IMPROVING SCIENTIFIC RESEARCH AT KENYA AGRICULTURAL RESEARCH INSTITUTE (KARI) – KAKAMEGA RESEARCH CENTRE: A RESEARCH METHODS APPROACH

BY

HILDAH NYENDE WAMBANI[1]

`A DISSERTATION SUBMITTED IN (PARTIAL)" FULFILMENT FOR THE DEGREES OF MASTER OF SCIENCE IN RESEARCH METHODS IN THE JOMO KENYATTA UNIVERSITY OF AGRICULTURE AND TECHNOLOGY.

June, 2011

[1] **E-mail: hildahwambani@yahoo.com Mobile:+254734775660/715567621**
Jomo Kenyatta University of Agriculture and Technology (JKUAT), P. O. Box 6200-00100, Nairobi, Kenya
OR KARI Kitale, P.O. Box 450-30200, Kitale, Kenya

DECLARATION

This dissertation is entirely my original work and has not been presented for a degree in any other university.

Signature_____ **Date**_____

HILDAH NYENDE WAMBANI

MSc Research Methods student

This dissertation has been submitted for examination with my (our) approval as University Supervisor(s).

Signature _____ Date: _____

PROF. KAMAU NGAMAU

(Dean – University supervisor)

Signature _____ Date: _____

DR. JOHN KIHORO

(*E-learning* Coordinator - University supervisor)

Signature _____ Date: _____

MR. MAURICE MUDEHERI (Biometrician - KARI Kakamega supervisor)

TABLE OF CONTENTS

ABSTRACT

MSc in Research Methods is a unique problem solving program that increases the capacity to plan, implement and generate quality research. Lack of current and necessary statistical skills in data analysis, interpretation and reporting and existence of weak review and data management frameworks contribute significantly to poor quality research. The main objective of the study was to improve the research undertaken at the institution through use of a research methods approach. A survey was carried out to identify the statistical needs of the scientists at the beginning of the attachment period. Interaction with scientists and technical staff, participation in research, attending planning meetings, holding trainings and provision of research approaches support were undertaken to improve the skills of scientists and technical officers involved in research. Capacity building in various data analysis software, quality assurance in data, research designs and presentation was also conducted. The results showed that seminars equipped the scientists and technical staff with the current skills required in the rapidly changing research environment. A large loss of data was due to inadequate data management procedures performed. It was observed that only a handful of statistical analysis for instance ANOVA, statistical packages and research designs were often used by scientists. A data management protocol was developed.

Key words: Research methods, data management, data analysis, research processes, consultancy

DEDICATION

My deepest devotion to my family especially Ecky Wambani (son), Ellah Mukoya (daughter), Mrs. Jenipher Asikoh Wambani (mother) for the understanding and love they bestowed, my late father Paul Wambani Makuba for the inspiration he infused in me and the entire research community who I hope will benefit from the research findings.

ACKNOWLEDGEMENT

My sincere gratitude to the Almighty God for providing good health and sound mind during the whole research period, my supervisors, Prof Kamau Ngamau, Dr. John Kihoro and Mr. Maurice Mudeheri for the technical and professional support they offered during the development of the dissertation, fellow students and friends for the encouragement and advice they generously availed to me.

I am entirely indebted to the Director KARI for providing a two year study leave and enabling environment to successfully undertake the MSc degree, the entire KARI Kakamega research community for providing an opportunity to work and interact with them, SSC- University of Reading, JKUAT and all the facilitators of the course for the constant guidance and literature they freely provided.

Finally, I appreciate RUFORUM for the financial support in form of a scholarship.

ABBREVIATIONS AND ACRONYMS

AEZ	Agro Ecological Zones
AMF	Arbuscular Mycorrhizal Fungi
ANOVA	Analysis Of Variance
APVC	Agricultural Productivity Value Chain
CBOs	Community Based Organizations
CIAC	Centre Internal Advisory Committee
CIMMYT	International Maize and Wheat Improvement
CRAC	Centre Research Advisory Committee
CTA	Technical Centre for Agricultural and Rural Cooperation
DAP	Diammonium Phosphate
DBMS	Database Management Systems
FARA	Forum for Agricultural Research in Africa
FAO	Food and Agriculture Organization
FSA-RET	Farming Systems Approaches to Research Extension and Training
GIS	Geographic Information System
IFS	International Foundation of Science
IMF	International Monitoring Fund
IMRaD	Introduction Methodology Results and Discussion
IPM	Integrated Pest Management

JKUAT	Jomo Kenyatta University of Agriculture and Technology
KARI	Kenya Agricultural Research Institute
KHCG	Kenya Horticulture Competitiveness Project
KLA	Kenya Library Association
LSD	Least Significance difference
MDG's	Millennium Development Goals
MMUST	Masinde Muliro University of Science and Technology
NARS	National Agricultural Research Systems
n.d	no date
NGOs	Non Governmental Organizations
OFSP	Orange Fleshed Sweet Potato
RCBD	Randomized Complete Block Design
RM	Research methods
RUFORUM	Regional Universities Forum for Capacity Building in Agriculture
SAS	Statistical Analysis Software
SIMLESA	Sustainable Intensification of Maize-Legume Cropping Systems for Food Security in Eastern and Southern Africa
SMART	Specific, Measurable, Achievable, Realistic and Time bound
SPSS	Statistical Packages for Social Science
SSA	Sub-Saharan Africa
SSC	Statistics Services Center, University of Reading
USAID	United States Urgency for International Development

LIST OF FIGURES

LIST OF TABLES

CHAPTER ONE

1 INTRODUCTION

1.1 Background

Research methods program

National Agricultural Research Systems (NARS) involved in agriculture and rural development in Africa constituting universities and national agricultural organizations, form the heart of research effort but lack the capacity to conceptualize, plan and implement effective research (Carter *et al.*, 1986). This is due to lack of appropriate research methods among many players. The limited capacity and inadequate support to the changing priorities of agricultural research has unfortunately weakened the quality of research and progress towards increasing food and nutritional security and alleviating poverty (CTA, 1997). This was well highlighted in the research proposal received by RUFORUM, International Foundation for Science (IFS) and recent publications on the importance of science and technology for development of African regions (Eicher, 2006; Johanson and Saint, 2007). As a result a Master in research methods program was initiated to promote more collaborative and high quality research which is necessary and sometimes mandatory for integrated solutions and approaches to development goals.

The program is unique in the sense that, it is problem-solving based (Usher *et al*., 1997; Savin-Baden 2000), it fills gaps between theoretical rigor and practical needs, it is regionally owned, has strong inter-linkage among fields of study and

innovative approaches. The program provides capacity building in East and Southern Africa through pooled efforts from various university members in Sub-Saharan Africa (SSA). This effort enhances the improvement of high quality research through development of graduate training programs that respond to the dynamic market needs and also strengthens innovative capacity (RUFORUM, 2007; Coe and Sonya, 2009). Research Methods program took 2 years, the first year involved rigorous course work while the second year was an attachment period for the professional to put into practice and sharpen skills acquired in the first year. The attachment activities were linked to on-going research. After training, the specialist in Research Methods was expected to offer support and guidance to scientists and technical staff. Currently, the program is piloted at Jomo Kenyatta University of Agriculture and Technology (JKUAT) and the University of Malawi (Bunda College) and would then be initiated in other universities. The Research methods professional was attached at Kenya Agricultural Research Institute (KARI), Kakamega.

Organization of Kenya Agricultural Research Institute (KARI)

KARI was established through the Science and Technology Act (Cap 250) in 1979 with core mandate to conduct, co-ordinate and co-operate with various stakeholders in the promotion of agricultural research and development activities countrywide (KARI 2005a, 2009a). The institute main focus is to strengthen Kenya's food security and income. KARI has 23 main research centers and 14 sub-centers (see Appendix A) located throughout the country (KARI, 2009b).

These provide an institutional framework that effectively manages and strengthens the government's agricultural research system, re-organize and consolidate agricultural research within the country. The organization has a total of 558 scientists, 855 technical staff and 2293 other support staff (KARI, 2005c).

KARI Kakamega was started in 1953 as a seed farm and later became one of the main KARI centres. The centre is located within Kakamega municipality, 1.5 km south-east of the town centre. The centre conducts adaptive, strategic and applied research using the Farming Systems Approach to Research, Extension and Training (FSA-RET). The centre generates and promotes knowledge, information and technologies that respond to client demands under different socio-economic circumstances. This is basically achieved through collaboration with other stakeholders involved in agricultural research for development. For example, Ministries of Agriculture, Livestock and Fisheries Development, NGOs, CBOs, farmer groups, provincial administration, regional and international research organizations and donor agencies.

The quality and relevance of research is ensured through mechanisms like Centre Research Advisory Committee (CRAC) and Pre-CRAC that are broad-based consultative research proposal review fora. The research undertaken at the centre falls in various categories; maize research (breeding, protection and agronomy), horticulture, oil crops, livestock, natural resources management (soil and water technologies), grain legumes, roots and tubers, finger millet and sorghum, post harvest value addition and nutrition, technology transfer, laboratory services,

biotechnology, socio-economic and biometrics. The latter offers research methods support to all the scientists and technical staff involved in the other research categories.

The biometrics department further ensures the quality of research through offering consultancy services; providing specific assistance in experimental design, data management, analysis and interpretation and editing of proposals, reports and publications. However, in most occasions the biometrician is consulted mainly during data management and analysis and not from the beginning and throughout the research project as expected. KARI Kakamega currently has a total of 36 scientists and 122 support staff (Muyekho, 2010).

The research tasks

During the one year attachment period, the Research methods professional undertook six tasks in the research process and developed a dissertation. The tasks included participation in research, attending team meetings, holding trainings through seminars or short courses, review of selected reports, proposal and papers and data management, analysis and reporting.

1.2 Justification

KARI has a mandate to carry out both national and regional research, which necessitate current application of research approaches by both the research scientists, and the technical staff who mostly handle trials established and data collection. The six research activities were presumed important in improving the quality of research at the centre. The research methods required to work in

emerging environment are new and changing rapidly hence various trainings are required in order to keep up with the pace (RUFORUM, 2007). The one year course work equipped the Research methods professional with skills to train scientists on how to handle new emerging issues in research. The ratio of the statistician to scientists is 1:36. This is overwhelming for the statistician and therefore need for training scientists in proper data management and analysis. The data management process usually takes a longer period than expected and no formal guidelines exist hence need to develop a protocol that guides both scientists and technical staff on the various procedures that ensure quality in data management. During the attachment the Research methods professional gave support to scientists and participated in various research projects.

1.3 Problem statement

Currently, the research method support given to scientists is inadequate in various organizations dealing with agricultural research for development and this has led to diminishing quality of research for agriculture in SSA (The World Bank, 2002; Watson and Crawford, 2003). KARI Kakamega is involved in research of different aspects at different levels by scientists who are supported by technical staff who lack current and adequate knowledge of research approaches just like most research institutions in developing world (Buysse, 2005).

A scientist should be supported from the beginning of the research project on which appropriate instrument to be used in data collection, how to obtain a sample size, which experimental design to be used, which kind of data to be collected,

how to develop SMART objectives and the appropriate software to be used for analysis. Very few scientists consult the statistical professional during the research project process and the technical staffs involved in data collection have inadequate basic knowledge on simple data analysis, data archiving and importance of data being collected. The main problems in data management arise due to the prolonged period between data collection and analysis and lack of knowledge on the activities undertaken in data management process to ensure quality. Moreover, there is no formal existing data management protocol to guide individuals during data handling (SSC, 2000).

This lack of research support means that, not enough reports and papers are produced since most data is not put into use as it should (Buysse, 2005). In most cases proposals are usually developed without involving all stakeholders, data is analyzed later after the research has long been completed and papers are developed in a rush to attend a conference resulting in low quality publications (Bloom *et al*., 2005; FARA, 2006).

Many proposals are developed and implemented per year and the review of the documents is done when a call for proposals is announced. The priority setting document that is developed after every five years is never reviewed at the end of the project to see if the yearly set priorities were achieved at the specified time. Support services from a professional are required throughout the year at the centre since ongoing research is at different stages.

Documentation of the attempt to improve research and build capacity provided insights that would enhance future provision of research methods support and training.

1.4 Objectives

Broad objective

To improve the agricultural research for development in KARI Kakamega through provision of required research methods and statistical support to research scientists and technical staff

Specific objectives

- To evaluate the current situation on research methods in KARI Kakamega
- To train scientists and technical officers on data management and analysis issues
- To develop a protocol on data handling and management
- To review six different documents at KARI Kakamega
- To document participation in research process at KARI Kakamega

1.5 Research questions

- What is the current situation on research methods at KARI Kakamega?
- Were scientists and technical officers trained in data management and analysis issues?
- Was a data management protocol developed?
- Were six different documents reviewed?
- Was the intended documentation achieved?

CHAPTER TWO

2 LITERATURE REVIEW

2.1 Research and Development

Agriculture is the backbone of overall growth for the majority of the countries in SSA region and is essential for poverty reduction and food security (FAO, 2003; World Bank 2003). While the agricultural sector is strategic to the long-term growth and development of most countries in SSA, the African governments' total budgetary expenditure on agricultural research and development has declined from an average of 8.4%, to 6.5% from 1984-1994 (IMF, 1996; Beintema and Stads, 2004), therefore hindering growth in most African countries. Failure to give sufficient priority to agriculture, research and development becomes one of the most serious errors in developmental strategy committed by a large number of African nations (Global Coalition for Africa, 1999) and hence limiting the attainment of the Millennium Development Goals (MDGs). There is a direct relationship between the country's level of technological and economical development on one hand and the amount of investment put in research activities of the country's institutions on the other hand (Mfumbusa and Mkamwa, 2005).

2.2 Research methods

It is a systematic process that attempts to solve the research problem logically adopting various steps and deals specifically with the manner in which data is collected, analyzed and interpreted (Pattron, 2008). It is important in understanding and evaluating the research of other researchers, plan and conduct

original research with minimum assistance (Gatali and Barihuta, n.d.) and imparts skills in enquiry, experimental design, data collection, measurement and analysis, interpretation and presentation (Coe and Sonya, 2009).

Earlier research by Coe (2003) showed that university research methods curricula have a responsibility to provide students with skills in problem identification, understanding of research methods principles, experience in how these have been applied in diverse settings, imagination to apply them to new and complex problems and confidence to practice the art of research design. This ensures effective research that makes significant contributions to science and development priorities for particular regions (Patel *et al.*, 2004).

