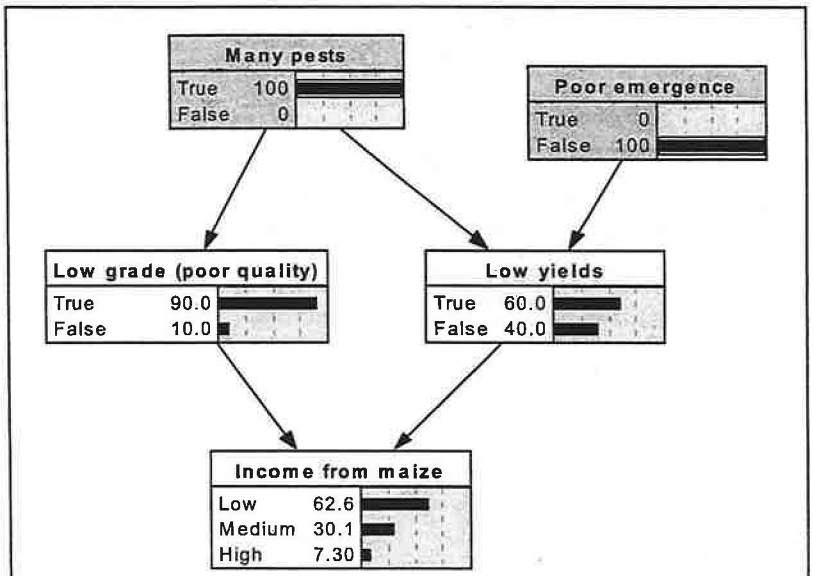


Figure 11

A fundamentally important idea in BBNs is "entering evidence", which allows us to make use of information that a particular event has actually occurred and recalculate all of the probabilities in the diagram. This is accomplished by means of the *updating algorithm* (Lauritzen and Spiegelhalter, 1988) and is supplied in BBN software. An alternative use for this is to explore "what-if" scenarios. By trying different states of nodes in the diagram, we can very easily examine the consequences of these possible realities.

For example, in Figure 11, we look at the consequences of *Many pests* without *Poor emergence*. We see that prospects for high income from maize are quite seriously affected.

Evidence can be entered in *any* combination of nodes so that many different "what-if" questions may be examined.

A particularly important "what-if" special case is to assume a particular result in the "end problem" node, and look at the probabilities of other nodes to gain understanding of the most likely combination of causes.

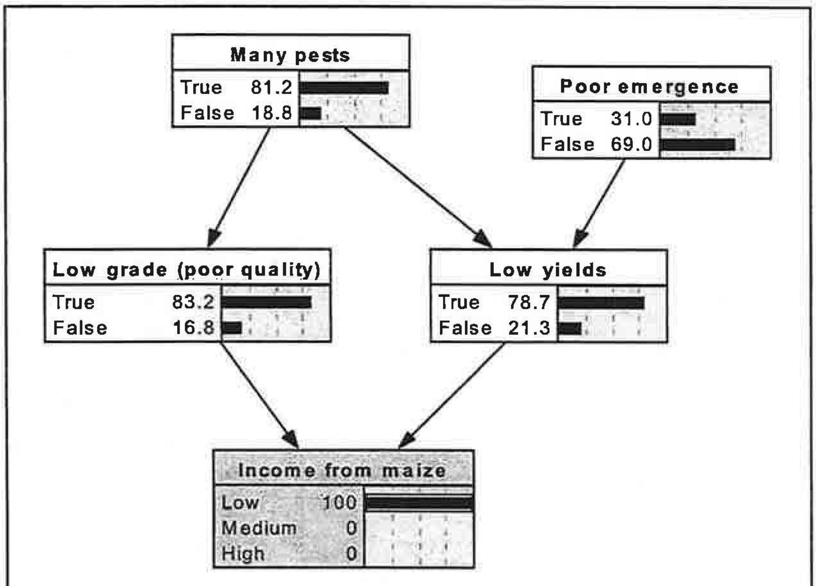


Figure 12

For example, Figure 12 shows the result of assuming that *Income from maize* is *low*.

We see that if this is the outcome observed, then there is a probability of about 81% that there were *many pests*, but only a 31% chance of *poor emergence*.

This ability to update probabilities *in all directions* in a network is one of the most powerful features of the updating algorithm.

Combining Several Bayesian Belief Networks

It is possible to build a pooled BBN from several individual ones in a similar way to what was described for scored causal diagrams. The basic idea of forming the set-theoretic union of nodes and links will work in the same way, with "imaginary" nodes and links added to each individual BBN to make them all similar to the pooled diagram. Weighted means of tables of conditional probabilities are attached to the nodes of the pooled BBN. There is, however, an important difference with scored causal diagrams. There, we found it convenient to insert zero scores on imaginary links and use these in the calculations as if they were ordinary scores.

A different technique is appropriate for BBNs. If a node *A* has an "imaginary" parent node *B*, then we construct *A*'s table of conditional probabilities so that the outcomes are *independent* of *B*. This amounts to saying that the state of *A* is not affected by the state of *B*, and can be thought of as being equivalent to assigning a zero score in a scored causal diagram.

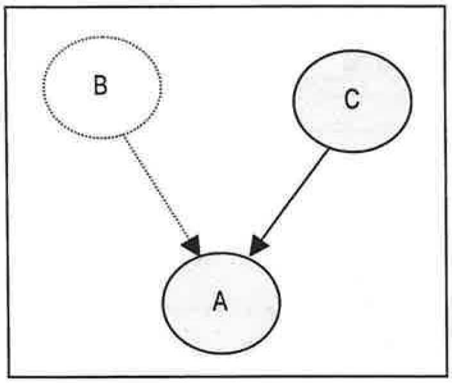


Figure 13

A simple example should make this clear. Suppose *A* has a "genuine" parent *C* and "imaginary" parent *B* has been added, as in Figure 13. Suppose that the table of conditional probabilities connecting *C* to *A* is:

	A	
C	*T*	*F*
T	0.9	0.1
F	0.2	0.8

To accommodate the introduced parent node *B*, we replace this by the table shown here. The conditional probability of *A* given the state of *C* is the same regardless of the state of the imaginary node *B*, so that $P(A \mid B, C) = P(A \mid C)$. This means that the introduction of *B* has no effect on the way that the "genuine" parent node *C* affects *A*.

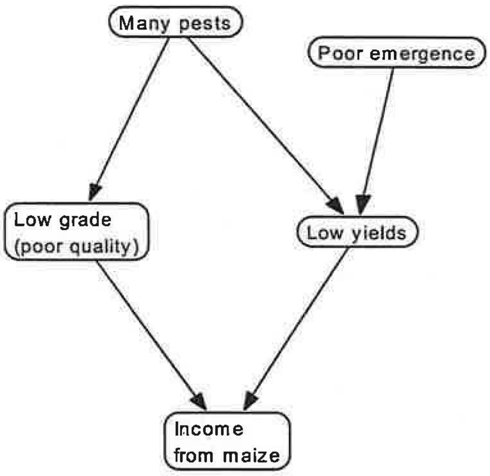
This is straightforward because the introduced node *B* has no parent nodes. If *B* were a child node of one or more parent nodes, then unfortunately there are difficulties in deciding how to assign conditional probabilities to it. In the special case in which *B* is a binary node with states *true* and *false*, then it would seem sensible to assign probabilities of 1 to the *false* state. In this way it will make no contribution to the *true* state in the pooled BBN. It is not yet clear how to deal with the more general case of a multi-state node and further work is required.

		A	
C	*B*	*T*	*F*
T	*T*	0.9	0.1
T	*F*	0.9	0.1
F	*T*	0.2	0.8
F	*F*	0.2	0.8

With these conventions, after constructing the pooled BBN, each individual BBN is augmented as required with nodes and links and the tables of weighted means of conditional probabilities can be calculated. To continue with the previous example, suppose three groups have constructed BBNs as follows.

Group 1 (8 participants):

Probabilities:



Group 1

Many Pests	
True	False
0.60	0.40

Poor Emergence	
True	False
0.25	0.75

	Low Grade	
MP	True	False
True	0.90	0.10
False	0.20	0.80

		Low Yield	
MP	***PE***	True	False
True	True	0.80	0.20
True	False	0.60	0.40
False	True	0.50	0.50
False	False	0.10	0.90

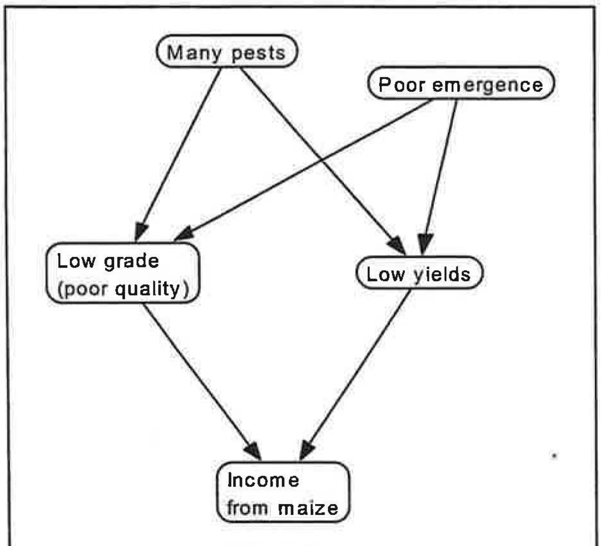
		Income from Maize		
LG	***LY***	Low	Medium	High
True	True	0.85	0.15	0.00
True	False	0.35	0.55	0.10
False	True	0.65	0.30	0.05
False	False	0.05	0.10	0.85

Group 2 (4 participants):

The nodes are the same as Group 1 but with a link $PE \rightarrow LG$ added. The probabilities were chosen as follows:

Many Pests	
True	False
0.70	0.30

Poor Emergence	
True	False
0.40	0.60



Group 2

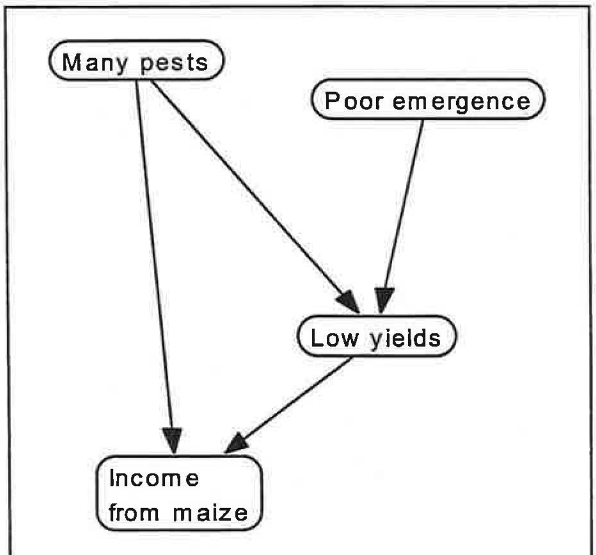
		Low Grade	
MP	**PE**	True	False
True	True	0.70	0.30
True	False	0.60	0.40
False	True	0.25	0.75
False	False	0.05	0.95

		Low Yield	
MP	**PE**	True	False
True	True	0.60	0.40
True	False	0.60	0.40
False	True	0.30	0.70
False	False	0.05	0.95

		Income from Maize		
LG	**LY**	Low	Medium	High
True	True	0.75	0.20	0.05
True	False	0.30	0.55	0.05
False	True	0.40	0.50	0.10
False	False	0.15	0.15	0.70

Group 3 (6 participants):

This group does not have node *LG*. The probabilities were:



Group 3

Many Pests	
True	False
0.70	0.30

Poor Emergence	
True	False
0.40	0.60

		Low Yield	
MP	**PE**	True	False
True	True	0.60	0.40
True	False	0.70	0.30
False	True	0.20	0.80
False	False	0.10	0.90

		Income from Maize		
LG	LY	Low	Medium	High
True	True	0.80	0.20	0.00
True	False	0.60	0.30	0.10
False	True	0.65	0.30	0.05
False	False	0.00	0.30	0.70

The pooled BBN, with all of the nodes and links in each of the three individual networks, is shown below.

The pooled BBN, with all of the nodes and links in each of the three individual networks, is shown below.

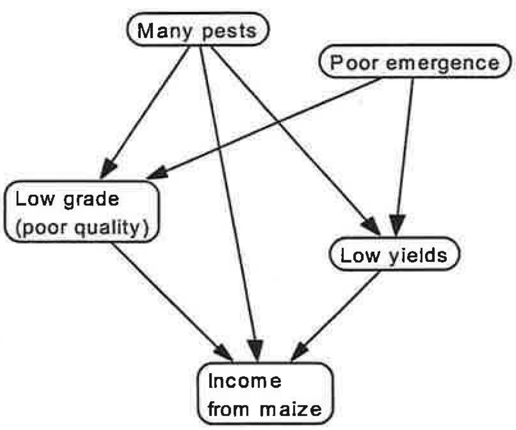
Using the rules described above, the following probabilities were calculated (weighted by the number of participants in each group):

Many Pests	
True	False
0.66	0.34

Poor Emergence	
True	False
0.33	0.67

		Low Grade	
MP	PE	True	False
True	True	0.56	0.44
True	False	0.53	0.47
False	True	0.14	0.86
False	False	0.10	0.90

		Low Yield	
MP	PE	True	False
True	True	0.69	0.31
True	False	0.63	0.37
False	True	0.36	0.64
False	False	0.09	0.91



Combined

			Income from Maize		
LY	MP	LG	Low	Medium	High
True	True	True	0.81	0.18	0.01
True	True	False	0.42	0.49	0.09
True	False	True	0.76	0.21	0.03
True	False	False	0.22	0.49	0.29
False	True	True	0.65	0.31	0.04
False	True	False	0.26	0.18	0.56
False	False	True	0.60	0.34	0.06
False	False	False	0.06	0.18	0.76

Measures of Distance Between Bayesian Belief Networks

If we regard the pooled BBN constructed as above from individuals BBNs as a sort of average, then a question which naturally arises is how to measure differences between the individuals and the average. As mentioned in the introduction, this is similar to fitting a regression model and examining the differences between particular data points and the model, i.e. the residuals. From a practical point of view, this could prove useful in identifying particular groups which have produced causal diagrams which are markedly different from the average, and further examination of these groups could yield interesting information. There are possibly several candidates for a measure of distance in this context. Here we suggest two measures. Further investigation is required to establish the statistical properties of these measures.

Each of these measures is a statistic which attempts to take account of the differences between the set of probabilities which define the BBN. The first, reminiscent of the traditional residual sum of squares in regression analysis is called the *Euclidean distance* (Jensen, 1996). If $p_{1j}, p_{2j}, \ldots, p_{nj}$ denote the complete set of probabilities attached to all of the nodes of the j^{th} BBN, and $\mu_1, \mu_2, \ldots, \mu_n$ denote the weighted mean probabilities, then we can define the Euclidean distance between the j^{th} BBN and the mean to be

$$Dist_j = \sqrt{\sum_{i=1}^{n} (p_{ij} - \mu_i)^2}.$$

The other measure proposed here is very similar to the deviance statistic resulting from fitting a loglinear model to a multi-dimensional contingency table. Measures like this have sometimes been labelled (rather meaninglessly) as "entropy" measures (Jensen, 1996). We call it the *deviance* measure, and define it by

$$Dev_j = 2 * \sum_{i=1}^{n} \{p_{ij}(\log p_{ij} - \log \mu_i) - p_{ij} + \mu_i\}.$$

For the above example, each of these measures has been calculated as an estimate of "distance" between the causal diagrams of each group and the pooled diagram. The results are shown in Table 3. (Note that the scales of the two measures are in no way comparable with each other.)

	$Dist_j$	Dev_j
Group 1	0.92	15.00
Group 2	0.68	15.85
Group 3	1.45	16.91

Both measures suggest that the causal diagram produced by Group 3 deviates more from the "average" than do the other two.

Further investigation is needed to establish whether statistical significance tests are feasible for these measures.

Discussion

One of the fundamental problems faced by the researcher attempting to draw inferences from the results of PRA activities is how to combine or pool the results of several separate exercises. Here we have attempted to show that there is a rational way of doing this with causal diagrams. The scored causal diagrams of Galpin *et al* are attractive in their relative simplicity, although there could be difficulties with them because of interactions and the problem of interpretation discussed above. Bayesian belief networks appear to offer interesting possibilities in their application to this problem, partly because of their inherent ability to deal with interactions, and partly because of their wide range of analysis that can be achieved with them.

However, the construction of a BBN is more demanding than a scored causal diagram and this could be a serious obstacle to their use in normal PRA work. Scored causal diagrams essentially provide information about the *margins* (row and column totals) of the tables of conditional probabilities required for BBNs. In certain circumstances, it may be possible for the facilitator to "fill in the gaps" based on his/her local knowledge to supplement information resulting from PRA work, to provide enough information for the researcher to proceed with a more formal analysis. Indeed, it is possible that the various elements of the probability content of a BBN can be provided by different stakeholders, and it can be a rational method of synthesis.

BBNs can have "utility nodes" as well as the simple event nodes used in the examples in this paper. These purpose of these nodes is to model the *consequences* of events, as opposed to their likelihood or probability, as we have been considering so far. The potential for these to extend the meaning of causal links beyond simple probability is a theme that needs to be explored further, as is the problem of eliciting utility or "importance" information.

Many of the ideas presented here are novel in this context and further work is needed to explore the usefulness of these methods in practice.

References

DeGroot, M.H. (1970) *Optimal Statistical Decisions*. McGraw-Hill, New York.

Galpin, M., Dorward, P. & Shepherd D. (2000) *Participatory Farm Management Methods for Agricultural Research and Extension: A Training Manual*. The University of Reading, UK.

Jensen, F.V. (1996) *An Introduction to Bayesian Networks*. University College London Press, UK.

Lauritzen, S.L. & Spiegelhalter, D.J. (1988) Local computations with probabilities on graphical structures and their application to expert systems (with discussion). *J.R. Statist. Soc*. B, 50, 157-224.

Norsys Software Corp. (1998) *Netica Application User's Guide*. Norsys Software Corp., Vancouver, BC, Canada (www.norsys.com).

O'Hagan, A. (1998) Eliciting expert beliefs in substantial practical applications. *The Statistician* **47** Part 1, 21-35.

Wilson, R.J. (1996) *Introduction to Graph Theory*. Longman, UK.

PART III

CASE STUDIES

Case Study 1

LARGER GRAIN BORER COPING STRATEGIES PROJECT

Neil Marsland[1], Peter Golob[2] and Savitri Abeyasekera[3]

[1] Natural Resources Institute
[2] Natural Resources Institute
[3] Statistical Services Centre, University of Reading

LARGER GRAIN BORER COPING STRATEGIES PROJECT

Background to Study

The Larger Grain Borer (LGB), *Prostephanus truncatus* (Horn), was first reported in Africa in 1981. The beetle, a severe pest of farm-stored maize and dried cassava was initially a major problem to farmers in western Tanzania. The devastation caused by the beetle was sufficient to induce those suffering its ravages to make a 1,000 km journey to Dar Es Salaam to pressurise the Government of Tanzania into taking action to counteract the problem. Thereafter, more than £10 million has been expended by various donors on research and control of this pest.

During the first decade after its initial establishment in Africa a very effective method was developed for its control. This involved the treatment of maize with a mixture of synthetic insecticides and although the efficient use of the chemical required a change in traditional storage practices, it became the template of control strategies across Africa.

During the last twenty years, LGB has spread to a number of countries throughout West and East/Central Africa. Research has continued to examine the biology and ecology of the pest and to devise methods of control, which are 'safer' and less environmentally sensitive than conventional chemicals. Workshops and conferences have been held throughout the period to evaluate progress and identify future needs. However, progress has always been regarded as the potential to control the pest rather than as a measure of the farmer's ability to alleviate the pest problem. However, at an East and Central Africa Storage Pest Management Workshop, held in Naivasha, Kenya in 1996, the general synopsis was that there was a need to evaluate how farmers had coped with the LGB problem during the past two decades in order to justify any proposals for further research. Accordingly it was recommended that a study be undertaken to evaluate farmers' reactions to LGB, to identify coping strategies and to ascertain whether the beetle should still be regarded as a major pest of primary importance.

