DFID CPP Project R8410

Improving the effectiveness of research within NARO, Uganda

Survey Report on

Assessing Statistical and Data Management needs of NARO researchers

September 2005

by

Savitri Abeyasekera and Ian Wilson
Statistical Services Centre,
The University of Reading

in collaboration with

The Monitoring, Evaluation and Policy Unit,
NARO, Entebbe, Uganda

Executive Summary

In April/May 2005, a postal questionnaire was used to elicit views of all NARO scientists and technicians about their current levels of skills in statistics and data management, and to learn about their future needs in these areas. The questionnaire was forwarded by the Monitoring, Evaluation and Policy Unit of NARO by e-mail to Directors of all NARO institutes and Managers of ARDCs with a request to copy it to all their scientists and technicians. Responses were received by 47% of the scientists and 50% of the technicians. Although this does not constitute full coverage, we believe that the findings reported below give a reasonable reflection of the true situation concerning NARO researchers' views of their needs.

Most scientists (95%) were of the view that they would use a statistician's assistance in their research work if one were available at their institute. The area where most respondents (87%) expected to involve a statistician was in questionnaire surveys, followed by livelihoods analysis (75%) and then by on-farm trials (69%) and on-station trials (68%). This perhaps reflects lower confidence amongst respondents in survey work compared to studies involving experiments. This finding is consistent with the statistical topics on which respondents had received training. About two-thirds of scientists and about 30% of technicians, had received training in biometrics or statistics, but none mentioned having had training in survey methods and only eight mentioned learning about sampling methods. Despite this, it is interesting to note that 63% of scientists and 31% of technicians, claim to have "some" or "lots of" knowledge about sampling methods. Perhaps this indicates a misconception that "sampling methods" refers only to the ability to select a sample appropriately once the method is specified, or thinking of sampling as representing just simple random sampling.

A surprising number of scientists (17%) said they had no knowledge of how to generate tables of counts and percentages which are most relevant to the analysis of survey data. This indicates that courses in biometrics and statistics, followed as part of a University degree probably did not include any survey components and maybe that the focus of much NARO research has changed and broadened since the times when staff were trained. Knowledge of other techniques was also rather weak amongst scientists with many having no working knowledge, or only a little knowledge, of 2-way analysis of variance (45%), of factorial designs (55%) and multiple regression (62%), despite these being areas featuring commonly in experimental agricultural and biological research.

Statistical software packages most familiar to scientists were SPSS, Genstat and MStat with about half the respondents having little to lots of knowledge in one of these packages. With the recent purchase of Genstat by NARO, those with Genstat experience will hopefully be able to share their skills with others in their Institutes. MStat is now generally regarded as an obsolete package, so familiarity with its use will not be of much benefit to NARO researchers and its further use should be discouraged. Whether or not knowledge of SPSS is useful will depend on whether NARO purchases this package in the future, or some other package such as STATA which has been incorporated into the teaching programme at the Institute of Statistics and Applied Economics in Makerere University.

Most respondents (74%) had been involved at some stage in planning a data recording sheet. However very few persons (25%) had had training in how to organise data in a spreadsheet so as to avoid errors during data entry. Procedures most commonly used for checking data quality were scrutinising data collection sheets, manual checking of paper records against computer versions and capturing errors during data analysis. Standard operating procedures for data validation, and devoting sufficient time to this activity seemed to be lacking. Only a few respondents (23%) were using Excel's substantial facilities for data validation although the basics of this software were familiar to most researchers.

The most important priority training area identified by (81% of) scientists was in *standard methods of statistics using appropriate software*. About 40-50% also identified *research data management*, *integrating qualitative and quantitative approaches*, and *interpreting results and reporting* as one of their top three training priorities. For technicians, training in both *research data management* (77%) and in *standard methods of statistics* (76%) were in the top three priorities, followed by *interpreting results and reporting* (59%). About two-fifths of technicians also listed *design of on-station and on-farm trials* as one of their top three priority training areas.

Of the 82% of respondents who expressed an opinion about *areas of concern regarding statistics and/or data management*, 71% recommended that they be given training in statistics and data management, while 22% recommended purchasing relevant software. About a fifth of respondents also recommended employing a statistician/data manager at their institute.

With respect to ***constraints to research in general***, over 95% of scientists and over 90% of technicians identified *software, training and support in statistics and data*

management as one of their top three constraints. This was followed by the need for *professional upgrading in their own discipline* (79%). About a third of respondents also gave *lack of computer access* and the *need for training in technical writing and time management,* as well as the need for general training as a further constraint.

The above findings clearly demonstrate NARO researchers' awareness of the limitations they face in conducting good quality research. There is clear evidence that they would welcome training programmes aimed at enhancing their skills in modern statistical approaches and in ways of managing their research data more effectively. The need for relevant software has been substantially dealt with, but adequate access to computers has yet to be addressed at some institutes and ARDCs. There is also a clearly expressed need to have the support of an Institute/ARDC based statistician. Recruitment of such statisticians under the new NARO structure is needed as soon as possible, together with appropriate plans to provide them with training opportunities and back-up support from locally based statisticians and biometricians, say at Makerere University, until they gain relevant consulting experience.

Assessing Statistical and Data Management needs of NARO researchers

The Monitoring, Evaluation and Policy Unit (MEPU) of NARO, in collaboration with the Statistical Services Centre (SSC) of the University of Reading, U.K., conducted a survey in 2005 of all NARO scientists and technicians about their current levels of skills in statistics and data management, and to learn about their future needs in these areas. This document reports the findings from this survey.

