

Farmer versus Researcher data collection methodologies: Understanding variations and associated trade-offs

Hannington Ochieng^{1,2*}, John Ojiem¹, Joyce Otieno²

¹Kenya Agricultural & Livestock Research Organization, Kibos Horticulture Centre, P. O. BOX 1490 - 40100, Kisumu, Kenya.

²School of Mathematics, Statistics, and Actuarial Science, Maseno University, P. O. BOX 333 - 40105, Maseno, Kenya.

*Correspondence to: hannington.odido@gmail.com

Abstract

The number of non-experts (such as farmers) participating in research activities has increased over the years, with the aim of them addressing their heterogeneous conditions. The situation has resulted in them being engaged in data collection through a process called crowdsourcing. The study examined the level of variation between data sets and the conclusions drawn from data collected using researcher (expert) and farmer (non-expert) methodologies, and also determined the associated trade-offs for using either methodology. The results showed a low convergence between individual observations of the methodologies on most variables with coefficients ranging from $|0.39|$ to $|0.60|$. However, there was stronger convergence in the conclusions drawn when the results were aggregated ($r > |0.80|$) for all the variables tested in this study. Therefore, expert and non-expert data were equivalent for average results. However, data may not be comparable for understanding variations in technology performance due to lack of precision in the subjective assessments of farmers relative to the objective measurements of the researcher.

Keywords: farmer, researcher, methodology, crowd science, crowdsourcing, participatory research

1. INTRODUCTION

Over the years, many agricultural technologies have been developed and introduced to smallholder farmers in most developing countries, including in Sub-Saharan African (SSA). Examples of these technologies include the introduction of new crop species and varieties,

and new agronomic and resource management practices. The technologies are aimed at improving diet and food security and to pave way out of poverty (*Simtowe et al., 2011*) in these regions. However, some of these technologies have been tested and evaluated by researchers in controlled conditions (on-station and researcher-managed on-farm trials), making their adoption into the smallholder farming systems slow/difficult due to the biophysical and socioeconomic heterogeneity that exists (*Adesina and Baidu-Forson, 1995; Nkonya et al., 1997; Doss, 2001; Deressa, 2009*). To realize the precision of the research and applicability and acceptability of results in these varied farmer conditions, researchers have adopted a participatory large N research approach. This approach enables farmers sampled from varied contexts to participate in the research processes through decision-making and by giving their ideas, through a concept known as crowd science (*Louv and Fitzpatrick, 2012*). Farmers may contribute to reducing the implementation cost brought about by the increased number of farmers by managing, evaluating the new technologies, and collecting data by crowdsourcing approaches (*Hand, 2010*). These implementation costs include input acquisition and distribution, management, and value in time, travel, data collection, and intrusion when organizing the research (*Hellin et al., 2008; Morris and Bellon, 2004*).

The farmers' involvement in data collection would entail using simple data collection methods such as ranking and rating, which is not as detailed and rigorous methods as researchers. Thus the significant challenges that may make researchers reluctant to integrate citizen science and crowdsourcing in their research activities include the format in which farmers collect data and data reliability. Even though the data collected may be ratings and rankings in a quantitative form, the usual methods of analysis may not be appropriate (*Coe, 2010*). There are analysis procedures that are appropriate for this kind of data sets, but most researchers are not comfortable using them. *Van Etten et al. (2016)* demonstrated that ranking by experts invariably matched those of Honduran farmers. \This means that data collected by farmers are reliable. However, there is a need to compare the data sets and the conclusions drawn from data collected using researcher and farmer methodologies and determine the associated trade-offs for using either methodology.

2. MATERIALS AND METHODS

2.1. On-farm trials

This study used datasets generated under a participatory on-farm trial to evaluate the adaptability of selected common bean (*Phaseolus vulgaris* L.) varieties in smallholder farming systems in Nandi County, western Kenya. The performance of seven promising bean

varieties (KK33, KK072, Alulu, KK15, KK16, KK194, and KK8) was evaluated in 45 farms during the short rains 2016 and long rains 2017 growing seasons, in plots measuring 9 m². Bean was planted at the recommended spacing of 50 cm by 10 cm, and fertilizer Triple Super Phosphate (TSP) was applied to all plots at the rate of 30 kg P ha⁻¹. All other standard agronomic recommendations were followed, except for the application of fungicides and pesticides.

2.2. Data collection methodologies

Researcher and farmer methodologies were used in data collection, which involved collecting the actual performance measurements for the researcher and rating the performance of bean varieties against several attributes identified by farmers.

