

African Crop Science Journal by African Crop Science Society is licensed under a Creative Commons Attribution 3.0 Uganda License. Based on a work at www.ajol.info/ and www.bioline.org.br/cs
DOI: <https://dx.doi.org/10.4314/acsj.v27i3.14>



REVISITING APPLICATION OF STATISTICS IN AGRICULTURAL RESEARCH IN SUB-SAHARAN AFRICA: ENTRY POINTS FOR IMPROVEMENT

T.L. ODONG, J.S. TENYWA and M. NABASIRYE

College of Agricultural and Environmental Sciences, Makerere University, P. O. Box 7062,
Kampala, Uganda

Corresponding authors: thomas.l.odong@gmail.com

(Received 8 May 2019; accepted 13 August 2019)

ABSTRACT

The importance of statistics in empowering the agricultural research process and sharpening interventions cannot be over-emphasized. Undocumented evidence points to misconceptions, misuse or underuse of statistics among agricultural researchers in sub-Saharan Africa (SSA); pointing to the possibility that the subject has been part of the causes the unfulfilled targets in the agricultural sector in the region. The objective of this study was to analyse and document weaknesses in statistical practice in agricultural research, with a view to identifying entry points for strengthening the performance of the sector for SSA to be able to achieve its set goals. A desk study involving 165 research articles published in the African Crop Science Journal over the period of 17 years (2000 to 2017) was conducted through a rigorous SWOT analysis for issues related to the use of statistics in the implementation of agricultural research in SSA. A checklist consisting of key elements related to study design; data collection, analysis and exploitation; and presentation, was used to guide the interrogation. Findings indicated that researchers generally made explicit description of treatment structures that fairly matched the study objectives and hypotheses (in the few cases where they were stated), with a few weaknesses in the description of factorial treatment structure. The Randomised Complete Block Design was most commonly used among the designs, with 3-4 replicates. However, there was hardly any justification for its use, as the blocking factors were never mentioned and thus their role in determining the precision of the results was difficult to determine. Analysis of Variance was the main method for data analysis, followed by correlations. The F-test and the associated P-values were the basis for decisions on treatment differences. Most researchers had problems with presentation and interpretation of P-values and significance level. Post adhoc tests mostly used the Least Significant Difference (LSD) for pairwise mean comparisons, with little consideration for the treatment structure, the number of treatments and the nature (qualitative or quantitative). Generally, estimates of treatment means were presented together with various measures of precision, in both tables and graphical forms. In several cases, LSD was used or misused interchangeably with standard error (SE) or standard error of difference (SED). Several statistical software were used for data analysis and presentation, with the main ones being SAS, Genstat and MSTAT-C. Key entry points for improvement heavily lie in human and infrastructural resource capacity improvement, most specifically in (i) periodic review of university and other tertiary institutions' curricula to provide sufficient time

allocation, physical space and relevant infrastructure for true hands on practice; (ii) more effective utilisation of the few statisticians available in the region, (iii) short term staff in-service retooling courses, (iv) sustained statistical service units wherever necessary, and (v) provision for periodic interactive statistician-researcher platforms (such as conferences and workshops) for sharing notes on challenges and achievements during implementation of their research programmes.

Key Words: Experimental design, P-values, SWOT analysis

RÉSUMÉ

L'importance des statistiques pour renforcer le processus de recherche agricole et affiner les interventions ne saurait être trop soulignée. Des preuves non documentées font ressortir des idées fausses, un usage abusif ou un usage insuffisant des statistiques parmi les chercheurs en agriculture en Afrique subsaharienne (ASS); soulignant la possibilité que le sujet ait fait partie des causes des objectifs non atteints dans le secteur agricole de la région. L'objectif de cette étude était d'analyser et de documenter les faiblesses de la pratique d'utilisation de la statistique en matière de recherche agricole, en vue d'identifier les points d'entrée permettant de renforcer les performances du secteur afin que l'ASS puisse atteindre ses objectifs fixés. Une étude théorique comprenant 165 articles de recherche publiés dans le journal *African Crop Science Journal* dans une période de 17 ans (2000 à 2017) qui ont été réalisées au moyen d'une analyse SWOT rigoureuse des problèmes liés à l'utilisation des statistiques dans la mise en œuvre de la recherche agricole en ASS. Une liste de contrôle comprenant des éléments clés liés à la conception de l'étude; collection, analyse et exploitation de données; et présentation, a été utilisé pour guider l'interrogatoire. Les résultats ont indiqué que les chercheurs avaient généralement décrit de manière explicite les structures de traitement correspondant assez bien aux objectifs et hypothèses de l'étude (dans les cas rares où elles avaient été énoncées), avec quelques faiblesses dans la description de la structure de traitement factoriel. Le dispositif en bloc (fisher) randomisé était le plus couramment utilisé parmi les dispositifs, avec 3 à 4 réplicats. Cependant, son utilisation n'était guère justifiée, car les facteurs de blocage n'étaient jamais mentionnés et leur rôle dans la détermination de la précision des résultats était donc difficile à déterminer. L'analyse de la variance était la principale méthode d'analyse des données, suivie par les corrélations. Le test F et les valeurs P associées constituaient la base des décisions sur les différences de traitements. La plupart des chercheurs ont eu des problèmes de présentation et d'interprétation des valeurs de p et du niveau de signification. Les tests post adhoc ont principalement utilisé la différence la moins significative (DMS) pour les comparaisons moyennes par paires, en prenant peu en compte la structure du traitement, le nombre de traitements et la nature (qualitative ou quantitative). En général, les estimations des moyennes de traitement étaient présentées avec diverses mesures de précision, sous forme de tableaux et de graphiques. Dans plusieurs cas, le DMS a été utilisé ou utilisé de manière interchangeable avec l'erreur standard (ES) ou l'erreur standard de différence (ESD). Plusieurs logiciels statistiques ont été utilisés pour l'analyse et la présentation des données, les principaux étant SAS, Genstat et MSTAT-C. Les points principaux d'amélioration dépendent essentiellement de l'amélioration de la capacité des ressources humaines et infrastructurelles, plus particulièrement de (i) la révision périodique des programmes d'études des universités et autres établissements d'enseignement supérieur afin de prévoir suffisamment de temps, un espace physique et une infrastructure appropriée pour une véritable mise en pratique; (ii) une utilisation plus efficace des quelques statisticiens disponibles dans la région, (iii) des cours de réoutillage en cours à court terme pour le personnel, (iv) des unités de services statistiques durables, le cas échéant, et (v) la mise en place de plates-formes interactives interactives de statisticiens et de chercheurs (conférences et d'ateliers d'écriture) pour partager des notes sur les défis et les réalisations au cours de la mise en œuvre de leurs programmes de recherche.