Effective research for development is limited by the ability of researchers to solve problems with realistic and valid approaches (RUFORUM, 2007: Kirway and Lema, 2003). However, the capacity of the researchers and lecturers in the SSA region can be improved through interventions which build on principles of self learning and discovery, wide consultation, critical and creative thinking and access to available teaching materials (Glasgow, 1997).

African scholars have observed and concluded that institutions in Africa need to develop new ways of teaching with tools that can be adapted to the training needs of African academicians so that they can re-orient their students to: think creatively, value team work and partnerships, recognize that they belong to a global world, and that whatever they do, should contribute to the improvement of the welfare of African's people. Unfortunately, African scholars and field workers

often lack reference materials that are based on local experiences (Gatali and Barihuta, n.d.)

2.3 Research methods activities

A seminar is a presentation delivered to an audience on a particular topic or sets of topics that are enlightening in nature. Training through seminars is important for capacity building in research methods and meets the statistical needs of the research project team members. It shapes the thinking of the employees which leads to quality performance in an institution (Knowles *et al.,* 1998). A trained employee is more efficient and productive. Trainings can be in form of gaining new skills or enhancing already existing knowledge. The various benefits of training employees are among others; increased productivity of employees, improved quality of work produced, increased knowledge and skills of employee, and optimized utilization of human resources and eventually increased attainment of both the individual and institutional goals.

Data management are steps or processes of handling data and they include designing data collection instruments, keeping the data sheets safe, entering data into computer files, transforming data, checking data for accuracy to guarantee quality, keeping track of incoming data, developing and documenting a database structure that integrates the various measures, maintaining records of the processing steps and archiving it for future use (Chege and Muray, 2004). Data management is important in the success of a research project since the process ensures observations made are valid, can be processed efficiently and will remain

available when need be (Chege and Muray, 2004; Muray *et al.*, 2002). Data must be managed effectively throughout the research process in order to maintain its quality.

There are various stages involved in data management (Fig 1) and each stage entail different tasks aimed at preserving data quality (Dasu and Johnson, n.d.).

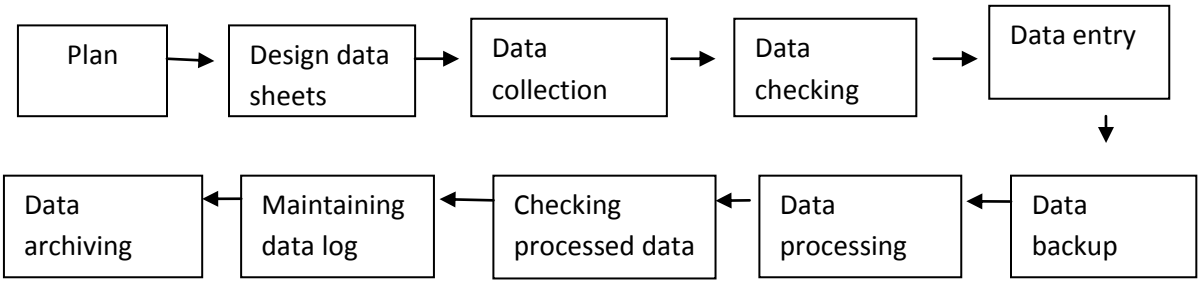


Fig 1: Data management processes

Source: Chege and Muray 2004 pg 194

Data analysis is a process that transforms raw data into results of the study presented as summaries that can be accessible to others through various forms for example reports, seminars and conferences (Susan, 2004). Key features of the data in the study are condensed using simple summaries in form of tables, graphs or charts. The process involves data preparation, performance of descriptive statistics and inferential statistics. During data analysis, questions, models and hypotheses are tested and inferences drawn from the sample data to more general population conditions. Usually, the researcher links each of the inferential analysis to specific research questions or hypotheses (Susan, 2004).

The main goal of data analysis is highlighting useful information and suggesting conclusions that support decision making (RUFORUM, 2007).

It involves in-depth discussion of the specific statistics to be used for example when using linear regression, one would find out the assumptions of the linear and constant variance and how to interpret the statistics. The sample size is also looked at critically to determine the number of participants required in a linear regression. It is important to have a statistical analysis plan that will either reject or not reject the null hypothesis formulated for the study.

Consultancy is the provision of expert advice on statistical matters like designing of a study, calculating sample size, managing data, analyzing, interpreting and presenting data, using statistical packages and publishing results from the knowledge, technical and non-technical skills (Derr, 2000), techniques and equipment of the consultant in meeting a specific need of the client (William and Woodward, 1994; Bangdiwala *et al*., 2002; Ader *et al.*, 2008).

Statistical consultancy enhances the ability to develop, select and apply appropriate methodology for a given problem in the real world (Jeske *et al*., 2007), it helps in choosing a correct analysis on the basis of informed judgment, it initiates the exploration of additional statistical methodology that is related to the clients problem and also cultivates a curiosity and a self sufficiency which are important during consultation (Russell, 2001). This improves the efficiency, effectiveness and relevance of a research project.

Consultation is usually recommended at the beginning of a study to help verify if the planned procedures and study size are adequate to address its objectives. It is further required throughout the project. At the initial stage the consultant may

suggest ways of maximizing available resources through the use of efficient study designs and data management techniques which can easily be incorporated in the research project unlike later in the project. The common observed practice is seeking expert advice at advanced stage of the project which may not be very beneficial to the entire project. A good consultation includes a write up of the problem under study, designing a research plan, statistical issues, summary of responsibility allocation, decision and recommendations (Belle, 2008).

Reviewing is an objective process of subjecting an author's scholarly work, research or ideas to the scrutiny of others who are experts in the same field (Ware, 2008). The process usually looks at relevance of the publication in line with the existing problem, if the conclusions and implications drawn by the author are rigorous and warranted and how this implication fit with other research done elsewhere. Reviewing aims at improving the quality of a publication. It enables a knowledgeable individual who has no direct participation in the research to make sound comments before the article is exposed to the highly critical review of the worldwide audience. During the review process, useful feedback that is open, complete and constructive is given to the author.

Team planning involves a range of activities designed to improve team performance, it focuses on bringing out the best in a team in terms of self development, positive communication, leadership skills and ability to work closely as a team in problem solving (Driskell *et al.*, 2006). Multi-disciplinary approach to research has ensured the environment for team working more prevalent when

carrying out research activities. The role of a research specialist starts right from the initial planning meetings and continues throughout the entire duration of the life of the research project for effectiveness. Team planning is designed to develop group members and their ability to work together effectively. Team planning improves team productivity, increases effective collaboration, helps in tracking individuals and team performances, increases members anticipation and ability to prepare for change. Team planning assists in identifying important issues and hence focuses on them and also aid in setting roles clearly for each individual ensuring team members understand clearly what is required of them. The process mainly facilitates in tracking and controlling costs. A good team work creates synergy (Humphrey *et al.*, 2009).

2.4 Skills gained

The rigorous one year course work prepared the Research methods professional with different effective research approaches for solving applied research problems. The Research methods professional acquired the basic principles of research design of survey/experiments and implementation of design applications. The course work assisted the professional in understanding the principles of agriculture and having the ability to discuss research methods in the context of the agricultural problems. The professional is familiar with issues relevant to measuring, monitoring and evaluation of development in agriculture and can describe any data set that is well-organized and of acceptable quality in ways that correspond to the

stated objectives of the study. The professional is able to perform statistical modeling in data analysis.

Applied mathematics for agriculture, has facilitated the professional to apply mathematical principles to real world problems, calculate cost-benefit analysis and formulate graphical equations. The professional has an understanding and working knowledge of spatial methods used in Geographic Information System (GIS), Genstat, R and Instat computer packages.

Various skills learnt from the course were communication and writing skills, searching for information, critical thinking, holding successful seminars and discussions, computing of some commands on computer software, leadership and organization capabilities, formatting a report professionally, technical and interpersonal skills.

CHAPTER THREE

3 RESEARCH METHODOLOGY

3.1 Theoretical and conceptual framework

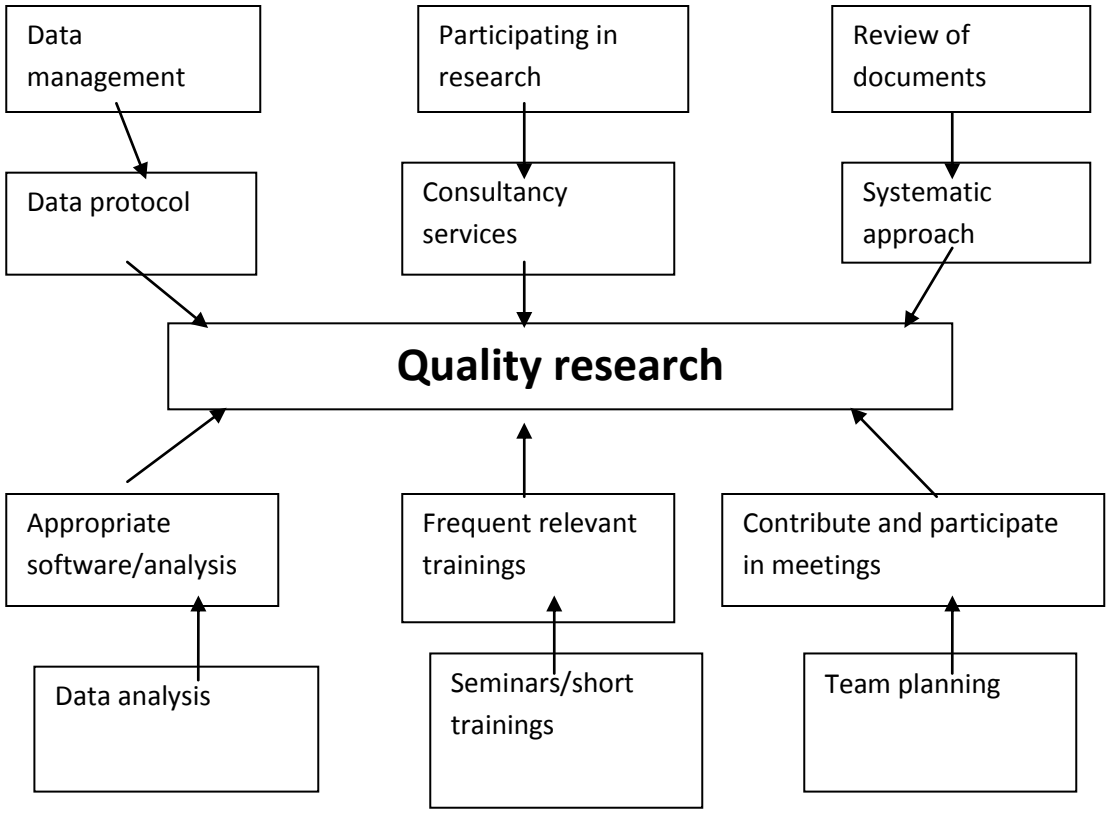


Fig 2: Conceptual framework for the research support process at KARI Kakamega

3.2 Research processes

3.2.1 Seminar

A statistical need assessment for scientists was done in November at the beginning of the attachment period in order to ascertain the actual needs scientists require to improve research. The semi-structured questionnaire (see Appendix B) was developed and pretested early November 2010 to assess the effectiveness of the tool. The target group was all scientists and the purposive sampling technique was used. During the administration of the tool, scientists were encouraged to seek further clarifications where need be. Part of the questionnaire content was marked as an exam.

Through consultation services, it emerged that data management process was poorly handled in various sections at the centre hence a seminar on quality assurance in data management was conducted. The target audiences were data entry personnel and scientists at the centre. Invitation letters were sent to them. After conducting the seminar a feedback from the trainees was collected by use of a seminar evaluation form (see Appendix C). The other seminars held were an initiative from individual scientists.

Information obtained from the consultation and semi-structured questionnaire was used to develop training manuals. Documentation was done on assessment needs leading to the training, materials used, practical and arrangements made in form of a CD.

3.2.2 Data management

The Research methods professional designed instruments for data collection. She also participated in data collection and organization and documented the role played in the process. Data files were prepared ready for analysis. A description of the data management problem and approach used to solve it was documented. A data and information management protocol was developed for use in future.

3.2.3 Data analysis

The Research methods professional was involved in analysis of various datasets from different sources. Prior to data analysis the following activities were performed; data was prepared, cleaned and organized ready for analysis, descriptive data analysis was executed and presented in the form of tables of means, pie charts, graphs and frequencies and inferential statistics took the form of answering questions, testing hypotheses and models. The process of contribution to the analyses and result generation was documented.

3.2.4 Participating in research

This process was achieved through forming working groups with scientists in helping with research planning and design. The Research methods professional made three on farm field visits and participated in research design layout. In-depth discussion series were incorporated in the consultation services provided to the scientists, students and technical staff. Contribution to review fora and research processes, challenges scientists presented, the way the challenges were handled and scientists' reception of the results were documented. Wide consultation from

books, internet resources, fellow students, supervisors and advisors was done in order to find answers for the questions and challenges posed. The current research methods were understood and approaches of improving the process was identified through proactive networking.

3.2.5 Review of proposals, papers and priority setting document

The Research methods professional undertook reviews from a research methods perspective on written materials produced by research projects in form of proposals, papers and reports. The process was systematic and scientific in nature seeking to improve the quality of the reviewed documents. The professional visited the library and went through the priority setting documents to assess if the implemented proposals met the set targets. This was achieved through comparing the achieved milestones to the prioritized researchable areas. The written reviews were discussed with and checked by the supervisors. The review process mainly looked at statistical issues.

3.2.6 Contributing to team planning

The research methods professional attended two planning meetings; participated in discussion about research methods. Observation and contribution made during the planning process and reception of the results were documented. Questions were asked to seek clarification and to have a deep understanding of the subject under discussion.

3.2.7 Data collection

Data was collected through a semi-structured need assessment questionnaire, seminar evaluation forms, face to face interviews, individual and group discussions, training and observations made through participation in the six research tasks. Data collected was on gender, age, designation, working experience in terms of number of years, current education level and attained education level during appointment period, statistical courses attended, specific skills acquired, scores attained from the statistical knowledge test, topics in statistics that one wishes to be trained, how scientists and technical staff handle research data, type of database used to store data, type of statistical analysis carried out on research data, statistical software used during data analysis and feedback responses from the training. More data information was obtained from the Research methods professional's contribution to research, library, reports, websites, organizational policies and other relevant documents.

