SOME STATISTICAL CONSIDERATIONS TO ON-FARM TRIALS IN KENYA

Peter M. Njuho

Follow this and additional works at: https://newprairiepress.org/agstatconference

Part of the Agriculture Commons, and the Applied Statistics Commons

This work is licensed under a Creative Commons Attribution-Noncommercial-No Derivative Works 4.0 License.

Recommended Citation

This is brought to you for free and open access by the Conferences at New Prairie Press. It has been accepted for inclusion in Conference on Applied Statistics in Agriculture by an authorized administrator of New Prairie Press. For more information, please contact cads@k-state.edu.
SOME STATISTICAL CONSIDERATIONS TO ON-FARM TRIALS IN KENYA

Peter M Njuho
Biometrics Section
Kenya Agricultural Research Institute, Headquarters
Kaptagat Road, P O Box 57811, Nairobi, KENYA.
E-mail: pnjuho@kari.org

ABSTRACT

The issue of design and analysis of on-farm trials is not clearly understood by agricultural researcher in Kenya. Since on-farm trials require participation and co-operation of the farmers who often differ in education level, chances of collecting unreliable and quite variable data are high. This paper highlights the importance of collecting quality data from on-farm trials, and in particular from the researcher designed and farmer managed trial type. Some complexities associated with the implementation, application of basic statistical principles, and analysis of on-farm trials are discussed. Questions that need to be considered priori to the implementation of any trial implementation and certain intervention measures, which incorporate farmers’ views, are suggested. The importance of making certain technical adjustments to minimize large variation resulting in on-farm trials is also discussed. The dilemma Kenyan researchers face when conducting on-farm trials, where the farmers’ level of education differ greatly are highlighted through examples and some scenarios.

Key Words: On-farm trials, Small-scale farmers, Technology.

1. Introduction

Agriculture is the prevalent economic activity among 80% of the Kenyan population. Thirty one percent of the gross domestic product (GDP) of Kenya originates from the agricultural sector. In Kenya, the majority of the producers are small-scale farmers who account for at least 90% of maize, potatoes and beans production. Most of the formerly mixed large farms have been sub-divided for settlement purpose into smaller units averaging 2 acres per family. These families, most of whom are resources poor, are expected to produce enough food to feed a population growing at 3.8 % per year (KARI Conference Proceedings, 1990). Intensive farming systems that require efficient technologies should provide solutions. Thus, development of well-evaluated technologies that are efficient and sustainable is a priority. To convince farmers to abandon their traditional farming systems in favour of new technologies is difficult unless there are benefits.

For quite sometime, researchers in developing countries assumed an understanding of farmers’ production requirements and conducted research without directly involving the farmers. This
Applied Statistics in Agriculture

approach led to low rate of adoption of the developed technologies. Therefore, some agricultural technologies made little impact on small-scale farmers’ production systems despite the high cost involved in their development (KARI, 1995). The production gap between the farmers’ actual on-farm yields and the predicted yields’ based on on-station results continued to widen. For example, an estimate of the gap between researcher and farmer maize (Zea mays L.) yields in Western Kenya was 2.9 tons per hectare (KARI, 1995).

On-farm trials provide an excellent way of comparing old technology to new technology under development before making a finally recommended for adoption. The importance of conducting on-farm experiments in developing countries cannot be overemphasised (Riley and Alexander, 1997; Coe, 1995; Okali and Farrington, 1994). The participatory approach to research where a farmer becomes a partner in formulation, implementation and evaluation of technologies is gaining ground. In 1987, Kenya Agricultural Research Institute (KARI), integrated a participatory approach to research into its system by creating a regional research division mandated to promote adaptive research (NARP I, 1986).