Mots Clés: Plan expérimental, valeurs P, analyse SWOT

INTRODUCTION

Substantial investment has been made in agricultural research and development to spur productivity of the sector in a bid to bridge the gapping production capacity to growing levels of food and nutrition insecurity, particularly in sub-Saharan Africa where population growth rates are among the highest globally. Regional and national targets have been set (e.g. Millennium development and more recently Sustainable development goals) to prompt and incentivise agricultural players to accelerate agricultural sector growth. Unfortunately, these and many similar efforts have generally led to limited or no meaningful agricultural transformation, as food insecurity and poverty continue to spiral strides ahead of production. One of the principal and yet overly neglected ingredients in research productivity is the effective application of statistics in research processes.

It is usually astonishing to be part of the audience discussing agricultural research findings in sub-Saharan Africa, involving statistics at conferences and in many research reports and publications, with misconceptions of statistics, including misuse and/or underuse during research design, implementation, data analysis and interpretation. Similar concerns have been raised by the Regional Universities Forum for Capacity Building in Agriculture (RUFORUM), especially lack of knowledge in areas of applied Statistics (Biometry) among many scientists in sub-Saharan Africa as a major hindrance to Agricultural research in Sub-Saharan Africa (RUFORUM, 2019). Unfortunately, a formal documentation and analysis to establish the gravity of the matter and establish priority areas to guide the development of strategic interventions are still lacking.

World over, statistics is widely accepted as a powerful tool in the scientific research process and decision making, especially during the current era of “big data”. This is most especially important in agricultural research,

whose goal is to better the performance of physical, social, cultural, economic, institutional, technological, political and psychological attributes among communities, all interacting upon each other. The agricultural sector and its challenges are thus a complex system that cannot be resolved by fractional or isolated approaches, without the intervention of multifaceted functional statistical procedures. Moreover, the usefulness of statistics in sharpening agricultural research interventions heavily depends on the extent to which the researcher adheres to standard principles and practices acceptable within the statistics profession. The objective of this study was to analyse and document status of weaknesses in statistical practice in agricultural research, with a view to identifying entry points for strengthening the performance of the sector for SSA to be able to achieve its set goals.

METHODOLOGY

A desktop study, with the consideration that published materials in African agriculture based journals can be a basis for identifying weaknesses in the use and practice of statistics in implementing agricultural research programmes in sub-Saharan Africa (SSA), was conducted through a SWOT (strengths, weaknesses, opportunities and threats) analysis. The process started with identification of up to ten journals popularly associated with SSA agricultural research publications. The journals were pre-evaluated against a set list of standard criteria, *viz.* lifespan, uninterrupted publication track record, diversity of agricultural sub-disciplines, and content of statistics in their publications among other considerations. Against this evaluation scheme, the African Crop Science Journal which was also advantaged to publish in both English and French languages was selected for this study.

African Crop Science Journal was launched in 1993 with the aim of providing a

quality forum for disseminating quality research results on tropical crop science, and enhancing their availability to researchers and development agencies in Africa and other developing countries (www.ajol.info/). Prior to use of the journal for this study, explicit permission was sought from and was granted by the Editorial Board.

For this study, we only considered original articles published between year 2000 and 2017 inclusive. The study period was for convenience broken down into four sets, namely 2000-2004, 2005-2009, 2010–2014 and 2015–2017. In order for this study to deliver meaningful recommendations, only articles based on field experiments conducted within SSA were considered.

A total of 732 articles were published within the study period (2000-2017), of which only 328 articles met the eligibility criteria described above. A total of 165 articles of the eligible articles were selected for the study to enable capture as much information as possible. The 165 articles were selected using stratified sampling using R statistical software, based on the field of study (Table 1). The number of articles from each field of study was based on proportional allocation. Classification by

field of study was made to ensure minimum overlapping between the groups.

The process of statistical evaluation was based on a checklist developed for the purpose. The key elements of the checklist included evidence of use and perception of statistics by the authors, nature of study objective and related hypotheses, study factors/treatments and their structures, methods of data collection and analysis; and pattern and quality of data interpretation. Literature from elsewhere was consulted to envision similar challenges, learn lessons and contextualise SSA challenges within the global agricultural research framework.