The DFID Crop Post Harvest Programme (CPHP) co-funded the study with the Rockefeller Foundation. In practice, the CPHP funded activities in Tanzania, which were co-ordinated by NRI, and Rockefeller funded activities in Kenya that were co-ordinated by CABI Bioscience. Initially, two training workshops were conducted to introduce staff to the loss assessment and RRA methodology, and to develop the design of the main surveys and questionnaires to be used. Thereafter, three teams conducted interviews with groups and individuals; two teams operated in Tanzania and another in Kenya. Each team spent a minimum of six weeks in the field, the period being extended as a result of adverse weather conditions due to El Ninõ effects. After data collation, a final technical workshop to discuss the results was held in Nairobi.

The Tanzanian survey was conducted only in villages that had suffered from LGB infestation. As most villages had experienced problems with this pest the villages, which were selected at random, were representative of the localities from which they were drawn.

Selected Summary of Results

Over the course of the last 15-20 years crop production has remained a cornerstone of household food security and the positions of both maize and cassava - in those areas where it has been historically important – remain unchanged, by and large.

The results of the survey present something of a puzzle. With the exception of a minority of farmers the Tabora and Kilimanjaro survey districts, *P. truncatus* does not appear to have influenced production, storage and marketing ***outcomes*** to any significant degree in any of the districts surveyed. It is true that there is considerable evidence that farmers in some areas have changed their storage behaviour. The key behavioural changes have been (i) a much more widespread use of ASD (Actellic Super Dust) and, in Kilosa district Morogoro, Actellic EC; (ii) a concomitant increase in the incidence of storing shelled maize (with the exception of Kilosa) and (iii) decrease in crib and platform storage. However, the changes have by no means been universal, and there are significant concerns particularly about the cost of ASD and its efficacy against the spectrum of storage pests found on the farm. In these circumstances, it is legitimate to wonder why *P. truncatus*, in association with other storage insects, has not had more of an impact on production, storage and marketing outcomes. One possible answer is that farmers have been overestimating the difficulties that they are encountering with pesticides and bags. This is possible, although problems with adulteration, cost and availability are well known. Another possible answer is that a combination of good protection measures in the past, together with the effects of the droughts of the last decade has prevented the build up of P. truncatus in the villages so that it no longer causes significant damage even to poorly protected grain.

In these circumstances, there have to be legitimate concerns over the prospects for the future, as if adequate protection measures are not taken. *P. truncatus* populations will soon increase, and losses will rise. When farmers have access to reliable insecticides and adequate storage facilities, the indications are that they will use them if they can afford to. The implication is, therefore, that measures be taken to tighten up the regulation of pesticides and ensure that they are more widely available. In addition, the question of cost should be considered, and if it is not tenable to reduce cost then this argues for an increased emphasis on low cost (perhaps botanical) protectants and on integrating other, less effective control measures to achieve adequate control. These issues merit further investigation.

Objectives of the Study and Required Methodology

There were two objectives for the study:

(i). To assess the role played by *P. truncatus* in determining changes in production, storage and marketing of the maize and cassava crop during the period between the time of the establishment of the beetle (mid to late 1980's) and today.

(ii). To assess the factors determining the role played by *P. truncatus* in these stages of the maize and cassava commodity systems, in particular the impact of insecticide treatment.

The specific information objective fell into the third type of the typology derived from Longhurst (1992) (see framework paper, Table 2)

"To obtain quantitative data with an understanding of processes causes (diagnosis); data could be used as benchmark data to assess trends, therefore method repeatable with high degree of confidence".

In order to achieve these objectives, a combination of sample survey and rapid rural appraisal (RRA) techniques was required.

Relationships Between Qualitative and Quantitative Components of the Study

Combining ideas and methods based on probability theory and statistical inference with those rooted in informal, participatory enquiry, can give benefits in terms of trustworthiness of data. Table 1 illustrates this.

Table 1: Relationship between: stages in research exercise, type of formal-informal combination, and improvements in trustworthiness.

(i) Design of study

Type	Explanation/Example	Function
Swapping	• Formal sampling procedures for informal work	• Reduced sampling error: better external validity for informal work
	• Use of mapping for formal work	• Reduced time and cost for household listing, sampling and enumeration.
Concurrent	• Correct use of different instruments for different variables within the same survey/ experiment	• Better internal validity for "qualitative" variables - belief, motivations etc. alongside better external validity for quantitative variables - rates, proportions etc. • "Enriching": Using informal work to identify issues or obtain information on variables not obtained by quantitative surveys.

(ii) Analysis

Type	Explanation/Example	Function
Sequential	• Analysis of formal outputs with informal approaches. e.g. testing null hypotheses; investigating unexpected outcomes.	"Refuting" or "Confirming": Verification of formal results through informal methods. • "Explaining": Using informal work to explain unanticipated results from formal survey work
Swapping	• Applying statistics to categorical and unbalanced data sets.	• Improved credibility of analytical conclusions from informal work. • "Exploring"?
	• Coding responses from informal work	• Enhances possibilities for aggregation, thus facilitating generalisation. • Enhances possibilities for stratification of sample for subsequent sample survey
Synthesising	• Blending the analytical outputs from informal and formal work into one set of policy recommendations. The outputs may be from Type A, B or C combinations.	• Higher quality policy recommendations

Design

In relation to design, there was some "swapping" : mapping was used to create village sampling frames, and also concurrent uses of contextual and non-contextual tools.

The function of participatory mapping to create village sampling frames was to reduce the time and cost required for household listing and enumeration. In addition, mapping served as a good ice-breaker. The maps were created on day 1 of the village studies. From this, households to be visited on day 2 were selected and enumeration areas delineated for individual team members.

Concurrent use of participatory and questionnaire techniques produced a blend of contextual information about changes in livelihood systems and food security over time with specific household level information on production, storage and marketing behaviour. The following table sets out the topics and tools used to tackle different aspects of the study.

Table 2: Concurrent use of research tools: Tanzania LGB study

Thematic area	Research approach
1. Changes in role of crop production in household food security strategies comparing 1985 with 1998.	RRA (Groups of men and women – some single gender groups - ranking strategies for 1985 and 1998)
2. Changes in farmers' perceptions of the importance of maize and cassava, comparing 1985 with 1998.	RRA (Groups of men and women – some single gender groups - ranking both crops against all other crops for 1985 and 1998)
3. Influence of P.truncatus on production, storage and marketing outcomes • Production levels • Role of P.truncatus in maize and cassava harvests • Role of P.truncatus in the choice of maize and cassava varieties • Role of P.truncatus in the duration of storage and volume of sales at farm level	Household sample questionnaire
4. Is P.truncatus still regarded as a problem? • P.truncatus in the context of major agricultural problems • P.truncatus in the context of other storage problems	RRA (Groups of men and women – some single gender groups - ranking strategies for 1985 and 1998)
5. Coping strategies for P.truncatus • Actellic Super Dust perceptions • Storage operations and structures	Household sample questionnaire

Note: The thematic areas appear in the table in the same order as they do in the final report.

Analysis

Attempts were made to apply statistics to the ranking data derived from the RRA exercises. Chi square tests and variants thereof were used to test for changes in rankings of the importance of *P. Truncatus* when farmers compared the situation at the time of establishment (in the mid – 1980's) with the present day. As an example, one exercise asked farmer groups to rank the importance of the pest in comparison to all other storage problems (a) at the time of establishment and (b) for the present day. The ranks were then compared and analysed using McNemar's test. Box 1 illustrates how ranking data for the past and present can be summarised.

Box 1: Perceptions of LGB

The final question of the group work looked at rankings of storage problems. One method of summarising the data is to consider the change in ranks between the past and present and these are summarised in Table 3.

Table 3: Comparison of change in ranking for importance of LGB in storage

		Present	
Past		Rank = 1	Rank > 1
	Rank = 1	24	13
	Rank > 1	2	4

Note: Any groups who had no ranking for LGB at one or both of the times were omitted.

The cells representing no change give no information about how the ranking of LGB has changed over the years. Only the bottom left and top right cells give information about change. McNemar's test (sign test in this case) can be used to test the null hypothesis of no change in attitude and the test gives a p-value of 0.0045 which indicates strong evidence for rejecting the null hypothesis. It is clear from Table 3 that there was a significant increase in the ranking, giving significant evidence for a reduction in the role of LGB as a storage problem.

Data Collection and Analysis Constraints

The study had to be designed around constraints of time, cost, terrain, weather, expertise, analytical and supervisory constraints.

Time

The size and coverage of the study was chosen with a time constraint in mind. It was required to present the results of the study at a conference in Beijing in July 1998. Working backwards from this, it was necessary to finish the field work by the end of February, so that there would be sufficient time to enter and analyse the data and write a survey report. In order to get the coverage required in the time required, survey teams could spend no longer than two days in each village. Table 4 sets out the basic components of the study and the timetable.

Table 4: Timetable for the study

Component	Responsibility	Timing
Initial visit by project leader to identify collaborators	Team leader	
Visit by NRI team to design and pre-test study components and train counterparts	Team leader (Post-Harvest specialist), Socio-economist	November 1997 (2 weeks)
Fieldwork	Field teams	December 1997 – March 1998 8 weeks
Coding and entering data from sample survey and RRA	UoG MSc. student	March – May 1998 5 weeks
Analysis of results and report writing	Socio-economist, NRI Entomologists	June – August 1998 6 weeks

Cost

Coverage of the survey was also chosen with cost constraints in mind. The total cost of the research was approximately £50,000. Given that three survey teams carried out the survey, the duration of the fieldwork could realistically be no longer than 2 – 3 months. Cost constraints also prevented a close supervision of the field survey work by the project socio-economist.

Terrain and weather

In certain areas, the terrain was known to be difficult. This, combined with the fact that the survey was taking place in the rainy season, meant that expectations of how much ground would be covered in the time available had to be modest.

Expertise and supervision

None of the team members were experienced in qualitative survey techniques, and levels of experience with questionnaires was also limited. This necessitated two weeks of training before the survey itself, however, even with this training, the level of tasks which could reasonably be expected of the teams was not sophisticated. The degree of sophistication and depth of investigation was also limited by the fact that it would not be possible for NRI staff to provide any field supervision.

Lessons Learned

As in all field based studies there were some methodological successes and some failures. A number of useful lessons were learned from both the successes and the failures.

Household mapping to create sample frame

The success of this exercise confirms that it is a simple and cost effective way of generating a sample frame whilst at the same time making actual planning of the enumeration process at the village level very easy.

Concurrent combinations

The study showed that the use of ranking in groups can be used to get a picture of a general context which helps in the interpretation of more specific household level data.

Statistical analysis of ranking data

The study demonstrated that statistical analysis can be applied to ranking data to strengthen the inferences drawn from visual inspection of the data. Problems with analysis of the data arose in comparison of ranking sets with different numbers of ranks: For example, we may have a situation in which 2 groups of farmers are asked to rank crops in terms of their importance to household food security. Group 1 may come up with 4 crops, whereas group 2 may come up with 8 crops. If the two groups are then asked to comment as to whether the rankings have changed over a given time period, then problems may arise in determining if there is any statistically significant differences in the changes recorded between one group and another.

Supervision

Another problem with some of the ranking exercises was esentially a supervision problem. The survey was carried out by two teams, one working in the northern regions, and one working in the southern regions. Although training had been given, one of the teams did not ask the correct questions on one of the ranking exercises. Consequently. it became very difficult to compare results across the two groups. A method was devised at the analysis stage to allow this to be done. The method converts ranks to scores, which can be controversial given the difference in meaning between a rank and a score.

More generally, quality of data collection was compromised because the project had to be managed from afar. Whilst the survey team leaders were both experienced field staff, they did not have the skills necessary to ensure uniformity and in depth investigation in all cases.

Design of survey instruments

The teams executed the group ranking exercises reasonably well. However, there were several confusions with respect to the questionnaire. These took considerable time to iron out as the data was being entered and analysed in the UK. In retrospect, the questionnaire was too open-ended and this led to confusion amongst some of the enumerators and some incompatible results, making analysis problematic on some questions. This was particularly noticeable when farmers were asked to provide explanations i.e. ***reasons for changes*** that had taken place. For example, the change from crib storage to sack storage may not have been reported in some cases as it would have been assumed by the enumerator that the person collating the data would know, automatically, that this must have occurred when the change from storing maize as cobs to that of storing shelled grain took place. On the other hand, some enumerators provided a long explanation as to why this had taken place, e.g. because a change of variety to HYVs necessitated quick sale to prevent damage in store, which meant storing loose grain, and therefore storing it in sacks for ease of access and convenience.

Time taken for collation and analysis of data

The time taken to collate and analyse the collected data was underestimated. Collation and analysis was compromised by capacity shortages at the UK end, and also by problems with the design of the survey instruments and some shortcomings perhaps in team training / supervision of field teams. Despite all this, however, sufficient analyses were performed to enable the main conclusions to be drawn within the required timetable.

Case Study 2

COMPUTERISING AND ANALYSING QUALITATIVE INFORMATION FROM A STUDY CONCERNING ACTIVITY DIARIES

S. Abeyasekera[1] and Julie Lawson-McDowall[2]

[1] Statistical Services Centre, University of Reading
[2] Farming Systems Integrated Pest Management Project, Malawi

COMPUTERISING AND ANALYSING QUALITATIVE INFORMATION FROM A STUDY CONCERNING ACTIVITY DIARIES

Introduction

Collecting information about daily activities of persons in households provides the means to learn about the division of labour between different groups of people, e.g. according to gender, or age, within some target area of interest. It also allows identification of constraints that households may face in learning and benefitting from technological innovations explored by development projects. The Farming Systems Integrated Pest Management (FSIPM) Project (1996-2000) in the Mombezi and Matapwata Extension Planning Areas in the Blantyre Shire Highlands in southern Malawi is one such project. Surveys and experimental work on-farm resulted in farmer assessments and bio-physical measurements that enabled a range of pest management strategies (PMS) to be evaluated. Since the ultimate suitability and adaptability of the PMS strategies would depend on farmers' resources and time availability to undertake particular IPM strategies, an understanding of daily sequences of activities within households was considered important.

A diary study was therefore undertaken during 1998. The information collected was quite voluminous and its computerisation became essential. This paper illustrates how a substantial body of qualitative information was handled in terms of coding and computerising the information, and demonstrates some simple procedures for data analysis. Recognising the data structure is emphasised as a crucial requirement to proper methods of data handling for subsequent statistical analysis.

Data Collection

The activity diary study was conceived within an ongoing case study of five household mbumbas, a mbumba being a cluster of households based on the unit of a mother, her adult daughters, their husbands and children. The five mbumbas included, respectively 3, 6, 4, 7 and 2 households. The number of persons within a household varied from 1 person to 9 members. Relevant data were collected by recording daily activities in a number of notebooks kept by a literate member of a household cluster group. The information was qualitatively recorded within four time periods per day to overcome the need to record actual times spent on each activity. Household members themselves suggested that four quarters was the natural division for their days' activities. Collecting the information qualitatively also meant that the cost of data collection was very low relative to similar studies involving direct observation by an outside researcher (Suphanchaimat, 1994). In the study here, the end of the recording period saw a total of activity information available for over 10000 person days at a cost of under £700.

Recognising the Data Structure with Respect to Qualitative Information

Appendix 1 gives an example of a typical day's diary of activities for two members from different household clusters. In computerising this information, the first requirement is to recognise any structure that resides in the data set. There is usually some form of structure, even if the data are essentially qualitative, since information is rarely collected in a wholly unorganised fashion. Samples are more typically chosen from certain major strata covering the population of individuals or households of interest. There is often a balance between the numbers of male and female headed households chosen. Wealth ranking exercises are common and such ranking divides the population into different wealth strata. Recognising these categories and any hierarchical sub-groupings (e.g. wealth ranking groups within villages and households within the wealth groups), helps to formulate the way in which the data can be organised and computerised.

For the case study being illustrated here, Table 1 shows the relevant data structure for one day of the diary records. This single day includes activities recorded in each of four separate time periods, i.e. morning, midday, afternoon and evening. The breakdown by individual is shown only for the first mbumba. The full data structure includes all days for which diary records have been made. The total number of records in the diary of activities for one individual member of a household is then 4 times the number of days for which data had been collected for that individual. The question then concerns the most appropriate way in which the information can be organised for computerisation.

Methodology for Computerising Qualitative Information

Generally, there are many advantages in using a proper database package (e.g. Access, dBase) to computerise and effectively manage large volumes of data. For computerising qualitative data, specialist software packages (e.g. NUDIST) do exist, but for studies that are relatively small, there is little justification for learning new software for this purpose. Procedures are needed that enable the analysis to be done quickly and efficiently. Our illustrations within this paper are therefore provided in terms of software with spreadsheet facilities. We have chosen Excel for this purpose since it is popularly used by many researchers, is readily accessible and is easy to use. It is nevertheless important to recognise that Excel has only limited database facilities and only a very limited range of procedures for statistical analyses.

For activity diaries, many activities may be recorded during a given time period of a day, and the numbers of activities recorded vary from person to person, from one time of day to another, and from day to day. The set of activities for a specific individual falling within one time slot in one day is referred to as a *multiple response*. They form a subset of all possible activities that a person might undertake. In situations where only a limited number of fixed activities are recorded, e.g. a fixed number of farming activities, it is possible to record each defined activity as a *dichotomous* variable, typically coded 0 and 1. If the activity is undertaken in the time slot concerned, then that activity is recorded as 1. Otherwise it is recorded as 0.

An initial step in planning the computerisation is to browse through a few of the diaries and make a list of the activities that individuals undertake. These activities may then be coded under a range of headings such as agricultural activities, household activities, etc. A full list of activities found in the FSIPM project clusters is given in Appendix 2. Once a draft list is ready, it is possible to plan the form in which the data should be entered.

Table 1. Structure of data for information collected from the activity diaries

Mbumbas	Households	Persons within hh	Morning	Midday	Afternoon	Late evening
1	1	1 2 3				
	2	1 . . 9				
	3	1 . . 7				
2	4 . . 7 9					
3	10 . . 13					
4	14 . . 20					
5	21 . . 22					
Total	22	91				

Essentially there are two levels of information. There is information collected about each person (age, sex, etc), so for the 91 persons included in the study, there are 91 rows of data, each row giving information corresponding to just one person. Each piece of information for a person occupies one cell of the spreadsheet. So several cells, all in one row are needed to contain all the different pieces of information for that person. For the diary study, the data at the "person" level for one mbumba is shown in Table 2. The identification code is always necessary to distinguish between different persons. In this study, the code has three components. The first gives the mbumba to which the person belongs (1-5). The second identifies the household within that mbumba, while the third and fourth digits correspond to the person number within the household. This information resides in a one file, separate from the file giving information on activities.

The second level of information corresponds to activities undertaken by each person at each of the four time slots in a day, repeated over many days. The bulk of the data resides here. The first step is to read through the activity diaries and note a code for each daily activity. Part of the data sheet used for recording purposes is shown in Table 3.

The data entry took place by allocating one column in an Excel spreadsheet for each activity undertaken during a particular time period of the day. Most individuals carried out 2 or 3 tasks within a specific time period. The maximum number of different activities that any one person did was nine. Hence 9 columns of data were needed to enter the codes for each activity. Figure 1 below shows a typical set of data after it was computerised. The numbers shown under the activity columns are codes for the range of different activities that household members do. Although sex and age information has been included in this file, this is not strictly necessary as the same information resides in the file containing person information. This can then be accessed by linking the two files via the person identification code.