Both researchers and farmers should benefit from on-farm trial studies. Researchers should benefit by publishing their findings and farmers by getting appropriate technologies. The statistician should study these data to understand the unique circumstances associated with on-farm trials. Normally, data from on-farm trials are inherently variable due to many farm related factors compared to on-station data. For example, in on-farm trials the researcher may give instructions for conducting the trial but the farmer decides to ignore or change the trial because of socio-economic problems; missing values are regularly encountered; farmers differ in level of education, management and income; and some farms though adjacent to each other may be very heterogeneous making it difficult to give a general recommendation to all farmers within the same region. Variability increases with greater non-researcher input, as less control can be achieved. The application of statistical principles that ensure collection of high quality data from on-farm trials and subsequent statistical analysis that leads to appropriate recommendations for sustainable production technology systems should be encouraged.

The objective of this paper is to highlight some basic statistical issues that scientists should consider when conducting on-farm trials. These issues may help our scientists understand how to handle complex data resulting from such trials. It also, tends to dispel the researchers’ notion that statistical principles have a minimal role in the design and analysis of on-farm trials.

2. Complexity in Designing, Implementing and Analysing On-farm trials Data

Most of KARI researchers took formal statistical courses while in college and are very conversant with studies conducted on research stations. However, only a few of them can relate statistical knowledge and skills learned from college to the more complicated design and analysis of on-farm trials. Designing and analyzing on-farm trials requires a broad
interpretation of traditional statistical principles to accommodate the complex factors surrounding the farmers. For example, consider a comparison of an improved and an old feeding technology using two cows on each of ten randomly selected farms. Because of different management, income, labour availability, etc., the ten farms may all differ from each other so much that they cannot be used as true replicates of the improved versus the old practice comparison. Additionally, researchers often erroneously assume that the unbalanced structure of complex participatory studies renders them unsuitable for statistical treatment (Riley, 1997).

Selection of true representative farmers remains a major problem to on-farm trials. The farmers we would like to select are experienced, potential users of the technology, willing to participate, etc. The classical statistical sampling techniques may not apply to such situations. The farmer participatory approach where extension and community leaders select the farms, though not full proof, is commonly applied (Stroud, 1993). A selection approach requiring further investigation is one where farmers from a given village are called to a "baraza" (meeting). At the baraza, with the help of a community leader, the researcher introduces trial modalities to the farmers and explains the type of farmers needed to participate. Then the farmers are asked to select among themselves those who fit the researcher's requirements. It remains unclear whether this approach could be considered random given that no known sampling procedure was applied. Different types of biases such as community, research and logistical are introduced into the process. We should aim at minimising these biases at all cost.

**Examples**

A statistician may provide a good design based on information provided by the researcher, but, in reality, implementation may become difficult. The following examples illustrate some randomization, design, and analysis difficulties encountered by Kenyan researchers when conducting on-farm trials.

**Example 1**

An on-farm trial study compares three different types of irrigation systems: drip, furrow and spot irrigation technologies. Drip irrigation is the new technology and the other two are the old technologies. Some farmers use furrow irrigation and others spot irrigation. Some studies elsewhere indicated drip irrigation to be superior over the others.

Practicability plus the cost of drip irrigation material makes it difficult to randomize the treatments. The drip irrigation requires laying down of water pipes, which are fixed to a bucket (water reservoir). The passing of pipes across plots could interfere with other two irrigation systems thus making it difficult to randomize drip irrigation.

**Example 2**
A researcher designed and farmer managed on-farm trial was conducted to evaluate six bean varieties at three different sites with twenty participating farmers per site. Certain restrictions were imposed on the participating farmers.

1. Three varieties were grown on each farm with two varieties being compulsory.
2. The third variety was selected from the remaining four.
3. Five farmers among the twenty were allowed to choose the same treatment.

An incomplete block design could fit the above requirements. However, the researcher had no idea what kind of design this was yet proceeded to conduct the trial. The trial falls under the category of researcher designed farmer managed type where the farmer is expected to implement the trial and record the data. The researcher remains an observer. In certain circumstance, the farmer consumed the product over time with the researcher recording the final yield at the end. Such data when analysed provide erroneous information on production potential of these varieties. Recording of some data that could be used later in adjustment in the analysis was not done. A simple analysis of variance without deeper consideration would be misleading.