2.2.1. Researcher methodology

Researcher methodology is a common practice in collecting data in researcher managed trials and involves collecting actual measurements e.g., weights and counts, with the aid of appropriate measuring tools and equipment. This methodology was applied to collect data on emergence, pests and diseases, days to maturity, and grain yield. Emergence was assessed at 14 days after planting (DAP) by counting the number of emerged plants and expressing as a percentage of the seeds planted in each plot. Pests and diseases (anthracnose, leaf rust, common bacterial blight, and bean common mosaic virus) whose symptoms appeared on the bean plants were assessed and recorded. The number of dead plants due to root rot was recorded cumulatively at 14, 21, and 28 DAP. Bean common mosaic virus (BCMV), anthracnose (ANTC), common bacterial blight (CBB), and leaf rust (LR) were scored as they appeared on a scale of 1-5, where 1 represents least diseased, and 5 represents most diseased. The scoring was based on CIAT bean program standard evaluation scale (*CIAT, 1985*). Damage by aphids was also scored on a scale of 1-5, as explained above. Days to maturity was assessed at 50% flowering. The plots were harvested, dried, and threshed separately. The grain weight of each plot was determined using a weighing scale and sub-samples taken for determining moisture content using an electronic moisture meter.

2.2.2. Farmer methodology

Before planting, farmers were purposively selected to participate in focused group discussions (FGDs) based on their involvement in previous on-farm research activities. The farmers were then divided into two equal groups (each with 12 participants). One group

consisting of farmers who have been participating in on-farm research activities for more than four years, while the other group consisted of new farmers who had not been involved in such activities. The discussions aimed at (a) finding out if farmers did research by themselves, (b) If they did, which variables did they consider when assessing crop performance? (c) Why are the variables important? (d) Does their participation in research activities influence their research ideas?

At the beginning of each trial season, each farmer was given an observation sheet containing the prioritized variables to be assessed. The importance of independent, individual assessment was emphasized, and farmers were requested to refrain from exchanging ideas about the varieties before evaluation to guarantee the independence of all data points (*Steinke et al., 2017*). Farmers largely complied with the request and also agreed to evaluate each attribute individually. They also developed a rating scale of 1-5, in which 1 = much better than current variety, 2 = slightly better than current variety, 3 = same as the current variety, 4 = slightly worse than current variety, and 5 = much worse than current variety. A current variety is one that the farmers use as a control. The performance of the new bean varieties was compared to current variety, KK8, a variety widely adopted and grown in the Nandi farming systems.

Farmers collected data on emergence, plant health, drought tolerance, maturity, pests & disease tolerance, growth habit, and grain yield (rating and volumetric). Emergence was assessed by each farmer at 14 days after planting (DAP). Plant health data was determined by assessing the color of the stalks/leaves between 28 and 35 DAP. The ability of the varieties to tolerate moisture stress was also assessed on occurrence. Growth habit, maturity, and severity caused by pests & diseases were rated at the onset of pod formation. This enabled the farmers to recall their experiences at the earlier stages of the trial and capture the incidences/severities of the late occurring pests & diseases on the plots. The bean plots were harvested, dried, and threshed separately. Grain yield was assessed in two ways. First, the farmers compared the yield of variety KK8 to that of the other varieties. Afterward, they used 250 ml tins to measure the volume of the grain produced and were able to approximate different levels, e.g., 0.75, 0.5, and 0.25.

2.3. Data processing and analysis

The data collected using the two methodologies (farmer and researcher) during the two experimental seasons were imported into R software package version 3.5.0, with RStudio interface Version 1.0.153. The datasets were then merged into one for processing and

analyses using different R procedures. Grain yield, maturity, and percent emergence data of Alulu, KK072, KK15, KK16, KK194, and KK33, collected using researcher methodology, and grain yield of the same varieties collected using farmer volumetric methodology were expressed relative to variety KK8 for each farm. This was achieved by subtracting the performance of KK8 from those of the other varieties. Cumulative Disease Index was calculated for anthracnose (ANTC), leaf rust (LR), common bacterial blight (CBB) and bean common mosaic virus (BCMV) on the farm i and season j for each variety (*Yamamoto et al. 2000*), equation 1:

$$DI_{ijk} = 100 * \left(\frac{\text{Total observed scores for different diseases}}{\text{Sum of maximum scores for each disease}} \right)$$

$$DI_{ijk} = 100 * \left[\frac{LR_{ijk} + BCMV_{ijk} + ANTC_{ijk} + CBB_{ijk}}{\text{Max}(LR) + \text{Max}(BCMV) + \text{Max}(ANTC) + \text{Max}(CBB)} \right] \quad (1)$$