Annex 1 provides a copy of the questionnaire used. Results from the questionnaire are expected to complement the findings reported in Sections 2 and 3 of the Draft Consultation Document offering first-phase views and recommendations from a Training Needs Analysis in Statistics and Data Management. These findings taken together will help to determine areas where training is most needed to achieve good quality research outputs.

The questionnaire was pilot tested with some scientists and technicians at the Coffee Research Institute in March 2005 and was subsequently modified. It was forwarded by e-mail to Directors of all NARO Institutes and Managers of the ARDCs, with a request to copy it to all research active staff in the Institute, receive their feedback, add any comments regarding the Institute's own vision and priorities as to its needs and then forward same to MEPU for onward transmission to Reading. The aim was to seek responses from all relevant staff so that training requirements within each institute, and across NARO as a whole, could be established more clearly, based on the perceptions of needs of the individuals and the Institutes/Centres concerned.

Responses were received from 94 scientists and 105 technicians, a total of 199 completed questionnaire returns. Of these 23% were from project funded staff. The scientists' response rate was 47%, while the technicians' response rate was 50%. These percentages include NARO staff from all research institutes and ARDCs. However, when results were restricted to those institutes and centres that <u>did</u> submit returns, the return rate for both scientists and technicians was about 56%. Annex 2 shows the details.

We are aware that up to half of the scientists and technicians did **not** return our questionnaires. The most radical assumption would be that **none** of them would **ever** use statistical approaches or support. This is not very plausible and this report instead assumes their responses are "missing at random"; i.e. those concerned were on courses, on leave, in the field or just too busy – or disorganised at the time to reply. Thus this

report assumes non-respondents would have given much the same answers, had they replied, as those who did respond.

Results from each component of the survey questionnaire are presented below, together with some discussion. Our conclusions are reported in the final section of this document.

1. Staff views on whether a statistician, if available, would be used

We asked respondents to consider a recently completed or current project and indicate whether a range of specified activities would be relevant to an activity or part of an activity of their project, and if so, whether they would involve a statistician in these activities if such a person was available near at hand. Nearly all respondents (95%) were of the view that they would. Of the 10 persons who didn't think they would use a statistician, 8 were technicians – mostly with no knowledge of statistics and therefore possibly unaware of how a statistician could help. The remaining two were scientists with considerable knowledge of statistical techniques and software. They probably felt they had enough knowledge for their own research purposes.

Table 1 summarises the findings. Here, the first two columns of percentages (adding to 100%) are restricted to those for whom the activity was relevant to the project being considered. It is seen that the percent of respondents who would use a statistician is high (> 68%) for most of the activities. However about half the respondents did not see a need to involve a statistician in participatory work with farmers or in laboratory studies. The latter appears to indicate that some respondents see their laboratory-based activities as involving straightforward designs and data analysis which they could comfortably handle. No doubt they are in some cases right, e.g. if the laboratory work is being done by a technician as a service provider within a bigger project. It is however surprising to find that a statistician's involvement is, for half the respondents, not recognised as potentially useful in farmer participatory studies. This is an indication that some NARO staff are unaware of statisticians' potential to assist with qualitative type enquiries and associated data collection procedures. Perhaps they have not encountered statisticians who are equipped, or interested, to help with such issues.

The final column of percentages in Table 1 shows that activities that are most likely to occur within a range of NARO projects are participatory work with farmers and on-farm trials (>80%), followed by on-station field trials and questionnaire surveys (>70%).

Other areas mentioned where a statistician would be involved include data entry, management, collection and/or analysis (9 responses), study design (3 responses), and sampling (2 responses), as well as more specialised areas such as cluster analysis, temporal and spatial analysis, estimation of genetic progress, farming systems analysis, and biodiversity studies, mentioned by individual respondents. Areas where a few individual respondents did not perceive a need to involve a statistician included technology invention, partnership studies, needs assessments and participatory economic analysis.

Table 1. Perception of needs for a statistician's involvement and relevance of activity area to current or completed project

	Activity relevant to current/completed project			Percent for whom activity is relevant to project (sample size=199)
	Would involve statistician (%)	Would **not** involve statistician (%)	Sample size	
a. Questionnaire Survey	87.1	12.9	140	72.4
b. On-farm trial	69.4	30.6	160	82.4
c. Participatory work with farmers	48.2	51.8	170	87.4
d. On-station field trial	68.4	31.6	152	78.4
e. Livelihoods analysis	74.7	25.3	99	51.8
f. Laboratory study	46.6	53.4	118	61.3
g. Other type of study 1	76.7	23.3	43	N/A
h. Other type of study 2	91.7	8.3	12	N/A

2. Current skills in statistics and data management

The level of skills that NARO staff have in statistics and data management were assessed by asking them about (a) training courses they have attended, when done, duration and software used; (b) their perception of their abilities in a range of statistical topics; (c) their practical knowledge in a number of areas; and (d) their knowledge of software.

2.1 Attendance at training courses

Most NARO scientists (93%) had had some training in statistics and/or data management, usually as part of a degree programme. Amongst the technicians this was much lower at 39% - unsurprising since technicians are less likely to have had first degrees, where statistics is generally taught. Of those respondents who had received some form of training, the training courses received by most (76%) were in the areas of biometrics and statistics, with many having followed more than one of these courses. Other frequent areas of training mentioned were: - dealing with data from participatory studies (persons who attended the SSC training in January 2004), data management,

research methods, experimental designs and the use of statistics software packages. Much of this training was through courses done in a single university term. Only 13% of the respondents had longer training of more than one University term. About half of the respondents had had their training in the past five years. However, only about 40% of those who received training had had the training based on a relevant statistical software package.