The Research methods professional accepted tasks from the supervisor and other staff in order to understand how the system worked and developed strategies for improvement. The data collected from the study was analyzed using Genstat 12th Edition a comprehensive statistical system that summarizes, displays and analyze data (Payne *et al*., 2003), SPSS a comprehensive full featured data analysis program that offers a variety of applications including data base management and reporting, statistical analysis and graphics and Microsoft Excel spread sheet which was used to generate pie charts and graphs.

CHAPTER FOUR

4 RESULTS AND DISCUSIONS

4.1 Trainings

Data collected from the semi-structured questionnaire showed that it was necessary for scientists to take self initiative in learning the required statistical skills in research. The major areas scientists needed training were; use of different statistical packages, appropriate use of different research designs and skills in various statistical analyses executed on research data. The responses from the questionnaire and seminar evaluation form are discussed under data processing.

A seminar on measures undertaken in data management process to improve data quality was conducted in May 2011 to a small group of 40 participants representing 28 scientists and 12 technical staff. Quality measures on several activities in data management process were decisively discussed. A group discussion was performed at the end of the training session highlighting the importance of having procedures put in place to ensure high data quality at the centre. During the seminar, it emerged that a lot of data had been lost due to the current data management practices. A critique discussion on the usual practice and its limitation was undertaken by the group. Generally, it was suggested that proper procedures in data management be adopted in order to improve the research undertaken at the center. It was evident that the research at the centre involved a diverse set of individuals from various disciplines who work together for a certain period of time, therefore training on principles of team building and dynamics was

highly essential for the smooth implementation of research projects. It was suggested that training on the current popular software packages was necessary since the previous old software were no longer in use. The challenges faced by the Research methods professional during administration of the tool were; some scientists lost the questionnaire, some gave little attention and filled it within the minimum time, a few misplaced the questionnaire several times, while some were too busy to be found. A lot of time was spent on following up a number of scientists who had the questionnaire.

During the interaction between the Research methods professional and the scientists, program heads, technical staff and scientists in management positions, issues concerning training were raised at different levels. There was clear need for gaining current information through trainings or short courses on various aspects of research. For example, efficient data management, appropriate use of available research designs, sound interpretation of various statistical terms, data analysis, use and access to different statistical software, effective consultancy services among scientists and between scientists and the biometrician and regular refresher courses on new issues for example climate change and how data collected from the new areas of research could be analyzed.

Several short trainings involving smaller groups and individuals were conducted at different times throughout the attachment period. Some scientists were trained on basic analysis using Genstat, R and SPSS. Four research agronomists were trained on analyzing data using Genstat; each scientist had their own dataset which were

imported from Excel spreadsheets. The following statistical analysis were executed on the data sets; t-tests which assessed the yield means of the two different varieties in each study and showed how statistically different they were, summary statistics of all the variables under the study that demonstrated the trend and patterns of the data and hence tentative conclusions drawn, Analysis of variance (ANOVA) for data collected in experiments laid out in Randomized Complete Block design (RCBD), Complete Randomized Design (CRD) and Split Plot Design (SPD). The analysis was achieved through procedure of general linear models and means were separated using least significance difference (LSD). ANOVA displayed initial tests of the treatments and provided an estimate of the residual variance and showed the source of variation. Regression analysis using simple linear and multiple linear regressions with or without groups was also performed and it illustrated how the response variable was related to the explanatory variables. Genstat discovery Edition 3 was installed in the scientist's laptops.

Only one scientist was interested in learning R. She was trained on basic introduction to R; this was simply because some scientists feared the command nature of the package. The following analyses were performed on the scientist dataset; imported data in R console, created vectors for the response (y) and explanatory (x) variables, plotted box plots to explore the data, a model and an ANOVA were fitted on the variables and finally graphs plotted. Data information was erased from the R console and a few packages were installed. The scientist

was particularly impressed with the statistically interactive graphs displayed. The two scientists who were trained in introduction to SPSS were basically having survey data sets. Data was imported from Excel, names of the variables were coded, the type of the variables were indicated clearly as strings or numeric, the coded variables were labeled to provide more information and deep understanding of what the codes represented, values for the variables were indicated. The data sets were subjected to the following analyses; descriptive (in form of frequency tables and tables of means), ANOVA and correlation.

Two short training courses entitled "Appropriate use of spreadsheets with illustration from Microsoft Excel; Introduction to effective use of database management system with illustration from Access and CSPro which are public-domain software packages for entering, editing, tabulating and mapping census and survey data, was conducted in the Livestock section. Milk production data was practically entered in Excel, Access and CSPro by the participants with assistance from the Research methods professional. The section collects data on a daily basis which consequently calls for a sound data management system. Examples of the type of data collected are milk production per cow, monitoring animals progress, calving to drying trends, breeding data (from birth to serving) and weight gain. From the above data, quarterly, monthly and centre reports are developed. The staff in Livestock acknowledged the importance of having data entered and stored in the spreadsheets and databases as a backup strategy. After the training sessions a letter was written to the centre director requesting for a computer, which was an

initiative from the technical staff with the Research methods professional's support. The staff was advised to have refresher courses on current issues in data management and most especially computer skills since most of them were less conversant with computer use. The training anticipated to increase effective data entry into computer files.

Two students; Kisiang'ani Sammy - Maseno University and Paul Muhati- Moi University both undertaking Bachelor of Science (BSc) Applied Statistics had a one week training by the student on attachment under the KARI supervisor's guidance. A different teaching approach from the normal practice was used. The students were instructed to read ahead and have a discussion with the facilitator later. In some sessions the two students were instructed to give presentations, which was followed by an in depth discussions. Each topic had a practical session. The following topics were covered: Basic elements found in a proposal, an exercise was given to the students to assess how well the proposals in CRAC and Pre-CRAC documents conformed to the stipulated guidelines. The two students found out that a number of proposals did not contain some of the required elements for example lack of clearly set research questions or hypothesis, indication of sample size without explaining the methods used to determine the sample size, a few of the abstracts lacked research problem, purpose of research and sampling techniques. Majority of the proposals lacked the significance, limitation and assumptions of the study to be carried out while in various proposals the problem statement was masked in the introduction.

The students with the help of the facilitator were taken through the various activities performed in data management process and were shown various types of checks that are normally put in place to maintain data quality. The students were taken through experimental designs and how to interpret the P value generated after data analysis. They entered data in Excel and prepared it ready for analysis, they analyzed data using Genstat and SPSS and made inference, they were shown the various ways of presenting results and the importance of archiving data obtained in different places.

4.2 Data administration

Loss of substantial proportion of research data is due to non-existence of a centralized system for data backup, archiving and documentation in most programs at the center. Data archiving is a major problem especially in the Livestock Section that is involved in daily data generation. The data are usually recorded in data sheets that are stored in files. Wear and tear of the files containing the data sheets and loss of the only copy available is a common occurrence. The hard and soft copies of data are stored mostly in one place, or kept by one individual. Retrieving and accessing such kind of data is difficult and time consuming. A seminar to that effect was held to sensitize both the scientists and technical staff on the importance of proper data management. A lot of time and resources are usually allocated to data collection but the same resources are inadequately distributed to planning, data analysis and data archiving. This usually compromises the quality of data produced at the end of the research process. In most cases the collected data is too

much, unnecessary and irrelevant to the study objectives. Data entry is performed by technical staffs that have little or no training in data entry therefore increasing errors in the data. In some instances the collected data is never analyzed.

The scientists and technical staff were advised to have planning meetings before commencing data collection, provide constant supervision during data collection, reduce data errors through minimal data transfer by use of pocket computers, design data sheets early enough, develop a sound data archiving system and train personnel handling data. The Research methods professional encouraged the Maize Section which has a sound data management system to give a seminar on how they manage their data.

The professional designed, pre-tested and administered a semi-structured questionnaire that she used in collecting her MSc data. She participated in designing questionnaires for collecting data in the Socio-Economic Section and prepared a data sheet for collecting information for the passion fruit project in Horticulture Section. The professional also participated in preparation of three data files of the Mc Knight finger millet project. The data was about evaluation of different finger millet varieties in four districts namely Teso, Bungoma, Siaya and Mumias (Table 1).

Table 1: 2009 Short rains PYT finger millet data collected from Alupe ready for analysis

Rep	Block	Row	Plot	Variety	Entry no.	Lodge%	Yield kgha-1
1	1	1	1	OKXG F5B5B12R8R2	27	2	1,613.1
1	1	2	2	IE 4115	95	2	1,367.2
1	1	3	3	GBK028044	81	6	509.7
1	1	4	4	U-15XNB F5BR12R4R1	9	2	798.9
2	1	1	1	OKXGEF61R10(R22)R3	50	0	1,468.6
2	1	2	2	OKXGEF61R8(R24)R3R	47	1	749.4
2	1	3	3	OKXGEF6B13R5(R9)R1	45	3	638.1
2	1	4	4	OKXEGEFB13(R2)R1R1	42	1	361.9
3	1	1	1	OKXGEF6B13R8(R24)R 3R	48	5	966.9
3	1	2	2	P6-4(7)	78	2	1,712.8
3	1	3	3	U-15XNBF5B1R12R4R1	8	3	1,286.4
3	1	4	4	U-156-224R7(R8)R3R2	38	3	805.6

The data set consists of 3 replications, 10 blocks, 10 rows, 100 plots and 17 variables in total (Table 1). The variables under study were yields (kg/ha), bird damage, number of days to physiological maturity, ear extension and shape, lodging (%), striga flow, number of days to 50% flowering among others.

Due to the existing problem of data management in the institution, a data management protocol was developed (see Appendix D). It is expected that the developed protocol shall be used as a systematic and complete tool to ensure efficient and effective project processes. The protocol was designed to ensure that

project procedures and results are well recorded and relevant information properly preserved and archived. The document is supposed to act as a necessary tool at the planning stage of a research project focusing attention on various strategies for data computerization, checking, organization and analysis and archiving. It is anticipated that the use of the document will contribute to improved data quality.

The library which is the main central place for data archiving at the centre is not well facilitated. Most of the librarian's duties concerning data backup and archiving are not performed adequately since the section has no budget of its own. The budget is drawn from the administration section that relies on 10% of the funded projects. The librarian was encouraged to collect and analyze data generated in the library and showcase the findings to scientists and other channels like Maktaba Awards sponsored by Kenya Library Association (KLA). He was further advised to write proposals for funding. It is the duty of the librarian to sensitize scientists and technical staff on new technologies developed, upgrade existing databases to conform to the latest rapidly developing technology and perform frequent client's needs assessment. The librarian was advised to acquire current skills and lobby for resources in order to execute his duties effectively.

4.3 Data processing

Data sets from research projects, MSc student, scientists, need assessment survey and seminar evaluation forms, were entered, prepared, cleaned, organized and analyzed by the Research methods professional. Drop down lists were created in some data sets during data preparation. During data cleaning and data validation,

Excel was used to generate standard decimal points and range of values required in the data to check for errors. The entries with extreme values and more decimal points were circled and later corrected with the consent of the data owner.

Five data sets from phase two finger millet projects (2010-2013) were organized and analyzed using Genstat. The finger millet trials consisted of varietal evaluation and fertility trials in four districts. The designs used in the trials were triple lattice with 100 entries, RCBD and split plot. The various statistical analyses executed on the data were; descriptive statistics in form of means of variables (yields, foliar blast, pest and disease incidence among others) and box plots to check for outliers. Analysis of variance (ANOVA) was used to check for varietal differences in terms of other variables such as yields. Means were separated by use of LSD if the ANOVA showed treatment differences and correlation of the variables was performed to identify the specific variable with significant effects on yields.

During data cleaning and organization, some of the errors observed and their effect on data information were documented. An example was taken from Mini-Core data 2009 long rains Alupe which had been entered by one of the section technical staff (Table 2).

Table 2: Various typing errors observed in the 2009 Mini-Core long rains finger millet data sets collected form Alupe region

Rep	Block	Row	Plot	Entry no.	Variety
1	3	7	27	79	VL149
1	3	8	28	78	RAV8
1	3	9	29	73	6350
2	9	6	86	69	6240
2	9	7	87	78	RAV8
2	9	8	88	23	2996
3	3	7	27	78	RAU8
3	3	8	28	48	4671

The variety names RAV8 and RAU8 are assumed to be different due to typing error while the entry numbers are the same (78) (Table 2). The errors found in data sets need to be rectified before analysis to avoid production of scientifically inaccurate information (Table 3).

Table 3: Generated summary statistics for Yields (Kg/ha^{-1}) for entry RAU 8 finger millet variety before error correction

Description	Values
Number of observations made	1
Number of missing values	0
Mean	517.5
Median	517.5
Minimum	517.5
Maximum	517.5
Lower quartile	517.5
Upper quartile	517.5
Standard deviation	0
Coefficient of variation	0

RAU8 which is assumed to be a different finger millet variety from RAV8 generated insufficient statistics values after data analysis (Table 3). This could easily mislead a scientist in making wrong interpretation (Table 4).

Table 4: Generated summary statistics for Yield (Kg/ha^{-1}) for entry RAV8 finger millet variety before error correction

Description	Values
Number of observations made	2
Number of missing values	0
Mean	208.5
Median	208.5
Minimum	123.1
Maximum	293.9
Lower quartile	123.1
Upper quartile	293.9
Standard deviation	120.8
Coefficient of variation	57.94

RAV8 finger millet variety is assumed to be different from RAU8 variety (Table 4). While in real sense the two finger millet varieties are the same. The summary statistics in Table 3 are different from those in Table 4 due to the typing error. The two Tables, 3 and 4 show the effects typing errors have on statistical values. A well entered research data generates statistically sound values after analysis (Table 5).

Table 5: Generated summary statistics for Yield (Kg/ha^{-1}) for entry RAV 8 finger millet variety after error correction

Description	Values
Number of observations made	3
Number of missing values	0
Mean	311.4
Median	293.9
Minimum	123.1
Maximum	517.5
Lower quartile	165.8
Upper quartile	461.6
Standard deviation	197.8
Coefficient of variation	63.5

Results generated after correcting errors (Table 5) which are slightly different from the ones in Table 3 and 4. Analyzing data without checking and correcting the errors eventually affects the quality of data information generated. Precaution should be taken in identifying and correcting errors before analysis. Typing errors (not exhaustive) were noted in the finger millet datasets during cleaning (Table 6).