For issues related to experimental design, the thrust was put on the treatment structure, types of experimental design used, number of replications, blocking factors, and clarity of the description of the experimental design. For data analysis, data types and analytical methods were examined, links between methods of analysis and study objectives, and statistical software used.

We also explored the options for presentation of results (illustrations: tables, graphics or narrative texts), types of information presented (e.g. means, ANOVA table, measures of precisions, etc.), and

TABLE 1. Number and distribution of the articles used in the study, clustered by field of study and year of publication

Field of study	Period of publication				Total
	2000-2004	2005-2009	2010-2015	2015-2017	
Agronomy	19	05	00	02	26
Crop improvement	18	10	14	13	55
Genetic diversity	05	04	12	04	25
Pest management	04	02	03	04	13
Physiology	01	00	04	01	06
Plant pathology	05	08	05	06	24
Soil sciences	07	03	02	00	12
Weed management	01	03	00	00	04
Grand total	60	35	40	30	165

appropriateness of the presentation in relation to study objectives/hypothesis and data types. All these were considered because they heavily bear on the ease and objectivity of interpretations of results in the context of the research problem in question. We used descriptive statistics (frequencies/percentages) to summarise the results. In some cases the responses were grouped into key thematic areas and summarised in tables and/or boxes to achieve the desired clarity.

FINDINGS AND DISCUSSION

Experimental design

Treatment and treatment structure. Most of the authors (89.8%) described the treatments and their structures explicitly. Succinct description of treatment structure not only makes it easier for other researchers to adopt relevant study procedures, but also makes it easier to link treatments and their structures to study objectives and to the emerging findings. Explicit description of treatment structure also facilitates Meta-analysis studies that are becoming very common in various study fields, including in agriculture. However, a few weaknesses in the description of treatment structure were observed (Box 1).

The deficiency in proper description of treatment structure, in some cases, was caused by the bulk of information loaded in one article. For example, in one of articles, up to six experiments were reported making it very difficult for all of them to have their structure described properly. This also made it extremely hard to follow the story line and link treatments and their structures to study objectives. Up to 5.8% of the articles showed lack of linkage between treatment structures and study objectives.

Types of experimental design used. Majority (78.4%) of authors used the Randomised Complete Block Design (RCBD), with or without split plot treatment arrangement. About 87% of the researchers used designs with blocking (RCBD, Split plot, Incomplete Block Design) (Table 2). This is not surprising because for field experiments, it is often difficult to obtain homogeneous experimental units. Furthermore, about 5% of the studies used Completely Randomised Design (CRD) despite the studies being field experiments. However, given that these are all field experiments, it is very difficult to justify the use of CRD.

Adequacy of the experimental designs. Majority (75%) of the studies were based on

Box 1. Weaknesses in the description of treatment structure observed from the reviewed research done in sub-Saharan Africa

- (i) Very brief description of treatment structure
- (ii) Choices of the treatment levels not properly justified
- (iii) Change of treatment structure from one season to another without justification
- (iv) Failure to identify and properly describe factorial treatment structure.

TABLE 2. Typical experimental designs used in agricultural research in sub-Saharan Africa

Experimental design	Cases (%) *
RCBD	64.0
RCBD split plot	14.4
Alpha lattices	8.1
CRD	5.4
Incomplete block	0.9

*About 7% of the articles did not mention the experimental design used

appropriate experimental designs. For field experiments the use of RCBD or other designs involving blocking in most cases is often sufficient. Nelson and Rawlings (1983) pointed out that because of its simplicity and efficiency, RCBD is the most popular design in agronomic research. For the rest of the materials used in the present study (25%) the designs used were either inappropriate (8.7%) or the narrative could not permit evaluation due to inadequate information provided (16.4%). Deficiencies in the experimental design used by researchers in SSA are presented in Box 2.

It was quite challenging to determine the appropriateness of the design used in some

cases due to poor description of the experimental design. For example, it was difficult to determine the appropriateness of blocking in cases where the blocking factors and plot sizes were not provided. In fact, up to 95% of the cases in which blocking was involved, the blocking factors were not articulated. Furthermore, about 12% of the articles did not report plot sizes. Information related to blocking factors and plot sizes is required to determine the suitability of the study design which will enable readers to determine validity of the results coming out of an experiment (Maindonald, 1992). Maindonald (1992) pointed out that description of the design as a "complete randomised design" or as "randomised block design" is insufficient, without the necessary narratives. Proper description of experimental designs and analytical procedures is essential for proper interpretation of results reported in the paper (Nelson and Rawlings, 1983).

Although blocking increases precision in an experiment, improper blocking usually inflates experimental error and results into wrong conclusions. The way out of improper blocking is to examine past performance in the field, as visual examination of the field may not reveal underlying spatial variability (other

Box 2. Weaknesses in use and description of experimental designs noted in field experimental processes in agricultural research in sub-Saharan Africa

- (i) Experiments not replicated or pseudo replicated
- (ii) Use of CRD in field experimentation without due justification
- (iii) Poor choice of factors for whole and subplots in split-plot arrangements (e.g. with only two levels for main plot factor, split plot design will result in very low precision even with four replications. In one case, it would have been better if harvest time were put on main plot for ease of operation)
- (iv) Using RCBD when handling several treatments with large plot sizes. For example with 89 varieties and plot size of 4.5 m x 5.0 m, it is hard to get homogenous blocks of size 2002.5 m² plus areas between plots

than slope) (Bowman (2001). Our experiences attest to the fact that researchers in SSA rarely use field history as input in designing their field experiments. The main blocking factor often used is slope. It is imperative that they incorporate information relating to previous use.