Example 3
Consider an animal feeding trial to compare two feeds against the farmers’ usual feed. The treatments were Lupin +Maize, daily meal and Control. The breeds were Friesian cross, Cross, and Ayrshire cross. The trial involved eleven farmers and lasted for five weeks. The response variable was milk yield measured in kilograms. The first week’s data represented the farmers’ milk production before the treatment and could be used as covariable.

This example demonstrates a typical animal feeding type II on-farm trial. The factors such as age, calving stage, breed, are known to increase variation yet the design could not allow for their control. Additionally, some treatments were not replicated and there were many missing data values, thus, making it difficult to have a meaningful analysis. In this situation, a good design would capture the known within farm variation but maintain simplicity so the farmers can easily conduct the trial while maintaining statistical principles.

Scenarios
The circumstances surrounding most researchers involved in on-farm trials in Kenya, where farmers’ level of education, income, farm sizes and labour availability differ greatly, are highlighted through the following scenarios.

Scenario 1
East Coast Fever (ECF) is a tropical disease associated with more than fifty percent of animal deaths in Eastern Africa. Ticks, which stick on animal bodies, are the disease vectors. A new management strategy to control this disease has been developed in KARI. The strategy involves vaccination followed by dipping of animals in a channel filled with diluted acaricide. This acaricide detaches ticks from the body minimizing infection.

An on-farm study was conducted to investigate on the best treatment combination. The treatment combinations were: vaccination and dipping, no vaccination and dipping, vaccination and no dipping, and no vaccination and no dipping (NVND). The trial was conducted without the scientific control (NVND) treatment because no farmer would agree to apply this important treatment. The factorial structure was incomplete without this treatment.

It was difficult to ensure the independence of these treatments because animals rubbed bodies as they grazed. In addition, some farmers dipped their animals that, according to the design, were not to be dipped. This study was meant to imitate the farmers’ interest in this management practice and it was difficult to impose controls. However, without these controls, it was not a statistical success.

**Scenario 2**

Suppose in a feeding on-farm trial a cross-over design is used to compare a new ration to the farmer’s practice with each participating farmer volunteering only one animal. A farmer provides the first ration to the animal as dictated by the design layout. Regardless of the initial ration, a decrease in milk yield occurs after changing to the second ration. Oftentimes, the farmers revert to the former ration that gave high milk yield. This move would be viewed by an economist as a good indicator of acceptability of the new technology. To a statistician, it is a flawed trial and very little can be done to salvage this data.

**Scenario 3**

In some on-farm variety trials, a farmer may harvest either some or all the plots and then put the yield in a common sack. Chances are high that the farmer did not take yield measurements on plot basis before putting the harvest into a common sack. This data is not usable for statistical analysis.

The above three scenarios illustrate common practices that complicate analysis of on-farm trials in Kenya. The traditional statistical approaches may be easily suggested for use in on-farm trials but their applicability can be difficult. Several non-experimental variables need consideration because they influence results but while fixing variables increases precision, it also reduces the relevance of the trial to “real life” farming practices. Researchers are torn between improving the farming production system of the small-scale farmers and maintaining statistical principles.
3. Some Elemental Considerations to On-farm trials

The basic statistical considerations associated with on-station experiments do apply to on-farm trials. What a “farmer” is and what the study objectives are must be clearly defined when conducting on-farm trials. These elemental statistical considerations are necessary to ensure that efficient and effective on-farm trials are used in generation and verification of technology.

As discussed earlier in this paper, randomization, design, treatments, scientific or practical controls, replications and experimental units are difficulty issues, which require careful scrutiny. There are large between and within farms variations in on-farm trials as suggested by high coefficient of variations recorded from various trials conducted in KARI (KARI, 1995). Replication, plot size and shape are the only elemental statistical components discussed below, in on-farm trials’ context.