Table 6: Additional typing errors observed during data cleaning of 5 finger millet data sets of phase two (2010-2013)

Rep	Block	Row	Plot	Entry No	Variety
1	2	1	20	87	ufm86
1	2	4	17	84	ufm 260
1	2	5	16	86	ufm 57
1	2	6	15	88	Seremi1
1	2	9	12	82	ufm 149
1	3	1	21	80	sarada
1	3	2	22	77	PR 202
2	10	6	95	88	SEREMI1
2	2	1	20	86	UFM57
2	3	2	22	84	UFM 260
3	1	4	4	80	SARADA
3	10	2	99	77	pr 202
3	10	3	98	87	UFM86
3	9	5	85	82	UFM149

Differently typed variety names have the same entry numbers (Table 6).

Conclusively, typing errors in data entry can be minimized by establishing a double data entry system early enough before data is collected; training data entry clerks and if resources allow have more clerks entering the same data for comparison.

A dataset from a questionnaire survey entitled “Household growing mushrooms” was analyzed using SPSS. The dataset was from an MSc student of Masinde

Muliro University of Science and Technology (MMUST). Data was collected from two divisions and comprised of 50 variables. The statistical analysis executed on the data were; descriptive analysis which took the form of tables of means, frequency tables and pie charts that showed the percentages of household profile, mushroom consumption, employment opportunities and household income, a graph showing the contribution of mushroom enterprise, livestock enterprise, cash crop enterprise and other enterprises to food security, correlation determining the existing relationship between the socio-economic and food security variables. Poorly coded and entered data not ready for analysis is difficult to comprehend (Table 7).

Table 7: Incorrect entries and codes of household growing mushrooms survey data set collected from Sabatia and Vihiga divisions of Kakamega district in 2007

Hp1	Hp2	Hp3	Hp4	Hp51	Hp52	Hp53	Hp61	Hp62
1	1	1	Jane			1		1
2	1	1	John			1	1	
3	1	1	Mary			1		1
4	1	1	Cathy	1				1
5	1	1	Petro			1	1	
6	1	1	Betty			1		1
7	1	1	Joan			1		1
8	2	1	Agnes	1				1
9	2	1	Mercy			1		1
10	2	1	Grace		1			1
11	2	1	Phyllis			1		1
12	2	1	Nelly			1		1
13	2	1	Grace	1				1
14	2	1	Debby			1		1

Hp1=Household code, hp2=Division, hp3=District, hp4=Name of respondent, hp51=single, hp52=divorced, hp53=married, hp61=male, hp62=female

A lot of time would be spent in order for an analyst to understand what the codes represent (Table 7). Hp51, 52 and 53 represent marital status and Hp 61 and 62 represent the gender of the respondents. This should have been entered in two separate columns representing marital status and gender. The data set was cleaned and organized ready for analysis by the Research methods professional (Table 8).

Table 8: Correct codes and entries of household growing mushroom survey data set from Sabatia and Vihiga divisions in 2007

HH code	Division	District	Name of Respond	Marital status	Gender
1	1	1	Jane	Married	Female
2	1	1	John	Married	Male
3	1	1	Mary	Married	Female
4	1	1	Cathy	Single	Female
5	1	1	Petro	Married	Male
6	1	1	Betty	Married	Female
7	1	1	Joan	Married	Female
8	2	1	Agnes	Single	Female
9	2	1	Mercy	Married	Female
10	2	1	Grace	Divorced	Female
11	2	1	Phyllis	Married	Female
12	2	1	Nelly	Married	Female
13	2	1	Grace	Single	Female
14	2	1	Debby	Married	Female

The same dataset is correctly entered, organized ready for analysis and easily understood (Table 8). It is important to code the variables in the data very well to avoid confusion. Most of the errors in Table 7 would have been avoided if some thought had been given to the layout of the data in the spreadsheets before data collection commenced. The student who owned the data should have taken responsibility and guided the data entry clerk.

Generally, most scientists store their data in forms that are not ready for analysis and occasionally store unclean or unorganized data. A lot of time is therefore taken

during analysis of such kind of data. It was observed that during data analysis most scientists preferred using ANOVA and rarely used statistical modeling. Excel spreadsheet was commonly used to store data by scientists even if the data structure was complex. In line with that, scientists and technical staff were advised to study their data well in order to store it in the appropriate database. Currently, the SAS package in use is out dated. The research institution should provide scientists with adequate and diverse statistical packages for analysis through purchasing the software. Scientists were encouraged to take initiative in downloading free and available packages such as R. The Research methods professional provided Genstat discovery and R statistical package to some scientists.

The responses from the semi-structured questionnaire entitled "Assessing the statistical needs of scientists at KARI Kakamega" consisted of 42 variables. The tool was given to 28 scientists but only 21 filled copies were returned. Mean frequency tables, graphs and pie charts were generated to assess the statistical knowledge of scientists. The first part of the questionnaire captured the background information of scientists involved in the study (Fig 3).

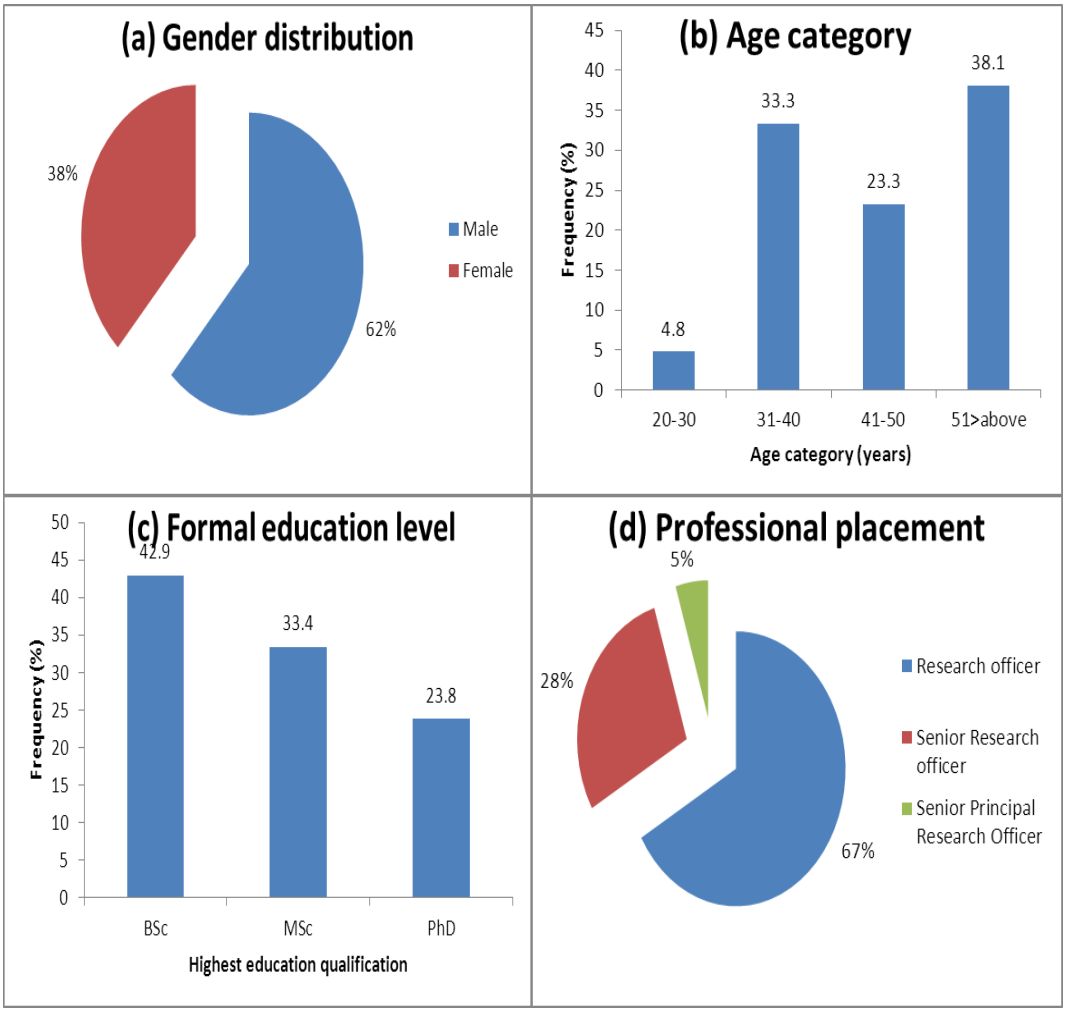


Fig 3: Key background information in terms of (a) gender distribution, (b) age category, (c) formal education level and (d) professional placement of scientists at KARI Kakamega

There are more male scientists (62%) than female scientists (38%) at KARI Kakamega (Fig 3). Majority of the scientists fall under the age bracket of 51 years and above which represents 38%, followed by 31-40 years, 41-50 years and 20-30 years age brackets representing 33%, 23% and 5% respectively.

The highest educational level attained after several years of appointment was BSc with 43%, followed by MSc and PhD representing 33% and 24% respectively. Most scientists are Research Officers (67%), while a few are Senior Research Officers (28%) and only one is a Senior Principal Research Officer.

A large proportion of the scientists were employed at BSc level (43%), while some at certificate level (19%) and one employed at PhD level. A number of trainings are usually offered to scientists through the research institution initiative. The results showed that majority of scientists had prior training in statistics as at 2011 (Fig 4).

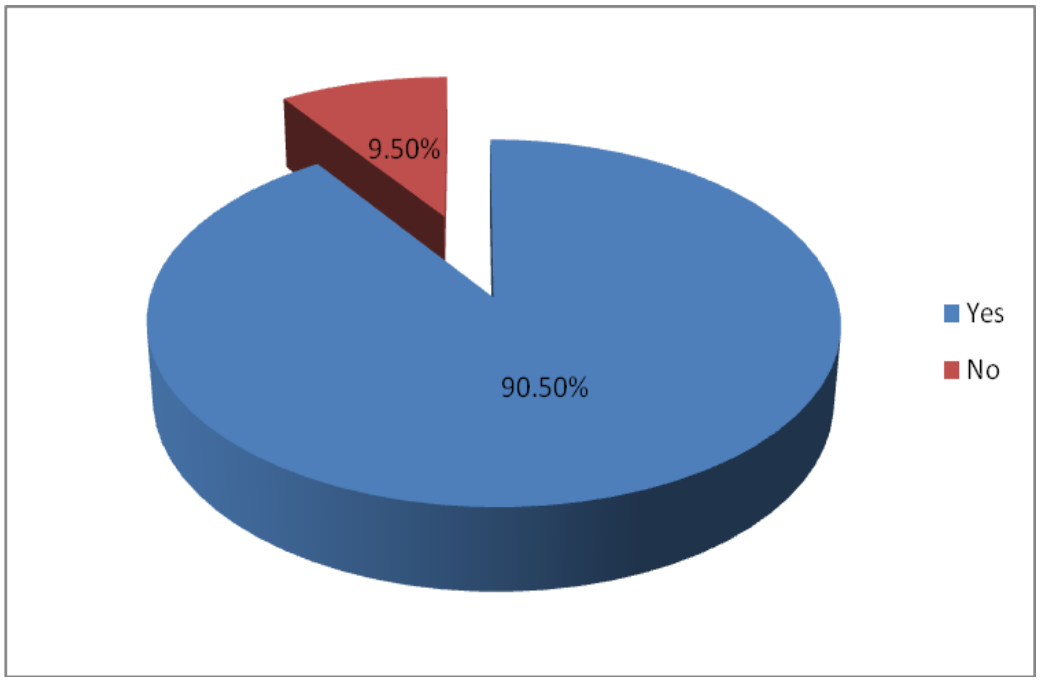


Fig 4: Percentage of KARI Kakamega scientists who had prior training in statistics

Majority of the scientists (90%) indicated that they had received statitical training while 10% had not received any statistical training (Fig 4). The training content was described at four levels by the scientists who received statistical training (Table 9).

Table 9: Description of the nature of statistical training received from KARI Kakamega scientists

Description	frequency (n=21)	frequency (100%) percentage
Training levels		
Basic	8	38.1
Intermediate	6	25.6
Advance	2	9.5
Specialized	5	23.8
Acquired skills		
Data entry	16	76.2
Data processing	16	76.2
Descriptive analysis	10	47.6
Inferential analysis	8	38.1
Modeling	7	33.3
Others skills		
Statistic programming	1	4.8
ANOVA	1	4.8

Majority of scientists (38%) believed the training they had was basic, 26% believed it was intermediate while 24% believed it was specialized and about 10% believed it was advanced in nature (Table 9). Data entry and processing were the

most acquired skills representing 76%, followed by descriptive analysis, inferential analysis and finally modeling representing 48%, 38% and 33% respectively. The other skills mentioned were statistical programming and ANOVA. The scientists working experience ranged from a minimum of 3 years to a maximum of 37 years. A large proportion of the research data is commonly analyzed by the statistician or scientist (Fig 5).

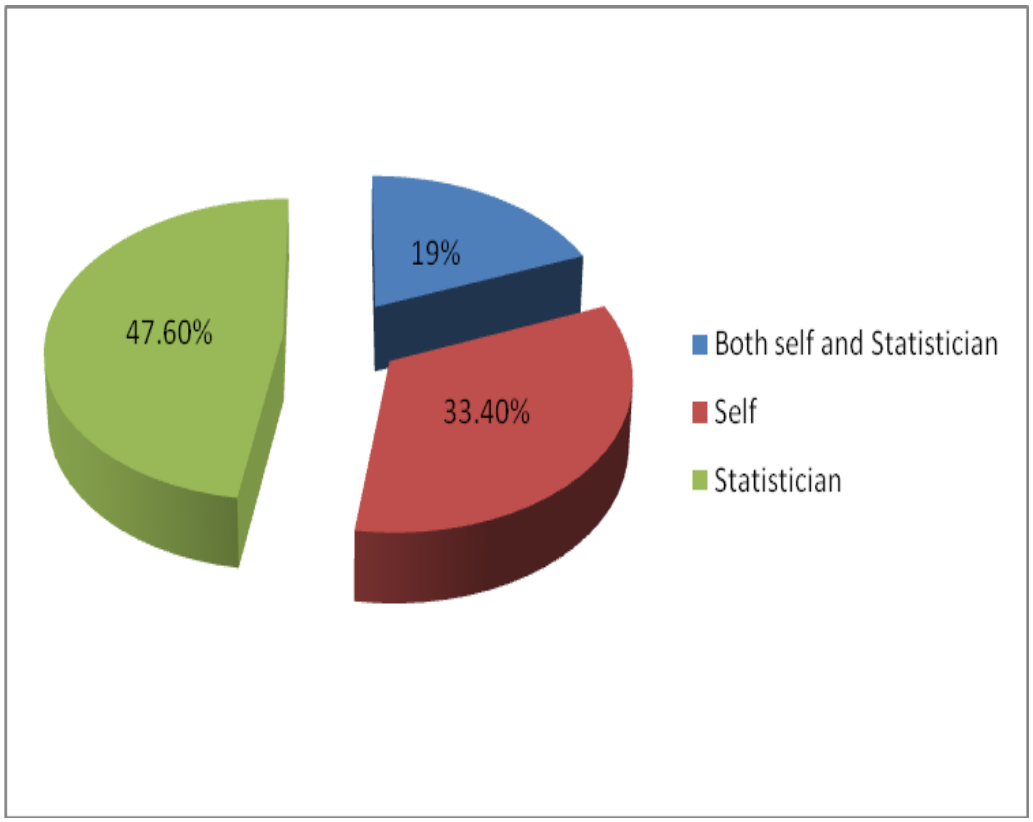


Fig 5: Individuals responsible of data analysis at KARI Kakamega

Almost half of the research data is analyzed by the statistician representing 48% while 33% data is analyzed by the scientists (Fig 5). A small percentage of research data is analyzed by engaging both the statistician and scientist. The

statistical analysis software scientists are familiar with and the type of analysis commonly executed on the research data were listed exhaustively (Table 10).