Table 2. Information about the case study household members

OVE-RAL L IDEN TIFIE R	CLUS -TER NO.	HOUSE HOLD NO.	PER-SON NO.	HOUSE-HOLD NAME	INDIVIDUAL NAME	AGE	SEX	RELATIO NSHIP TO H'HOLD HEAD
CLUSTER: MUTHOWA								
1101	1	1	01	Muthowa	Mr Muthowa	70	M	head 1
1102	1	1	02	Muthowa	Mai Machemba Muthowa	60	F	spouse 2
1103	1	1	03	Muthowa	Musowa Bulaya	17	M	child 3
1204	**1**	**2**	**04**	**Naluso**	**Mr Naluso**	**37**	**M**	**1**
1205	1	2	05	Naluso	Agnes Machemba Naluso	38	F	2
1206	**1**	**2**	**06**	**Naluso**	**Elaton Naluso**	**21**	**M**	**3**
1207	1	2	07	Naluso	Christopher Naluso	16	M	3
1208	1	2	08	Naluso	Esther Naluso	14	F	3
1209	1	2	09	Naluso	Juma Naluso	10	F	3
1210	1	2	10	Naluso	Jimmy Naluso	7	M	3
1211	1	2	11	Naluso	Victoria Naluso	6	F	3
1212	1	2	12	Naluso	Felisita Naluso	3	F	3
1313	1	3	13	January	Mr January	45	M	1
1314	1	3	14	January	Esther Machemba January	35	F	2
1315	**1**	**3**	**15**	**January**	**Roderick Mkwezelamba**	**20**	**M**	**3**
1316	1	3	16	January	Joyce Mkwezelamba	10	F	3
1317	1	3	17	January	Charles Mkwezelamba	7	M	3
1318	1	3	18	January	Dyson Sipili	4	M	3
1319	1	3	19	January	Chrissie January	1	F	3

Table 3 - Data sheet for coding diary activities

DATE			DATE			DATE		
1			1			1		
2			2			2		
3			3			3		
4			4			4		
1			1			1		
2			2			2		
3			3			3		
4			4			4		

Figure 1. An example of the way data on activities are organised.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
1	PERSON	SEX	AGE	DATE	TIME OF DAY	ACTIVITY 1	ACTIVITY 2	ACTIVITY 3	ACTIVITY 4	ACTIVITY 5	ACTIVITY 6	ACTIVITY 7	ACTIVITY 8	ACTIVITY 9	ACTIVITY 10
2	4517	0	28	01.03.98	1	15	12	20	14	16					
3	4517	0	28	01.03.98	2	13	12	20	16	17					
4	4517	0	28	01.03.98	3	13	20	21	11						
5	4517	0	28	01.03.98	4	29									
6	4516	1	35	01.03.98	1	27	11								
7	4516	1	35	01.03.98	2	24	11	17							
8	4516	1	35	01.03.98	3	21	4								
9	4516	1	35	01.03.98	4	29									
10	4517	0	28	02.03.98	1	5	11								
11	4517	0	28	02.03.98	2	12	20	16	14	17					
12	4517	0	28	02.03.98	3	7									
13	4517	0	28	02.03.98	4										
14	4516	1	35	02.03.98	1	5	11								
15	4516	1	35	02.03.98	2	21	11								
16	4516	1	35	02.03.98	3	7	11	23							
17	4516	1	35	02.03.98	4										

An alternative method for computerising the activity information would have been to have had one column allocated for each activity. So for example, activity code 1 above could refer to "Cultivating/checking the field". This would be one column and entries in that column would be *dichotomous*, i.e. just two possibilities, coded as 1 or 0 according to whether or not the individual had done the particular activity on that day, within the time period under consideration. This would have led to 47 columns of data rather than the nine columns above, the 47 columns corresponding to codes 1 to 47 used for the 47 different types of activities undertaken. Further,

- the analysis becomes more complicated, even with respect to producing simple summaries;
- data entry becomes more time consuming (even utilising the *split* facility in Excel) since the correct column for the activity has to be located before the entry is made.
- the data file becomes quite large with many empty cells for all the rarely occurring activities.
- data quality checks become more difficult since an entry in the wrong column for a particular activity can only be identified by returning to the original diary notebooks.

In this particular study, information for each month was entered on a separate sheet in an Excel workbook. There was a separate workbook file for each author, i.e. the literate member who was responsible for recording daily activities for a specified number of household members. In total there were 12 authors.

Data Analysis

Only a small component of the analysis undertaken is presented in this report. The aim is to indicate possible types of data summary. The analysis was carried out using the statistics software package SPSS (version 9) as this software has good facilities for dealing with *multiple responses*, e.g. when several activities are undertaken within each quarter day time period. SPSS also has extensive facilities for data analysis.

An initial look at the data files for each month revealed several incomplete records for specific days of the month or for particular individuals. This was largely due to some inefficiency in recording procedures. Given the complexity and depth of detail in the information collected, it was decided to restrict the data, within each month, to persons for whom information was available in all four quarters of the day for the same specific number of days in a month. This reduced the total number of data items to numbers shown in the last two columns of Table 4. The information is split into four person groups, i.e. boys, girls, men and women, where boys and girls are identified as those who are 16 years of age or younger. The most complete data can be seen to come from the months of February to June 1998 covering activities of more than 30 persons for all days of these months except in June.

Table 4. Persons and days for which complete records are available

Month	Boys	Girls	Men	Women	Total no. persons	No. of days of complete records
January	2	4	12	13	31	21
February	2	4	12	15	33	28
March	2	5	17	16	40	31
April	5	6	16	18	45	30
May	8	10	22	18	58	31
June	8	10	22	18	58	29
July	0	1	5	6	12	25
August	0	1	8	7	16	31
September	0	1	8	8	17	30
October	0	1	8	8	17	31
November	0	1	8	8	17	30
December	0	1	11	9	21	29

Overall summary of all activities

As a first step in the data analysis, the multiple responses within each quarter day, corresponding to activities that had taken place in that quarter, were totalled. This gave the number of quarter-day units in which the activity took place, totalled over all persons and all days specified in Table 4 for a given month. However, interpreting this information across months was difficult because of the variation in the number of persons and dates.

The last 2 columns in Table 4 and the total number of activities are shown in Table 5, together with additional information concerning just one activity, namely the activity "Drawing Water" (see Appendix 2, code 13). This table illustrates one way in which the information could be summarised in a more meaningful form to allow comparisons across months and within months. The penultimate row of Table 5, i.e. the average number of "drawing water" instances per person day, can be used for month to month comparisons. It is also useful to see how often an activity is undertaken (e.g. row e of Table 5) from amongst all activities in that month (row c of Table 5), and whether its relative importance amongst activities in a month varies across months. For "Drawing Water", the results appear in the final row of Table 5.

The apparent increase in the proportion of time spent on drawing water in the late months of the year is more likely to be associated with the increased proportion of women in the sample at this stage rather than any other factor. This shows the importance of EITHER keeping a consistent cohort OR analysing boys, girls, men and women separately.

Table 5. Using *water drawing* as an activity to illustrate different forms of summary

Activity	January	February	March	April	May	June	July	August	September	October	November	December
a =Number of persons	31	33	40	45	58	58	12	16	17	17	17	21
b =Days of complete records	21	28	31	30	31	29	25	31	30	31	30	29
c =Total no. of all activities	4260	6591	9070	10176	15211	14335	3020	5371	5399	5701	5523	6360
d =Person days = (a × b)	651	924	1240	1350	1798	1682	300	496	510	527	510	609
e =*Drawing water* instances	294	437	693	673	1304	1152	222	433	488	540	528	520
Average per person day=e/d	0.45	0.47	0.56	0.50	0.73	0.68	0.74	0.87	0.96	1.02	1.04	0.85
%time drawing water=100e/c	6.9	6.6	7.6	6.6	8.6	8.0	7.4	8.1	9.0	9.5	9.6	8.2

Activity summaries for major tasks

Information similar to that shown in the final row of Table 5, are shown in Table 6, totalled over major categories of tasks (categories as given in Appendix 2). It may be seen that for some tasks, there is a clear difference between the first six months of the year and the last six months of the year. This may either be due to the cohort of persons involved in these two periods being different, or it may be an indication of real changes in activity at different times in the year. Latter is likely to be a substantial part of the explanation for the proportion of activities on agricultural tasks and livestock caring being higher later in the year, while proportion of instances where rest or social activities take place is correspondingly lower in the July to December period.

Table 6. Number of instances when activities took place under each activity category, expressed as a percentage of the total number of activities undertaken in the month.

Activity category	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
Agriculture total	8.5	7.5	9.4	8.7	4.6	4.7	7.8	11.2	13.7	12.1	14.5	15.9
Livestock total	1.6	1.4	1.1	0.6	0.7	1.3	1.9	2.9	3.8	3.5	3.3	2.8
Domestic activities	13.5	13	15.8	13.1	15.7	15.2	22.3	21.2	21.2	22.3	22.1	17.8
Food processing	26.8	30.7	30.1	30.6	30.8	30.7	32.4	29.9	29.8	30.3	28.7	28.4
Other domestic work	8.0	8.0	9.3	8.6	13.5	13.4	14.2	14.0	10.6	12.8	11.6	10.5
Marketing	2.7	2.0	0.6	0.6	1.4	1.1	0.3	0.4	0.2	0.8	0.6	1.9
Rest/social/sickness	25.6	24.3	22.2	27.8	24.1	23.7	15.5	16.0	13.4	15.8	15.5	19.1
Off-farm business	2.9	2.8	2.7	1.8	2.4	2.4	2.5	2.6	2.8	1.7	2.4	2.1
Obligatory duties	1.1	2.3	1.9	1.3	2.3	1.6	2	0.5	2.6	0.3	0.4	0.3
School	1.1	1.7	2.2	1.4	2.2	2.5	0.2	0.2	1.3	0.0	0.0	0.0
Doing ganyu	0.8	0.4	0.5	0.6	0.1	0.2	0	0.1	0.2	0.0	0.1	0.6
Other social duties	1.5	1.5	1.0	1.6	1.4	1.1	0.8	0.7	0.5	0.6	0.3	0.5
Absent without reason	1.5	1.5	1.0	1.6	1.4	1.1	0.8	0.7	0.5	0.6	0.3	0.5

Frequency of activities on a day by day basis

Summaries discussed in 5.1 and 5.2 above refer to quarter-day frequencies. An alternative is to consider the number of days in each month when major tasks were undertaken by (say) adults in the sample, i.e. persons more than 16 years of age. Note that this does not imply a whole day was spent on the activity, but that some time in the day was devoted to the activity.

The available data cannot be used directly for this purpose since data for each month are based on a different number of adults. Moreover these days are based on all days of the month for most months, but only for a smaller number of days in January (21 days), June (29 days), July (25 days) and December (29 days). Person days in these months must therefore be scaled to the correct number of days in that month before further analysis.

In January for example, 135 days concerned involvement in house cleaning activities by 25 adults. But complete data records for this month are available only for 21 days.

Hence an estimate of the total number of days in January when cleaning was an activity during the day

$$= \frac{135}{21} \times 31 = 199.29$$

But this number of days concerns 25 adults in the sample. Hence the average number of days in the month when cleaning was an activity for an adult

$$= \frac{199.29}{25} = 7.97 \approx 8 \text{ days.}$$

This summary measure shows how often specific activities were undertaken in a month. Table 7 gives the corresponding results split by gender for the months of January to June. Days spent on agricultural activities are seen to be higher for men than for women, particularly with respect to planting of crops. Livestock caring seems to be predominantly a male activity, whereas picking/chopping firewood, drawing water, cleaning, preparing food and other food processing related activities were tasks mainly undertaken by the women. There is some indication that more days are spent by women than by men selling market produce at the market or at home.

Conclusions and Lessons Learned

The main purpose of this guide has been to demonstrate how qualitative data arising from activity diaries can be computerised via a suitable coding exercise. Some indications are given of how the resulting data may be summarised. However, interpreting results from studies of this type must pay due attention to the data collection method and the quality of the data. The latter depends largely on the accuracy and care given by authors in recording the information. Of the 12 authors, only one was a female, since literate members tended more frequently to be the male members in the household. Any possible bias arising from the gender of the author cannot be ascertained from this study.

Further care is needed in comparing results across the months of the year, since the cohort of people to whom the results apply differ in the different months. In particular, very small sample sizes were obtained in the months from July to December. If the cohort which corresponds to these months differs in any substantial way from the larger cohort in the first six months of the year, then any observed differences between the two six month periods may be just a spurious effect. Some of the results do indicate such a difference, but it would be difficult, because of the above reasons, to judge whether or not this reflects a true difference in activities undertaken in the two time periods.

Retrospectively, we recognise some clear lessons that this work has provided. It was clear that the work undertaken was ambitious in expecting high quality and complete data from a study which required very detailed information on daily household activities over a considerably lengthy period of time. The benefits of better quality data far outweigh the benefits of added information on every single day of the month for all months of the year. A suitable chosen set of sampling dates within each month of (say) about 10-12 days, inclusive of one week-end, would have been sufficient to learn about the variation in household activities from month to month. The intensive nature of the data collection process, although very cheap, also meant that it was not possible to supervise the work effectively to have greater confidence on the accuracy of the data. Some data which were very questionable were omitted as a result at the data coding and entry stage. The data analysis was also lengthy and complicated because of the lack of complete records. A smaller study might well have led to better quality data and would have been less costly in terms of the analysis time spent by ex-patriate staff.

Finally, it should be noted that the reported results are all descriptive summaries of the information collected. There is no basis for applying formal statistical inferential procedures since the information has been collected from a case-study of a group of cluster households. However, the results do provide some indication of the range of activities undertaken by resource poor households and the division of labour between different activities and gender groups.

References

Suphanchaimat, N. (1994) Household record keeping as a means of understanding farmers decision making. In Chapter 22 of Tools for the Field: Methodologies Handbook for Gender Analysis in Agriculture. Edited by Feldstein. H.S. and Jiggins, J. IT Publications Ltd.

Table 7. Average number of days in each month when a particular task was an activity for a man or woman (January – June)

Activity	January		February		March		April		May		June	
	Men	Women	Men	Women	Men	Women	Men	Women	Men	Women	Men	Women
1. Cultivating	4.7	2.3	4.7	2.3	4.1	2.6	2.6	0.5	1.8	0.6	3.8	2.1
2. Planting	1.6	0.9	3.3	1.1	5.6	1.9	2.9	0.6	0.8	0.2	0.8	0.1
3. Tilling/ridging	3.1	0.7	4.9	0.5	5.7	1.2	3.3	0.6	0.7	0.1	2.1	0.7
5. Weeding	3.8	0.3	2.3	0.5	2.9	1.7	4.1	1.9	1.1	0.9	0.4	0.1
6. Banking	3.9	4.2	0.1	0.3	0.2	0.0	0.1	0.0	0.1	0.0	0.0	0.0
7. Fertilizing	2.2	1.2	1.6	1.0	2.5	0.3	2.2	0.1	1.4	0.3	0.4	0.1
8. Using pesticides	1.4	0.0	0.5	0.0	0.6	0.1	0.4	0.1	0.5	0.0	0.7	0.0
9. Harvesting	0.7	0.7	1.3	1.9	2.2	3.1	6.4	7.4	2.4	2.4	3.1	3.0
11. Livestock caring	6.3	0.5	6.1	0.5	4.2	0.6	3.3	0.2	3.1	0.6	3.4	1.5
12. Picking/chopping firewood	0.5	2.4	0.8	3.5	1.5	4.7	0.6	3.8	0.7	3.8	0.5	4.0
13. Drawing water	3.2	19.3	3.2	15.2	5.3	17.5	5.4	17.1	9.8	17.9	8.7	15.3
14. House cleaning	2.7	12.8	2.4	10.8	3.6	15.3	4.2	14.9	4.9	16.5	4.7	13.0
16. Preparing food	2.1	24.2	4.3	20.7	4.1	24.4	3.3	23.2	4.0	25.5	5.9	22.2
18. Going to milling machine	0.0	1.7	0.0	1.3	0.3	1.5	0.1	1.1	0.2	1.4	0.1	1.0
19. Pounding etc	0.0	0.5	0.3	1.3	1.4	4.2	2.6	7.6	0.8	4.4	0.6	3.3
24. Maintaining/making tools	0.1	0.0	0.3	0.0	0.8	0.1	0.6	0.0	1.1	0.1	1.2	0.1
26. Marketing home produce	3.2	7.4	3.0	3.9	0.9	1.7	0.5	1.5	3.4	3.8	2.9	2.8
31. Business	2.7	4.3	2.3	3.9	1.8	2.4	2.7	1.1	5.1	2.1	4.5	1.7
43. Caring for children	0.1	1.6	0.3	2.3	0.5	1.6	0.6	1.7	0.5	3.4	0.3	2.5
46. Watering crops	0.1	0.0	2.8	0.7	4.5	0.9	6.3	0.6	7.6	2.1	6.9	2.1

APPENDIX 1. An example of a typical day's diary of activities for members of two household clusters

Cluster: Kalonga
Recorder: Namangwiyo
Member: Alekeleni
Date: 25.05 98.

Morning

- Went to draw water for domestic purposes
- Went to work on a tobacco estate

Afternoon

- Cooked and ate relish and nsima
- Sweeping around the house

Late Afternoon

- Drew water
- Cleaned plates and pots
- Cooked relish

Evening

- Prepared nsima and then went to sleep

Cluster : Muthowa
Author : Naluso
Member: Estere
Date: 18.06.98.

Morning

- Drew water
- Cleaning her face
- Cleaned plates and pots
- Cooked porridge
- Cooked beans

Afternoon

- Cooked nsima before the beans were ready

Late Afternoon

- Went to get relish; black jack leaves and she cooked them ready for lunch
- Drew water
- Had a bath

Evening

- Prepared nsima for the whole family
- Placing utensils(pots, plates, spoons) and other properties(chairs, axe, hoes) accordingly

APPENDIX 2. Codes for the different activities undertaken by household members

CODE	AGRICULTURAL CROP ACTIVITIES
1	Cultivating/checking the field
2	Planting/transplanting crops
3	Tilling/ridging
4	Sowing seeds/preparing seeds/materials for planting/maintaining nursery
5	Weeding
6	Banking
7	Fertilizing a field/dimba or applying manure/watering crops
8	Applying pesticides/removing infected plants
9	Harvesting or hauling crops from the fields, e.g. maize, cassava, potatoes, etc
10	Buying agricultural inputs/collecting starter packs and other benefits
46	Watering crops
11	**LIVESTOCK**
	Feeding animals / Dipping animals
	Gathering animal feed / Cleaning a khola / Leading livestock to a khola
	DOMESTIC ACTIVITIES
12	Gathering firewood and chopping wood
13	Drawing water
14	Sweeping, washing, weeding, cleaning, chasing mosquitoes, killing insects etc.
	FOOD PROCESSING
15	Harvesting green crops for relish
16	Preparing food, making porridge
17	Eating
18	Going to a milling machine
19	Pounding, winnowing, unsheathing, shelling, drying maize
20	Making fire
	OTHER DOMESTIC ACTIVITIES
21	Bathing self or child. Shaving.
22	Ironing, visiting a taylor, paying debts
23	Building/looking for building materials, fencing
24	Making/maintaining farm tools
25	Shopping at the market or grocery

CODE	MARKETING
26	Selling market produce at the market or home
	REST/SOCIAL ACTIVITIES/SICKNESS
27	Attending church activities
28	Chatting, visiting friends, families, places, going out for social activities
29	Playing, resting, sleeping, basking in the sun, playing games
30	Seeking health services, medicine from the bush, suffering/sick
	OFF FARM BUSINESS ACTIVITIES
31	Doing business, baking, checking on debtors, tinsmithing
32	Supervising labourers, paying labourers
33	Reporting for regular work
34	Seeking employment
	OBLIGATORY SOCIAL DUTIES
35	Attending a funeral rite
36	Escorting friends/visitors
37	Preparing medicine
38	Looking after the sick, visiting the sick
39	Hearing cases
	SCHOOL
40	Attending school / related activities
41	**DOING GANYU**
	OTHER SOCIAL DUTIES
42	Assisting in emergencies, attending interviews
43	Caring for the children/delivering a baby
47	Taking a message/seeking postal services/writing
45	**ABSENT WITHOUT REASON**
46	Watering crops (also listed under agricultural activities)

Case Study 3

GENERALISING RESULTS FROM MATRICES OF SCORES

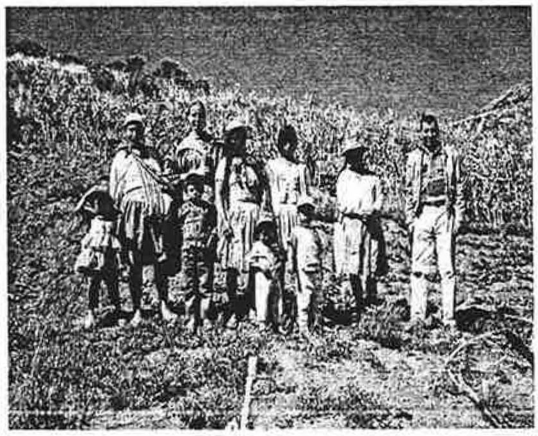
C.E. Barahona, Statistical Services Centre, University of Reading

GENERALISING RESULTS FROM MATRICES OF SCORES

An example

In 1998 we joined a field visit to the village Los Pinos organised as part of the research project "Participatory improvement of soil and water conservation in the temperate valleys of Santa Cruz, Bolivia" (DFID research project R6638). The objective of the trip was to visit trial plots where farmers where experimenting with improved hillside management techniques and to carry out a participatory assessment of the trials.