3.1 Replications

Replication in on-farm trials is always at two levels: within and between farms. Whenever farms are considered random components, the mean squares term for farm by treatment interaction in analysis of variance is used as the denominator in computing the F-ratio testing for the treatment effects. The degrees of freedom for this interaction is not affected by the number of replications within a farm but does increase as the number of farms is increased. With an increase in degrees of freedom, there is usually an increase in the power of the test; therefore, the number of farms included in the study should be maximized. Whenever farms are considered experimental units, the replications within them are generally sub-samples and not replications. Increasing the number of farms maximises the degrees of freedom for the farm by treatment interaction. However, there should always be at most two replications within a farm to safeguard against losing the whole farm in case of missing data.

The large soil heterogeneity in small-scale farms within the same region can be attributed to different cultural practices and the farming history of each farm. In on-farm trials, this heterogeneity can be taken into consideration in trial planning by using various blocking techniques suggested by Mead (1988).

3.2 Plot Size and Shape

It is difficult to separate plot size and plot shape considerations when conducting on-farm trials. The two main practical considerations in deciding on plot size are ease of management and variability resulting from field irregularities. In general, an increase in plot
size decreases the variability of an experiment by increasing the precision of yield measurement from a single plot. However, the experimental area should not exceed 20% of the farmer’s cropped land (Stroud, 1993). Additionally, mechanical aspects, type of technology, land area, labour, available funds for field trials and good judgement should dictate plot size to use in on-farm trials. For example, use of hand tools is often more compatible with the use of small plots; whereas, larger plots are necessary when the crop is planted, cultivated and harvested with standard farm machinery. Cochran (1957) presented a variance component example that showed long narrow plots are superior to square plots, on the average, even if the direction of the gradient is not known. Therefore, relatively small, long narrow plots are recommended for trials conducted on small-scale farms.

4.0 Consideration Before the Study Begins

To collect high quality data and reliable research results from on-farm trials requires development of a comprehensive plan priori to the start of the trial. In view of the issues discussed earlier in this paper, the using of the following plan will aid the researcher in understanding and developing proper statistical methodologies in designing on-farm trials and analyzing resultant data.

1. Clearly define the objectives of the trial and identify one or two principal response variables to be measured.
2. Carefully select the treatments to provide the most efficient possible answers to the questions prompting the trial. Understand the treatment structure. For instance, if you are not interested in the full factorial set of treatments, then use only those treatments of interest. What are the considerations of having more than one treatment per farm?
3. Clearly, define the control treatment. Should it be the recommended practice or farmers’ own practice or the average practice? Is it a practical control or a scientific control? Is there any justification for having a scientific control?
4. While considering both cost and precision, determine objectively the number of farms selected and number of replications per farm to be used. Sampling methods must take into account any heterogeneity that may exist between and within the farms.
5. Do not arrive at the site with an exact field plan. After visiting the site and identifying sources of variability, decide on the design structure. The farmer’s knowledge about the variation in the farm should be used in determining blocking factors and block size.
6. Develop proper understanding between the farmer and researcher. Characterise socio-economic factors surrounding the farmer. Consideration of the degree of farmer participation is crucial since it could effect variability and likelihood of collecting quality data. Record all those characteristics of the farms and environment expected to have differential effects on the experimental units.
7. List all relevant response variables to be measured. Consider ease of collection.
8. List risks and assumptions associated with the trial. What are the consequences of losing a certain percentage of information?
9. The available human and financial resources as well as the length of the trial should be used as a frame for deciding on the most efficient statistical methodology to use.

We should try to and get away from rigidity. That means, we should allow for re-designing of the trial as time goes on and as more information is collected. For example, sequential designs which allow for intervention as a trial progresses can be applied or one might add an extra factor to an existing design in a convenient way.