Table 10: Type of software used and analysis performed to research data by KARI Kakamega scientists

Variable description	frequency (n=21)	frequency (100%)	percentage
Analysis software used			
Genstat	6	28.6	
R	1	4.8	
SAS	13	61.9	
Excel	10	47.6	
SPSS	13	61.9	
STATA	2	9.5	
Others			
LIMDEP	1	4.8	
Sigma plot	1	4.8	
Type of analysis performed			
Descriptive statistics	13	61.9	
ANOVA	18	85.7	
Regression	11	52.4	
Modeling	3	14.3	
Correlation	14	66.7	

SPSS and SAS packages as the most commonly used statistical analysis software, followed by Excel and Genstat, while R is the least used package (Table 10). Other packages scientists mentioned were LIMDEP and Sigma plot. ANOVA is the

frequently executed analysis on data by scientists followed by correlation, descriptive statistics, regression and the least was modeling. Part of the tool was administered as an exam to scientists to gauge their statistical knowledge. Above 50% mark indicated the ability of scientists in understanding statistical terms, performing proper data management procedures, executing required data analysis and interpretation (Fig 6)

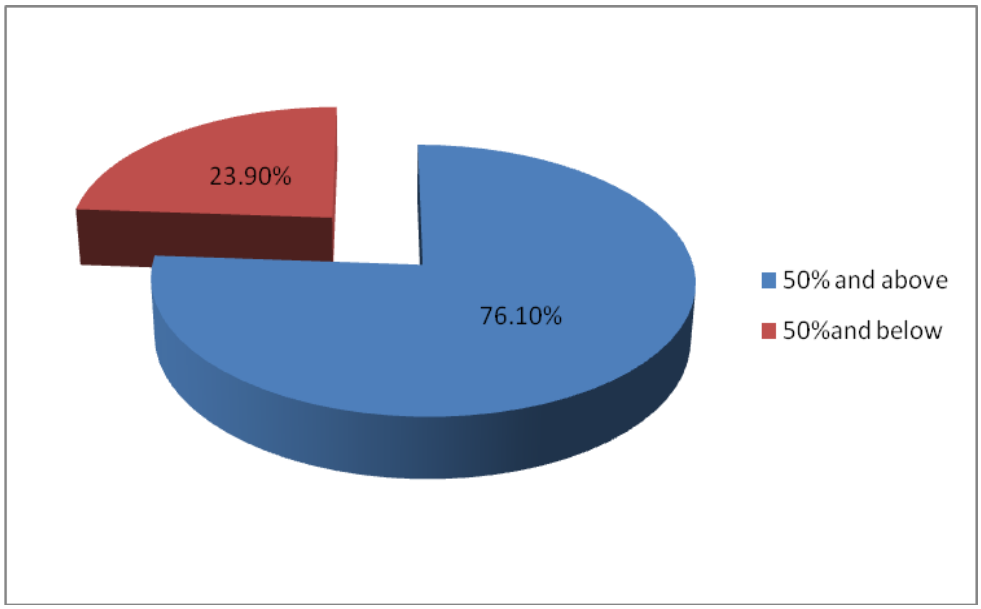


Fig 6: Statistical exam performance by KARI Kakamega scientists in 2011

Overall 76% scientists scored above 50% while 24% scored below 50% (Fig 6). The highest mark scored in the given exam was 70%, the minimum 34% with an average of 56% and median and mode mark of 58%. This clearly portrays that most scientists only possess basic statistical understanding insufficient to critically evaluate and interpret the statistics information generated from the data collected.

All the scientists indicated that they usually seek statistical advice for their research project when designing their studies which was the correct answer required. Fifty two percent thought the main contribution of statistical analysis to their research was to help them explain the reasons for variability in their data while 48% thought that statistical analysis was performed in order to back up theories with results from the analysis technique. Eighty one percent defined statistics as a discipline that deals with describing data and making generalizations, while 14% defined statistics as an exact science. Fifty eight percent of the scientists gave valid reasons to why the normal distribution is important in statistical analysis while 42% gave inapplicable reasons to why normal distribution is important in statistical analysis. Fifty seven percent stated that the standard error was more useful than standard deviation for interpreting estimates of population characteristics while 43% disagreed with the statement. Eighty one percent defined a standard error as a measure of precision while 10% defined it as a measure of significance and the other 10% as a better measure of spread than standard deviation. The statement "increasing the sample size usually reduces the standard deviation" was indicated to be true by 52% of the scientists while 48 % thought the statement was false.

Scientists further indicated the research designs they were familiar with and stated the designs they commonly used during research project implementation with valid reasons to why they used the particular designs (Fig 7).

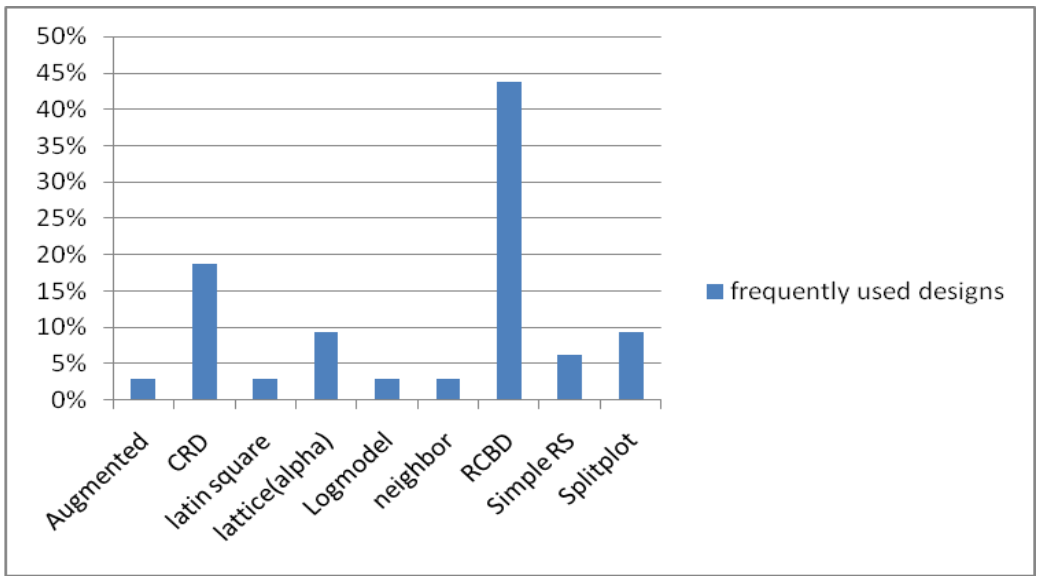


Fig 7: Commonly used research designs by KARI Kakamega scientists

RCBD, CRD, split plot and lattice (alpha) designs as the most commonly used research designs by scientists while the least used were Augmented, Simple Random Sampling, nested design, nearest neighbor and log model (Fig 7). The list of the research design was compiled from the scientist's response.

Choice of specific research designs by scientists depended on; 1) the appropriateness of the tasks under study, 2) the type of study – experiment/survey, 3) the number of treatments in the study, 4) provision of controlled environment, 5) objectives of the study, 6) ability of the research design to minimize bias/reduce errors due to environmental effects, 7) the simplicity in laying out the design in the field, 8) suitability of the design for agronomic trials, 9) provision of recommendations in the shortest time possible and 10) the capability of the design to provide good results. During the questionnaire administration, scientists were

asked to list the topics they would prefer to be taught in a statistical training (Fig 8)

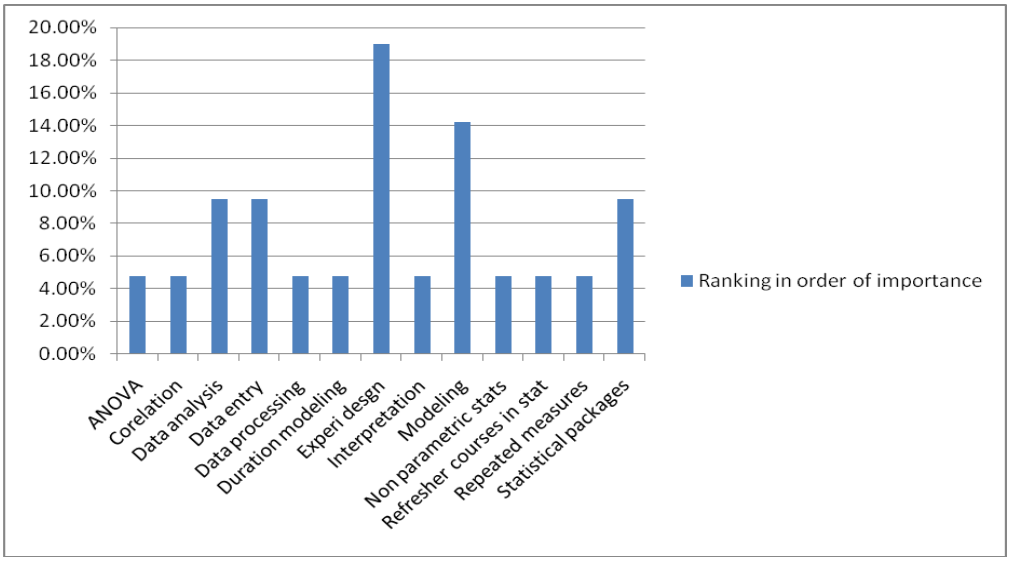


Fig 8: Preferred research topics in a statistic training course

Experimental designs, modeling, knowledge on use of diverse statistical packages, data entry and analysis were the most preferred topics in a statistical training course (Fig 8). The least preferred topics were ANOVA, correlation, data processing, duration modeling, and interpretation of results, non-parametric statistics and repeated measures.

Since the process of seminar evaluation was not compulsory, only 36 out of 40 participants were involved in the evaluation exercise and the responses analyzed (Fig 9).

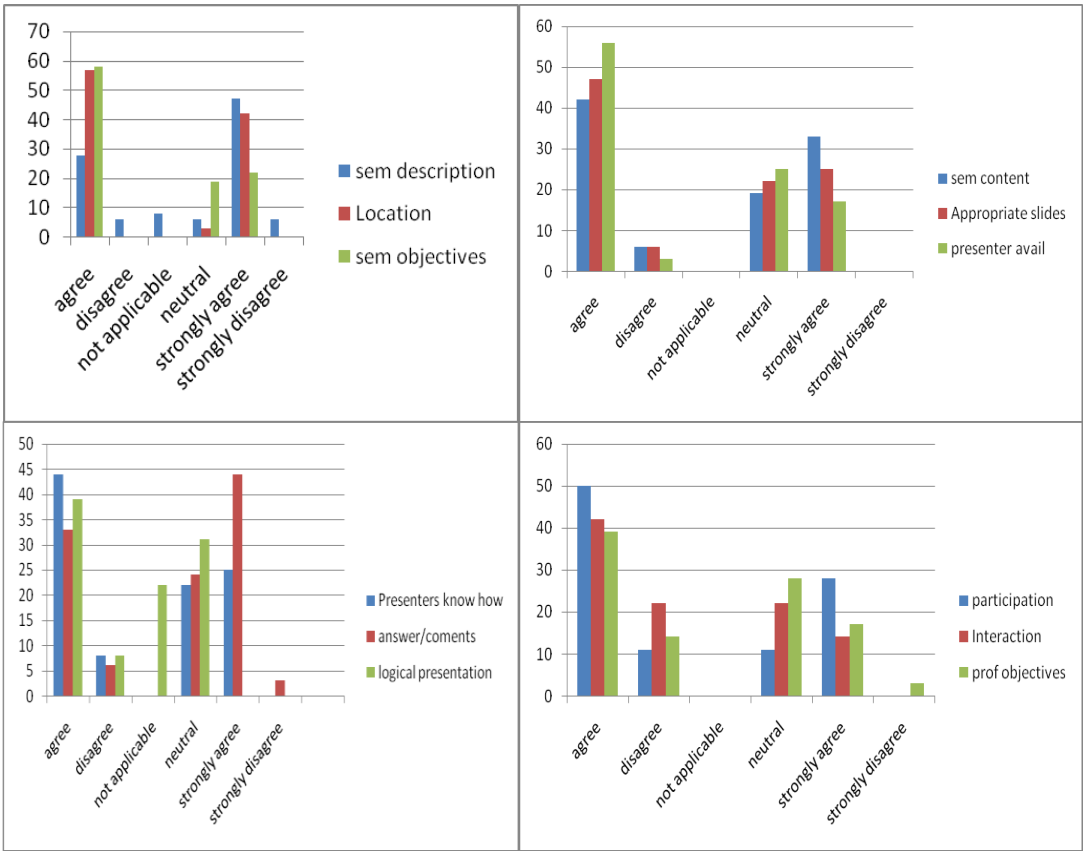


Fig 9: KARI Kakamega Scientist and technical staff feedback responses from quality data management seminar held in May 2011

A large proportion of scientists and technical staff strongly agreed on clarity of seminar description (47%) and answers/comments provided by the presenter (44% (Fig 9). Majority of the scientists agreed on location acceptability (57%), logical flow of seminar objectives (58%), relevant organization of seminar content (42%), appropriately organized slides (47%), availability of presenter (56%), presenter's knowledge on the topic presented (44%), logical presentation (39%), encouraged participation (50%), beneficial interaction (42%) and meeting of professional objectives (39%). While two were not satisfied with the seminar description and one was not satisfied with the clarity of answers/comments provided by the presenter. Most of the scientists thought the seminar was effective in most topics

especially data processing and planning, while three did not respond to the question and four thought it was not applicable. Some participants indicated that the seminar could be improved in data archiving and data quality, though majority indicated that the topics were well covered. Seventy eight percent of the scientists believed others would benefit from the seminars while 22% thought the seminar would not be beneficial to others. The participants suggested more time and more practical exercises to be incorporated in the seminar if it was held again.