Farmer in a plot with live contour barriers



Group of farmers and researcher with crops and live barriers in the background

The main tool for the participatory assessment of the techniques was a matrix of scores. To construct the matrix, members of the community attending the workshop were split in two groups by gender. After having visited the plots in the field, the methodology used for the construction of the matrix in each of the subgroups was as follows:

1. To list the techniques for soil and water conservation under trial in the community.

2. To brainstorm about the criteria on which they wanted to assess the techniques.

3. To construct an empty matrix with the list of techniques at the top and the criteria down on the left hand side column.

4. To discuss and assign scores (using pebbles) from 0 to 5 to each technology for each one of the criteria listed. A high score indicates that the technology is considered good. Each cell in the matrix was scored independently of the others, i.e. each group discussed for each combination of criteria and technology what score was to be given regardless of the other criteria or techniques under trial.

Eventually each group produced a matrix of scores that allowed the participants to assess the techniques they were trying out in their plots and to learn about the diversity of criteria generated by male and female members of the community when listing desirable characteristics of the soil and water conservation techniques.

The discussion produced the following immediate results:

- Provided information to the participants (community and visitors) about the techniques under trial relevant to the local conditions and

- Enabled participants to carry out the analysis of the performance of the technologies in their own environment and under their own criteria. This would contribute to the villager's and researchers' understanding of the techniques and their role in improving soil and water conservation in the village.

From the point of view of the researchers these results undoubtedly contribute to the achievement of the project purpose stated as "To develop and promote economically viable land, soil and water management practices through participatory diagnosis, experimentation and evaluation of improved techniques based on both indigenous and scientific knowledge" (Lawrence, 1996). However in the project memorandum the researcher also points out that "The research methods and techniques developed in Bolivia will be of value to development agencies and researchers working in other areas". To achieve the latter the researchers will need to be able to generalise what is learnt in the villages that participate in the study to other areas not currently included in the study. The research project covers three localities in the temperate valleys of Santa Cruz: Los Pinos, Chacopata and Pozuelos (Eid, 1997).

I am aware that my short involvement in the project (one visit only) does not provide me with enough information to judge its overall success. To assess the research project is not the objective when writing this paper, nor do I claim to have enough information to attempt to do so. My objective is to use this example to discuss what can be done to achieve generalisation of results such as those obtained in Los Pinos. In addition I will discuss the methodological implications of generalisability on the design of research projects.

Why Do We Claim That Statistical Methodologies Lead to Generalisable Results?

It is broadly accepted that statistical methodologies have the potential to generate results that can be applied to a wider population than those directly taking part in the research process. Barahona (2000) points out that the main reasons that support this claim are:

1. Information is obtained from an appropriate number of study units, not from one study unit only.

   The main problem with making generalisations from results of one study unit (a case study) is the difficulty in determining the level of uncertainty of the findings if applied to a different group of study units.

2. The study units selected for the study are assumed to be representative of the target population.

   To achieve this we require a clear definition of the units that make up the population of interest, I will refer to this population as the "target population".

This is usually well done in most research projects through a list of characteristics of those who would benefit from the research outputs, the specification of a problem affecting a population, the definition of a geographic or ecological area, definition of biophysical characteristics of the units, or a combination of these or other definitions. The study units eventually could be defined as farmers, groups of farmers, villages, agro-ecological environments, etc. The resulting target population is the collection of all the study units that fulfil the required characteristics.

Once the target population is defined, the selection of those taking part in the study itself should take place. This stage is crucial in so far as generalising results is concerned and will determine whether it is possible to assume that the study group is representative of the target population.

Statistical methods rely on random selection of the study units to get a representative study group. It is important to point out that the principle of random selection does not mean the indiscriminate use of simple random sampling but the use of random selection of units at appropriate levels.

3. All study units provide information based on the use of a consistent methodology.

   For example in traditional survey work the use of this consistent methodology takes the form of a questionnaire with carefully defined questions that are to be asked without alteration to all selected respondents. However we do not argue for the use of questionnaires for all research purposes and acknowledge that they have limitations and weaknesses. The important point here is that all the study units should provide the same type and quality of information. This depends on using a well-defined, consistent methodology for collecting the information throughout the study.

4. What is said about the population is acknowledged to be imprecise. However the extent of that imprecision can be estimated. In other words, we are aware that the results are not exactly right, but we can estimate how wrong they are likely to be.

   In this context precision should be interpreted as the degree to which our conclusions would hold if the case study were to be repeated in different study units within the same target population. Intuitively, if the results were to show high variability between different study units, then the precision of the results would be considered low. On the other hand if the results were consistent across study units their precision would be regarded as high.

The assumption that the study group is representative of the target population, together with the information about the uncertainty of the findings are the two main points that allow researchers to claim that their results are generalisable. Eventually these two points will allow users of the results to decide whether those results are applicable to their own reality too.

A more detailed discussion of the points above can be found in the theme paper "Sampling issues in participatory research".

Towards generalisation of participatory results

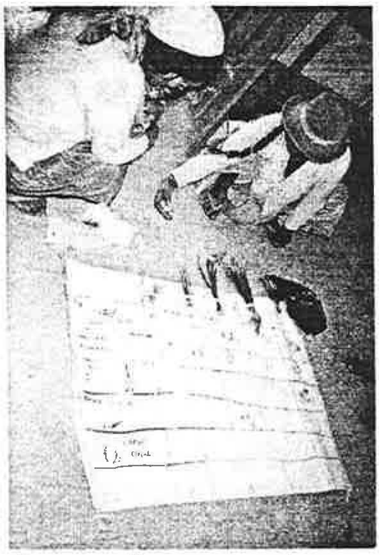
Going back to the matrices of scores in Los Pinos, we will discuss methodological considerations that research teams using matrices of scores should take into account to maximise the potential to produce generalisable results.

Why matrices of scores?

These are well-established participatory tools that produce both numerical and non-numerical results. The non-numerical results have been discussed at length by practitioners of participatory methodologies elsewhere and we will not deal with them in detail. To discuss the information that can be collected with this tool, let's first examine the matrix itself.

The matrix of scores is a device with a flexible but well-defined structure that imposes a series of rules to the discussion and organises the analysis made by the participants. It does this without seriously restricting participation. The rules imposed to the discussion and analysis in Los Pinos were as follows:

1. The techniques were predefined by the techniques on trial in the community and were to be listed at the top of the matrix in the first row.
2. The criteria for the assessment of the techniques were to be discussed and agreed upon by the members of the group. Eventually the criteria were to be put on the first column of the matrix.
3. No limit was imposed to the number or type of criteria that could be included in the matrix. This was because capturing the diversity of criteria for assessment of the techniques was one of the research objectives.
4. The scoring system was predetermined to be from 0 to 5, with 0 being a low score and 5 a high score, the higher the score the more benefit or desirability of the technique under the corresponding evaluation criterion. The group was to discuss and agree on the score to be given to each combination technology by criteria.



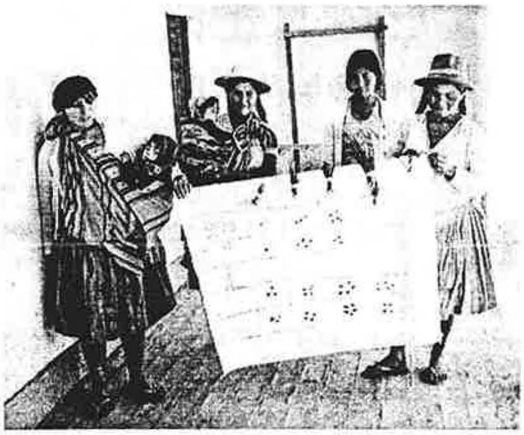

Groups of women and men completing the matrices of scores separately.

The result of imposing these rules to the methodology for the discussion is that regardless of where it takes place the information will be of the same nature and suitable for cross-site analysis.

The set of techniques at the top of the matrix might change slightly from community to community and the criteria used to assess the techniques will change form group to group but this does not affect the ability of the researcher to integrate the results from several communities.

It is important to notice that we are focussing on the researcher as the main user of this information. However this emphasis does not affect the use and benefits that the farmers will get from their analysis at local level and the understanding of the benefits of the technologies for the communities.

The information collected at Los Pinos is summarised below (after some rearrangement to put the techniques in the same order for both groups).



Group of women with their finished matrix

Group of men with their finished matrix

Group of women

	Tech1	Tech2	Tech3	Tech4	Tech5
Produces fodder for sheep	2	3	5	5	5
Produces fodder during the dry season	3	4	4	5	5
Source of Firewood	0	0	0	0	5
Retains soil in the contour	3	4	5	5	5
Grows fast	3	4	5	4	5

Group of men

	Tech1	Tech2	Tech3	Tech4	Tech5
Good for cattle	3	3	5	4	1
Retains soil in the contour	1	1	4	2	5
Reduces diseases in the crops	0	0	4	0	5
Grows fast	1	1	0	1	4
Tolerates draught	1	2	4	2	2
Tolerates frost	0	0	0	0	0
Good yield for fodder of cows and oxen	1	1	4	1	5

The use of matrices of scores as described above is a first step towards the achievement of generalisable results in studies such as the one used in this example. It is however not enough to attain this objective. Another requisite is to be able to measure variability and precision.

Data analysis

Let's take the results from the group of women in Los Pinos. Two main results that can be read from the matrix are:

a. Women in the village value soil and water conservation techniques according to five criteria: 1) produces fodder for sheep, 2) produces fodder during the dry season, 3) source of firewood, 4) retains soil in the contour and 5) grows fast.

b. When using these criteria to assess the techniques under trial they prefer technique 5 as it receives the maximum score all the time.

To generalise these results we would need to answer the following questions:

- Can the researcher conclude that the criteria used in Los Pinos are a comprehensive list of desirable characteristics for the soil and water conservation techniques to be developed and promoted among the target population?
- Would technique 5 be consistently preferred by farmers in the target population?

With the results from Los Pinos it is not possible to answer these questions. The researcher would need to run similar participatory exercises (using the same methodology!) to gather information about variability and precision.

To exemplify how to manage and summarise results from several participatory exercises, results from 20 discussions carried out with other groups of women will be used (these data are simulated and do not come from the work of the research team in Bolivia). The choice of 20 groups for this example is arbitrary; how to choose an appropriate sample size is discussed in detail the theme paper "Sampling issues in participatory research".

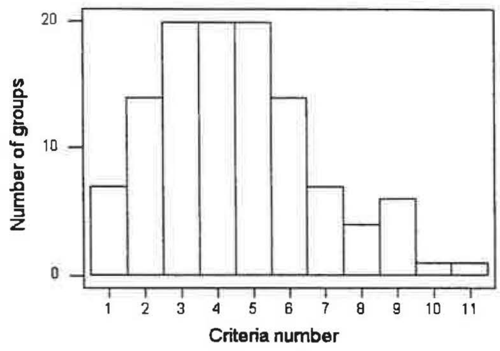
Which evaluation criteria are of relevance for the target population?

We will first focus our attention on the analysis of criteria. For simplicity and generalisability criteria will be referred by numbers 1, 2, 3, ... etc. A set of hypothetical results for 20 groups are summarised below.

	Group																				
Criteria	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	Total
1		✔			✔			✔	✔			✔						✔		✔	7
2	✔	✔	✔	✔		✔	✔			✔	✔	✔	✔		✔		✔	✔		✔	14
3	✔	✔	✔	✔	✔	✔	✔	✔	✔	✔	✔	✔	✔	✔	✔	✔	✔	✔	✔	✔	20
4	✔	✔	✔	✔	✔	✔	✔	✔	✔	✔	✔	✔	✔	✔	✔	✔	✔	✔	✔	✔	20
5	✔	✔	✔	✔	✔	✔	✔	✔	✔	✔	✔	✔	✔	✔	✔	✔	✔	✔	✔	✔	20
6	✔	✔		✔	✔	✔	✔	✔	✔		✔	✔		✔	✔				✔	✔	14
7					✔	✔	✔	✔		✔	✔									✔	7
8	✔			✔					✔											✔	4
9						✔	✔	✔				✔	✔							✔	6
10				✔																	1
11			✔																		1

Each tick indicates that the corresponding group chose the criterion during their discussion. For example group 1 chose criteria 2, 3, 4, 5, 6 and 8.

The information can be presented in a histogram as follows.



If the study group could be assumed to be representative of the target population the researcher can now start to make qualified statements about what the groups of women in the target population are likely to regard as important criteria. From the example it could be argued that the recorded criteria can be split into three groups:

Criteria consistently used by all groups	3, 4 and 5
Criteria used by more than 50% of the groups but not by all of them	2 and 6
Criteria appearing with lower frequencies	1, 7, 8, 9, 10 and 11

Understanding why criteria 3, 4 and 5 were consistently used would help in the selection of technologies that are likely to be of relevance to the communities for whom the benefits of the research are intended. This analysis must be informed by the qualitative information gathered by the researcher during the participatory exercises.

Criteria 2 and 6 might be relevant to communities under specific conditions. An exploration of communalities and differences between the groups that used these two criteria and the groups that did not use them might give indications about communities with specific requirements and lead to specific conclusions for particular sections of the target population.

Criteria with low frequency of occurrence are likely to reflect local requirements. These may be relevant to the specific group that used them but their importance cannot be said to be applicable to the target population.

Which technique do the discussion groups prefer?

Another issue that might require generalisation is the comparison of techniques. Often this is better discussed taking one criterion at a time. However, in some cases the researcher might want to find out if a technique is preferred when assessed on multiple criteria. In those cases the set of criteria should be carefully considered so that the results are of practical use. If such a set can be defined a summary of the scores needs to be defined, often the total or the average of all the scores over the criteria are used. Alternatives to this are discussed in Theme Paper 4.

We deal with the case of the comparison of technologies on the basis of one criterion only. The procedure for comparing technologies across sets of criteria could be similar. For example let's assume that "provision of fodder for sheep" is a criterion consistently selected by all the groups of women. Hypothetical results of 20 discussion groups for this criterion are presented below.

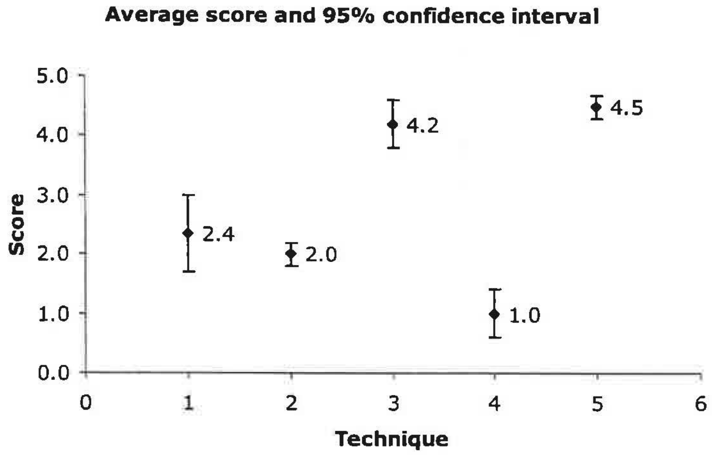
Scores for criterion "provides fodder for sheep" for all groups

Group	*Tech 1*	*Tech 2*	*Tech 3*	*Tech 4*	*Tech 5*
1	2	1	5	0	4
2	5	2	3	2	5
3	1	3	5	0	5
4	1	2	5	1	5
5	1	2	5	1	5
6	2	2	3	0	5
7	5	2	3	1	5
8	4	3	4	0	5
9	3	2	3	0	4
10	2	1	3	1	4
11	1	2	4	2	4
12	3	2	4	2	4
13	1	2	5	2	5
14	2	2	5	0	5
15	1	2	5	1	5
16	4	2	4	2	4
17	4	2	5	1	4
18	1	2	3	1	4
19	2	2	5	1	4
20	2	2	5	2	4

The analysis of these data could start with simple descriptive statistics of the scores for each technique. For example

Technique	N	Mean	95% confidence interval for the mean
1	20	2.4	(1.7, 3.0)
2	20	2.0	(1.8, 2.2)
3	20	4.2	(3.8, 4.6)
4	20	1.0	(0.6, 1.4)
5	20	4.5	(4.3, 4.7)

These results are easier to interpret if displayed graphically. The figure below shows the average score given to each technique together with a bar that represents the 95% confidence interval for that average.



For example for technique 1, the average score is 2.4 and the 95% confidence interval varies from 1.7 to 3.0. This should be interpreted as follows:

> The 20 groups gave technique 1 an average score of 2.4. However we know that if we were to ask all the members of our target population to score technique 1, the average would be different. Using the answers from our study group it is possible to say that the interval 1.7 to 3.0 has a chance of 95% of covering the average we would get from the whole target population.

Confidence intervals are a way of showing the precision of our estimated average. For example, the wider the interval the lower the precision of the study group average in estimating the target population average.

If we look at the relative position of the confidence intervals for the techniques, some of them overlap on the vertical scale. For example the confidence intervals for techniques 4 and 5 do not overlap and it can be said that the population average scores for these two techniques would be different. In contrast, techniques 3 and 5 have confidence intervals that overlap. This means that their population averages are likely to be about the same value and it would be difficult to argue that the average population scores for these two techniques are different without further analysis. Notice that these statements are made about the target population and not only about the groups that took part in the study.

This is a preliminary analysis of the data that gives an idea about the precision in the estimation of the average scores. This analysis is rather crude and can easily be improved upon by using simple statistical models such as the analysis of variance. An analysis of variance will estimate a common variance for the population of scores under each technique. It will then be possible to carry out hypothesis tests to compare mean scores for techniques or groups of techniques. The analysis of variance relies on a series of assumptions that need to be satisfied before accepting its results. For further discussion about the assumptions underlying the analysis of variance see theme paper "Ranking and Scoring".

More detailed analysis might be required to investigate the reasons for the preference of a technique by certain groups. For example the researcher might have additional information about the local conditions for each of the groups, for example: ethnic composition, predominant type of tenancy of the land, level of degradation of the soil, wealth status of the village, etc. These might be considered as factors (in statistical jargon these might be factors or covariates) that would explain the variability in the scores given to each technique. It is possible to formulate statistical models in which the score is a function of all or some of these explanatory factors.

When using statistical models researchers should bear in mind that "All models are wrong, but some are useful" (Box, 1979). Their usefulness depends on the validity of the assumptions underlying the model and how well the model fits the pattern observed in the data. The theme paper "Eliciting farmers perceptions through the use of ranks and scores" mentions models that can be applied to scores such as the analysis of variance or proportional odds models.

Important assumptions

There are some assumptions we have so far accepted without questions. These are:

1. The researcher is interested in making statements that apply to a conceptual 'population'.
2. The information collected is reliable and has been carefully registered by the researcher.
3. The study group is adequately large and is representative of the target population.
4. There has been an element of randomness in the selection of the study units.

Assumption 1 has to do with the purpose of the research. We do not argue that all work done in partnership with communities should lead to generalisable research results. However if the researcher wants to achieve this goal some of the points discussed above might be useful. Farmers might be interested in generalisable results too, particularly when the research is promoted by groups of farmers with limited resources.