5. Technical Adjustments to On-farm Trials

The probability of having results that are more variable or missing data with on-farm trials is higher than with on-station experiments. Data variability could be caused by natural conditions or generated by farmers unknowingly. However, researchers conducting on-farm trials often assume that the farmers will closely follow instructions or how to implement the trial. Based on this assumption both the researcher and the statistician should always be prepared for any eventualities. The instructions given to the farmers must always be compatible with their current practices.

Statisticians are capable of making technical adjustments at the data analysis stage if information on exactly what happened with the trial is provided. Need for technical adjustments can be minimized if the researcher followed these steps.

1. Improve the level of understanding between the farmer and researcher concerning trial modalities. There should be some flexibility during the trial in order to obtain the best information.

2. Consider use of covariables. It is extremely important to record all information about particular influences on the trial. Perhaps this information could be used to adjust for data variation. There should be proper coding of farmer characteristics and activities. Availability of this information to the statistician could allow for use of regression analysis or generalised regression analysis.

3. Have a reasonably large number of farmers involved in the trial. Remember some of them may be dropped later if they are not representative of the target population or if too many data are missing. The expected variability should be estimated and considered before the start of the trial.

The statisticians should assure researchers that regardless of the difficulties and inconsistency in conducting a study, there are methods to extract some information from on-
farm trial data. Thus, data should never be discarded. Still, more information could be extracted if understanding the pre-trial considerations discussed in Section 4 and following the above steps minimize these unexpected outcomes.

6. Highlights on Design and Analysis Issues related to On-Farm Trials

In most cases, researchers tend to think about the experimental design after the data are collected and now need analysis. Such an oversight complicates data analysis and can influence decisions drawn from the analysis. Quite often, the statistician spends time cleaning the data, adjusting for variations unaccounted for by the design and building a model after data collection. Results from studies conducted in such a way often fails to address the original study objectives and researchers find it difficult to publish such work. If researchers clearly understand the experimental design used in the trial, statistical analysis can be very easy. Addressing the following issues should guide researchers in planning their trials.

6.1 Design

Design identification is as important as selecting the treatments for testing in a trial. Therefore, a researcher should always be in a position to explain the experimental design. The treatment structure, design structure, experimental unit, and randomization method used in a trial should be clearly explained in order to devise a proper analysis. Various experimental designs which, with care, can be applied to on-farm trials are discussed in many statistical textbooks (Cochran and Cox, 1957; Mead, 1988; Milliken and Johnson, 1984). The commonly used design structures are completely randomized design, randomized complete block design, Latin square design, and incomplete block design. All these designs have their own characteristics with advantages and disadvantages. The basic design concepts require consideration of blocking, use of covariables, treatment structures and confounding whenever applicable. Design of any experiment involves identification of appropriate experimental units, objectives in form of specific questions and use of single-degree-of-freedom in the analysis where possible.

An unbiased estimate of variance is obtained whenever an appropriate design is used and necessary adjustments are made to treatment means. It is not always possible to obtain an unbiased estimate of variance in on-farm trials. The degree of uncertainty for not detecting a real difference in treatment means would be high in the case of an overestimated variance. Conversely, a chance of detecting a false difference in treatment means exist in case of an underestimated variance. A false difference in treatment means would imply that the new technology is superior to the farmers´ technology, which would not always be the case. On this basis, farmers would abandon their technology in favour of one less desirable. If the new technology were falsely rejected, we would then be rejecting a technology that could eventually improve the production system of the farmer. Knowledge of the experimental
design and statistical analysis before the study began would provide information about the likelihood of these risks.

6.2 Protocol Implementation

Implementation of researcher designed and farmer managed on-farm trial is challenging and requires extra care. The researcher may plan the trial well but things might become unprofessional when the farmer gets involved. A well-planned trial would lead to results that are less biased. It is necessary to discuss with the farmer the problem under investigation and to consider his attitude towards the proposed solutions.