Number of replications. The number of replications observed in the present study ranged from one (no replication) to six. Majority (56.1%) of studies reported three replications, followed by four replications (20.6%) and two replications (9.4%). Three studies reported one replicate (one of which indicated pseudo replication as replicate). About 9% of the reviewed studies did not disclose the number of replications used. None of the studies provided justification for the number of replications used. Based on the number of treatment combinations and the number of replications, some articles had very few (3, 4 and 8) error degrees of freedom (Table 3), signifying low level of precision and statistical power.

The number of replicates has a direct, highly predictable, repeatable and tangible effect on precision and the ability to detect differences among treatments (Raudonius, 2017) and Casler (2015). According to Casler (2015), if all the other aspects of the experiment (design and execution) were done properly, the failure to detect differences among treatment means will most likely be due insufficient and/or improper scale of replication. Experiences show that most researchers arbitrarily choose three or four as the number of replications since those are the most frequently used numbers

in previous studies. Statistical packages such as GenStat have facilities for determining the number of replications and statistical powers but are rarely used by researchers.

Sampling for data collection. More than 60% of reviewed studies did not describe the procedures for sampling during data collection. For studies which articulated sampling methods, there were variations in the approaches used. Sampling approaches used in the reviewed studies included (i) selecting plants randomly from the whole plot, (ii) selecting plants randomly from the middle/central rows, (iii) making observations on all plants in the middle or central rows, (iv) sampling along a diagonal of the plot, (v) sampling an area of a plot, and (vi) making observations on all the plants within the plot. The number of plants sampled was various but ranged 2 to 50 plants. Understanding how sampling is done during data collection is important in determining appropriate error term for hypothesis testing or measures of precision associated with different estimates. It should be noted that when sampling is done within the plots, then in addition to the usual experimental errors there is also sampling error which needs to be taken into consideration when doing data analysis. As pointed out by Nelson and Rawling (1983), unless different levels of samplings are included in the linear model, there is a high risk of using the wrong error term in the ANOVA. What is striking is that despite the variation in the sampling approaches used, these variations were not considered in model specification for data analysis.

TABLE 3. Number of field studies in sub-Saharan agricultural research, with less than 10 error degrees of freedom

Number of articles	Number of treatment	Number of replications	Error degrees of freedom
1	4	2	3
3	3	3	4
3	5	3	8

Data analysis

Statistical analysis methods. Majority (82.9%) of the studies had analysis of variance (ANOVA) as the main method for data analysis; followed by correlations (12.6%). Plant breeding and diversity studies (12.6%) proceeded from ANOVA, to estimation of genetic parameters (heritability, combining ability, genetic correlation, and genetic variance component). Other methods of statistical analyses included descriptive statistics, Regression, Chi-square, Z-test, t-test, AMMI, MANOVA, Path analysis, GLM, Mixed model/REML, GGE/HA-GGE, PCA, Generation mean analysis, Contingency test and Mann-Whitney test. It is clear from the present study that the used advanced methods of data analysis such as mixed model/REML, is still minimal among agricultural research scientists in SSA. The nearly universal use of ANOVA by agricultural researchers is not surprising, given that it is one of the areas covered in most courses offered to agriculture students in most universities in SSA. Although ANOVA is a powerful technique for understanding variation patterns in data, excessive dependence on it can be a handicap to researchers (Nelson and Rawlings, 1983) since it acts as a deterrent for trying other analysis methods. Researchers should be encouraged to explore other options, such mixed models which handle more complex scenarios where ANOVA assumptions are hardly applicable.

Statistical packages used for data analysis.

About 24% of reviewed studies did not mention statistical packages used for data analysis in their research. The three main statistical packages used included SAS (29.7%), GenStat (27.0%) and MStatC (10.8%). The remaining articles (about 10%) used statistical packages that included Statistical Package for the Social Sciences (SPSS), Breeding View, GGE Biplot, Minitab, R, PB Tools, XLStats, Agrobases Generation II and Statgraphic. However, most studies lack

appropriate description and citation of the package used. According to Arafin *et al.* (2016), computer programmes used for statistical analysis should be described, specifically, the name of the programme, the version as well as the specific add-on packages to enable other to repeat the analysis. However, it should be noted that proper description of statistical analysis package/programme is not a substitute for proper description data analysis method itself. As pointed out by Kramer *et al.* (2016), authors and reviewers should recognise that statistical software is a means of implementing statistical analysis, not a statistical method in itself. Our experiences indicate that the choice of statistical packages is determined mainly by previous researcher training and what is available within the research institution. We found it rather strange that a researcher uses GenStat to perform ANOVA and then goes on to perform correlation coefficient using SAS. This might also be an indication that what is described in the methodology may actually be different from what is practiced by researchers. This could also be an indication of “cut and paste” of methodology sourced from elsewhere.