2.2 Perception of knowledge of statistical techniques

As expected, knowledge of statistical techniques, in terms of feeling able to understand and use the method without help from anyone else, was higher amongst scientists than technicians. Generally few technicians had much statistical knowledge beyond descriptive statistics (see Table 2). One surprising outcome was the claim by some to having 'little' to 'lots' of knowledge about sampling methods when they didn't know any of the statistical techniques appearing in Table 2 (6 persons) or knew only about simple descriptive summaries and graphs (20 persons). This is likely to be because of a possible misunderstanding on their part in thinking that "sampling methods" just means being able to do a sample selection once the method has been specified, rather than seeing or thinking about sample design as a key part of research design.

Amongst the scientists, most (88%) felt that they knew about basic summaries, but a surprising number (17%) claimed to have no knowledge of how to generate tables of counts and percents, most relevant to analysis of survey data. Perhaps this is because courses in biometrics for agriculture and biology students rarely include any survey components. Knowledge of other techniques was also rather weak amongst scientists with many having *no working knowledge* or only *little knowledge* of a two-way anova (45%), of factorial designs (55%) and multiple regression (62%). These are areas which feature commonly in agricultural and biological research. Only a few had *some* or *lots of knowledge* of more advanced methods such as logistic regression (9%) and multivariate techniques (13%), or of the less often used non-parametric methods (18%).

Table 2. Knowledge of statistical techniques (%'s per type of respondent)

Areas in Statistics	Scientists (n=94)				Technicians (n=105)			
	None	Little	Some	Lots	None	Little	Some	Lots
Tables of counts & percents	11	12	34	44	46	23	26	6
Tables of means & summaries	6	5	37	51	38	30	27	6
Standard errors	13	22	40	25	65	22	11	2
Simple Linear Regression	13	27	35	26	74	15	9	2
Chi-square tests	18	31	36	15	83	11	5	1
Comparing 2 means using t-tests	16	27	36	21	79	12	8	1
One-way analysis of variance	15	19	37	29	69	18	11	3
Two-way analysis of variance	25	20	30	26	80	10	9	2
Factorial designs	36	19	32	14	83	11	4	3
Multiple regression	32	30	32	6	87	11	2	0
Sampling methods	11	27	49	14	37	32	24	7
Non-parametric methods	47	35	16	2	93	5	1	1
Logistic regression	61	31	9	0.0	95	5	0	0
Multivariate analysis	54	33	11	2	91	6	3	1
Graphs using Excel	6	11	39	44	61	19	13	7
Graphs using other software	37	21	32	10	76	19	4	1

Responses to current levels of knowledge of statistical techniques were also summarised by the authors by giving scores to respondents' claimed knowledge levels:- 0 for 'none', 1 for 'little', 2 for 'some' and 3 for 'lots'. The average score for knowledge in statistics for scientists was 1.47 i.e. between 'little' and 'some' with 95% confidence limits of 1.3 to 1.6, while for technicians the average score was 0.43, i.e. between 'none' and 'little', with confidence limits of 0.34 to 0.52. Since these scores are based on a 0-3 scale, these results again emphasise the low level of knowledge of statistical methods amongst NARO staff.

One specific problem where *most* staff have limited knowledge is that there are few good examples to follow. Many staff would not know how much they do not know. Those who reckon they know *lots* may be better-informed than immediate colleagues yet still have serious gaps.

2.3 Perception of practical knowledge in key subject areas

The survey questionnaire also asked respondents to assess their own level of practical knowledge in a number of areas. The results are shown in Table 3. Over 80% of scientists and more than 44% of technicians had *little* to *lots* of practical knowledge in all areas specified except those with respect to *livelihoods analysis* and *natural resources*

monitoring. Retrospectively, responses to this question are not very useful since the level of practical knowledge would depend on the types of projects that respondents have so far got involved with, in their work. Of course the use of scales such as 'none, little, some, lots' is also affected by the self-perceptions and attitudes of the individuals concerned.

Table 3. Practical knowledge in specific subject areas (percentages per type of respondent)

Subject area	Scientists (n=94)				Technicians (n=105)			
	None	Little	Some	Lots	None	Little	Some	Lots
On-farm studies	12	18	50	20	27	35	34	4
Farming systems research	19	34	35	12	32	42	25	1
Farmer participatory approaches	7	26	44	23	18	31	47	5
PRA/PLA	10	29	38	23	38	25	33	4
Survey methodology	4	28	50	18	31	41	24	4
Sample size determinations	5	36	47	12	43	39	16	2
Impact assessment studies	18	43	35	4	55	27	18	0
Livelihoods analysis	36	34	22	7	58	29	13	0
Project monitoring & evaluation	14	40	35	11	42	31	25	3
Natural resources monitoring	44	35	18	3	72	21	7	0

2.4 Knowledge of statistics and computer software

Respondents were also asked to rate their level of competence in the use of a range of statistical and computing packages, again on a 0-3 scale with 0 representing 'none' and 3 representing 'lots'. The results are summarised in Table 4. It may be observed that amongst scientists, the most familiar packages are SPSS, Genstat and MStat with about half the respondents having 'little' to 'lots' of knowledge in one of these packages. MStat is a nearly obsolete package no longer being developed or supported, and is extremely limited in its capabilities. It tends not to encourage good statistical practice.