A commitment to implementation of the protocol should be required before a trial is started. Trips to the trial sites should be encouraged to facilitate determining the number of blocks, block orientation, and types of measurements recorded. The coding format of qualitative information should be decided at this stage and made uniform across all farms.

6.3 Data Collection

The need for quality data in statistical analysis is as important as the need for raw materials in a manufacturing industry. Therefore, consideration for the type and method of data collection should be made at the planning stage. Researchers should always have answers to the following questions.

1. Who should collect the data? Is it the researcher, technician, farmer, or all? Why?
2. How often should the data be collected? Daily, weekly, monthly or yearly? Why?
3. Should this data be collected? Is it pertinent to the purpose of the study? Why?
4. Are the data qualitative or quantitative? What is the scale for each measurement?
5. If coding or scoring is used, is it uniform within farm? Between farm?
6. Are measurements made on-farm or on-station? Why?

Data collection is costly and time consuming; therefore, a well thought out format for data collection should be developed before the trial start. It would be disappointing to conduct a long-term trial where inadequate or inappropriate information is collected. If the farmer is the sole collector of data, a simple data recording form should be developed. Frequency of data collection should be minimized to avoid recording errors and use of field notebooks should be encouraged.

Relevant information needs to be recorded. This might include date of harvesting, level of management, date of calving, age of animal, level of education, and acreage. The researcher can collect physical and social on-farm characteristics. However, the farmer should be requested to collect this type of data if it is impossible for the researcher to do so.
Computer facilities are now available and researchers need to be familiar with spreadsheet facilities, such as Microsoft Excel, Lotus 1-2-3, Database, etc. for the data management. The statistical packages such as SAS® and GENSTAT® can easily access data from these application packages.

6.4 Data Analysis

We conduct statistical analysis to extract relevant information from data to make inferences about a target population. Clear definition of the experimental design, type of response variables and how these variables relate to each other, experimental unit, pre-planned treatment comparisons that relate to the trial objectives, assumptions made, fixed and random effects, etc., allow selection of an efficient and effective statistical method for data analysis. In summary, the researcher must develop a protocol addressing design and analysis. Many researchers ignore this aspect of the study plan and have no clear analysis plan for the data. This lack of planning results in the use of various forms of statistical analysis after collecting the data that may not be consistent with the study’s objectives.

The analysis of on-farm trial data, in particular the researcher designed and farmer managed trial, requires very careful consideration as to whether factors are fixed or random effects. The inferences drawn from such analyses differ. Njuho (1997) discusses in detail issues related to possible impact that experimental factors have on analysis of variance when effects are considered either fixed or random. Application of PROC MIXED procedure in the SAS® system is essential in the analysis of on-farm trial data. For example, unavailability of a common control or a common experimental design on each participating farmer should not be a problem when conducting and analyzing on-farm trials (Njuho, 1995) using SAS®.

7. Summary and Conclusions

The traditional statistical procedures used in designing and analyzing on-station experiments are also used for on-farm trials. However, certain non-statistical considerations should be made for on-farm trials in order to minimize the large between and within farm variations, which sometimes have coefficient of variation of over 100%. Statistical principles when properly applied to on-farm trials lead to reliable inferences, which eventually lead to making appropriate recommendations. Statistically, we must ensure that as much information as possible is collected about the pattern and potential causes of variation between trial units.

There is a great need to understand the social and economic factors surrounding the farmers before conducting the trial. As demonstrated by the three scenarios, no amount of data manipulation will result to meaningful solutions from trials when the farmer does not follow the protocol.
Acknowledgements

The author thanks Ms Veronica Taylor, Department of Statistics, Kansas State University for presenting this paper on my behalf and assisting in editing. The views of the anonymous reviewer are appreciated. The director, KARI is thanked for granting permission to present.

References