Data analysis methods. One of the weaknesses identified in this review is poor description of data analysis methods. A number of reviewed studies (26%) had none, skeletal or uninformative descriptions of data analysis methods. The most common phrases used in the reports included “Statistical analysis was done using SPSS and MStatC”, “Data was subjected to statistical analysis”, “Data was subjected to standard ANOVA”, “Data was subjected to a t-test or ANOVA as appropriate”, “Hierarchical ANOVA as done”, “Yield data subjected to ANOVA to generate treatment means, determine significant differences between means”, “Data analysis involving analysis of variance was performed”. Some analyses were not described in the methodology section, and yet results from such analysis were presented in the results

section and further discussed in the articles. Raudonius (2017) made a similar observation that insufficient information is often provided on how the statistical analyses were performed. Statistical/data analysis methods should be described with enough detail to enable a knowledgeable reader with access to original data to verify the reported results (Mairdonald, 1992; Lang and Altman, 2015; Arafin *et al.*, 2016; Kramer *et al.*, 2016).

Assumptions of ANOVA and data transformation. Analysis of Variance (ANOVA) is often the main tool of data analysis for experimental work. However, there are assumptions that need to be met for validity of results from ANOVA. Of the majority of the reviewed studies that used ANOVA, only 22.8% indicated that they checked for assumptions of analysis of variance. Without checking the assumptions underlying a statistical method, one is prone to performing incorrect analysis. According to Piepho and Edmonson (2018), a good statistical analysis addresses the aims of the experiment and produces accurate and efficient information that provides a simple and parsimonious description of the results, while taking a proper account of the error structure of the design. Less than one third (28.6%) of the studies that checked ANOVA assumption, performed data transformation to correct for the assumption of ANOVA. The types of transformation performed included square root transformation, Arcsine and log. One article did not specify the type of transformations. Although the justification for performing transformation centred on the usual assumption of the analysis of variance, some authors stated them inappropriately. Examples of justification for data transformation included “Conform to ANOVA assumption, constant variance”, “Correct for normality”, “Correction for requirement of normal distribution”, “Correction for heterogeneity of variance” and “Normalised the situation”. The variables subjected to transformation included

percentages, counts and scores (McDonald, 2014).

The use of approaches such as generalised linear models, mixed models and non-parametric statistics to handle violations of the assumptions of traditional ANOVA was rare in the reviewed studies. This is not surprising given that most statistical courses offered to students of agriculture are elementary in nature. We scanned through several course outlines of statistics courses offered to both undergraduate and postgraduate students within East and Southern Africa and noted little or no mention of mixed and generalised linear models. These approaches need to be popularised among scientists in sub-Saharan Africa either through short courses or incorporating them in the curriculum.

Post ANOVA/Post Adhoc test. Of the studies that used ANOVA, 73% applied post adhoc test. The majority (82.1%) of the studies that applied post adhoc test used Least Significant Difference (LSD); followed by SNK (7.46%), Tukey (2.99%) and Contrast (2.99%). Dunnet test, Bonferroni test and t-test were each used in one reviewed case. The challenge of reporting of LSD was noticed, especially in studies with factorial treatment structure. Some researchers reported LSD for main effects instead of the one for interaction (two-way table); or reported LSD even when F-test does not show significance; yet they claimed to be using Fisher’s protected LSD. In the cases where significant interactions are detected, two-way or multiple-way tables or figures then associated LSDs should be presented; whereas where no significant interaction is detected, one-way tables should be presented with appropriate LSD.

The present study identified cases in which planned contrasts (comparison or polynomial contrasts) should have been more appropriate than LSD. For examples there were number of studies with quantitative treatment levels for which polynomial contrast would have been more appropriate. As pointed out by Piepho

and Edmondson (2018), unstructured analysis of quantitative levels of treatments is uninformative about any possible underlying model of treatment factor effects.

There were a number of studies with objectives and description of treatment structure clearly suggesting planned comparison, for which comparison contrast would have been the most suitable post adhoc analysis method. Also, there was evidence of erroneous use of LSDs when making many pairwise comparisons of treatments. The use LSD values for making comparison of more than treatment degrees of freedom has been discouraged due to increase in type I error (Kuehl, 2000). Kramer *et al.* (2016) analysed the problem of statistical use in a horticulture journal and identified incorrect means separation procedures as one of the statistical problems common among articles published. Because the choice of method of treatment comparisons affects the conclusions about treatments, researchers must be explicit about what mean separation method is used and provide the rationale for its choice.

REPEATED MEASURES AND MULTIVARIATE ANALYSIS

Although 25% of the reviewed study materials indicated that repeated measurement was employed during the data collection, none of the articles performed analysis of repeated measures. Successive measurements on the same plot (more specifically on same plant) are likely to be serially correlated, and this means that for a reliable and a more efficient analysis of repeated measures data, proper account of the serial correlations should be taken between repeated measurements (Piepho *et al.*, 2004). Kramer *et al.* (2016) identified inappropriate analysis of data from multiple dependent variables, on the same unit of observations, as the most common problem in the use of statistics in horticultural research.

Multivariate analysis methods such as Principal Component Analysis are better suited

for handling correlated variables. From the few reviewed cases in the present study, there is a clear indication of lack of knowledge in the use of multivariate analysis methods. When correlation among measured variables is disregarded in statistical analysis, important information among dependent variables is lost (Sainani, 2010). In majority of the reviewed research materials reviewed, more than one variable were measured on the same plot or plants, but they were often analysed independently without considering the dependency between them. Given the multivariate nature of data collected from agricultural experiments, the use of multivariate analysis methods needs to be encouraged among agricultural researchers in the region to be able to make correct inferences.