Knowledge of Genstat amongst some staff will be of great assistance to many of the NARO institutes in view of NARO having recently purchased an institutional licence for Genstat so that it can be distributed to all institutes and ARDCs. Amongst survey respondents, several even said they have 'lots' of knowledge of Genstat, viz. 6 from SAARI (2 of these being technicians), 3 from NAARI and 1 each from KARI, FOSRI and Abi ARDC. It would therefore be reasonable to involve these persons as resource personnel in any training that is done within the institutes to upgrade staff skills in the use of Genstat software. Similar comments apply to SPSS – a package particularly useful for survey data analysis. However, in addition to being rather expensive, this is

currently not widely or legally available at NARO. Such a survey data analysis package would be extremely useful to complement facilities that Genstat provides, but NARO has still to decide whether SPSS or Stata is the better choice, given that

(i) there is a bigger existing user base for SPSS, but

(ii) Makerere University (ISAE) has now moved to using Stata

(iii) Stata is relatively cheap and can be purchased outright, whereas SPSS is much costlier and probably involves annual licensing.

Table 4. Competence in statistics and computer software (%'s separately for scientists and technicians)

Software package	Scientists (n=94)				Technicians (n=105)			
	None	Little	Some	Lots	None	Little	Some	Lots
SPSS	45	26	26	4	89	8	4	0
Stata	87	10	2	1	97	3	0	0
SAS	64	17	14	5	94	5	1	0
Minitab	83	10	4	3	98	2	0	0
Mstat	51	11	21	17	75	12	11	2
Genstat	50	22	17	11	90	4	5	2
Excel	2	16	51	31	37	31	23	9
Access	57	29	12	2	80	12	8	0

Competence in use of the database package Access is low, but this is not of great concern if there can be just one or two competent persons at each institute/centre who can design and effectively use a database in a package such as Access, for archiving data from all research at that institute or centre. Competence in Excel on the other hand, is high amongst scientists and reasonably adequate amongst technicians. This indicates a high level of at least basic usage of Excel, the likely software package for entering and organising data from research activities. However, of those who had little to good knowledge of Excel, only 35% of scientists and 20% of technicians had had any training in using a spreadsheet to avoid errors during data entry (see also section 3 below). The ability of all staff to make full use of good features in Excel for data entry, validation and management would be highly desirable.

3. Practices and training in data management

About 74% of respondents had been involved at some stage in planning a data recording sheet. Comments on this aspect were given by about 20% of the respondents. Of those who had been involved in planning a recording sheet, some (11 respondents) commented that their experience was little, it was done under guidance, or done informally, and that they would welcome gaining more skills in this area. Nearly all comments from those who had **not** been involved in planning a recording sheet, were to request training in this area. Other comments were largely to indicate where or when the activity happened, or to specify the area of research (e.g. on-farm, on-station trial) which required this activity. Two persons suggested considerations that were needed during the process of preparing a data recording sheet, i.e. the need to consider treatments, numbering, etc., associated with the study, and the need to consider just the important variables so that time was not wasted collecting non-useful information. These are worthwhile observations.

Few persons (25%) had had training in how to organise data in a spreadsheet to avoid errors during data entry. About a third of such persons also gave further comments, mostly to indicate that they still needed more guidance and training in this area. Comments were also received from about 22% of those who had not had training, again largely to indicate that they needed this training. Some other comments given were "I'm teaching myself", "knowledge by working with established staff", "tried to learn from those who know", "no formal training but have read literature from SSC", "no practice", and "this is important to avoid wasting time".

Table 5 shows a summary of multiple responses given to the question regarding procedures currently in use or recently used, to check data quality. Procedures most commonly applied for checking data quality were scrutinising the data collection sheets (67%), manual checking of paper records with computer versions (39%) and capturing errors during data analysis (33%). Although Excel is familiar to most staff, it is interesting to note that relatively few use Excel's facilities for data validation.

About 15% of respondents gave comments on the above issue. Eight of them asked for training in this area. Others mainly expanded on the responses they gave, while three indicated that senior officers were responsible for data quality. One said he/she collected qualitative data, another admitted to not checking the data for errors, while two highlighted the lack of knowledge, software and computers.

The above is an important entry point for efforts to upgrade the overall quality of NARO's data-oriented research. Clearly the issues involved affect a great many individuals and projects, many of the issues are not intellectually inaccessible, and good data entry and data management is a corner stone of ensuring overall quality and efficiency.

Table 5. Percent using specific procedures for checking data quality

Procedure used for checking data quality	Number using procedure	Percent of persons using procedure (n=187)
Does not apply to the work I do	6	3.2
Scrutinising the data collection sheets	125	66.8
Using Excels facilities for data validation	42	22.5
Random checks of the computerised data	44	23.5
Not sure since someone else's responsibility	25	13.4
Manual checking of paper and pc records	73	39.0
Capturing errors using EDA	35	18.7
Capturing errors during data analysis	62	33.2

The request for information on the number of days spent each year on data collection, data validation, data entry, data management and doing simple summaries and graphs, drew many responses that did not specify the requested number of days. This is likely to have been difficult to specify for a number of people, and some gave either percentages for each item, or the frequency with which they did the activity, e.g. routinely, monthly, etc. The data were therefore re-coded to indicate whether the activity took little, some, or lots of time. The results are shown in Table 6. It is assumed that the non-responses (from the 94 scientists and 105 technicians who completed the questionnaire), correspond to those for whom the activity is not relevant because it's not part of their work.