Assumption 2 applies to all areas of research but it is particularly important in those cases when the information gathering exercise takes place under difficult conditions for data recording and storage.

Assumptions 3 and 4 have to do with the suitability of the information for statistical analysis. We recommend considering these points in the light of the discussion presented in the theme paper "Sampling issues in participatory research".

References

Barahona, C.E. (2000). "Statistical issues arising from the use of Participatory Farm Management methods" in Participatory Farm Management methods for agricultural research and extension: a training manual by Galpin, M. Dorward, P. and Shepherd D. University of Reading. Appendix 1.

Box, George E.P. (1979) "Robustness is the Strategy of Scientific Model Building" in Robustness in Statistics. Eds. R.L. Launer and G.N. Wilkinson, Academic Press, p. 202.

Eid, M. (1997). "Estado del trabajo de investigación de conservación de suelos y agua en laderas de los valles mesotérmicos de Santa Cruz" in Estrategias para prácticas mejoradas de conservación de suelo y agua en los sistemas de producción de ladera en los valles andinos de Bolivia. Memoria Segundo Taller Ed. Sims, B.G. Silsoe Research Institute. p 65-70

Lawrence, A. (1996) Participatory improvement of soil and water conservation practices in hillside production systems in the Andean Valleys of Bolivia. Research and Development Grant Application and Project Memorandum.

Acknowledgments

I wish to thank Anna Lawrence, Senior Research Associate at the Centre for Natural Resources and Development, University of Oxford and Miguel Eid, researcher at the Centro de Investigación Agrícola Tropical in Santa Cruz, Bolivia, for allowing me to join their visit to the village of Los Pinos in Santa Cruz, Bolivia.

Case Study 4

ON-GOING EVALUATION OF FRP PROJECT: "SUSTAINABLE MANAGEMENT OF MIOMBO WOODLAND BY LOCAL COMMUNITIES IN MALAWI" (R6709)

Neil Marsland[1], Simon Henderson[2] and Bob Burn[3]

[1] Natural Resource Institute
[2] Natural Resources Institute
[3] Statistical Services Centre, University of Reading

ON-GOING EVALUATION OF FRP PROJECT: "SUSTAINABLE MANAGEMENT OF MIOMBO WOODLAND BY LOCAL COMMUNITIES IN MALAWI" (R6709)

Background to Study

The FRP project *Sustainable Management of Miombo Woodland by Local Communities in Malawi* (R6709), which is implemented directly by the Forestry Research Institute of Malawi (FRIM), began in October 1996 and is due to finish in September 1999. R6709 is an extension of a previous FRP project (R4599) implemented collaboratively by the University of Aberdeen and FRIM (1992-1995).

In summary, R6709 is applied, action-orientated, NFM research. Its Purpose is to produce "techniques for sustainable management of forest resources by local people". This is to be achieved through the development and promotion of "co-management" by local communities and government of indigenous *miombo* woodlands situated on forest reserves, estate and customary land.

In broad terms, the project's strategy is based on two main lines of investigation. First, research into social/economic aspects of woodland utilisation and management by local people has been undertaken to improve understanding about the range of products demanded by them and the arrangements necessary for successful community management. Second, silvicultural/biophysical research is attempting to determine the potential *sustainable* supply of these products from the woodland resource. The results of this twin track approach will be used to design management prescriptions that focus on the production of woodland products demanded by local people but that adhere to the requirement for sustainability.

The average annual decline in forest biomass in Malawi is currently estimated to be around 3.5% per year as a result of both clearance for agriculture and resource degradation through overexploitation. Official estimates (Forestry Department) indicate that wood consumption is increasing at a rate of 9.25% p.a. The bulk of this demand is for fuelwood. Wood is the main source of fuel for cooking and heating water in 98% of rural households and meets virtually all other energy using activities. Together, charcoal and firewood provide 94% of the total energy consumed in urban households (USAID 1997). The vast majority of rural households are dependent on natural forests for construction materials and for a variety of foodstuffs which supplement household diet. In addition, households with land shortages obtain some 30% of household income from activities based around forest resources on public land (USAID 1997).

Uncontrolled cutting of green wood on publicly-owned, protected areas (including forest reserves) is officially prohibited. Poles can be pre-bought and then cut under the supervision of Forest Guards or alternatively bought directly from the authorities while firewood (dead wood) can be collected on payment of a licence fee. In principle, non-timber forest products (NTFPs) can now be freely collected though some confusion persists, in part because certain products (e.g. grass thatch) attracted fees in the past and also because the status of NTFPs collected for commercial purposes is somewhat unclear. Notably, the NEP recognises communities' rights to benefit from sustainable utilisation of natural resources on all public and customary land, though the implications for issues of access and charges have not yet been fully resolved.

To date (i.e. 1998/1999), *Sustainable Management of Miombo Woodland by Local Communities in Malawi* has worked primarily in two forest reserves where co-management operations are being piloted with surrounding communities: Chimaliro forest reserve (Kasungu/Mzimba districts) located on the border between the central and northern regions; and Liwonde forest reserve (Machinga district) in the southern region. In addition, experiences from a community's own efforts to manage an area of woodland (38ha) on customary land (Mangwere Hill) are also being studied. Areas within both Chimaliro and Liwonde forest reserves have been selected on the basis of site conditions and taking into account the wider environmental functions of the woodlands. These have been demarcated into blocks for management purposes and assigned to groups of villages (see below).

	Block size (ha)	Villages per block (No.)	PSPs per block (No.)
Chimaliro (152 km^2)			
Block I	18	3	14
Block II	118	3	13
Block III	74	3	12
Liwonde (274 km^2)			
Block I	416	3	10
Block II	288	4	10
Block III	468	2	10

These blocks represent the focus of co-management activities, where silvicultural interventions will take place (according to agreed management plans) in order to generate poles, firewood and other wood products for use by the participating communities. In return, the communities must provide labour for forest management (e.g. boundary marking, firebreak maintenance, controlled early burning, supervised harvesting and patrolling). Co-management is also expected to legitimise communities' rights of access to the reserves more generally, for the collection of non-timber forest products (NTFPs) and dead wood products. The regulatory framework is provided by the co-management constitutions, drawn up with each community involved, while detailed operations will be guided by block management plans.

The Evaluation Challenge

A number of considerations guided the development of the evaluation approach. The first and most important point to make is that there were two objectives behind the evaluation. The main objective was to field test an on-going evaluation methodology *which could be used in whole or in part across all FRP projects*. The second - *and very much secondary* - objective was to provide an acceptable on-going evaluation of R6709 itself. All of the remaining points in this section reflect the need to devise an approach that fulfilled these two objectives.

First, to be of maximum value, the approach needed to provide information that was both useful for internal research management purposes and meaningful to external audiences. In order to meet these twin objectives, the approach needed to report on *actual* events, situating these within an impact-orientated framework rather than the activity-focused framework conventionally addressed by monitoring. Given that R6709 is long-term in nature yet both internal and external audiences require timely results, *on-going evaluation* (i.e. monitoring progress towards impact) offered the most feasible solution.

Second, for the approach to form the basis of an evaluation *strategy*, it needed to be applicable across different projects and facilitate at least qualitative comparison of performance. However, different projects, undertaking different types of research under different conditions will inevitably require different assessment methods. The approach, therefore, must select evaluation criteria that are widely applicable and can be applied consistently but also permit methodological flexibility within the overall framework provided by the criteria.

Third, given the long-term nature of much NFM research, the approach to on-going evaluation necessarily relies on leading, intermediate indicators of progress towards impact. These indicators, however, do not measure "impact" in an absolute sense but rather are relative measures that identify change over time. On-going evaluation, therefore, must be undertaken periodically. This need for repeat assessment implies that the approach needed to be relatively inexpensive, but also that the depth of analysis possible might be constrained by available resources.

Finally, the usual uncertainty surrounding the outcome of RNR research in general is compounded for natural forest management research by data shortages, social and institutional issues and long research lags. For the results of an on-going evaluation exercise to be credible, therefore, the key factors that affect the likelihood of achieving impact must be identified and their "riskiness" explicitly incorporated into the assessment process.

The Evaluation Approach

In the light of these considerations, the study attempted to develop an alternative approach to evaluation, based loosely on the principles of the "Balanced Scorecard", an approach used widely in industry in developed countries. Similar to the Balanced Scorecard, the proposed approach comprised four components:

(1) *Internal perspective*
(2) *Client perspective*[4]
(3) *Test of research effects*
(4) *Uptake network*

The inclusion of components (1), (2) and (4) explicitly recognises that all FRP projects share three key characteristics: they have internally established targets (i.e. Outputs), they have clients who are expected to make use of results, and they rely on the actions of external actors for the results of research to be applied more widely. Furthermore, adequate performance against Outputs, satisfaction on the part of the clients and wider application of research recommendations, (whether directly in a productive process or indirectly by informing decisions that influence subsequent actions) are considered to represent the "lowest common denominator" of successful research[5]. While they are necessary rather than sufficient conditions for impact to be realised, the advantage of these criteria is that they are shared by all projects, even the most problematic for evaluation.

Component (1): Internal Perspective

The objective of the internal perspective is to assess current and likely future performance of research against stated Outputs, based on the OVIs identified the log-frame.

Component (2): Client Perspective

The objectives of this component are to (a) assess whether significant levels of dissatisfaction exist with the research process to date; and (b) assess the degree of commitment to the future implementation of research.

Component (3): Test of research effects

This component of the evaluation seeks to:

(a) determine whether there is evidence of positive change among beneficiaries resulting from research;
(b) qualitatively assess the significance of any changes;
(c) assess local perspectives regarding the costs and benefits of participation in the research;
(d) determine whether there is evidence to suggest that particular groups are performing significantly better than others (according to gender, wealth and education).

[4] The term "clients" is used to describe the target audience for the results of research, in the context of its developmental objectives, and does not refer to the Programme or donor. In the case of applied, participatory research (such as R6709), the primary clients are considered also to be the intended beneficiaries.
[5] Research "success" unless otherwise specified refers to achievement of the developmental aims of the project/Programme and not to narrower "scientific" definitions that, for example, may be limited to the testing of hypotheses.

Component (4): Uptake network

FRP's interest in wider impact, coupled with the growing awareness that such impact is dependent on factors *external* and *subsequent* to research, suggests the need for more systematic assessment of future prospects. This need is considered all the greater for NFM research. In the majority of cases, predicted efficiency gains indicate *a priori* the need for *widespread* adoption in order for research to "pay-off". In addition, the long-term nature of NFM research means that FRP must consider prospects beyond the typical project funding window (three years) in order to allocate research funds effectively.

The evaluation sought to predict up-take by making use of a Bayesian belief network approach using the Netica™ software developed by Norsys Software Corporation. Bayesian techniques are more normally associated with decision-making problems under uncertainty or identification of causes of an event that has already occurred (e.g. in medical diagnosis). Use of belief networks for the purposes of on-going evaluation is believed to be relatively novel. However, in the time available, the study could not fully assess current applications of these techniques, most noticeably at ICRAF[6]. Readers interested in finding out more about Bayesian belief networks are referred to the theme paper entitled "Quantifying and Combining Causal Diagrams".

Data Collection and Analysis

A formal household questionnaire survey was conducted within project villages at both Chimaliro and Liwonde sites (150 households included in each site). In addition, four control villages (i.e. not participating in the project) were selected at both Chimaliro and Liwonde from areas *adjacent* to the project areas. 25 households were interviewed in each of the control villages (i.e. total of 100 households at both sites). The major difference between project and control questionnaires was that the latter omitted questions that referred directly to "co-management" or the research.

[6] ICRAF is using belief networks in a range of land use assessments. Examples include estimating deforestation risk, adoption potential of agroforestry technologies, desegregating population and other census data to match with remote sensing data and (work in progress) using hyperspectral data for soil analysis.

Table 1: Overview of data sources by component

Evaluation component	Indicators	Formal Q'nnaire Survey	Informal RRA Exercises	Key Informant Interviews	Baseline/ 2ndary Data
I. Internal perspective	Progress to date against Outputs			✓✓	✓✓
	Prospect for achievement of Outputs			✓✓✓✓	
II. Client perspective	Awareness/participation among communities	✓✓✓✓			
	Information/explanation	✓✓✓✓			
	Implementation arrangements	✓✓✓✓			
	Advantages/ disadvantages	✓✓✓✓			
	Willingness to continue	✓✓✓✓			
III. Test of research effects	Change in availability of forest products	✓✓	✓✓		✓
	Influence of research	✓✓	✓✓		✓
	Perception of Cs & Bs of participation in research	✓✓✓✓			
	Distribution of benefits	✓✓✓✓			
	Indirect & non use goods and services		✓✓✓✓		
	Acceptability of restrictions on use		✓✓✓✓		
	Expectation of future benefits	✓✓	✓✓		
	Opportunity cost of forest land	✓✓✓✓			
	Financial capital		✓✓✓✓		
	Social/human capital		✓✓✓✓		
	Transforming processes		✓✓✓✓		
IV. Uptake network	Research timeframe			✓✓✓✓	
	Technical success			✓✓✓✓	
	Meet client expectations			✓✓✓✓	
	Wider applicability			✓✓	✓✓

Relationships between Qualitative and Quantitative Components of the Evaluation

Combining ideas and methods based on probability theory and statistical inference with those rooted in informal, participatory enquiry can give benefits in terms of trustworthiness of data. Table 2 illustrates this.

Table 2: Relationship between: stages in research exercise, type of formal-informal combination, and improvements in trustworthiness.

(i) Design of study

Type	Explanation/Example	Function
Swapping	• Formal sampling procedures for informal work	• Reduced sampling error: better external validity for informal work
	• Use of social mapping for formal work	• Reduced time and cost for household listing and sampling.
Concurrent	• Correct use of different instruments for different variables within the same survey/ experiment	• Better internal validity for "qualitative" variables - belief, motivations etc. alongside better external validity for quantitative variables - rates, proportions etc. • "Enriching": Using informal work to identify issues or obtain information on variables not obtained by quantitative surveys.

(ii) Analysis

Type	Explanation/Example	Function
Sequential	• Analysis of formal outputs with informal approaches. e.g. testing null hypotheses; investigating unexpected outcomes.	"Refuting" or "Confirming": Verification of formal results through informal methods. • "Explaining": Using informal work to explain unanticipated results from formal survey work
Swapping	• Applying statistics to categorical and unbalanced data sets.	• Improved credibility of analytical conclusions from informal work.
	• Coding responses from informal work	• Enhances possibilities for aggregation, thus facilitating generalisation. • Enhances possibilities for stratification of sample for subsequent sample survey
Merging	• Blending the analytical outputs from informal and formal work into one set of policy recommendations. The outputs may be from Type A, B or C combinations.	• Higher quality policy recommendations

Design of study

(i) Informal to formal: Use of village mapping to generate sample frame

The sample design for project households was based on a single-stage cluster sample within each of the stratified substrata, with villages as clusters. Project villages were stratified first by association with particular the co-management blocks in each reserve and then by proximity to the reserve (i.e. near and far). Because of time and resource constraints, a systematic sampling method was used to select households within the selected villages. The sample frame was generated through a process of village mapping, with villagers marking out the number and location of each dwelling unit in the village, together with the name and sex of the household head. All the names and numbers were recorded by the RRA field teams and a systematic sample was taken. This process was found to be useful for three main reasons. First, it served as an initial ice- breaker, allowing the RRA team to interact with members of the village. Second, and more importantly perhaps, it provided a very rapid and accurate way of generating a comprehensive sampling frame for selected villages. Characteristically, the whole process would take between 1 and 2 hours for Chimaliro and 1 to 3 hours for Liwonde. The process was lightly longer in Liwonde than in Chimaliro owing to the larger village sizes in Liwonde. Finally, the existence of an accurate village map helped greatly in planning the actual enumeration and dividing tasks up between enumerators.

(ii) Formal to informal: "Tagging" RRA respondents to random sample

In the case of the RRA exercise, three groups were selected in each of the six villages. Originally, intention had been to directly link the RRA results with the questionnaire results by ensuring that the RRA respondents were a sub group of the questionnaire respondents. This was to be achieved through rapid socio-economic analysis of questionnaire responses by field teams immediately after coming back from the questionnaire interviewing. Field teams had been trained to analyse responses and group respondents into three wealth groups on the basis of predetermined criteria. Through being able to identify members of wealth groups and cross-referencing to questionnaire data, it was hoped that the strength of triangulation between the RRA and the questionnaire responses could be strengthened. In addition it was hoped that the "tagging" of RRA to questionnaire data would generate interesting and potentially important combinations of depth and generalisability. However, due to non-availability of large number of questionnaire respondents during the RRA exercise, FRIM field staff effectively abandoned this approach and just relied on a combination of opportunity sampling (simply asking villagers at random whether they would be prepared to be involved in the exercise, and then classifying according to the predetermined criteria) and by selection through the chief - i.e. asking the chief to select groups of poor, medium and wealthy on the basis of the predetermined criteria. While this is viewed as a major disappointment from the point of view of the evaluation methodology, it is not felt to negate the results, just temper the strength of conclusions that may be drawn.

(iii) Informal to formal: Use of key informants to generate initial values in Uptake Pathway Network (Beyesian network)

The views of key informants were used directly in the assessments made under components (1) (internal perspective) and (4) (uptake network). In the case of the former, the lead researcher on R6709 provided the necessary information while in the latter, informants comprised FD HQ staff members familiar with the research and who have detailed knowledge of the forest reserves in Malawi, the lead researcher on R6709 and R6709's UK liaison consultant who has wide experience of forestry in Malawi. In component (1), secondary data (essentially project reports) were used to inform the discussions and subsequent assessment. In component (4), secondary data were combined more explicitly with the subjective assessments of key informants to project potential adoption of research.

Key informant interviews were used to inform assessment of the potential adoption ceiling. Opinions were obtained from two FD HQ staff members familiar with the research and who have detailed knowledge of the reserves, the lead researcher on R6709 and R6709's UK liaison consultant who has wide experience of forestry in Malawi. Each informant was (separately) asked to assign a score to each reserve as follows:

1 = "high chance that co-management can be successfully implemented";
2 = "good chance that co-management can be successfully implemented";
3 = "less confident that co-management can work";
4 = "little chance of co-management working".

Informants' scores were then averaged and the results regrouped as follows:
"High" = average score 1 - 1.75
"High-Medium" = average score 1.76 - 2.5
"Medium-Low" = average score 2.6 - 3.25
"Low" = average score 3.26 – 4.0

Analysis

(i) Sequential: "confirming"

In order facilitate understanding of potential impact on "natural capital", it is necessary to compare the most important forest products identified by respondents with those most likely to be influenced by research. The questionnaire survey asked respondents to identify the most important forest products (unprompted) and frequencies of response were used to rank the products. In the RRA exercise with project villages in Chimaliro, participants were again asked to identify the most important products and assign a score to each (with 10 being the most important). Table 4 shows that the results of the RRA confirmed those of the questionnaire.