DATA REPORTING AND INTERPRETATION

Significant levels. Results reporting in general and reporting of significance level in particular was identified as one of the main shortcomings of the studies reviewed. In fact 37% of the studies reviewed did not correctly articulate the significance levels. Generally, they were characterised by incomplete reporting of significance with emphasis put only on the P-values or significance level. The traditional reporting of significance level necessitated inclusion of test statistics, degrees of freedom and exact P-value (e.g., $F(2, 24)=1.3, P=0.25$ instead $P>0.05$). In most articles, exact P-values were not reported, but P-values were reported as inequality (e.g. $P>0.05$ instead of $P=0.07$ or $P=0.7$). In some articles the P-values were presented in tables but not referred to in the main text. Box 3 presents weakness in the presentation of significance levels, observed in the reviewed published research materials.

According to Lang and Altman (2015), the problem of poor statistical reporting is long-standing, wide spread, potentially serious, concerns mostly basic statistics, and yet is

Box 3. Weaknesses in the presentation of “levels of significance” observed in reviewed research articles

- (a) Wrong presentation/interpretation of P-values and significance level. In one reviewed article significance levels were quoted as 0.5 and 0.1 instead of 0.05 and 0.01 respectively. Another article reported that “treatment difference were significant ($P>0.05$)”. Other articles used inappropriate phrases such as “significance of mean squares” when reporting the significance level.
- (b) Reporting of the significance levels did not take into consideration the factorial treatment structure. The significance of the interaction and main effects were not report correctly.
- (c) There was mix up in the presentation, what was presented in the Table was different from what was reported in the text. In some cases, reporting of significance level was mixed up with mean separation

largely unsuspected by most readers of the biomedical literature. Dransfield and Brightwell (2012) also alluded to misuse of statistics in the field of medicine; while Colquhoun (2014) dealt with grave issues with misinterpretation of P-values.

Although the misuse of P-values and/or significance level has been reported extensively, its use in agricultural research has not raised much controversy. The decision by Basic and Applied Social Psychology in 2005 to ban the use of P-values in articles appearing in the journal, prompted the American Statistical Association (Wasserstein and Lazar, 2016) to publish a policy statement on the use of P-values in research publication. In our opinion, P-value remains an important statistical tool in agricultural research, but should be used appropriately. Journals need to come up with more elaborate guidelines on presentation of results to ensure uniformity and eliminate inappropriate presentation.

Data presentation format. Majority (97.11%) the authors presented results in tabular form, with 58.65 percent using tables only while 38.46% used both tables and other forms of data illustrations. Only 3% used figures alone. Otherwise, it is our considered opinion that the most suitable data illustration

should be one that impacts the greatest and fastest impression of the effect of the treatments/factors under study, to the readership. In other words, an effective data illustration should be one that substitutes for bulk of the text that would otherwise be required to provide a vivid impression about the effects of the treatments. Key challenges observed with tabular and figure presentations are presented in Box 4.

From this study, it is clear that most authors lack the capacity to envision the most relevant datasets that would be sufficient to deliver a precise and news worthy research message just sufficient for a single journal publication. Accordingly, authors are faced with the difficulty of fitting the bulk of results achieved by a research project into the limited space allowable by a journal. These point to the need for capacity strengthening in scientific writing skills, and writing for peer reviewed journals in particular. There were tables that deserved splitting into two or more tables to achieve the required clarity.

Several authors especially in areas of plant breeding presented detailed analysis of variance table, even including formula for calculating expected sum of squares. However, most of the information presented in the ANOVA table

Box 4. Challenges met by sub-Saharan African agricultural researchers with use of data illustrations for presenting research results

- (a) Poor choices of table types (one-way or multi-way) to present especially data from factorial experiments. There were cases where one-way table were appropriate (when only main effect were significant), but the authors presented two or three-way tables, and in other cases two or three way tables were presented instead of one-way. The significance of the different effects (main effects and interactions) was not used to inform the tabular and graphical presentation of results in 44.1% of the reviewed articles.
- (b) Presenting means for several variables in the same table, including variables whose means were declared not significantly different by F-test, made understanding the table difficult. Crowding multiple columns and rows in a single table neither saves time, space nor simplifies datasets for most readers. In fact in most cases, columns of technically unrelated parameters are paraded together in a single table, at the expense of smooth chronological flow of the text threading up through related parameters. Plant breeders were the most frequently haunted culprits of this vice.
- (c) Text and table presentations were not synchronised. Sometimes what were presented in the table or figure was different from what was presented in their associated narrative texts.
- (d) Poor arrangement of values in tables making comparison very difficult and misleading. This included variations in decimal places even within a single parameter dataset
- (e) Using unsuited illustrations for representing continuous and discrete variables; for instance bar graphs for plant height and line graphs for number of oranges
- (f) Using main factor effects mean separation tools on illustrations intended to display interaction effects
- (g) Tables to present data that would better be presented by line graph in order to show trends
- (h) Tables containing information that are not referred to the text. Most often information such as degrees of freedom, sum of squares, mean sum of squares and formulae for expected mean sum of squares, F-values and P-values were presented without reference made to them in the narrative text. In fact, 41% of the reviewed articles did not make reference to some of the information presented in tables or figures.

was not alluded to in the discussions; thus making such data tables redundant. As much as the actual values of sum of squares or mean sum of squares are important for comparing the different sources of variations in plant breeding, reference should be made to them in the discussion to justify their presence in the data illustration.