Table 6. Percent of time spent on data related activities

	Scientists				Technicians			
Data related activity	**n**	**Little**	**Some**	**Lots**	**n**	**Little**	**Some**	**Lots**
Collecting the data	88	27	58	15	98	15	57	28
Checking the data for errors	84	88	10	2	75	72	27	1
Doing the data entry	84	68	29	4	81	56	36	9
Organising and managing the data	83	63	34	4	73	63	29	8
Doing simple data summaries and graphs	88	57	40	3	64	75	23	2

The summaries in Table 6 are based on crude estimates of time spent according to responses given to this question, but it is nevertheless interesting to note that scientists generally appear to spend more time doing data collection than concerning themselves

with data quality. In general most scientists seem to involve themselves in all aspects of the data collection and management process.

4. Priorities for training

Respondents were asked to identify the top three priority areas they would choose if offered training courses in statistics and data management. The results are summarised in Table 7a for scientists and Table 7b for technicians. The final column in the table shows the percentage of respondents who identified the training area as either their 1st, 2nd or 3rd priority. Responses from about 9% of respondents were excluded from the analysis where the question had been misunderstood and incorrectly answered.

The results are fairly clear. Over 70% of the scientists specified training in *standard methods of statistics using appropriate software* as one of their top three priorities. This is followed by *research data management, integrating qualitative and quantitative approaches* and then by *interpreting results and reporting*. There is also some interest, by about 23% of the respondents, for training in the use of GIS and impact assessments.

Amongst technicians, training in *research data management* was considered a priority, alongside *standard methods of statistics using appropriate software*. The third priority for technicians was for training on interpretation of results and reporting. It is interesting to note that many technicians (41%) also listed design of on-station and on-farm studies as a clear fourth priority but this rates low on scientists perception of training needs. This finding is rather disappointing since the importance of good study design cannot be over-emphasised. Maybe this is an example where most scientists have limited knowledge and don't know how much they don't know, whereas technicians know they don't know enough. Our experience with NARO researchers is that on-farm designs in particular are set up in ways similar to those for on-station trials whereas such studies need to closely link ideas and concepts underlying the design of both experiments and surveys. It is time to be more imaginative and use improved designs fitting more closely to study objectives, which can be analysed using modern statistical methods.

Table 7a. Top three priority areas for training for scientists

Areas for training	1st Priority (n=84)	2nd Priority (n=83)	3rd Priority (n=81)	% identifying area as one of top 3 priorities (n=94)
Standard methods of statistics using appropriate software	45.2	22.9	12.3	80.7
Research data management	22.6	14.5	11.1	48.2
Integrating qualitative and quantitative approaches	4.8	22.9	19.8	47.0
Interpreting results and reporting	8.3	19.3	14.8	42.2
Use of GIS	6.0	7.2	13.6	26.5
Impact assessments	3.6	7.2	16.0	26.5
Design of on-station & on-farm studies	6.0	1.2	11.1	18.1
Design of surveys	3.6	3.6		7.2
Writing protocols	0.0	1.2	-	1.1
Other*	0.0	0.0	1.2	1.2
Total	100.0%	100.0%	100.0%	-

* One person gave analysis of molecular variation as the 3rd priority.

Table 7b. Top three priority areas for training for technicians

Types of training	1st Priority (n=93)	2nd Priority (n=93)	3rd Priority (n=90)	% identifying area as one of top 3 priorities (n=105)
Standard methods of statistics using appropriate software	32.3	26.9	17.8	76.3
Research data management	34.4	26.9	16.7	77.4
Integrating qualitative and quantitative approaches	1.1	5.4	10.0	16.1
Interpreting results and reporting	5.4	26.9	27.8	59.1
Use of GIS	2.2	0.0	7.8	9.7
Impact assessments	0.0	0.0	1.1	1.1
Design of on-station & on-farm studies	21.5	10.8	14.4	46.2
Design of surveys	1.1	3.2	4.4	8.6
Writing protocols	0.0	0.0	0.0	0.0
Other*	2.2	0.0	0.0	2.2
Total	100.0%	100.0%	100.0%	-

* One technician gave repeated measures as the 1st priority, another gave bioinformatics as the 1st priority.

5. General issues in statistics and data management

The questionnaire also sought to determine areas of concern that NARO staff expressed in response to open questions regarding statistics and/or data management needs at their institute, and what recommendations staff would suggest to overcome such concerns. Answers were received by 82% of the respondents to each of these questions. The results are summarised in Tables 8a and 8b.

Table 8a. Concerns about statistics and data management needs at institute/centre

Concerns	Number giving responses	Percent of persons expressing concern (n=162)
Need skills in statistics & data mgt	107	66.0
Need relevant software	41	25.3
Lack of a statistician/data manager	30	18.5
Inadequate computers and access to them	19	11.7
Need technicians trained	20	12.3
Need discipline specific training	4	2.5
Poor data management	9	5.6
Lack of books on stats/data mgt	3	1.9
Problems with hardware maintenance	1	0.6
Other	6	3.7

Table 8b. Recommendations suggested to overcome concerns

Recommendations	Number giving responses	Percent of persons giving recommendation (n=163)
Give training in statistics & data mgt	116	71.2
Purchase relevant software	35	21.5
Employ statistician/data manager	32	19.6
Provide adequate computers	14	8.6
Give technicians training	13	8.0
Give discipline specific training	5	3.1
Improve data management	3	1.8
Acquire books/training guides on stats	4	2.5
Someone to handle hardware maintenance	2	1.2
Other	4	2.5

It may be observed that the need for skills in statistics and data management and the need for training in these areas feature highest, these being mentioned by more than 65% of respondents in various forms, e.g. some asked specifically for training in data analysis and interpretation, or study design, or software training or training in data collection, data entry and data management. The need to purchase relevant software

was recommended by about 20% of respondents, followed by the need for a statistician/data manager (about 19%). About 10% of respondents also mentioned lack of computers or lack of access to computer use as a concern as a recommended need for their institute. Other concerns and recommendations perceived by respondents included the need for discipline specific training, the need for better data management, lack of books and problems of hardware maintenance.