4.4 Contribution to research

Pre- CRAC and CRAC forums are some of the mechanisms the institution has established to ensure quality and relevance of research is of scientific acceptable standards. The two fora meetings were held on 2nd to 4th March 2011 and 13th to 15th April 2011, respectively. The Research methods professional was invited to participate and made the following observations; Most of the findings in the reports and papers contained results without statistical evidence, some did not have statistically analyzed data, others had incomplete materials and methods that were not exhaustive for replication, several had inadequate references to provide evidence to the results obtained and almost all the presented documents lacked P-values, hypotheses, research questions and sound objectives. At the end of the session the Research methods professional with the guidance of the supervisor made a presentation entitled "The basic components found in a High Quality Scientific Proposal" It was observed that a handful of designs had been commonly used while some designs were never mentioned to have been used. Most of the

scientists did not have valid reasons to why they used particular designs. This lead to a short presentation entitled “Factors affecting the use of a particular research design”. It emerged that most scientists assumed they knew the information.

Some reports had untitled tables that were not numbered. Some presenters consistently shifted from one slide to the next due to the mixed up slides. All the presentations made were in power point form. Individuals were advised to acquire presentation skills and diversify in other forms of presentations apart from power point. The scientists and extension officers attending CRAC were advised to download the free available presentation packages from the internet and an example of visual mind maps was given. Some new proposals had indicated the proposed funding agency but the principal investigator had not made prior arrangements to make a formal request to the identified agency; furthermore the start date of the project was overdue (January 2011). Due to the competitive nature of the process, early preparation was recommended. Some progress reports lacked the expected outputs that are compared to the results obtained in order to assess the percentage achieved and describe way forward.

The Research methods professional advised the scientists to update their statements in order to portray the current real life situation. For example sweet potatoes, cassava, arrow roots were mentioned to be a poor man’s food, while in real life situation the food commodities are largely consumed by the rich whereas the poor sell them in order to buy other alternative products like bread. Additionally the said products have become scarce and very expensive. The

scientist were encouraged to be open minded and stop thinking as they did a few years back and embrace the changing situation in research. There was poor attendance of various stakeholders in the two forums especially the private sector which led to inadequate representation of the Agricultural Productivity Value Chain (APVC). APVC is an important tool in developing holistic research technologies that enhance the integration of the knowledge and contributions of all relevant stakeholders. The forums have been adopted for a long period of time, the administrators were advised to change to a more current and effective approach that would encourage the participation of all stakeholders.

The Research methods professional participated in practical layout of different research designs in the field, which involved 2 activities in the maize program that had used RCBD, and Alpha lattice designs and 2 activities in finger millet program that had used triple lattice design and RCBD. The mentioned designs for the two programs had been provided by the funding agency called International Maize and Wheat Improvement Center (CIMMYT) and Mc knight foundation. The professional provided assistance to an MSc student of University of Nairobi undertaking her research at KARI Kakamega in laying out research design for her project. The experiment titled Effect of Arbuscular Mycorrhizal Fungi (AMF) inoculation and P fertilization on soil P and P uptake by Orange Fleshed Sweet Potato (OFSP) comprised of the following variables, AMF inoculums which consists of 3 single commercial strains, 1 single indigenous strain, a combination of the commercial strain and a control, 3 fertilizer levels at 0 kg/ha, 20kg/ha and

40kg/ha. The experiment was replicated 4 times. Since the student had not indicated in the proposal the type of design to be used, she was advisable to use Split-Split-plot design because of the number of variables in the experiment and variability in the farm. The main plot was fertility levels, sub plots were the 2 varieties of OFSP and sub-sub-plots were the 6 inoculums.

The professional participated in the layout of research design in a horticultural project entitled development of Integrated Pest Management (IPM) and disease control options for passion fruit. The experiment involved six farmers of Lugari district. The experiment constituted of 1 passion variety and 5 treatments in form of 5 different fertilizer levels. The treatments were tithonia, no fertilizer application, farmyard manure, chicken manure and Diammonium Phosphate (DAP) applied at planting. The farmers were used as replicates giving a total of 6 replications for the trial. The professional advised the team to use a RCBD design in which fertilizer levels represented the unit plots. Randomization was done through number coding of treatments in the trials, each number represented a treatment. The numbers were then written on pieces of paper which were picked at random and allocated to each plot. Randomization of one site was independent of the other sites and the replications were not adjacent in all the trials. The common practice with scientists is randomization layout of one site used in the other remaining sites. The replications are usually adjacent in all the sites for instance replicate 1 adjacent to 2 which are adjacent to 3. The required number of plants per treatment was 10 but since land was a limiting factor in some farmers the plot sizes

were reduced, some replications were curved to fit the available land. In some farms, one replication was located in a different place and in others the total number of plants per treatment was reduced to 8 since data was to be recorded on a sample of trees not the entire population.

During the participation in research through offering consultancy services to scientists it emerged that a lot of tasks are given to the statistician at peak seasons for example during pre CRAC, CRAC, conference or workshop presentation and mainly at data analysis and review stages. While the statistician is rarely consulted in conceptualization, data collection and preparation of data ready for analysis. Scientists were advised to take an initiative and learn to perform some statistical analysis while the statisticians just fine tunes if need be. Poor interpretation of outputs obtained from the data analyzed was observed among scientists and this could be improved through enhanced P values interpretation training.

4.5 Review of documents

An internal peer review system known as Center Internal Advisory Committee (CIAC) has been established at the research institution. The system can be strengthened by incorporating a diverse internal review structure that assists scientists in lobbying for research funds as well as guiding them in performing the expected research. The Research methods professional reviewed six documents in total (Table 11).

Table 11: Reviewed documents during the attachment period- April 2011

No	Authors	Title of the article reviewed	Type of article
1	Salaysa, B.D.S., Ajanga, S and Odongo, O. M	KARI-Kakamega report on priority setting exercise carried out in October and November 2005	Priority setting document
2	Kisiangani Paul Nyongesa (MMUST MSc student)	Contribution to Oyster mushroom cultivation to food security in western Kenya	MSc proposal and MSc thesis
3	R.N. Magoti, F.M. Matiri and S.C. Amboga	Assessment of the potential domestic market for cut flower and mapping for improved production in Mount Kenya region	Publication meant for a journal
4	S.S. Inzaule, J. Awino, N. Makete, W. Mutebi, E. Egadwa and P.Aseyo	Effects of varieties, inorganic fertilizer and organic fertilizers on bacterial wilt incidence and yield in tomatoes	CRAC progress report
5	J. O. Achieng, S. Ajanga and D. Gemenet	Participatory evaluation of drought tolerant maize varieties in western Kenya	CRAC proposal
6	Patrick Ndemba and Eunice Onyango	An investigation in the applicability of the "Burke-Lit win" model for affecting organizational change and measuring individual performance	CRAC proposal

The prioritized researchable areas highlighted in the priority setting document were not well represented in the proposals and progress reports during CRAC (Table 11). A large percentage of the research in the proposals presented was on the agronomy theme that took lower ranks during prioritization, while the highly ranked crop breeding research took a small percentage (1st document). The constraints identified in the different regions were not included as part of the researchable areas hence making the document inadequate in providing solutions for the specific regions. Furthermore, some regions had adequate information

while others had inadequate information therefore presenting unbalanced identification of researchable areas. The authors never gave valid reasons to why they thought ranking was the best method to be used to collect data. The flow in the executive summary was not logical and there was some disconnection in some parts of the section. The document had inadequate references to provide evidence to the claimed statements, some statements were repeated and the table texts were placed above the table instead of below as required. More collaborators should have been involved to represent the value chain adequately.

The MSc proposal and thesis documents were not well formatted using MS Word Styles (2nd document). The documents lacked an appropriate table of contents, properly reported statistical methodology with regards to Introduction, Methodology, Results and Discussion (IMRaD) guidelines. The tables in the document were inconsistently numbered and inadequately titled. The collected data was not adequate and sufficient to meet the stated objectives or support claimed conclusions. The author of the paper to be published in a journal was inconsistent in writing the paper and there was no logical flow of the results of the study (3rd document). The document had self explanatory tables and figures well presented, the reported analyses were appropriate and results well suited for the purpose. The paper had inadequate references and general the statistical acceptance of the paper was low since the result percentage of some tables was not totaling to 100% as expected.

The CRAC progress report on tomatoes lacked analyzed data and references of previous work to provide support to the conclusions made (4th document). The statistical significance of the results was not indicated. The CRAC proposal (5th document) lacked basic sections for a proposal such as justification, problem statement, hypotheses, research questions, expected outputs, work plan and budget. Some activities in the proposal were presumed to have been performed in February 2009 while the proposal was presented in April 2011. The document lacked clear indication on what sort of data was to be collected and the actual methodology to be used.

Finally, the last reviewed document lacked a table of content, clear objectives, hypothesis, research questions, expected output and inadequate description of the research methodology to be used (6th document). Inadequate relevant information was provided from the previous studies and the author provided insufficient justification to why there was need for a new study. The proposal contained inconsistent logical flow; some sections were too long having inappropriate information and lacked valid description and a list of all the statistical tests to be used in the analysis.

Generally, adequate time should be allocated to developing proposals and papers in research; it was evident that obvious mistakes were common in most of the documents reviewed. Guidelines need to be followed strictly to improve the quality of document. Formulation of scientifically sound objectives at the proposal development stage would guide the scientists in collecting the required data. The

internal peer review members should be trained constantly to update themselves on current statistical issues. The reviewed documents have been written on a CD.

4.6 Participation in team planning

Most of the project teams at the research institution hold planning meetings that clearly state the different roles individuals should perform before a research project is implemented. But in some programs team planning is never held at any research project stages instead activities are performed as need arises. Such situations can lead to omission of important issues or delayed execution of some activities. Holding planning meetings in all research projects was recommended.

Through invitation, the Research methods professional attended two planning meetings held at the research institution. The first meeting was held on 26th January 2011. The meeting involved United States Agency for International Development (USAID) food security team, Kenya Horticulture Competitiveness Project (KHCG) and KARI Kakamega scientists. The main objective of the meeting was to assess the food security levels in Nyanza and Western Province and hence identify gaps for more research. The research institution was required to state categorically its contribution to food security in the two regions. During the discussion, existing opportunities for collaboration were identified. The professional participated in the identification of research gaps and prioritization of researchable areas. The USAID team accepted to fund the various outlined research activities for 5 years. The expected output from the project was improved food security through changed farmer's livelihoods.

The second planning meeting involved scientists and technical staff working in the Sustainable Intensification of Maize-Legume Cropping Systems for Food Security in Eastern and Southern Africa (SIMLESA) project. It was held on 27th January 2011. The team members strategized how they would work effectively and planned for the various activities. The professional participated in the allocation of duties and setting time frames for each activity. The professional emphasized the importance of having several copies of the duty allocation document available to all participants, in order to ease access to information. During participation in the planning meeting of the professional was given a task to critically analyze the questionnaire content. The baseline survey questionnaire for Kenya was long and contained over 100 questions which were to be administered within one day for each farmer. The professional recommended the questionnaire to be divided into different sections and administered at different times to each farmer if resources allowed. This would reduce fatigue of both the farmer and administrator and therefore obtain significant information in the long run. Lengthy questionnaires would lead to the farmer/ respondent giving inaccurate information consequently compromising the quality of data produced.

During the one year attachment period, the professional took an initiative and acquired data analysis skills in SPSS package that is mostly used for survey data and data analysis skills in SAS package which is powerful statistical software designed to provide complete, comprehensive set of tools that can meet the data analysis needs of an entire organization.

4.7 Experiences and lessons learnt

The following are the experiences and lessons learnt by the professional during the one year research attachment at the research institution

- A practical experience in the whole process of a research cycle; data collection, design lay out, participation in research, data management, analysis, reviewing, planning among others
- It is imperative to make a list of all activities to be performed in a given period of time and further indicate deadline for each activity to make work more effective and efficient
- It is important to have a pen and a note book all the time to note down the smart ideas that come up when one least expects
- Always take a break from the busy schedule to refresh the brain.
- Learning is a process that never ends. Individuals should always be ready to learn and never make assumptions of knowing everything.
- Quality information is obtained from the data collected when both the scientists and a professional take adequate time going through the data to understand its structure in order to produce the required output
- The professional acquired skills in how to perform a combined site analysis using Genstat
- Understanding and appreciating the diverse background of different team members leads to effective working conditions

- Developed networks are important in exposing one to available opportunities
- Success is the ability to exploit all available options for a problem
- It is essential to consult widely before providing a solution to a particular problem
- Experienced some difficulties in developing the dissertation since the seminar section research methodology was almost similar to the results and discussion.
- Proof reading one's own documents helps a lot in increasing the quality of the documents
- For scientists and technical staff to have up to date information on their finger tips they need to adopt a continues reading culture

CHAPTER FIVE

5 CONCLUSIONS AND RECOMMENDATIONS

5.1 Conclusions

Quality is an important prerequisite in research and can be achieved if the various processes in a research cycle are executed systematically. Multidisciplinary approach in research is one of the tools which when effectively utilized by scientists and technical staff involved in research, enhances research quality. Current skills always support valuable performance in any research institution. With the rapid development of technology that usually surpasses the type of skills individuals possess, continuous refresher courses become a necessity in order to increase the scientists and technical staff's competence. Scientists should take an initiative to venture into online trainings and include a training component in the proposals developed for funding.

Allocation of adequate time to each research activity in a project avoids rushing in some processes for example data analysis. This would subsequently lead to reduced pressure load usually placed on the statistician during peak seasons. Quality research can be maintained in a research institution through having a functional laboratory, availability and access to information, well-organized feedback mechanisms to data generators, timely availability of funds and integrating APVC in all the developed proposals.