Most authors reported the values of coefficient of variation (%), and Least Significant Differences (LSD), but reference was never made to them in the narrative texts. Although it is important for the authors to present as much information as possible, the authors need to articulate the relevance of each piece of information before including it in the table or figure. Crowding bar graphs with several standard errors of means often acts as a distraction to readers rather than providing useful information.

Overall, the research reports evaluated suggested a need for improvement with regard to statistical analysis and results presentation. The improvements ranged from simple removal of redundant information to major re-organisation of result sections.

Measures of precision and data quality. The most common measures of precisions used by authors in this study included the Least Significant Differences (LSD) (46.2%), standard error (SE) (30.8%) and standard error of difference (SED) (5.1%). About Twenty one percent of the reviewed articles used LSD with either SE (16.7%) or SED (3.8%) as measures of precision. The use of more than one measure of precision creates redundancy and should be avoided. However, it should be noted that most authors used LSD for hypothesis testing rather than as a measure of precision. Twenty four percent of reviewed materials did not provide any measure of precision, except CV% that measures the quality of data/data generating process rather than precision (Bowman, 2001). In the present study, this statistic (CV%) appeared in tables (ANOVA, means), just because statistical

software provided it as part of the ANOVA outputs; as such it was not referred to elsewhere in the text.

Overall, there is indecision among researchers as to whether to use SE or SED as measures of precisions when presenting results from experiments. According to Stern *et al.* (2004), standard error (or confidence interval) is more appropriate as a measure of precision for a survey designed to estimate population parameters; whereas standard error of difference is suitable for experiments designed to compare treatments. Given that of most the experiments in agriculture are intended for comparison of treatments, the use of SE was found to be rather inappropriate. The standard error of difference would, therefore, be the most appropriate measures of precision for results from experiments. We also noted that the standard error of means provided in most articles were based on individual treatment variances, rather than the pool variances from the analysis of variance. The analysis of variance assumes equal variances, thus justifying the use of pooled variance. As much as it is important to present table or figure of means with appropriate measures of precision, care should be taken that additional information such as measures of precision provided should not lead to crowding of the tables or graphs and distracting the readers from picking the key message being communicated. A case in question is the CV%, which is often presented among results without making reference to it.

Synthesis and proposed entry points for improvement. The strong points identified from the reviewed articles include: explicit description of treatment structure, treatments studied adequately addressed the study objectives and the results were properly used to support the argument or discussion. Majority of the authors described well how the studied variables were measured. Key areas of weakness identified include: Poor choice and description of experimental design,

description of data processing, data analysis and presentation of results. As pointed out by Kramer *et al.* (2016), the possible reasons for the weakness identified include: i) Researchers being unable to cope with changes in both agricultural and statistical sciences ii) inadequate interaction between researchers and statisticians/biometricians iii) Inadequate teaching/learning of statistics/biometry at the universities and iv) weakness in the journal review process.

The Findings from this study are not only essential for direct improvement agricultural research at implementation level but also at training level where the basic principles of statistics are imparted. Like any other field, the agricultural research landscape has been changing over time and researchers need to keep abreast with those changes. This review also provided us with opportunity to determine whether agricultural researchers are also coping up with the changes in the field of applied statistics/biometry, which is a key tool for agricultural research. Statistics/Biometry has also been changing drastically over the years especially advances in computer sciences. With increased computing power it is now easy to run most software on laptops and not mainframe computers. Our focus on the weaknesses of the articles does not in anyway indicate a negative bias towards the articles published in African Crop Science Journal or SSA researchers. In general, most of the reviewed articles articulated the main messages from the study clearly with exceptions of few cases. Several articles were excellently written and can be used by journal as samples to point out how a good paper should be like. Nevertheless, considering that the articles published by African Crop Science Journal went through a rigorous review process and statistical ineptness was partly the reasons for the high rejection rate of up to 90%, the problems to address related to suboptimal statistical practice in the agricultural research sub-sector are huge, definitely greater

than African Crop Science Journal is able to display.

This paper raises an infinite range of concerns for improvement of the use and practice of statistics for achieving effective agricultural research objectives. In our opinion, the major concerns identified relate to human and infrastructural capacity shortfalls. Here, we highlight proposed key entry points to the most critical areas to spur effective operations of this vital sector in SSA.

POLICY RELATED RECOMMENDATIONS

Universities and tertiary institutions

- (a) There is need for periodic review and updating of agriculture related training curricula for universities and tertiary education institutions to cater for the “non-traditional areas” and more recent advances in statistics, such as mixed models, generalised linear models, multivariate analysis, non-parametric statistics
- (b) Pre-requisite courses or bridging statistics courses are necessary in training programmes involving regional inter-university students mobility/exchange to even out the diversity of backgrounds that often drag effective implementation of training programmes at targeted universities
- (c) Provide sufficient time, infrastructure (space, appropriate software and hardware) and human resource to cater for sufficient hands-on practical skills to the increasingly large number of students within relevant SSA institutions. Use of “*Problem based approach*” delivery methods, including use of pertinent case studies, is recommended
- (d) Provide in-service retooling in applied statistics, for graduate student supervisors. These could be implemented as workshops or short courses particularly during off-active-semester periods