There were also many other interesting comments of concern or recommendations made by individual respondents. Some of these are listed below.

Scientists' comments

- Lots of research going on but data not put in a form that can be easily accessed by the user;
- Filing system should be modernised for easy accessing of the information;
- A biometrician should be as important as a socio-economist since both are cross-cutting;
- There is need to manage/assemble information for people of different interests to access at one point;
- Abrupt/unplanned on-farm data collection;
- Statistics and data management are key in research, and institutes should provide the training since not much is done at university.;
- Lack of funds to attend international conferences;
- Each scientist manages the data and its hard to reconcile/use the data when they are away.

Technicians' comments:

- Funds should be released timely for data collection;
- Few available pcs are old and used by senior officers;
- Availability of a statistician whenever we want him;
- For quality control and assurance, technicians should be involved in scientific data interpretation.

6. Constraints to the effectiveness of research

Our final question was aimed at putting the information from the questionnaire into perspective by asking respondents to identify the three priority constraints that they believe affect their own research. The results are summarised in Table 9 and very clearly demonstrate that software, training and support in statistics and data management is the primary constraint for more than two-thirds of the respondents. Further, this constraint is the 1st, 2nd or 3rd priority constraint for over 95% of the scientists and over 90% of the technicians.

Of course it is acknowledged that the context of the survey drew particular attention to statistical and computational issues, but the very strong response prioritising this item far exceeds any likely bias in the responses. There can be no doubt that NARO staff attach real importance to these issues.

The second priority constraint, closely related to the above, identified the need for more professional updating/upgrading courses in their own discipline. Lack of computer access and difficulties with sharing computers is the third identified constraint, followed closely (and more importantly for scientists) by the need for training in technical writing, time management and other general forms of training. These results emphasise very clearly that lack of skills and training in statistics and data management are perceived by both scientists and technicians as a serious stumbling block to the effectiveness of their research.

There were several other constraints identified by both scientists and technicians. These were the lack of funds to attend international conferences (19%), inadequate data or lack of funds to collect more data (9%), the need for more reference books, library facilities and internet access (9%), the holding of too many meeting or conflicting demands by senior staff (7%) and the need for more administrative and organisational support (6%). Two respondents also mentioned lack of funds to do research.

Table 9. Constraints to research effectiveness as perceived by scientists and technicians

	Scientists				Technicians				All respondents
Type of constraint	**1st constr. (n=85)**	**2nd constr. (n=85)**	**3rd constr. (n=77)**	**One of top 3 constraints (n=85)#**	**1st constr. (n=95)**	**2nd constr. (n=92)**	**3rd constr. (n=85)**	**One of top 3 constraints (n=95) #**	**One of top 3 constraints (n=180)**
Need for software, training and support in stats & data management	70.6	16.5	10.4	96.5	69.5	20.7	4.7	92.6	95.0
Need for more professional updating in own discipline	17.6	44.7	16.9	77.6	18.9	40.2	24.7	80.0	78.9
Lack of computer access or difficulties with sharing computers	3.5	5.9	11.7	20.0	9.5	22.8	14.1	44.2	32.8
Need for training in technical writing, time management & others	-	10.6	15.6	24.7	1.1	7.6	29.4	34.7	30.0
Lack of funds to attend international conferences	-	4.7	16.9	20.0	-	4.3	15.3	18.5	18.9
Inadequate data and lack of funds to collect more data	2.4	5.9	6.5	14.1	1.1	1.1	2.4	4.2	8.9
Need for more reference books/library/internet access	-	5.9	7.8	12.9	-	2.2	3.5	5.4	8.9
Too many meeting or conflicting demands by seniors	1.2	2.4	10.4	12.9	-	-	2.4	2.4	7.2
Need for more admin. or organisational support	1.2	3.5	2.6	7.1	-	1.1	3.5	4.3	5.6
Other*	3.5	-	1.3	4.7	-	-	-	-	-

- Lack of funds to do research (2 responses); No statistician at institute; subscription to journals
- The 4th column of numerical figures for scientists and technicians is not exactly equal to the sum of the previous three columns since some respondents did not clearly indicate their 2nd and/or 3rd constraints, leading to a reduction in sample sizes.

7. Conclusions

This report has highlighted scientists' and technicians' views on statistics and data management issues concerning their research. It gives results that clearly demonstrate respondents' having some awareness of their limitations with respect to statistics and/or data management. Of major note is the identification of lack of skills and the need for training in statistics and data management as a primary constraint limiting the effectiveness of NARO research. Emphasis was given to this constraint ahead of the need for professional updating/upgrading in their own subject discipline.

The survey results also highlighted the lack of sufficient practical content, e.g. through the use of appropriate statistical software, in the University courses staff had attended, sometimes a long time back. This limits researchers' ability to use statistical software confidently in their research. It was therefore not surprising to find scientists identifying training in standard methods of statistics using appropriate software as the area in which they would most like to receive training. Technicians were also interested in this type of training, but expressed an even greater preference for training in research data management. It would therefore seem appropriate to begin any future training with general training in standard statistical methodology, primarily for the scientists, and a separate course in research data management, primarily for the technicians. This would provide a complementarity of skills amongst all NARO researchers, with subsequent courses concentrating on other training areas of interest as expressed by scientists and technicians.