Table 4: Comparison of responses from questionnaire with RRA exercise: Importance of different forest products

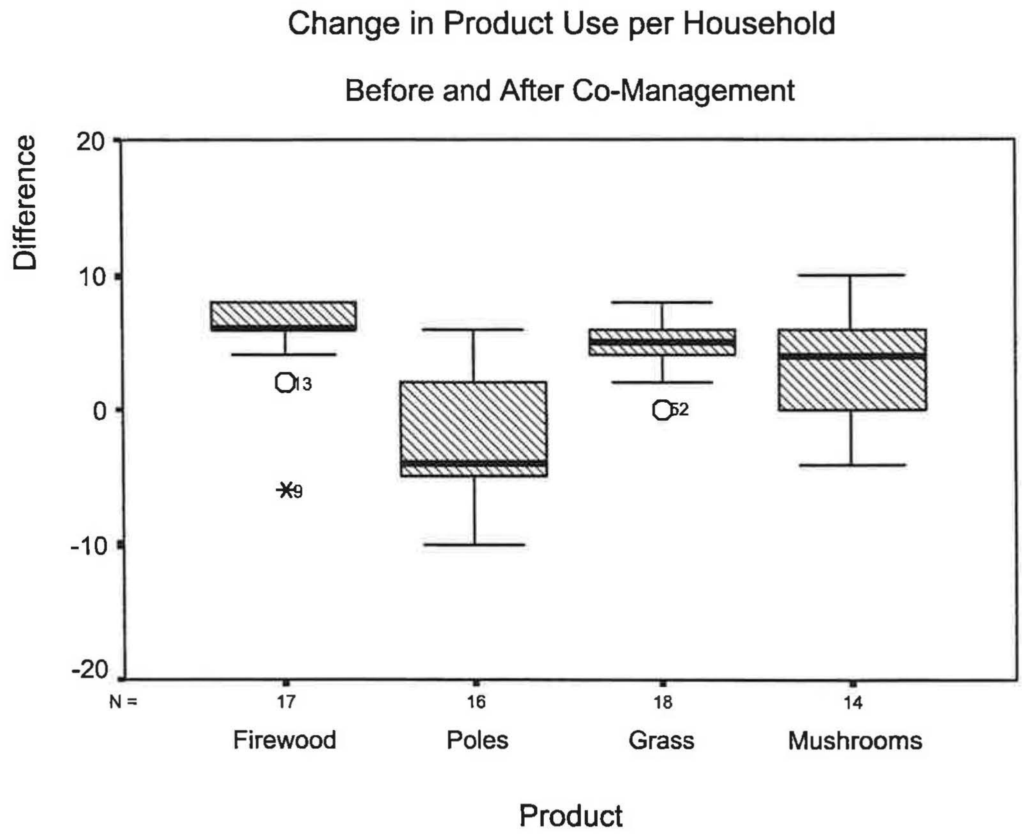
Product	Questionnaire survey		RRA exercise
	%	Rank	Rank
Firewood	94	1	1
Grass/thatch	84	2	2
Mushroom	70	3	3
Poles/timber	58	4	4
Rope fibres	28	5	5
Medicine/herbals	24	6	6
Fruits	22	7	7

(ii) Swapping: Use of statistical packages and statistics to analyse RRA data

RRA data was entered into excel spreadsheets and analysed using SPSS and simple excel tools. Box and whisker plots were used to identify more clearly a number of relationships including: changes in product use; changes in sources of forest products; changes in importance of products gathered from reserve as perceived by respondents; balance between domestic consumption and cash uses of woodland products, and; differences in consumption of products across wealth groups. Some simple statistical testing was done (f-test; Duncan's multiple range test) to explore differences in scores. An example of the type of analysis done is given in the text Box (next page)

Change in consumption of forest products

Taking all villages together, it appears that very large majorities in the sample survey felt that access to all products apart from poles had improved under co-management. This pattern is repeated in findings from the RRA. Here, groups were asked to estimate the magnitude of the change in forest product use before and after co-management.



ANOVA
Difference

	Sum of Squares	df	Mean Square	F	Sig.
Between products	606.538	3	202.179	16.350	.000
Within products	754.324	61	12.366		
Total	1360.862	64			

Difference
Duncan's multiple range test

	N	Subset for alpha = .05	
Product		1	2
Poles	16	-2.25	
Mushrooms	14		3.43
Grass	18		4.89
Firewood	17		5.41
Sig.		1.000	.136

F test results confirm that there are significant differences between the means of the products, and a Duncan's multiple range test confirms that the F test result is accounted for by the difference between the mean of the response for poles and all the other means.

Design Objectives and Operational Parameters / Constraints

Introduction

At each site, the study was designed to evaluate with-project effects vs. without project effects as well as before vs. after project comparisons. The key instruments were the RRA (Chimaliro only) and the sample survey (both sites). On one level, the design of the study was shaped by a 2 x 2 matrix:

	With (project villages)	Without (non-project villages)
Before	Baseline data	Baseline data
After	Sample survey (plus RRA in Chimaliro only)	Sample survey

The "before" information came from baseline studies done in Chimaliro. The study sought to cover the with - without comparison by surveying both project and non-project villages at both sites. The decision not to include a RRA component for the Liwonde site was made on the assumption that the additional information gained through an RRA would not justify the extra cost involved owing to the fact that co-management was much more recently established in Liwonde than in Chimaliro.

Time

Both the sample survey and the RRA had relatively tight windows in terms of fieldwork. This limited the number of villages that could be sampled. Out of a total of 18 project villages, 12 were sampled. Time taken from design of survey and RRA tools to analysis of results was approximately 4 months (beginning of November 1998 to end of February 1999). Details are in Table 5 below

Table 5: Stages in the evaluation

Component	Responsibility	Timing
Initial visit by project leader to identify collaborators	Team leader	September 1998 (2 days)
Visit by UK team to design and pre-test evaluation components and train counterparts	Statistician, Team leader (Economist), Socio-economist	November – December 1998 (3 weeks)
Fieldwork for RRA	Chimaliro team	December 1998 2 weeks
Fieldwork for sample survey	Chimaliro and Liwonde tams	December 1998 – February 1999 4 weeks
Coding and entering data from sample survey and RRA	Reading university data entry staff	January – February 1999 3 weeks[7]
Analysis of results	UK team	February 1999 2 weeks
Report writing	UK team	March – April 1999 3 weeks

[7] Most of this work was not costed

Money

The total cost of the evaluation was just under £ 50,000. It should be noted that if this exercise were to be repeated, the actual costs would be considerably less, as large amounts of time were spent developing the methods. In crude terms, one could probably expect a reduction of the order of £15,000 to £25,000 if the evaluation was to be repeated on the Liwonde and Chimaliro sites in subsequent years.

Lessons Learned

The evaluation was successful in some ways and not – so – successful in others. A number of useful lessons were learned from both the successes and the failures.

Household mapping to create sample frame

The success of this exercise confirms that it is a simple and cost effective way of generating a sample frame whilst at the same time making actual planning of the enumeration process at the village level very easy. Details of how to go about doing village mapping to obtain a sampling frame will be included in the best-practice guidelines to accompany these case studies.

Supervision

The exercise highlighted the problems that can occur with inadequate supervision of rapidly trained field staff. The "tagging" problem noted under 4.1 (ii) above, probably would have been avoided if a member of the NRI / Reading team had been available to supervise the RRA. Due to time and money constraints, however, this was not possible. This highlights a thorny issue: how to achieve the optimal balance between supervised and non-supervised fieldwork given tight budgetary and time constraints. There is no easy answer to this. In retrospect, the tasks demanded of the RRA team was at the limit of their capabilities, given the amount of training that they had. The team did very well, but was unable to cope with some of the unforeseen circumstances. Given that constant supervision by a member of the NRI / Reading team was out of the question, problems could have been reduced if either the training period was longer; and / or field supervisors were more experienced and / or the tasks were simplified.

The elicitation process for Bayesian networks

It would appear that the development of a systematic approach to eliciting the views of institutions and those in them is much less well developed, or at least practised than the battery of techniques that exist for deriving the opinions of farmers. Having said this, there are some straightforward techniques of semi-structured interviewing, ranking and scoring that can be applied very easily. In retrospect, the elicitation of the opinions of key informants would have been improved if these tools had been applied more systematically. The elicitation process can be improved also by deriving some more tailor-made techniques, perhaps borrowing from the well-established body of market research. In all of this it is important to bear in mind that the process of elicitation will obviously depend on the degree of participation by primary and secondary stakeholders. This in turn will be shaped by the objectives of the exercise and by time, money and expertise constraints.

Application of statistics to ranking data

Through entering the ranked data into SPSS, it was possible to strengthen the conclusions of the exercise by representing relationships in a clear and consistent descriptive form e.g. box and whisker plots, and by undertaking further analysis using statistical tests.

Case Study 5

THE USE OF STATISTICS IN PARTICIPATORY TECHNOLOGY DEVELOPMENT – THE CASE OF SEASONAL FEED SCARCITY FOR GOATS IN SEMI-ARID INDIA

Conroy C, Jeffries D, and Kleih U
Natural Resources Institute, University of Greenwich

THE USE OF STATISTICS IN PARTICIPATORY TECHNOLOGY DEVELOPMENT – THE CASE OF SEASONAL FEED SCARCITY FOR GOATS IN SEMI-ARID INDIA

Background

Scientists have acquired a tremendous amount of knowledge about the feed resources and nutrition of ruminants, both large and small (Acharya and Bhattacharyya, 1992). Despite this, the adoption of technologies developed by researchers, for enhancing fodder production and improving grazing management systems, has been poor (*ibid.*; Sidahmed, 1995). This is partly because feed technologies have often been developed without the involvement of the intended users, and without an adequate understanding of their farming systems and constraints.

Since October 1997 the Natural Resources Institute (NRI) and BAIF Development Research Foundation (BAIF) have been collaborating on a research project entitled "Easing Seasonal Fodder Scarcity for Small Ruminants in Semi-Arid India, through a Process of Participatory Research". The project is working in two districts of Rajasthan (Bhilwara and Udaipur) and one district of Gujarat. BAIF had been undertaking a goat development programme in parts of Bhilwara for several years prior to the research project.

The project uses a participatory approach to technology development (PTD) in identifying constraints faced by livestock keepers and in experiments to overcome them. The following sections highlight the different phases in the research programme, the use of statistical tools in analysing data collected through participatory on-farm trials, and lessons learned.

Farmer Participation in Feeding Trials

The research programme uses three distinct phases in identifying goat-keepers' constraints and addressing the latter[1], namely:

Phase 1: Understanding livestock systems and constraints;
Phase 2: Deepening understanding of problems to be addressed in trials;
Phase 3: Participatory on-farm trials.

RRA (Rapid Rural Appraisal) tools and techniques such as semi-structured interviewing, diagramming and mapping were used in phase 1 in order to understand the livestock systems and constraints as perceived by goat-keepers. If the latter highlighted feed related problems, the research team tried in phase 2 to deepen their understanding, in particular through participatory problem tree analysis and participatory herd histories.

Provided feed related constraints were identified as important to the goat-keepers during the course of the previous phases, in phase 3 trial villages were then selected

[1] Corresponding to Type B of quantitative / qualitative combinations in the Methodological Framework, i.e. "Sequencing".

primarily on the basis of three criteria. They: (a) were located in an area where BAIF was working, and ideally, already had dealings with BAIF; (b) contained goat keeping households that were below the poverty line; and (c) contained enough goats (at least 50 and preferably more) in the relevant class to enable a meaningful trial to be undertaken.

This case study paper concentrates on phase 3, showing how statistical tools were used in analysing quantitative data obtained through participatory trials.

The process of designing, monitoring and evaluating the feed supplementation trials was intended to involve goat-keepers actively. Table 1 illustrates the four different modes of farmer participation in agricultural research, and Table 2 provides an indication of the degree of goat-keeper participation in the trials.

Table 1: Four Different Modes of Farmer Participation in Agricultural Research

1. Contract	2. Consultative	3. Collaborative	4. Collegiate
Farmers' land & services are hired or borrowed: e.g. researcher contracts with farmers to provide specific types of land	There is a doctor-patient relationship. Researchers consult farmers, diagnose their problems and try to find solutions	Researchers and farmers are roughly equal partners in the research process & continuously collaborate in activities	Researchers actively encourage & support farmers' own research & experiments

Source: Biggs, 1989.

Table 2 Indications of the Degree of Goat-Keeper Participation in the Trials

Trial – number, supplement & year	Overall mode of participation*	Was a Priority Need Addressed?	Who Decided Nature of Treatment?	Joint Evaluation?	Is treatment likely to be adopted?
1. UMG – 98 (Bhavnagar)	1 - 2	X[2]	R	✓	X
2. PJ pods & barley – 98	2	✓	R, with G-Ks' agreement	✓	✓ (with modification)
3. Barley – 98	2	?	R, with G-Ks' Agreement	✓	X
4. UMG – 98	2	?	R, with G-Ks' agreement	✓	X
5. PJ pods & barley – 99	1 - 2	✓	R	✓	✓ (with modification)
6. PJ pods – 99	3	✓	R/G-K jointly	✓	?
7. Barley – 99	2	?	R/G-K jointly	✓	?

Source: Conroy and Rangnekar (2000a)
Code:
1 = Contract 2 = Consultative 3 = Collaborative.
R = Researchers. G-Ks = Goat-keepers
UMG = Urea/molasses granules; PJ = *Prosopis juliflora*
X = no, ✓ = yes, ? = not sure.

[2] The research in Bhavnagar subsequently (in 1999) focused on addressing water scarcity in the dry season, which the goat-keepers had identified as their main constraint.

As Table 2 illustrates, the predominant mode of participation was consultative, with collaborative participation occurring in one case (out of seven), and elements of contract participation in two cases. This reflects that participation in research is a gradual process, and collaborative and collegiate modes of participation require time and the building of trust between the parties involved.

In all of the trials it was the researchers who identified the type of supplement to be used. However, this was based on the findings of participatory constraints analyses and knowledge of livestock-keepers' experiences with similar technologies in other localities. In most trials, the participants appeared to agree that the proposed treatment was a sensible one, and contributed 33-50% of the cost of the treatment. In Trials 6 and 7 (of Table 2) goat-keepers were more actively involved in determining the treatment, in the latter case having the major say in the daily quantity.

Statistics Related Issues in On-farm Trials[3]

The trial participants were goat keeping households in the villages who belonged to scheduled castes or tribes (the poorest groups). They were divided between the treatment and control in such a way that there would be roughly equal numbers of goat-keepers and of trial does in each group.

> ***Experimental design***
>
> The purpose of this experiment is to compare the effect of supplementation to standard practice. It is important to ensure that the experiment is measuring this difference and not some other difference caused by a biased selection of participants. For example one group of farmers might not be as 'poor' as the other group or one group might be better educated.
>
> These problems can be minimised by using random selection procedures. There isn't a prescriptive solution to random sampling and a common-sense approach usually suffices. One particularly useful approach is to stratify the goats to ensure for example that each treatment group has the same age distribution. However it is still useful to record parameters that might influence the results, such as temperature or initial weight of goat. These variables can then be used in an analysis to give a better interpretation of differences between the two groups of goats.
>
> Goats are more suitable for on-farm trials than large ruminants as their life cycle duration is shorter. This allows trials to be conducted on an annual basis and to generate results within a few months.

Some animals were sold during the course of the trials and were not monitored subsequently. The initial and final (some were sold or died) numbers of does in each group are shown in Table 3.

[3] This section describes trials 2 and 5 of Table 2.

Table 3 Numbers of Goats in the Trials

	Treatment		Control	
	Initial	Final	Initial	Final
1998	25	24	25	23
1999	56	50	63	55

Replication

There is no answer to the question of 'How many goats should be used'. The primary consideration is the likely amount of variability amongst the goats. An estimate of this can often be obtained from previous experiments or relevant literature. The basic rule is that higher variability requires more replication for the detection of differences between control and supplementation. A minimum for this type of experiment would be six goats in each group. However this assumes that all the goats which enter the experiment, survive. This is clearly not a viable assumption in this experiment and should be considered when deciding on sample size.

This experiment has been conducted over two years, with enough replication to analyse each year as a separate experiment. It is not advisable to rely on using a combined analysis over years to provide adequate replication, if the way how the trial is conducted differs from one year to another. For example, if there were a difference in the protocol of feed supplementation it would be difficult to justify a combined analysis.

For each trial at least one local person (not necessarily goat-keepers participating in the trials, but usually someone living in the village, including literate adolescents), was trained to undertake the monitoring of the trial animals. Farmers were encouraged to record the data themselves but they were not convinced of the value of doing so from their viewpoint in that "they could see changes in their goats' performance anyway".

The monitors visited the participating households every 15 days during the supplementation period and the kidding season. Records were made of: breeding activity (including heat, number of services and conception); the health and condition of the does; and the number of kids born and, in a few cases, their birth weight. In addition, as part of a twin-track approach, every month or so the BAIF researcher had a group meeting with participants to discuss with them how the trial was progressing[4]. At the end of the first trial, in December 1998, the researcher convened an evaluation meeting with participants from both the treatment and control groups. A similar meeting was undertaken for subsequent trials.

[4] Corresponding to Type C of quantitative / qualitative combinations in the Methodological Framework, i.e. Concurrent use of tools and methods from the different traditions: "Mixed Suite".

Monitoring, evaluation and data management

This is a vital part of an on-farm trial and resources should always be made available for these procedures. Without adequate data management the whole trial can be lost.

Every effort should be made to speed up the collection of data by field staff and the entry of data in a computerised form. This helps to reduce errors and speed up the dissemination of results back to the farmers.

Where local staff do not have access to a computer, or don't know how to use one for data analysis, they should be encouraged to do simple forms of analysis (e.g. using a calculator to estimate mean weights, and plotting a graph manually to show changes over time for both trial and control groups).

Selected Results

The treatment had the desired effect, with does in the treatment groups having higher conception rates than those in the control groups. Another benefit in the first trial (in 1998) was that there were no late conceptions in the treatment group: earlier conception is beneficial in that it results in earlier sales (hence a quicker return on investment) and/or higher prices. The conception data are summarised in Table 4.

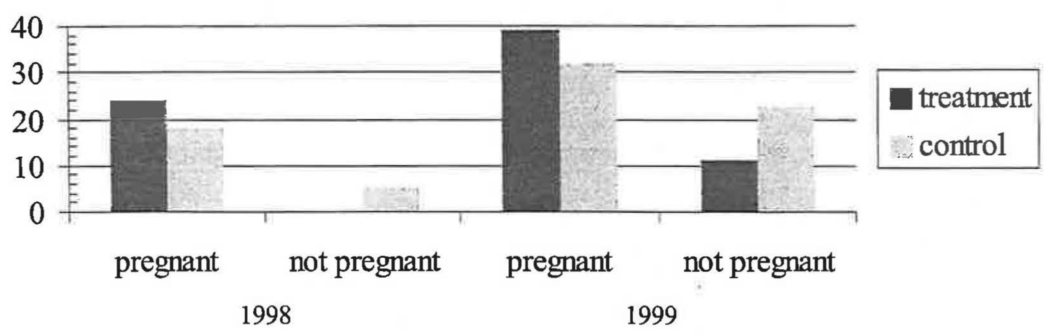
Table 4 Conception Data for Mature Does

	1998		1999	
	Pregnant	Not pregnant	Pregnant	Not Pregnant
Treatment	24	0	39	11
Control	18	5	32	23

Graphical presentation

Many readers find graphical representations of the data easier and quicker to interpret. For example the data in Table 4 can be represented as a bar graph, which more clearly emphasises the important differences.

Conception data for mature does



Although it is clear that does in the treatment group have a higher conception rate, it is useful to quantify this difference statistically.

The 1998 data were tested using an exact chi-squared test. This gave a p-value of 0.022, giving significant evidence that the conception rates are different for the two trial groups. It is clear from the cross-tabulation that this is caused by the 100% conception rate for the treatment group. The results were not conclusive, however, as there were only three goat-keepers in the treatment group, one of whom owned 13 of the 24 mature does. The difference could, therefore, have been related to inter-owner differences and non-experimental variables.

In order to eliminate this factor, the trial was repeated in 1999 in another village, with larger numbers of goat-keepers (13 and 14 in the treatment and control groups respectively) and goats. The conception rates (see Table 4) were again different between the treatment and control groups. An asymptotic chi-squared test gives a p-value of 0.055, indicating that at the 5% level there is no evidence to reject the null hypothesis of equal pregnancy proportions for the treatment and control groups. However, the p-value is very close to the 5% level, and a pragmatic interpretation is that there is some evidence that those animals taking the supplement have a higher conception rate, but further experimentation is necessary to quantify the strength of this inference.

Analysis of cross-tabulations

The two cross tabulations (for 1998 and 1999) in Table 4 are an extremely common method of presenting data and there is an extensive literature on their statistical analysis. The aim of the analysis is to test for independence between the rows and columns. For example in this case to test whether there are any differences in conception rate between the treatment and control.