Research institutions

- (a) Provide in-service retooling in applied statistics, for active researchers through short courses, particularly during off cropping seasons.
- (b) Utilise statisticians in the development and implementation of group or individual research projects
- (c) Statistics should be prioritised in the mainstream of agricultural research, by establishing strong support units with full human resource capacities and sustainable funding streams; emulate CGIAR Centers
- (d) Support exclusive national or regional platforms (such as annual conferences and seminars) for sharing practices and predicaments of use of statistics in agricultural research. Such forums will provide the opportunity for an interface between statisticians and routine researchers within the SSA region

2. Statisticians/Biometricians

- (a) Establish a forum for practicing statisticians/biometricians to compare notes and streamline service delivery to clients within the agricultural sector in SSA.
- (b) Leverage from the limited statisticians available in SSA through promoting programmes like staff exchange, staff mobility, such as the Intra-ACP/Intra-Africa operationalised by the European Union and regional initiatives such as those by Regional Forum for Capacity Building in Agriculture (RUFORUM)
- (c) Statisticians ought to be proactive and more innovative in responding to calls for research proposals, rather than waiting for invitations from other research partnering teams. Several concerns raised in this paper can be packaged into substantive research projects centred directly on statistics principles and practices

3. African journals

- (a) Provide an elaborate guide to authors and reviewers with regards to description of critical statistical areas such as treatments and design, methods of data collection and analysis (including correct software)
- (b) Periodically invite review articles with a thrust on statistics so as to obtain independent statistician's view
- (c) Periodically invite articles targeting different aspects on use of statistics, especially highlighting advances in statistics
- (d) Consider including at least one statistician on the journal's editorial or advisory board
- (e) Include statisticians in the review process/editorial board systems

ACKNOWLEDGEMENT

The authors acknowledge the permission granted by the Editorial Board of the African Crop Science Journal to use their journal as the study specimen. The authenticity of the information contained in this article is the responsibility of the authors and not African Crop Science Journal.

REFERENCES

Arifin, W.N., Sarimah, A., Norsa'adah, B., Najib Majdi, Y., Siti-Azrin, A.H. and Kamarul Imran, M. 2016. Reporting statistical results in medical journals. *Malaysian Journal of Medical Sciences* 23:1-7.

Bowman, D.T. 2001. Common use of the CV: A statistical aberration in crop performance trials. *The Journal of Cotton Science* 5:137-141.

Colquhoun, D. 2014. An investigation of the false discovery rate and the misinterpretation of P-values. *Royal Society Open Science*. 20141:140216. DOI: 10.1098/rsos.140216.

- Dransfield, R.D. and Brightwell, R. 2012. Statistical mistakes in research: Use and misuse of statistics in biology. 16 Nov. 2015. <http://influentialpoints.com/Training/statistical_mistakes_in_research_use_and_misuse_of_statistics_in_biology.htm>.
- Kramer, M.H., Paparozzi, E.T. and Stroup, W.W. 2016. Statistics in a horticultural journal: Problem and solutions. *Journal of American Society of Horticultural Sciences* 141(5):400-406. Doi: 10.21273/JASHS 03747-16
- Kuehl, R.O. 2000. Design of experiments: Statistical principles of research design and analysis, 2nd ed. Pacific Grove, CA: Duxbury Press.
- Lang, T.A. and Altman, D.G. 2014. Statistical analyses and methods in the published literature: The SAMPL guidelines. In: Moher, D., Altman, D.G., Schulz, K.F., Simera, I. and Wager, E. (Eds.). Guidelines for reporting health research: A User's Manual. Oxford, UK: John Wiley & Sons. pp. 264– 274.
- Maindonald, J.H. 1992. Statistical design, analysis, and presentation issues. *New Zealand Journal of Agricultural Research* 35: 121-141
- McDonald, J.H. 2014. Handbook of Biological Statistics (3rd Ed.). Sparky House Publishing, Baltimore, Maryland. pp. 140-144.
- Nelson, L.A. and Rawlings, J.O. 1983. Ten common issues of statistics in agronomic Research and reporting. *Journal of Agronomic Education* 12: 00-106.
- Piepho, H.P. and Edmonson, R.N. 2018. A tutorial on the statistical analysis of factorial experiments with qualitative and quantitative treatment factor levels. *Journal of Agronomy and Crop Science* 204:429-455
- Piepho, H.P., Buchse, A. and Richter, C. 2004. A mixed modeling approach to randomized experimental design with repeated measures. *Journal of Agronomy and Crop Science* 190:230-247. <https://doi.org/10.1111/j.1439-037X.2004.00097.x>
- Raudonius, S. 2017. Application of statistics in plant and crop research: Important issues. *Zemdirbyste-Agriculture* 104(4): 377 - 382. DOI 10.13080/z-a.2017. 104.048.
- RUFORUM (2019, March 18th): Fact Sheet Masters in Research Methods. Retrieved from https://educationdocbox.com/College_Life/75430781-Fact-sheet-masters-in-research-methods-no-0001-september-2015-issn.html
- Sainani, K. 2010. The importance of accounting for correlated observations. *PM R*. 2010 2(9):858–861. doi: 10.1016/j.pmrj.2010.07.482.
- Stern, R.D., Coe, R., Allan, E.F. and Dale, I.C. (eds). 2004. Good statistical practice for natural resources research. CABI, Wallingford. 388pp.
- Wasserstein, R.L. and Lazar, N.A. 2016. The ASA's statement on P-values: Context, process, and purpose. *The American Statistician*, DOI:10.1080/00031305.2016.1154108