5.2 Recommendations

- Follow up mechanisms to be put in place to assess how effective the trained personnel are sharing the knowledge gained to others as expected of them
- Special trainings to be initiated at the centre level other than waiting for trainings initiated from the headquarters
- With the current diverse changes in research refresher courses are mandatory to those who participate in research
- Computer literacy in technical staff requires to be enhanced
- Establish an effective centralized systematic procedure for data management
- Improve on knowledge sharing through visiting other centers and establishing monthly seminars at the centre level
- Incorporate new emerging issues for examples divisions of regions into counties
- Scientists were encouraged to consult and read extensively in order to make scientifically sound defense and judgment on the proposal, papers and reports developed
- Mentoring young scientists to be incorporated in the research system
- Establish a central system for communication in order to update the managerial staff of all the communication done among scientists through

the internet. The scientists can further copy the communications to the respective managerial staff

6 REFERENCES

Ader, JH, Mellenberg, JJ & Hand JD 2008*, Advising on research methods: A consultant's companion*, Johannes Van Kessel, Huizen, Nertherlands, V 23:40-43.

Bangdiwala, SI, Amarillo, ML, Ughade, S, Rodriguez, MN, Singer, JM, Munoz-Navarro, SR, Komoltri, C, & Cumsille-Garib, JF 2002, *Teaching consultancy through direct experience in research: the approach of the Department of Biostatistics at the University of North Carolina at Chapel Hill*. In B. Phillips (Ed.), *Proceedings of the Sixth International Conference on Teaching Statistics*, Cape Town: International Statistical Institute. CDROM

Beintema, NM & Stads, GJ 2004, *Investing in Sub-Saharan African Agricultural Research: Recent Trends, IFPRI 2020 Conference Brief No. 8*, Website http:// www.asti.cgiar.org (accessed 20th December 2010)

Belle, VG 2008, *Statistical rules of the thumb*, second edition, John Wileys and Sons Inc., Hoboken, New Jersey:217-219

Bloom, D, Canning, D & Chan, K 2005, *Higher Education and Economic Development in Africa*, World Bank, Washington, DC.

Buysse, W 2005,Working paper draft 1: *Towards a conceptual framework for teaching research methods and organizing research support; Experiences from the project 'Capacity strengthening in research methods', working with researchers of NARS in East and Central Africa*, IBS SUSAN 2005 conference. Addis Ababa, Ethiopia, 12-16 December 2005.

Carter, RL, Scheaffer, RL & Marks, RG 1986, "*The Role of Consulting Units in Statistics Departments*, The American Statistician, V 40: 260-264.

Chege, GW & Muray, PK 2004, *Data management and analysis*, in Patel, B.K., Muir-Leresche, K., Coe, K & Hainsworth, S.D(eds), GEAR :*Graduate Environmental and Agricultural Research*, A Revised Edition of The Green Book :*A Guide to Effective and Relevant Graduate Research in Africa*, ISBN 9970-866-00-1 The African Crop Science Society 2004: 192-207

Coe, R 2003. 'Put it in Numbers: Research Methods in Agriculture and Rural Development', in Patel, B. K., Maina, M., Hagmann, J., Woomer, P. L. and Adipala, E. (eds), *Curriculum Development and Transformation in Rural Development and Natural Resource Management*. Forum Working Document No. 7. Proceedings of a Strategy Workshop conducted at the Rockefeller Foundation's Bellagio Center in Italy: 12-16 November, 2001:44-48

Coe, R & Sonya, L 2009, *A Course Guide (MSc in Research Methods)*, https://www.reading.ac.uk/ssc/media/.../A_**Guide**_for_Students.pdf(accessed (*on18th December 2010)*

CTA, 1997, *Strengthening biometry and agricultural research, Summary and recommendations of a CTA/University of Hohenhein workshop, 1996 CTA (Technical Centre for Agricultural and Rural Cooperation),* Wageningen, The Netherlands

Dasu, T & Johnson, T (n.d.) *DQ overview Research in Data Quality (DQ)* http://www.dataquality-research.com (accessed 3rd February 2011)

Derr, J 2000, *Statistical Consulting: A Guide to Effective Communication*, Duxbury& Pacific Grove, CA.

Driskell, JE, Goodwin, GF, Salas, E & O'Shea, PG 2006, What makes a good team player? Personality and team effectiveness, *Group Dynamics: Theory, Research, and Practice*, V 10: 249–271

Eicher, CK 2006, *The evolution of agricultural education and training: Global insights of relevance to Africa.* Department of Agricultural Economics, Michigan State University, East Lansing, Michigan

FARA, 2006, *Agricultural research delivery in Africa: an assessment of the requirements for efficient, effective and productive national agricultural research systems in Africa*. Executive summary and strategic recommendations, Accra, Ghana, Accessed on-line at http://www.fara-africa.org/files/NARS%20Assessment%20exec%20sum.pdf on 19 March 2007, FARA

FAO, 2003, *FAOSTAT*, and Website: http://www.faostat.fao.org/default.htm (accessed 20 December 2010).

Gatali, C and Barihuta, JC (n.d.) Research methods courses at the National University of Rwanda: Overview and evaluation of the current situation in terms of organization and teaching, htpp://www.ruforum.org/sites/default/files/ResearchMethodsResources (accessed 1st December 2010)

Glasgow, NA 1997, *New curriculum for new times, a guide to student-centered, Problem-based learning,* Thousand Oaks, CA, USA - London, UK - New Delhi, India, Corwin Press.

Global Coalition for Africa, 1999, "Promoting Agricultural Productivity and Competitiveness in Sub-Saharan Africa," Paper of the 1999 Meeting of the Economic Committee of the Global Coalition for Africa Nairobi, Kenya. April 9-10, 1999.GCA/EC/N.2/4/1999 (accessed 26th November 2010)

Humphrey, SE, Morgeson, FP, Mannor, MJ 2009, *Developing a Theory of the Strategic Core of Teams: A Role Composition Model of Team Performance*, Journal of Applied Psychology

IMF Government Financial Statistics Yearbook, 1996 in Harsch, "Africa Strives to Revitalize Agriculture" Africa Recovery, United Nations October, 1997 : 11 **http://www.un.org/ecosocdev/geninfo/afrec/vol11no2/overview.htm,** (accessed 3rd Dec 2010)

Jeske, DR, Lesch, SM & Deng, H 2007, *the Merging of Statistics Education, Consulting and Research; a Case Study,* Journal of Statistics Education V 15, www.amstat.org/publications/jse/v15n3/jeske.html, (accessed 15th January 2011)

Johanson, R. & Saint, W 2007, Cultivating knowledge and skills to grow African agriculture Report to the World Bank, June 2007

KARI, 2005a, *KARI at a glance booklet,* KARI Information & Documentation services (IDS), Nairobi, Kenya

KARI, 2005b, Strategic Plan; *Agricultural Innovations for sustainable development; KARI optimal staffing requirement by centre*, annex V: 40, June 2005, KARI Information & Documentation services (IDS), Nairobi, Kenya

KARI, 2005c, Strategic Plan, *Agricultural Innovations for sustainable development; KARI optimal staffing requirements by Cadre as at January 2005*, annex IV: 39, June 2005, KARI Information & Documentation services (IDS), Nairobi, Kenya

KARI, 2009a, *KARI Service Charter*, KARI Information & Documentation services (IDS), Nairobi, Kenya

KARI, 2009b, Human resource strategy 2005-2009; *Research Programs and networks of research centers*, chapter 1: 4, KARI Information & Documentation services (IDS), Nairobi, Kenya

Kirway, TN & Lema, NM 2003, *System approach a training manual*, Volume II, Interpress of Tanzania Ltd, Dar es Salaam

Knowles, SM, Holton, FE, Swanson, AR 1998, *The adult learner: the definitive classic in adult education and human resource development*, 5th Edition, Gulf Pub. Co, Houston

Mfumbusa, B & Mkamwa, T 2005, *Strategies to improve research quality at St. Augustine University of Tanzania (SAUT),*an essay organized by The Great Re Search

Muray, P, Garlick, C & Coe, R 2002, *Research Data Management,* ISBN 92 9059 1463, Kul Graphics Ltd printers, Nairobi, Kenya

Muyekho, FN 2010, *Research Programs and Achievements booklet*, KARI Information & Documentation services (IDS), Nairobi, Kenya

Patel, BK, Muir-Leresche, K, Coe, R & Hainsworth, SD 2004, The Green Book*: A guide to effective graduate research in African agriculture, environment, and rural development*, The African crop science Society, Kampala, Uganda

Pattron, DD 2008, *Research methodology*, http://www.experiment-resources/research.com (accessed 23rd December 2010)

Payne, RW, Baird, DB, Cherry, M, Gilmour, AR, Harding, SA, Kane, AF, Lane, PW, Murray, DA, Soutar, DM, Thompson, R, Todd, AD, Tunnicliffe, G, Wilson, R, W & Welham, SJ 2003, *Genstat®for Windows™ Introduction (7th Edition)*, ISBN-1-904375-08-1, VSN International, Wilkinson House, Jordan Hill Road, Oxford, UK, http://www.vsn-intl.com/

RUFORUM, 2007, *Masters in Research Methodology*, www.ruforum.org (accessed 15th November 2010)

Russell, KG 2001, "*The Teaching of Statistical Consulting*," in Probability, Statistics and Seismology: A Festschrift for David Vere-Jones, :20-26, edited by Dayley, DJ, Applied Probability Trust, Sheffield, UK

Savin-Baden, M 2000, *Problem-based learning in higher education, Untold stories*, Philadelphia, PA, USA, The Society for Research into Higher Education and Open University Press

SSC, 2000, *Case Study No. 6 – Good practice in data management*, University of Reading, UK http://www.reading.ac.uk/ssc. (Accessed on 30th December 2010)

SSC, 2010, *Data Flow*: *Organizing action on Research Methods and Data Management*. http://www.reading.ac.uk/ssc. (Accessed on 30th December 2010)

Susan, JR 2004, '*Analysing the data'*, in Patel, BK, Muir-Leresche, K, Coe, R & Hainsworth, SD (eds), GEAR: Graduate Environmental and Agricultural Research, A Revised Edition of The Green Book: *A Guide to Effective and Relevant Graduate Research in Africa* : 207-220, ISBN 9970-866-00-1 The African Crop Science Society 2004

The World Bank 2002, *Constructing Knowledge Societies: New Challenges for Tertiary Education*. Washington, USA, The World Bank

Usher, JR, Simmonds, DG, Boud, D & Feletti, GI 1997, *Industrial enhancement through problem based Learning*; *The challenge of problem-based learning,* London, UK; Stirling, USA, Kogan : 143- 233.

Ware, M 2008, Peer review: *Benefits, perceptions and alternatives*, Journal of Management Inquiry, V 12: 344–351

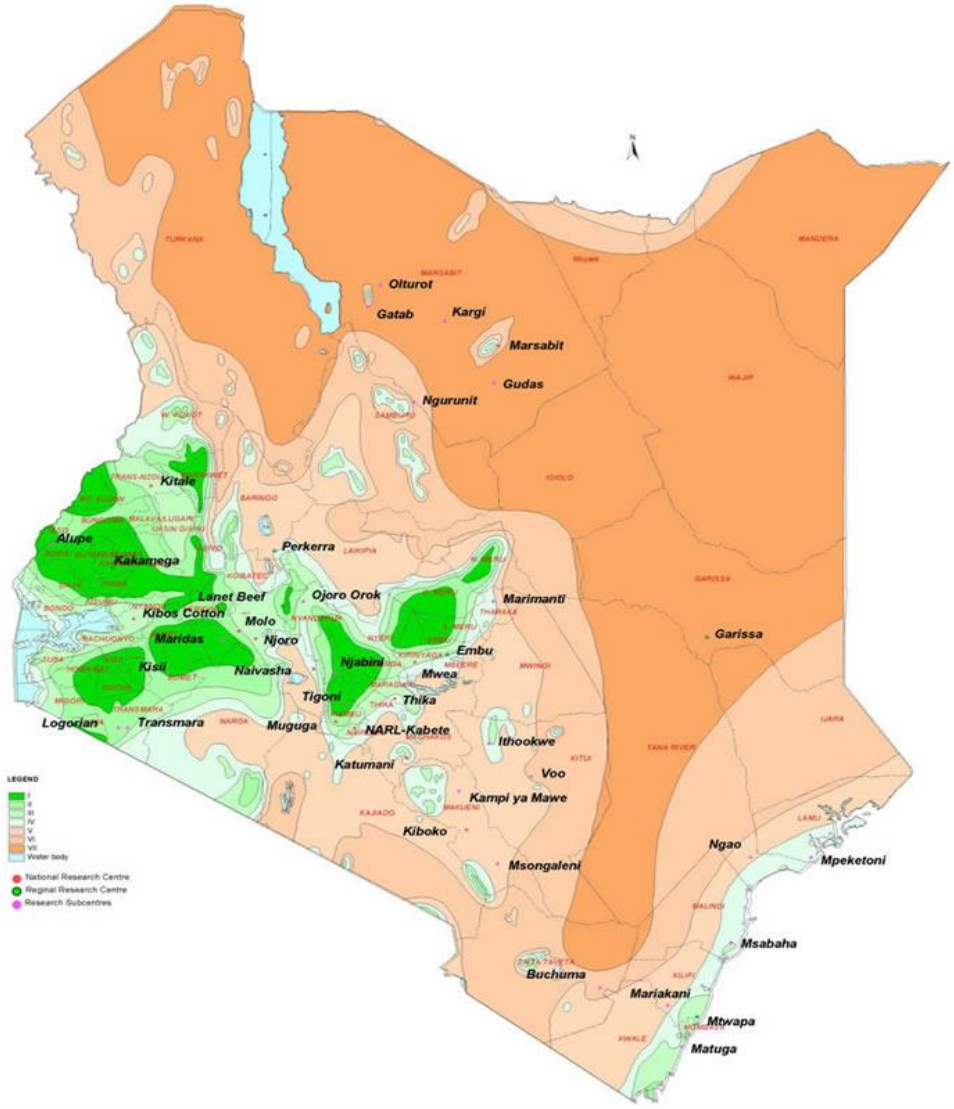
Watson, R & Crawford, M 2003, *Strategic approaches to science and technology development*, Washington, USA, the World Bank

Williams, APO & Woodward, S 1994, *The Competitive Consultant: A Client-Oriented Approach for Achieving Superior Performance*, Macmillan, Basingstoke

World Bank 2003, *World Development Indicators 2003*, World Bank, Washington, DC

7 APPENDICES

7.1 Appendix A: KARI center network map



7.2 Appendix B: Semi-structured needs assessment questionnaire

Assessing the statistical knowledge of scientists at KARI Kakamega

(This information will be confidential)

Section A: Background information

1. Gender □Male □ Female
2. Age □20-30yrs □31-40yrs □41-50yrs □51yrs and above
3. Designation (RO, SRO, PRO….)

   _____
4. Working experience (no of years)

   _____
5. Education level when employed □Certificate □Diploma

   □Bachelors

   □Masters □Doctorate □ other (specify)

_____

6. Highest education level attained □Bachelors □Masters

□Doctorate □Other (specify) _____

7. Have you had any prior training in statistics □Yes □ No

If yes, what was the level of training received? (Multiple choices possible):

□Basic □Intermediate □Advanced □ Specialized

8. What specific skills did you acquire from the training? (Multiple choices possible)

□Data entry □Data processing □Descriptive analysis

□Inferential analysis □Modeling □Other (specify)

_____

Section B: Test your statistical knowledge

9. Who analyses your data?

□Self □Biometrician/Statistician □hired/project consultant

10. Which data analysis software are you familiar with? (Choose more than one)

□Genstat □ R □SAS □Excel □SPSS □STATA □Others specify_____

11. Which analysis do you execute on your data? (Choose more than one)

□Descriptive summaries □ANOVA □Regression □Modeling

□Correlation □Others specify_____

12. Which research designs are you familiar with?

_____

_____

13. Which ones do you use and Why?

_____

_____

14. When should you first consider seeking statistical advice for your research project? (Choose one answer).

□When designing your study

□When collecting data

□When analyzing data

□When applying an analysis technique to support your hypotheses

15. The main contribution of statistical analysis to your research project is (Choose one answer).

□To help you to explain the reasons for variability in your data

□To obtain the *p* values

□To back up your theories with results from an analysis technique

16. Statistics is ...

(Choose one answer)

□An exact science

□A discipline that deals with describing data and making generalizations

□An applied art which mixes computers with graphs to give some nice results

□A bit dodgy, torture numbers and they will confess to anything

17. A researcher has an experiment with 36 pots, where each pot can contain a different number of plants. The height of each plant is measured. If you were asked to organize the measurements in an electronic file for

statistical analysis, how would you arrange the data? You would ... (Choose one answer).