The most common form of the test is the chi-squared test. This is an asymptotic test and as a rule of thumb it should not be applied to tables with cell counts of less than four. Clearly the cross tabulation for 1998 has a cell count of zero and the test is not appropriate. However in these cases a numerical test can be used, which measures how unique the actual table is compared to all possibilities with the same row and column totals [Sprent].

For the 1998 data in Table 4 a numerical (often called an exact) chi-squared test gives a p-value of 0.022 and hence significance at the 5% level. The standard chi-squared test gives a p-value of 0.016, which in this case is similar to the exact result. Note this is not always the case and if in doubt it is advisable to check by using the numerical method. The data from 1999 can be analysed using the conventional chi-squared test, as the cell counts are all 'large'.

Like many statistical tests the chi-squared test is a hypothesis test. Hypothesis tests are interpreted using p-values and usually tested at the 5% level. (This means that if the p-value is less than 0.05 (5%) there is evidence for rejecting the null hypothesis).

A more detailed analysis of the data can be undertaken by considering the number of kids born to each doe. The combination of higher conception rates and higher twinning rates results in higher kidding rates (number of kids/number of does, expressed as a percentage) in the treatment groups, as can be seen from Table 6.

Table 6 Kidding Rates (percent)

	1998	**1999**
Treatment	116.6	100
Control	78.3	69.1

Table 6 summarises the data, but a more detailed description can be obtained by giving the number of kids produced by each group of does as shown in Table 7.

Table 7 Distribution of kids born

	1998		**1999**	
	Treatment	Control	Treatment	Control
No kids*	0	5	11	23
Single	19	16	28	26
Twins	4	1	11	6
Mean	1.2	0.8	1	0.7

* Note this row refers to those does that were not pregnant. The two does that aborted have been dropped from the sample for this analysis.

It is clear that the mean number of kids born is higher for the treated does in both years. A Mann-Whitney test (adjusted for ties) was used to compare the two treatments and control group for both years.

Non-parametric tests

The most common test for comparing the mean values of two different groups is a t-test, but the t-test assumes a continuous normal distribution for the observations. The distribution of the number of kids cannot even be claimed to be continuous and it would be difficult to justify the use of the parametric t-test for this data.

The Mann-Whitney test is a non-parametric equivalent of the t-test. However the Mann-Whitney test quantifies whether two distributions differ in location (i.e. mean or median). For this application the medians of the treatment and control groups for both years are all 1 and in this case the difference in location, equates to a difference in means.

The Mann-Whitney test is based on calculating ranks for each observation [Sprent] and is available in many standard statistical packages, but for data like this the test should be adjusted for tied ranks.

The Mann-Whitney test, (adjusted for ties) for differences between the mean number of kids in the treatment and control group gave a p-value of 0.02 in 1998 and 0.01 in

1999. At the 5% significance level there is clear evidence that the mean number of kids per doe is higher in the treatment group for both years.

Numerical methods

When there are many ties the Mann-Whitney test can give misleading results. There is an alternative type of non-parametric tests based on permuting the data. This is an example of a randomisation test and as computing speed continues to increase these tests are likely to become more common.

The basis of the test is a random allocation of the responses to each group to quantify how rare the observed data is (Manly, 1998). For each randomisation the difference between the means is measured by the standard t-statistic. Since the data do not satisfy the necessary assumptions for a t-test, the t-distribution cannot be used to quantify how extreme the observed data are. As an alternative the randomisation distribution of the t-statistic can be used to quantify the significance of the mean difference for the observed data.

Randomisation tests have the advantage that they are easily programmed and no specialist statistical software is required. Although many permutations have to be used (1000 or more), this should not be a problem as the results of tests are not required in real-time.

Randomisation tests were performed with 1000 permutations being used for each year. For 1998 the test gave a p-value of 0.024 for a difference of 0.35 between the treatment and control groups and for 1999 the p-value was 0.026 for a difference of 0.3. These tests confirm the inference of the Mann-Whitney tests and reinforce the outcome that treated goats are significantly more fecund than the goats from the control group.

Confidence intervals

The mean differences between the treatment and control groups have been estimated and found to be significant, but the accuracy of the differences has not been established. A concise and informative method for showing the accuracy of the mean difference is to give an interval for the estimated true difference between the mean of the treatment and control groups. This is called a confidence interval and the 95% confidence interval is usually quoted. The 95% confidence interval can be interpreted as being an interval, which has a 95% chance of containing the true difference between the means.

There are techniques for calculating confidence intervals using the Mann-Whitney test, but they are difficult to apply and do not always give useful results. An alternative technique is to use the randomisation distribution to numerically estimate the 95% confidence intervals (Manly, 1998).

Lessons learned

The case study demonstrates the role statistics can play in Participatory Technology Development. A three phase approach was adopted in identifying goat-keepers general constraints, deepening understanding of feed related problems, and undertaking participatory on-farm trials.

Consultative participation was the most common mode of participation encountered during the course of the research, with scope for more collaborative participation during subsequent trials. There is strong evidence of goat-keepers' interest in the supplementation trials demonstrated by the fact that the goat keepers who participated in the 1998 trial decided to test *Prosopis juliflora* pods alone in a subsequent trial, without any material support from the project.

Based on quantitative data collected by trained monitors, both trials provide clear evidence that the treatment results in does producing significantly more kids than they would otherwise have done. It would be a mistake, however, to rely on quantitative data alone. Trial participants' views are also of paramount importance, which is why these trials adopted a monitoring system that had both, quantitative and qualitative components. For example, goat-keepers sometimes prefer to have one kid rather than twins, as it may have a better chance of survival. They could, therefore, regard twins as a disbenefit rather than a benefit, but this point would not be captured by quantitative data, and hence the results could easily be misinterpreted.

The recommendation domain will be determined by: (a) the geographical distribution of *Prosopis juliflora*, which is quite widespread, particularly in semi-arid, saline areas; and (b) the extent to which goat-keepers in those areas are experiencing this kind of problem.

References

Acharya, R.M. and Bhattacharyya, N.K. (1992) Status of small ruminant production. Paper presented at the Vth International Conference on Goats, New Delhi.

Conroy, C. and Rangnekar D.V. (2000a), Constraints facing goat-keepers and ways of addressing them through a participatory approach: some experiences from semi-arid India, BAIF/NRI Goat Research Project Report No2, Natural Resources Institute, Chatham, UK.

Conroy, C., Rangnekar, D.V., Sharma, M. and Vadher, M.H. (2000b) *Use of Prosopis Juliflora Pods/Barley Supplement to Improve the Reproductive Performance of Does.* Paper for VIIth International Goat Conference.

Manly B.F.J. (1998) Randomization, bootstrap and Monte Carlo methods in biology, Chapman Hall.

Rao, N.S.R. and Reddy, M.R. (1983) Utilization of Prosopis juliflora pods in the concentrate feeds of cattle and sheep. Indian Journal of Animal Sciences, 53(4): pp367-372.

Sidahmed, A (1995) Livestock and Feed Development and Improvement Research Needs in West Asia and North Africa. In: Gardiner, P and Devendra, C (eds) *Global Agenda for Livestock Research: Proceedings of a consultation.* International Livestock Research Institute: Nairobi.

Sprent P. (1993) Applied nonparametric statistical methods, Chapman Hall.

Singh, S.P. (1995) Favourite Agroforestry Trees. Udaipur: Agrotech Publishing Academy.

Case Study 6

WEALTH RANKING STUDY OF VILLAGES IN PERI-URBAN AREAS OF KUMASI, GHANA.

D Jeffries[1], H Warburton[2], K Oppong-Nkrumah[3], E Fredua Antoh[4]

[1] Natural Resources Institute, University of Greenwich
[2] Natural Resources Institute, University of Greenwich
[3] Kwame Nkrumah University of Science and Technology, Ghana
[4] Kwame Nkrumah University of Science and Technology, Ghana

WEALTH RANKING STUDY OF VILLAGES IN PERI-URBAN AREAS OF KUMASI, GHANA.

Background to Project

A high rate of urban growth is a common phenomenon in many countries. As cities expand the areas surrounding them are often subject to rapid changes: agricultural land may be converted to housing and commercial use, job opportunities for the peri-urban inhabitants may shift from those based on agriculture to more urban pursuits. Kumasi, the second largest city in Ghana, is one such city that has experienced rapid growth.

In 1997 a research project, the Kumasi Natural Resources Management Project, was started with an aim of investigating the effects of urbanisation on the use of natural resources and on the livelihoods of the people living in the peri-urban areas of Kumasi. The project was funded by DFID under the Renewable Natural Resources Research Strategy (RNRRS). The overall objectives of this project were to research ways of improving the management of key natural resources in the peri-urban area in ways that would enhance the sustainability of the resources and contribute to the livelihoods of the local inhabitants.

As part of the project, research was conducted into the means of livelihoods and access to resources of peri-urban inhabitants in order to find out how urbanisation was affecting local people, and to investigate who gained and who lost from the urbanisation process. After initial studies, work focused on 4 peri-urban villages, all located within 20 km of Kumasi. These were Apatrapa, Aburaso, Duase and Swedru.

Choice of Research Tools

Throughout the project a mix of qualitative and quantitative research methods were used. An initial village characterisation survey had identified three main village types within the Kumasi peri-urban area. These are the "urban", "peri-urban" and "rural"[5] villages. The four villages under study were therefore chosen from these clusters with at least one from each cluster.

Detailed research was undertaken in each village. This included a household survey and a wealth ranking exercise.

The household survey

There was very limited secondary data available on villages in the peri-urban area. For example, the most recent population and occupation census dates from 1984 when economic conditions in Ghana were very different from the current situation. Therefore a structured questionnaire survey was undertaken to find out the current situation regarding population, household and occupation structure, and to investigate the prevalence of factors such as the loss of agriculture land – an issue that had been raised in previous qualitative research. The specific objectives were to provide:

[5] These are relative terms – all the villages are within 40 km of Kumasi

- Quantitative data on types of houses and household structure;
- Quantitative data on individuals' access to resources and facilities, occupations and job location;
- Specific information on access to farmland and any loss of farmland due to urbanisation.

The household survey consisted of a structured questionnaire designed to collect selected information on every house, household and individual in the four villages. The topics covered information on the type of houses, rooms, facilities available and number of households living in each house. For each individual member of a household, information was collected on their gender, age, education, occupation, job location, years in the village, access to farmland, land taken for development.

The questionnaire was pre-tested then carried out by a team of four enumerators. The data was entered into a Microsoft Access database for analysis. Data entry screens were designed to aid the researchers in entering the data. Information from 365 houses and 5, 480 individuals was entered and checked.

Wealth ranking study

One of the main concerns of the project was to ensure that the situation and concerns of the poorer and more vulnerable people within the community were understood and taken into account in any proposed development of strategies to enhance natural resource use planning. Inequalities in wealth affect access to every aspect of people's lives such as services, education and health and the ability to respond to changes. Households of differing wealth will have different problems and needs as well as varying ability to adopt or adapt proposed technologies. In farming communities it affects availability of inputs and ability to adopt proposed technologies. We also wanted to assess who benefited and who lost out during the process of urbanisation, and the reasons for this. The wealth ranking study was designed to find out about local perceptions of wealth and poverty, and the relative wealth levels within the selected villages.

Why Wealth Ranking?
For a researcher, finding out about wealth and incomes within a community is difficult. People are often not willing to provide information on incomes. In addition, in communities such as the study villages, people are often self-employed with multiple sources of income and differing levels of support from family members, so it is difficult to calculate their income accurately. Questionnaires across the whole community are very time-consuming and can still miss important factors affecting wealth. Community group approaches can miss the poorest people, who often have low levels of involvement in community affairs and may be less likely to express their opinions in discussions.

The concept of wealth ranking is based on utilising local knowledge about people's levels of wealth. Local people who live and work in the same village and who can observe others over a long time period may be a better judge of levels of wealth than an outsider. Also, in all societies, the local people have their own concepts of wealth, which are not only dependent on cash income. There are socially established indicators of well being, for instance, the royal status of a person.

In a wealth ranking exercise, key informants from the local communities rank their fellow villagers into wealth categories. The informants decide on their own definitions of wealth and wealth categories. The wealth ranking exercise therefore helps to bring out the complexities and realities of wealth and poverty, rather than using definitions predetermined by the researchers.

The specific objectives of the study were:

- To find out what characteristics the local communities use in judging relative wealth of individuals.
- To establish the relative wealth of individual households within the communities within several broad categories of wealth.
- To identify the poorest and most vulnerable in the communities, and
- To provide a basis for selection of households and individuals for case studies across the whole range of the socio-economic spectrum.

In addition, there were objectives with reference to the methodology:

- To investigate whether wealth ranking could provide useful results (qualitative or quantitative) when used in the context of peri-urban villages, which contain a mixture of long-established and recent inhabitants.
- To investigate whether information from the wealth ranking could be used with information obtained from the questionnaire survey of the same households.

Methods

In each village, in conjunction with village leaders, the researchers identified three key informants who were said to have extensive knowledge of the village. The researchers requested that at least one woman be included in the informants.

Each informant was interviewed separately, and asked for their own ideas of wealth, and what makes one person better off than another. They were given a set of cards containing the names of all the household heads in the village and asked to arrange the cards into groups of similar wealth. (The list of names was derived from the household survey). The informants could use as many or as few groups as they wished.

Because there were a large number of household heads in each village, the informants needed sufficient time to investigate the status of the household heads[6]. The informants were therefore given four days to sort the households into ranks, so that they had adequate time to clear issues of uncertainties with regards to the actual status of a household. The informants were however advised not to discuss rankings with others so that they could render independent rankings. The team later held discussions with each informant about the basis for the categorisation of the households.

---

[6] Many wealth ranking exercises are completed in one day, but usually with a smaller number of households

On the fourth day the team met the informants and discussed the rankings and indicators of wealth. The informants went through the ranking process again in the presence of the researchers and, in some cases, made further divisions between the wealth ranks during the discussions.

Limitations.
In most ranking exercise, there are normally 'unknown ranks' where informants do not know all the members of the community well enough. However the four days given to the informants allowed them some time to inquire about the various household heads and their wealth status. This may have reduced of unknowns in the ranks, but there may still be some households where the informant knew little about their wealth status even though they assigned them to a wealth category.

There is also the temptation to play down the wealth status of the informants themselves and their families. This situation could however not be eliminated as the informants are given a free hand to do the ranking themselves.

The informants initially perceived the exercise as difficult. According to them, it might not be possible to assess all the household heads in the communities and also there was a risk that some people might be angry over their wealth status being discussed by the informants, as generally people are not prepared to disclose their wealth. Following the exercise, the informants did not report having experienced any difficulties in practice.

Results from the ranking exercise were entered into an Excel spreadsheet.

Findings

Each informant identified a different number of wealth strata. The number of strata ranged from 4 to 6. They ranked the households from 1 for the poorest strata up to the richest strata. Figure 1 shows the ranking scale chosen by informants from Aburaso.

Figure 1. Aburaso: Groups of ranks identified by Informants.

Informant 1	Informant 2	Informant 3
Very rich Rich Moderately rich Poor Very poor	Very rich Rich Poor Very poor	Very rich Rich Moderately rich Poor Very poor

The informants also described the wealth categories they had used. Figure 2 shows the results from Aburaso.

Figure 2. Aburaso: Indicators of Wealth by Informants

Informant 1	Informant 2	Informant 3
Very rich Own a house Own a private car Businessman Royal (access to sell land) Commercial transport **Rich** Cocoa farmers Traders/store operators Master artisans Vegetable farmers Mechanics Drivers Salary workers **Moderately rich** Masons Carpenters Hairdressers Dressmakers Kiosk operators **Poor** Hawkers Petty traders Cooked food sellers Construction workers Cobblers Food crop farmers **Very poor** Unemployed Disabled Aged Casual labourers Small scale farmers (usually on building plots	**Very rich** Own a house Businessman Royal (access to sell land) **Rich** Manager in a company Own a commercial vehicle Cocoa farmers Own a private car Traders/store operators Master artisans Vegetable farmers Mechanics Drivers Salary workers **Poor** Masons Carpenters Hairdressers Dressmakers Kiosk operators Hawkers Petty traders Cooked food sellers Construction workers Cobblers Food crop farmers **Very poor** Unemployed Disabled/aged with little/no support from family/ children Casual labourers Small scale farmers (usually on building plots	**Very rich** Own a house Businessman Manager in a company Own a commercial vehicle Own a private car **Rich** Cocoa farmers Person with support from children outside the country Traders/store operators **Moderately rich** Master artisans Vegetable farmers Mechanics Drivers Salary workers Shoe and leather works Masons Carpenters Hairdressers Dressmakers **Poor** Hawkers Petty traders Cooked food sellers Construction workers Large food crop farmers **Very poor** Unemployed Disabled/aged with little/no support. Casual labourers Small scale farmers

Although there was a variation in the number of ranks used by each informant, and the criteria used for each strata, there was a high degree of agreement between informants in all the villages over the characteristics of rich and poor. For example, food crop farmers and casual labourers were always assigned to the poor or very poor categories; vegetable farmers and skilled trades were ranked above food crop farmers; cocoa farmers, transport operators and business men assigned to the rich categories.

Using Quantitative Methods to Analyse the Wealth Ranking Data

Tests for consistency

One question that researchers wanted to determine was whether the ranks assigned by the informants were generally consistent for the same household. Although wealth ranking is based on subjective and relative criteria, if a significant number of households are ranked very poor by one informant, but very rich by another, then it would throw the ranking process into question. It might indicate that the informants had not understood, or had not wanted to rank households as requested by the researchers, or that informants did not know the wealth of many of the households. It might be expected that there would be less consistency in the larger, more mixed villages like Apatrapa, due to less knowledge about other villagers, than in a smaller, more rural village like Swedru.

Although the ranking scales are different, they can still be assessed for consistency. Spearman rank correlation can be used to compare pairs of informants and Kendall's coefficient of concordance [2] gives a measure of consistency between the rankings (0 no association to 1 exact correspondence). Figure 3 gives the results for all villages and it is clear that there is high level of correspondence between the informants and it can be assumed that the rankings of each informant are consistent. Differences in consistency between the villages do not correspond closely with the size of degree of urbanisation of the village.

Figure 3 Consistency measures for the informants' rankings

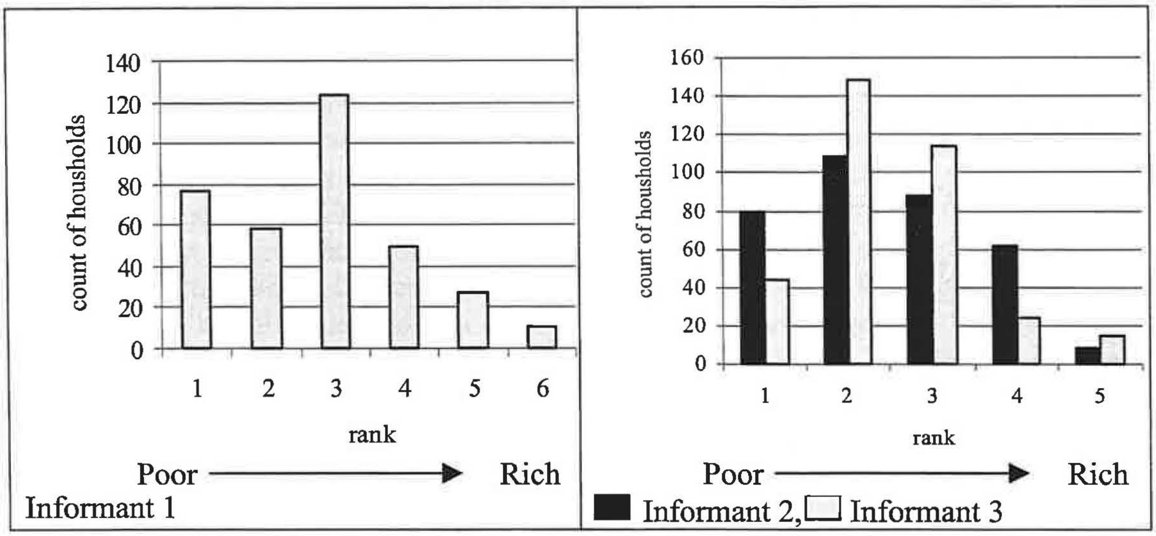
Village	Kendall's coefficient of concordance*	Spearman rank correlation*
Apatrapa	0.83 (p-value < 0.001)	**Informant 1** 1 **Informant 2** 0.78 1 **Informant 3** 0.73 0.74 1 **1 2 3**
Aburaso	0.79 (p-value < 0.001)	**Informant 1** 1 **Informant 2** 0.60 1 **Informant 3** 0.66 0.78 1 **1 2 3**
Duase	0.77 (p-value < 0.001)	**Informant 1** 1 **Informant 2** 0.58 1 **Informant 3** 0.75 0.63 1 **1 2 3**
Swedru	0.85 (p-value < 0.001)	**Informant 1** 1 **Informant 2** 0.83 1 **Informant 3** 0.72 0.76 1 **1 2 3**
	A p-value of less than 0.05 shows that at the 5% level the coefficient is significantly different form zero.	At the 5% level all correlations are significantly different from zero.