The recent purchase of Genstat software will largely overcome concerns expressed by experimental scientists and technicians about the lack of statistical software at their Institutes. However, there needs to be a commitment by researchers to acquire skills in the use of this package, as soon as possible, by setting aside a day or two of their time for this purpose. Tutorial materials for learning Genstat are available and these can be provided to NARO research active staff as self-teaching materials.

NARO researchers also expressed concern about the lack of access to computers. This needs to be a priority issue to be addressed by NARO as soon as possible since it is crucial to the production of high quality research. Along with this is the need to ensure there is good virus-checking software and a system in place for assistance with virus-related and other problems.

The survey aimed to seek the views of **all** research active staff, including externally funded project staff. However, responses were received from only about half of them. It is possible that those who did respond were more interested in issues of statistics and data management than those who didn't. In this case, the sample of respondents' comments cannot be regarded as being a perfect representation of all NARO researchers. The findings reported here may therefore appear to have a slight bias towards those who had greater interest in statistics and data management. We assume however that the non-respondents are "missing at random", i.e. that they would have given much the same answers, had they replied, as those who did respond.

It must also be recognised that there are areas of fundamental research such as molecular genetics and some accurate laboratory-based procedures, which do not necessarily generate complex data structures, and we would not expect much interest in statistics from those who specialise in such work. However extensive data generating activities are common and, with growing trends towards interdisciplinary research, providing researchers the necessary skills to handle the large volumes, and complex structures, of data they generate in an effective and efficient way becomes of paramount importance.

ANNEX 1: SURVEY QUESTIONNAIRE

Assessing Statistical and Data Management training needs in NARO Research Projects

This short questionnaire aims to give you the opportunity to express your views on statistical and data management issues concerning your research and to state areas where training would most help you to achieve good quality research outputs. Responses will be analysed by the Statistical Services Centre at the University of Reading, U.K., and will feed into recommendations to NARO as to researchers' and technicians' training needs.

We would be very grateful if you would kindly complete this form and return it to the Director of your institute for onward transmission to MEPU, then to the consultants in Reading. Please feel free to attach additional sheets to this questionnaire if you need more space. Thank you in anticipation for completing this questionnaire.

Dr. Savitri Abeyasekera, Principal Statistician
Statistical Services Centre, The University of Reading, U.K.
8th April 2005

0. Your personal details

Name	Designation (Technician, Res. Asst, RO, etc)	Institute (KARI, NAARI, Mukono ARDC, etc)	Are you NARO staff or project funded	Area of work (entomology, fish biology, forest products, etc)

1. For a recently completed or current project, please provide:-

Title: _____ Start/End dates: _____

Brief description of your role in the project:- .....

.....

.....

Indicate whether you expect any of the following to be an activity or a part of an activity in the project referred to above. Tick **all** those applicable.

(a) Questionnaire Survey ☐ (b) On-farm trial ☐

(c) Participatory work involving farmers ☐ (d) On-station field trial ☐

(e) Livelihood analysis ☐ (f) Laboratory study ☐

(g) Other type of study (please specify) ☐ (h) Other type of study (please specify ☐

....

If a statistician was available at your institute, would you expect to involve the statistician in any of the areas ticked above? **Yes / No** (circle answer)

If you answered "Yes" to the above question, please list (by using the appropriate letters (a), (b), ..., etc. from above) the areas where you would expect to involve the statistician.

.....

2. Your current level of knowledge in statistics and practical application areas

(a) List titles of all courses in statistics and/or data management you have followed at university or through a short-term training programme.

Title of course	When done (year)	Duration,e.g. one university term, 3 weeks, etc.	Software used in the course (if any)

(b) Please rate, on a 0-3 scale (0=none; 1=little; 2=some; 3=lots) your **current** level of knowledge of statistical techniques below in terms of being able to understand and use the method in your present day work without help from anyone else. (This will help us assess the starting point for any statistical training).

Tables of counts and percentages	☐	Chi-square tests	☐	Multiple regression	☐
Tables of means and other summaries	☐	t-tests	☐	Logistic regression	☐
Excel graphs	☐	Non-parametric tests	☐	Multivariate/cluster analysis	☐
Graphs using software other than Excel	☐	1-way ANOVA	☐	Sampling methods	☐
Standard errors	☐	2-way ANOVA	☐	Other (please specify)	☐
Simple linear regression	☐	Factorial analysis	☐	.....	

(c) Please rate, on a 0-3 scale (0=none; 1=little; 2=some; 3=lots) your **current** level of **practical knowledge** in the following areas (This will help us assess the areas where a statistician's input may be needed).

Design of on-farm studies	☐	Survey Methodology	☐	Project monitoring & evaluation	☐
Farming systems research	☐	Sample size determinations	☐	Natural resource monitoring	☐
Farmer participatory approaches	☐	Impact assessment studies	☐		
Participatory rural appraisal (PRA/PLA)	☐	Livelihoods analysis	☐		

(d) Please indicate your level of competence in the use of the software packages listed below using a 0-5 scale (0 = None; 1=little competence; 2=some; 3=lots).

SPSS ☐	SAS ☐	MSTAT ☐	Genstat ☐
STATA ☐	Minitab ☐	Excel ☐	Access ☐

(e) If there were training courses in statistics/data management, identify **3** areas of priority (1=1st priority, 2=2nd priority, 3=3rd priority) that you would like to see covered in the training (you may write in your own topic of interest if you wish). Note that it may not be possible to cover all 3 areas in one training course.