□Put one pot per row; on each row the various plants within the pot in sequential column

□Put one plant per row, repeating the pot number for plants in the same pot

□Put the pot numbers as headings for the columns, and plants on the rows

□Calculate the average plant height by hand and enter the average for each pot

18. The number sequence 1, 2, 3, 4, 5 has a mean of 3 and variance of 2.5. Without using a calculator, what is the mean and variance of the sequence -2, -1, 0, 1, 2? (Choose one answer).

□3 and 2.5 □0 and 2.5 □5 and 3 □0 and 2

19. In hypothesis testing, statistics such as Chi-squared, *t* and *F*, should be _____ in order to reject null hypothesis.

(Choose one answer).

□Large □Small □about 5 □uncorrelated

20. Which of the following statements are valid as a reason why the Normal distribution is important in statistical analysis? (Choose at least one answer).

□It is a distribution followed by many naturally-occurring variables

□Many statistical models are based on assuming the data follow a Normal distribution

□It is a symmetrical distribution

□The sample distribution of the mean approaches a Normal distribution as the sample size increases

21. "Standard errors are more useful than standard deviations for interpreting estimates of population characteristics." This is... Answer:

□True □ False

22. "A 99% confidence interval for a population mean is a range where the mean of the population falls 99 out of 100 times." This is... Answer:

□True □False

23. In practice, you need to know the following probability distributions (Choose one answer).

□When analyzing data, it doesn't really matter what the theory says. We believe in practice!

□The Normal distribution - it comes up a lot

□A small range of distributions relevant to your area and your data

□Every probability distribution

24. A standard error can best be thought of as...

(Choose one answer).

□A measure of precision

□A measure of significance

□A way of seeing your mistakes

□A better measure of spread than the standard deviation

25. You have 225 observations collected using simple random sampling. The mean of your sample is 3 and the standard deviation is 2.5. Without doing a formal hypothesis test, can you conclude that the population mean is very likely to be different from zero? (Choose one answer).

□No, I need to carry out a Chi-square test to answer the question

□ No, I need to carry out a *t*-test to answer the question

□Yes, from the information, I know that the appropriate test will lead me to a conclusion,

□No, I need to carry out an ANOVA test to answer the question, I don't know

26. Increasing the sample size usually reduces the standard deviation.

Answer:

□True □ False

If you were to attend training course in statistics, which topics would you like to be taught? List in order of importance to your needs (from the most to the least important)

_____

_____

_____

Thank you for your time.

Adopted from SSC and modified

7.3 Appendix C: Seminar evaluation form

Date: 10th May 2011 - Data management procedures for quality assurance

	Strongly Agree	**Agree**	**Neutral**	**Disagree**	**Strongly Disagree**	**Not Applicable**
Seminar description was clear						
Location was acceptable						
Seminar objectives were logical and clear						
Seminar content was relevant and organized						
Slides were organized and appropriate						
Presenter was enthusiastic and available						
Presenter was knowledgeable						
Presenter provided clear answers/comments						
Presentation was logical and clear						
Presenter encouraged participation						
Interaction with others was beneficial						
Seminar met my professional objectives						

In what areas was the seminar particularly effective?

In what areas could the seminar use improvement?

Would others you know benefit from this seminar? Yes No

Additional comments and suggestions:

Adopted from SSC

7.4 Appendix D: KARI Kakamega Data management Protocol

This guide is not intended to be definitive but it forms the basis on how the data should be managed. A well written and complete protocol is essential for high quality results; it technically evades problems during the implementation of activities and makes publishing and document retrieval easier. It is hoped that the provided guidelines will aid researchers and all the technical staff dealing with data to plan the data management aspects of their research projects from the onset.

Basic components found in a data management protocol:

Title of activity – This should be very descriptive and concise

Activity leadership – It includes the name of the principal scientist, research assistants, technical staff and all individuals responsible for the activity

Institution (s) responsible for the running of the study – Indicate all the institutions involved

Activity description

- Background and justification of the activity under study – State the type of information expected from the outputs of the activity, justify the decisions about how the activity is designed, explain reason for the study and any other relevant background information
- Objectives – This should be simple, specific, stated in advance and not after the research is done, avoid over ambitious objectives that cannot be adequately achieved within the specified time limit

- Material and methods – This is the most important part and should be written in full details to enable replication. It includes:
    - Location where the activity will be carried out, replications
    - Type of study – Is the study an experiment or survey
    - Study designs – Describe the type of design for example cross sectional study, the design layout, field plan, sample size, description of variables used, size of plots in the farm. Choice of design should be explained in relation to the study objective. Provide a thorough description of all the study design procedures used in a logical format.
    - Materials used – Explain exclusively the treatments, participants or subjects in the study.
    - Responsibilities –State clearly who is responsible for managing the trial, preparing the farm (marking and plot layout), planting, weeding, data collection, supervision and all other activities in the study

The above information has been purposely included for ease of understanding the data and it also serves as Meta data of the data.

PRE-FIELD DATA MANAGEMENT

- **Planning**

  Planning is very important for the logical flow of all the activities to be implemented during the data management process. Planning takes into

account the research objectives, resources and skills available. Document all the planning procedures well and ensure all stakeholders are involved during the planning stage.

- **Designing data entry sheets and collection instruments**

  Appropriate field data entry sheets and collection instruments need to be designed early enough before data collection begins. In survey studies, the collection instruments (questionnaires, checklists among others) should be pretested and restructured if need be. The data entry sheets and instruments should be kept in a safe place. Develop and document a data structure for the data under study.

- **Naming, organizing and managing data files**

  - It will be useful in addressing these issues of naming and organizing data files at an early stage of planning for easy archiving process. Have a sound and clearly defined file or directory naming schemes. Appropriate database system should be used in relation to the type of existing data structure. Describe how the data will be maintained throughout the management cycle
  - Train all staff involved in data collection, entry, processing and archiving
  - Installation of data checking system before the actual data collection commences, example of such kind of system is double data entry.

FIELD DATA MANAGEMENT

- **Data collection**
    - State the type of data to be collected
    - Specify who, how and when the data will be collected
    - Specify what data will be collected at each particular time period
    - Pilot testing all aspects of the data collection especially for the survey study
    - Document all the data collection procedures and instructions
    - Outline the standard measurements to be used in the study and specify clearly the precision (number of decimal places) to which all the measurements are recorded
    - Indicate and document the procedures for data coding, monitoring and verification
- **Data checking**

  Data checking for accuracy, quality and errors is done during data collection (raw data); while the data is being entered in the computer and after the data has been entered. Trained staffs who understand the nature of the data should be involved in the data checking. Data should be collected and recorded carefully. Checks can be incorporated during data collection for example number of plants harvested in the field to verify the quantity harvested and also use of photographs to record the status of a plot or row. Some statistical packages have special facilities installed for data checking

hence can be used. If the software allows one to build in more checks for example yield ranges and yields more than 50 tones/ha circled. Specification can be made of how data checking will be done. After the data entry, checks can be done through simple initial analysis for example extreme values, box plots, scatter plots, tables of summaries, this shows the minimum and maximum values, compares groups of data and displays outliers, trends and means of each treatment.

POST FIELD DATA MANAGEMENT

- **Data entry**
  - Data should be entered in the field after observations are made in appropriate record sheets for example field books, record books or data-logging devices
  - Build facilities in the data collection sheet for recording notes about the plot or farm, this secondary information is usually valuable at the data analysis stage to explain any curious trends of the data
  - Data entry should be done by skilled personnel
  - Data entry should be prompt and complete
  - Data should be entered in its raw form and in the same order as they were collected
  - No hand calculations should be done prior to data entry
  - Clearly state the data entry procedures
  - Exercise minimum transfer of data that brings along transfer errors

- State how data will be computerized in well organized computer files
- Study carefully the structure of data collected in order to identify the appropriate system for data entry. Excel is used when the data set is simple (one level) while a database management systems (DBMS) is used for complex data sets (more than one level). An example of a DBMS is CSPro, Access and EP Info.

- **Data back up**
  - Captured data must be backed up to safeguard against ruin for example loss or burning of existing copies. This can be in form of hard or soft copies that are located in different places.
  - State where and how the data collected is going to be backed up. For example use of excel sheets, database, records or data sheets
  - An agreed strategy for regular file back up to be documented for example establish a routine for making more copies of the data regularly.
- **Data processing**
  - Outline clearly the proposed statistical methods to be used for analysis
  - State clearly the type of software to be used and its ability to perform the type of analysis desired for example summary data transformed to graphics

- Specify the format the data will be organized to facilitate analysis
- Specify which exploratory data analysis to be carried out; descriptive summaries for all variables, tables, graphs and charts
- Prepare an analysis plan that specifies the statistical analysis procedure to be used; determine the response variables, explanatory variables and specify the method of analysis for instance analysis of variance and who will carry out the analysis
- Describe how the data will be examined and statistically analyzed to answer the objectives
- Provide a brief sample size calculation

- **Maintain a data log**

  The master copy of the data needs to be maintained by establishing an audit trail to track changes made on the data. Ensure incoming data and data status of the report is kept for ease tracking. This is important since recollection of activities done early is always poor at a later stage. All entries done should be dated and information on which copy the entries represent (first, second or final copy) be provided. The master copy changes according to the different stages of data management and it is advisable to document and develop a consistent version-numbering system. Final outputs for presentation should be correct, consistent and complete in line with the final version of the master copy of the data.

- **Documentation of Meta data**

  Meta data are data about data and provides specific information that is made available to users in order to improve their understanding of the data. Comprehensive and complete metadata helps users to make informed and full use of data and minimizes the likelihood of misuse. It is documented in hard and soft copies

- **Monitoring and safety**
  - Safety measures need to be put in place early enough
  - The security, privacy and confidentiality of the data should be put in place
  - Establish procedures for data archiving and ensuring data security
  - Develop a flexible data access method to authorized individuals
  - Data custodian needs a list of individuals with authority to access data, when security is an issue as shown in Table A below:

Table A: Authorized individuals to access the institutions information

Name of individuals with access to data collected	Job Title	Organization

- Indicate and agree upon when archiving begins and how it will be maintained
- Establish who owns the data and information generated from it
- Document agreement on all rights to the data between collaborators
- Documents archived should be checked, corrected and of reliable data that can be subjected to scientific scrutiny without raising any doubts in the minds of subsequent researchers
- Archiving takes different forms for example soft and hard copies (including the hand written records)
- Establish a frequent data monitoring schedule

GENERAL/ROUTINE DATA MANAGEMENT

- **Data quality control**
  - Specify quality control procedures and if they have been used before it should be referenced
  - At this stage various checks need to be put in place to ensure the data collected is of high quality and maintained throughout the project phase. Mostly enter the data as soon as it is collected when the mind of the staff is still fresh for remembrance.
  - Indicate which software will be appropriate for data entry, validation, management and archiving
  - Who, how and when the data recording sheets will be prepared
  - Training of all the participants in data collection

- Establish procedures for checking data collection forms for completeness and accuracy
- Setting validation procedures on the computer
- Data entry-who will do and how
- Establish procedures for checking computer entries (manually or double data entry)
- Exploratory data analysis to check for any errors
- Feedback mechanisms especially from manager to data collectors
- Establish procedures for storing raw data files (who, when, where and how)
- Establish procedures for backing up files and updating the master copy of the data
- Describe who is responsible for the appraisal of data quality and how often this will be done
- It is very important for any research institution to have a good policy guideline on quality assurance in data management

- **Responsibility**

  There are various tasks associated with data management process and it is important to allocate responsibilities for each task. This should be agreed upon and hence documented for future use. This helps in tracking the progress of activities.

If there is no data manager, identify persons responsible for the following tasks:

- Data management
- Preparation of data collecting forms
- Designing data entry forms
- Data collection
- Supervision of data collection process
- Checking data after collection
- Data coding
- Entering the data and checking after computerization
- Data ownership and collaboration
- Archiving the data sets and metadata

CONTACT DETAILS FOR PRINCIPAL INVESTIGATOR:

Title: _____

Name: _____

Position: _____

Highest qualification: _____

Full postal address: _____

Postcode: _____

Telephone: _____ Email: _____

Further information can be found in the two references below

The University of Reading Statistical Service Centre, 1998, Data Management Guidelines for Experimental Projects, Biometrics Advisory and Support Services to DFID

I.M. Wilson and S. Abeyasekera, 2006, Writings Research Protocols: A statistical perspective