* All coefficients and correlations have been adjusted for ties.

Comparisons of ranks

The distribution of the ranks for the informants in Apatrapa is shown in Figure 4.

Figure 4 Rankings from each informant for Apatrapa



The distributions have similar shapes, confirming the agreement shown in Figure 4. As expected the distributions are skewed with more poor households than rich. This indicates that the median and the inter-quartile range in conjunction with the mean are the most appropriate summary statistics. Figure 5 shows these statistics for each of the villages, where the poorest households have been assigned a rank of 1 and the richest a rank of 4, 5 or 6 depending on the scale.

Figure 5 Mean, median, upper and lower quartile rankings

	Informant 1*				Informant 2*				Informant 3*			
	lower	med	upper	mean	Lower	med	upper	mean	lower	med	upper	mean
Aburaso	2	3	4	2.8	2	2	3	2.5	1	2	3	2.3
Apatrapa	2	3	4	2.8	2	2	3	2.5	2	2	3	2.5
Duase	1	2	3	2.3	1	2	3	2.0	1	2	3	2.0
Swedru	2	2	3	2.4	2	2	3	2.5	1	2	3	2.1

* Note this refers to the appropriate informant in each village, i.e. different informants were used in every village and the table represents 12 informants.

This table should be interpreted with great care due to the different rankings used by different informants in each village. Even if the informants use the same number of wealth strata, their definitions may differ. For example, in Aburaso informants 1 and 3 use the same scale (see Figure1), but informant 1 appears to rank households as richer than informant 3 (see Figure 2).

Comparisons across villages are more difficult than within a village as even though the ranking scales can be the same the meanings may be different. Limited empirical comparisons are possible across village, but great care should be taken with their interpretation.

Analysing the Wealth Ranking and Household Survey Data to Identify Factors Affecting Wealth Ranking

One of the objectives of the project is to discover which are important wealth indicators. One way of doing this would be to compare the wealth ranking data with the data collected for each household in the household questionnaire survey. In order to do this, the wealth ranking data has to be combined into a single score in some way. Because the informants did not choose the same scale this makes a combined ranking of wealth very difficult. One option is to use the results only from informants using the same number of ranks, and use the average score. The major disadvantage of using data from only two informants is that the information collected form the third informant is wasted. However since one of the informants always uses a different scale it is not straightforward to combine the rankings.

One method for combining informants that use different ranks is to convert the ranks to scores for all three informants and then average across all informants. Figure 6 shows two possible procedures for averaging over ranks of each informant.

Figure 6 Converting ranks to scores.

Method 1

Convert ranks to scores by dividing the rank for each informant by the number of ranks and then average for all informants. For Aburaso Informants 1 and 3 use five ranks and their scores range from 0.2 to 1. Informant 2 uses four ranks and the scores range from 0.25 to 1. The scores have therefore been aligned as shown below.

Informants 1 & 3		Informant 2
Very rich	←	Very rich
Rich		Rich
Moderately rich		Poor
Poor	←	Very poor
Very poor		

Method 2

An alternative approach is to align the extreme ranks for Informant 2 with those of Informants 1 & 3 as shown below.

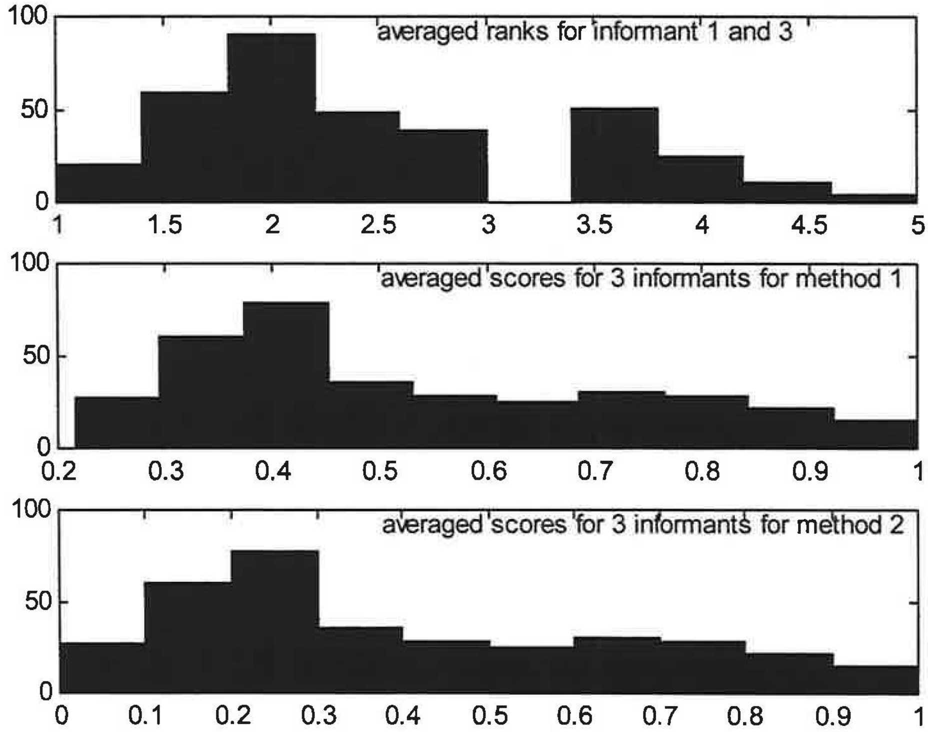
Informant 1 & 3	Informant 2
Very rich ←	→ Very rich
Rich	Rich
Moderately rich	Poor
Poor	→ Very poor
Very poor ←	

This is achieved by subtracting 1 from each rank and dividing by (number of ranks – 1), which converts both scoring systems to 0 to 1.

Both these techniques make the assumption that the wealth rankings are equi-spaced. It is very difficult to verify whether this assumption is valid, but it does give a method for combining the information from the three informants.

Figure 7 compares the distribution of the averaged ranks for two informants with the distributions of the combined scores of all three informants using methods 1 and 2 for Aburaso.

Figure 7 Comparison of averaged ranks and two scoring systems.



It is clear that the distributions for methods 1 and 2 are very similar, although they have slightly different ranges. They also represent the main features of the averaged ranks.

Factors Affecting Wealth Ranking

The different methods of combining the wealth ranking data were used in turn in analysing the data. The following example is of an analysis of data from Aburaso using regression analysis.

Prior to the regression analysis other forms of analysis such as chi-squared and canonical variate analysis and trend analysis had been used to look at associations of factors such as education level, possession of facilities, gender and dependancy ratios with wealth rank. Results from these were used in selecting the explanatory variables used for the regression analysis.

The first stage in a regression analysis is to compile a list of the explanatory variables and their description as shown in Figure 8. This section uses only data from Aburaso, from which one record was deleted due to missing values. For regression techniques to be effective there must be no missing values in the explanatory variables.

Multiple regression [3] is a complex procedure, particularly when there are many variables and the use of statistical packages does not ensure a valid model.

The idea of multiple regression is to try and select a subset of explanatory variables (any variable form Figure 8) that can give a significant and meaningful fit to the response (the combined wealth ranks).

It has already been shown that there are relationships between the variables, but an important assumption of multiple regression is that the explanatory variables should not be linearly related. Meaningful groupings of the variables as in Figure 8 (the relational structure of the database should reflect these groupings) makes it easier to spot which variables might be correlated. Scatter plots and correlation coefficients should also be used. In this manner the number of variables can often be greatly reduced.

The response variables in this case are the average ranks. One of the major assumptions of analysis of variance is that the data is continuous. This is clearly not the case and at this stage the multiple regression can only be thought of as a tool for selecting subsets of variables likely to influence wealth ranking. When suitable models have been derived, more consideration is required before these can be claimed as predictive models.

When performing multiple regression it is important to differentiate between qualitative variables (e.g. fridge ownership) and quantitative variables (dependency ratio or counts) and most packages have facilities for this.

Figure 8 List of explanatory variables and their types.

Variable	Type	Description
Hhsize	Count	Number of people in household
Adult15	Count	Number of adults of 15 and over
Child14	Count	Number of children of under 15
OAP	Count	Number of adults of 65 and over
Adfem15	Count	Number of adult females 15+
Admale15	Count	Number of adult males 15+
Female	Count	Number of females (all ages)
Wrkadult	Count	Number of adults of working age (15-64)
Depratio	Continuous	Dependency ratio: (Child14+OAP)/Wrkadult
HHHsex	Qualitative (2 levels)	Gender of household head
Maritals	Qualitative (7 levels)	Marital status of household head
Ethnic	Qualitative (11 levels)	Ethnic origin of household head
HHHage	Count	Age of household head
Adultsrm	Count	Number of adults per room
Childrm	Count	Number of children per room
Peoplerm	Count	Number of people per room
HHHEduc	Qualitative (6 levels)	Educational level of household head
HHHapprent	Qualitative (2 levels)	Apprenticeship training
SSS	Count	number with secondary school education
JSS	Count	Number with junior or school education
Primary	Count	Number with primary school education
Noeduc	Count	Number with no education
Apprent	Count	Number with apprenticeship training
Student	Count	Number of students in household
Unemployed	Count	Number of unemployed in household
Casulab	Count	Number of casual labourers in household
Farmers	Count	Number of farmers in household
Minfrmer	Count	Number of farmers (second occupation) in household
Houstype	Qualitative (3 levels)	Compound, villa or simple
Dwaterin	Qualitative (4 levels)	Type of water source in house
Toiletin	Qualitative (6 levels)	Type of toilet in house
Electric	Qualitative (2 levels)	1 if Electricity in house
Backyard	Qualitative (2 levels)	1 if backyard round house
Roomnum	Count	Number of rooms in house
Radio	Qualitative (2 levels)	1 if present
TV	Qualitative (2 levels)	1 if present
Fridge	Qualitative (2 levels)	1 if present
Cycle	Qualitative (2 levels)	1 if present
Motorbik	Qualitative (2 levels)	1 if present
Car	Qualitative (2 levels)	1 if present
Yrsvill	Count	Number of years lived in village
Residst	Qualitative (6 levels)	Residency status
Hometown	Qualitative (14 levels)	Hometown of household head
Roomten	Qualitative (6 levels)	Type of room tenancy
Rent	Continuous	Monthly room rent
Majocc	Qualitative (72 levels)	Major occupation of household head
Minocc	Qualitative (27 levels)	Secondary occupation of household head
Villfarm	Binary	Household head has a farm in the village
Farm1typ	Qualitative (6 levels)	Type of main farm
Farm1ten	Qualitative (6 levels)	Tenancy of main farm
Farmotyp	Qualitative (3 levels)	Type of farm outside village
Farmoten	Qualitative (6 levels)	Tenancy of farm outside village
Workplac	Qualitative (41 levels)	Place of work
Landloss	Continuous	Land lost by household

It is difficult to assert that the ranks or scores are continuous and one of the major assumptions of any regression analysis is continuous data. At this stage multiple regression [3] can only be thought as a tool for selecting subsets of variables likely to influence wealth ranking. The data is also skewed, but regression analysis is generally robust to departures from normality.

Figure 9 gives appropriate multiple regression models for each of the responses for Aburaso.

Figure 9 Multiple regression models.

Average ranks for informant 1 and 2
Average rank = 2.0 + *gender of household head* + *radio* + *number of years lived in village* + *type of outside farm* + 0.23 * number of adults over fifteen +0.007 8* dependency ratio – 0.21 * number with no education – 0.43 * casual labourers.
Method 1
Average score = 0.37 + *gender of household head* + *radio* + *number of years lived in village* + *type of outside farm* + 0.05 * number of adults over fifteen +0.0016* dependency ratio – 0.043 * number with no education – 0.12 * casual labourers.
Method 2
Average score = 0.20 + *gender of household head* + *radio* + *number of years lived in village* + *type of outside farm* + 0.06 * number of adults over fifteen +0.0020* dependency ratio – 0.056 * number with no education – 0.15 * casual labourers.

At the 5% level each model contains the same significant terms and the inference from each model is similar, although of course the coefficients are different. The terms in italic are qualitative variables and to complete the models the estimates of their effects must be given. Figure 10 gives the effects for the qualitative variables, where the effect for the first level of each variable has been set to zero.

Figure 10 Table of effects.

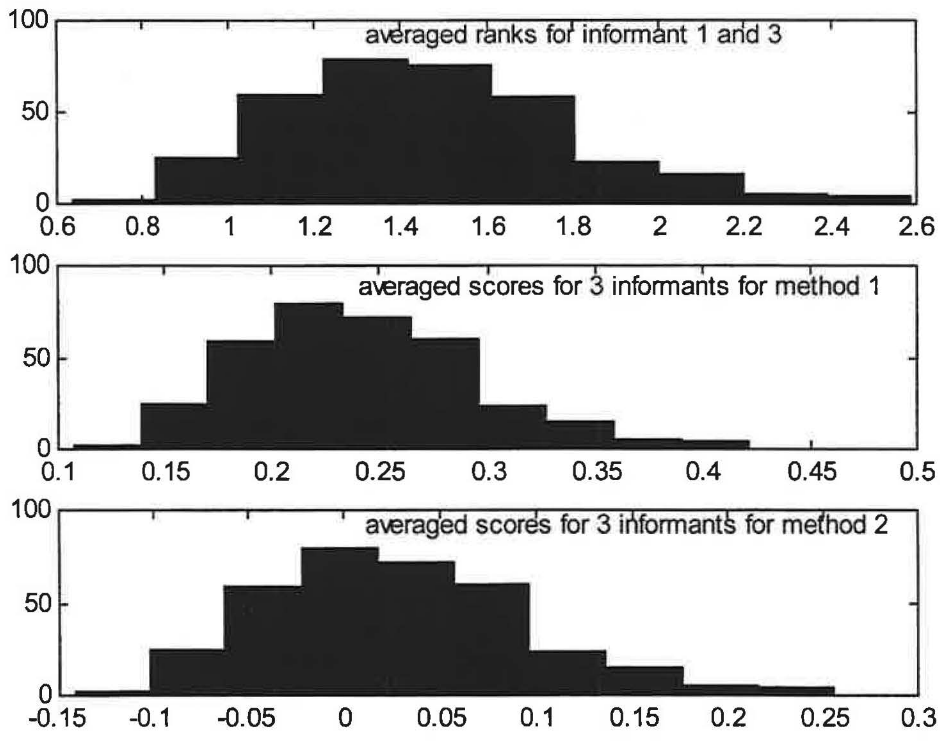
	Average ranks	**Method 1**	**Method 2**
effect	**Value**	**value**	**value**
Female household head	0	0	0
Male household head	0.17	0.043	0.05
No radio in household	0	0	0
Radio in household	0.43	0.09	0.11
<1 year in village	0	0	0
1-5 years in village	-0.28	-0.09	-0.11
6-10 years in village	-0.08	-0.07	-0.10
11-20 years in village	-0.86	-0.22	-0.29
>20 years in village	-1.53	-0.25	-0.32
Whole life	-0.43	-0.10	-0.13
No outside farm	0	0	0
Food crop	-0.06	0.002	0.001
Tree crop	0.50	0.09	0.12

Again the values vary, but the inference from each model is similar. The inference from fitting this multiple regression can be summarised as:

- Male headed households and households with radios have greater wealth
- Newcomers to the village are perceived to have greater wealth
- Of those households with outside farms, those with tree crops are the wealthiest
- The richer the household the larger the number of adults over 15 years.
- High numbers of household members without education or high numbers of casual labourers lower the wealth ranking.
- Wealth ranking increases as the dependency ratio increases.

The regression model has so far been employed to select sets of variables that are useful wealth indicators, but the predictive power of these models has not been considered. Figure 11 shows the fitted values for each of three models. One immediate problem is that the fitted values can be outside the permitted range of ranks or scores.

Figure 11 Histograms of fitted values.



The distributions are similar and all are highly skewed to low values, which does reflect a clear trend from Figure 6. The representation of the higher wealth rankings is poor.

This illustrates that although the models in Figure 9 are useful for variable selection, they are not accurate predictive equations. The predictions might be improved by considering interactions and including terms up to the 10% significance level. Although it might transpire that there are terms that have not been measured but that are important criteria for wealth ranking and it might not be possible to obtain a satisfactory predictive model.

Lessons Learned

The importance of the wealth ranking study is not based only on the ranks assigned by each informant. The process of wealth ranking is as important as the final ranks, as it is during this process that the researcher is able to learn about perceptions of poverty. There is a potential trade-off between making the results easier to analysis statistically, by, for example, predetermining the number and criteria of the wealth strata used by the informants, and allowing the informants to categorise as they wished. If the primary purpose of the study is to discover more about aspects of wealth and poverty and people's perceptions, then predetermining the criteria to be used will defeat the purpose. On the other hand, if the researchers already know sufficient about the wealth criteria they wish to use, and know that they want to relate the results to other data sources, then restricting the wealth strata will make the data easier to analyse statistically.

Researchers need to be clear from the outset what their priorities are. In this case, investigating perceptions of wealth was a main objective: analysing the data statistically was an additional method, so we have to accept the limitations of the analysis.

If researchers do not restrict the strata used then extreme care is needed in combining the results of different ranking scales. It is extremely difficult to justify combining ranks of different scales. An additional problem with ranking data is that it does not conform to any distributional assumptions and the inference from parametric methods cannot be assumed to be accurate.

However, as long as caution is used in interpreting the results, it is possible to derive useful information from the qualitative-quantitative analysis. We were able to investigate the consistency of the ranks assigned, and this does provide some additional assurance about the wealth ranking process.

In analysing the wealth ranking data with the questionnaire data from the household survey it is evident that for a complex survey such as this with many explanatory variables there is no one prescriptive technique that will provide all the answers. There is no one "correct" way of combining wealth ranks and the regression analysis illustrated here is one among many different techniques that can be used depending on the type of data.

However, the fact that the regression model using three alternative methods of combing wealth ranks yielded some similar results, does indicate that the significant variables identified may be important factors related to poverty, and are worth further investigation. For example, the results indicate the relationship between the female-headed households and the lower wealth ranking categories.

In summary the analysis in this case study has provided summary statistics for the comparison of informants and indications of which characteristics are relevant for wealth assessment. Adding the quantitative analysis to the wealth ranking study has been useful in strengthening hypotheses based on qualitative analysis. As long as the analysis is not pushed too far and inferences drawn which are not supportable (for example, in comparing wealth levels), it can yield useful results.

References

[1] http://www.rdg.ac.uk/ssc/dfid/booklets.html

[2] P. Sprent, 'Applied non parametric statistical methods'; Chapman & Hall, ISBN 0 412 44980 3

[3] R. Mead & R. n. Curnow, 'Statistical methods in agriculture and experimental biology'; Chapman & Hall, ISBN 0 412 24240 0

[4] W.J. Krazanowski, 'Principles of multivariate analysis' ;Clarendon Press, ISBN0 19 852230

[5] T.P.Ryan, 'Modern regression methods'; Wiley, 1996.

[6] Natural Resources Institute (June 1999), *Kumasi Urban Natural Resources Studies,* Kumasi Natural Resources Management Research Project (R6799), Chatham. Project report for DFID/NRSP.