1. Design of on-station and on-farm studies ☐	4. Standard methods of statistical analysis using appropriate sofware ☐	7. Writing protocols ☐
2. Design of surveys ☐	5. Integrating qualitative and quantitative approaches ☐	8. Use of GIS ☐
3. Research data management ☐	6. Interpretation of results and reporting ☐	9. Impact assessments ☐
10. Other (specify) ☐	11. Other (specify) ☐	12. Other (specify) ☐
.....	.....	.....

3. Practices and training in data management

(a) Have you ever been involved in planning a data recording sheet? Yes / No (circle answer)

Comment further if you wish.....

(b) Have you ever had training in how to organise data in a spreadsheet to avoid errors during data entry?

Yes / No (circle answer)

Comment further if you wish.....

.....

(c) Specify procedures you currently use, or have recently used, to check data quality, i.e. to ensure data recordings are correct.

☐ Does not apply to the work I do	☐ Not sure since this is someone else's responsibility
☐ Scrutinising the data collection sheets	☐ Manual checking of paper records with computer version
☐ Using Excel's facilities for data validation	☐ Capturing errors using exploratory summaries and graphs
☐ Random checks of the computerised data	☐ Capturing errors during data analysis.

Comment further if you wish.....

.....

.....

(d) In the last year, approximately how many days have you spent on each of the tasks listed below. Add comments alongside if you wish.

Collecting the data.....

Checking the data for errors.....

Doing the data entry.....

Organising and managing the data .....

Doing simple data summaries and graphs.....

4. General issues in statistics and data management

Please comment on any areas of concern you have regarding statistics and/or data management needs within your institute.

What recommendations do you have to overcome concerns you have expressed above?

Please attach additional sheets to this questionnaire if you wish to give further comments that might help us in assessing training needs in statistics and data management for yourself or others in your institute.

5. Perspective on other Constraints

The present exercise, of which this survey is part, aims to recommend systematic improvement in the availability of statistical software, training and support in statistics and data management at NARO Institutes. There may well be other constraints to your effectiveness in the research system which are more important for you. We would like you to put the information from this questionnaire *into perspective* by identifying your own 3 priority constraints below using codes 1=main constraint; 2=2nd constraint; 3=3rd contraint. You may include additional constraints that you believe affects the research.

Software, training and support in statistics & data management ☐	Need for more professional updating/upgrading courses in your own discipline ☐	Need for technical writing, time management and other general training ☐
Lack of computer access or difficulties with sharing computers ☐	Need for more reference books/library/internet access ☐	Too many meetings or conflicting demands on your time by your seniors ☐
Inadequate data and lack of funds to collect more data ☐	Need for more admin. or organisational support ☐	Lack of funds to attend international conferences ☐
Other (please specify): ☐	Other (please specify): ☐	Other (please specify): ☐
----------	----------	----------

THANK YOU FOR COMPLETING THE QUESTIONNAIRE. PLEASE PASS IT NOW TO YOUR DIRECTOR.

Annex 2: Survey Returns Summary

Assessing Statistical and Data Management Training Needs in NARO

		Cadre status*			Forms returned			Scientists	Technicians		Returned Director
Institutes	ID numbers	Scientists	Technician	*Total*	Scientists	Technician	*Total*	Percent returned	Percent returned	Overall % returned	Manager Forms
KARI	001-037,119,120	45	49	*94*	16	23	*39*	36	47	41	yes
NAARI	038-077, 121	24	52	*76*	19	22	*41*	79	42	54	yes
LIRI		14	12	*26*	0	0	*0*	0	0	0	
FIRRI	122-139#	25	25	*50*	9	8	*17*	36	32	34	yes
FORRI		12	7	*19*	0	0	*0*	0	0	0	
FOSRI	078-081	5	1	*6*	3	1	*4*	60	100	67	yes
AEATRI		7	4	*11*	0	0	*0*	0	0	0	
SAARI	140-175	22	22	*44*	16	20	*36*	73	91	82	yes
CORI	176-188	7	10	*17*	7	6	*13*	100	60	76	yes

		Cadre status			Forms returned			Scientists	Technicians		Returned
ARDCs	ID numbers	Scientists	Technician	*Total*	Scientists	Technician	*Total*	Percent returned	Percent returned	Overall % returned	Manager Forms
Abii	082-089	6	4	*10*	4	4	*8*	67	100	80	yes
Mukono	189-200	13	6	*19*	6	6	*12*	46	100	63	yes
Mbarara	090-092	4	3	*7*	3	0	*3*	75	0	43	yes
Kachwekano	093-101	7	3	*10*	6	3	*9*	86	100	90	yes
Bulindi	102-111	4	7	*11*	3	7	*10*	75	100	91	yes
Kitgum	112	0	1	*1*	0	1	*1*		100	100	yes
Ngetta	113-118	4	4	*8*	2	4	*6*	50	100	75	yes
Serere		1	0	*1*	0	0	*0*	0		0	
Bulegeni		0	1	*1*	0	0	*0*		0	0	
	Totals	**200**	**211**	***411***	**94**	**105**	***199***	**47%**	**50%**	**48%**	
	Totals for institutes that returned forms	*166*	*187*	*353*	*94*	*105*	*199*	***57%***	***56%***	***56%***	

* Cader status includes project funded staff but not those currently away from the institute, e.g. on study leave.
\# ID123 was an additional form filled by the same scientist completing ID124