Cumulative disease index (DI) for each variety was also expressed relative to variety KK8. The variety outputs for each of the researcher variables (percent emergence, maturity, disease index, and grain yield) were determined by calculating their mean performances across farms, while those for farmer variables were determined by calculating the percentages of farmers who rated each variety better than, same as or worse than variety KK8. Pearson's correlation was used to assess the strength of the relationships between the researcher variables and farmer variables, and between the researcher variable outputs and farmer variable outputs. The results were then presented using scatterplots. In addition, the farmer ratings for grain yield, maturity, and emergence were converted to binomial, where 0 represents performance worse than KK8, and 1 represents performance better than KK8. The binomials were then modeled against the researcher findings using logistic regression as a general linear model (glm), and the findings graphed to show their distributions.

3. RESULTS AND DISCUSSIONS

3.1. Results of the Focused Group Discussions

FGDs confirmed that farmers do their research, and further revealed that farmers also have variables of interest that they use in evaluating bean varieties. However, the order of importance of the variables differed by farmer. Also, the findings showed that the longer the

farmers participated in research activities, the more organized their evaluation criteria were. However, the variables and the reasons for evaluation were the same in both groups. Farmers explained during the FGDs that they evaluated crop varieties in four stages: establishment, vegetative, harvest, and post-harvest stages. Farmers prioritized nine variables, namely emergence, plant health, drought tolerance, maturity period, pest & disease tolerance, growth habit, and grain yield, which mainly include phenotypic traits. The finding was in agreement with Cleveland and Soleri (2007), who stated that farmers select crops majorly due to their phenotypic expressions. Grain yield is assessed by rating and volumetrically, while the others are by rating. Variables from the post-harvest stage, such as taste, were not included since they involve a strong element of subjective assessment or preference (Steinke *et al.*, 2017). In any case, the researcher always uses farmers to evaluate taste. It was also noted that farmers employ a comparative rating on most of the variables when assessing performance. The participants further clarified that they do paired comparison in which the new variety is checked against the popular variety for the information to be meaningful. Quantitative assessment is only adopted on grain yield, which is mainly measured on a volumetric basis using tins of different sizes, e.g., 2 kg tins (widely known as *gorogoro*, in Kenya).

Table 1. Traits farmers and researchers use in bean varietal evaluation, their measurements, their importance, and the conclusions drawn

Bean stage	Trait	Farmer methodology	Researcher methodology
Establishment	Emergence	Assessing days to emergence (potential for early maturity), and gaps (potential for a high stand at harvest). Prefer fewer days to emergence (food security), and fewer gaps (high yield) Influenced by: Seed dormancy, rotten seeds or eaten by pests (insects, squirrel)	Important for good plant population for analysis
Vegetative	Growth habit	Assess if variety is erect, bushy or climbing tendencies Influences how the variety can be integrated into farming systems (pure stand or intercrop and other benefits) Upright, bushy varieties can be intercropped, help control weeds. Climbing varieties are suited for pure stand planting	Assess if variety is erect, bushy or climbing tendencies Bushy varieties are more susceptible to pests and foliar diseases that may come from the soil.
	Pest & Disease	Assesses the damage caused Indicates performance potential in areas with a high risk of pest & disease Variety showing least damage is the best Always considers the severity and do not consider interactions	Assess the level of susceptibility or resistance of a variety to pests and diseases Use scores with varied scales
	Vigor	Assesses the size and color of the stalks/leaves Indicates the ability of a plant to resist pests and diseases. Indicates performance potential Thick and dark stalk/leaves show survival capacity hence high yield expected Influenced by: disease, pest or moisture stresses	
	Plant height	Assesses growth rate: approximated using (farmer) body parts Indicates the response to stresses and stages for agronomic management e.g., weeding, applying fertilizer, etc. Stunted growth is bad: stressed plant (low yields)	Measure heights using scales. Indicate the growth rate
	Flowering intensity	Number of flowers Shows potential for high yield Pods are developed from the flowers, so the more the flowers, the more the pods Influenced by: disease, pest or moisture stresses	
	Drought	Assesses reaction to moisture stress Variety showing least symptoms of moisture stress is preferred due to their potential for good yield in a dry season Influenced by: disease or pest stresses	Measures daily minimum and maximum temperature, and the rainfall for a region Used to explain the variation in performance
Harvesting	Maturity	Assess color change of plant parts, e.g., pods & leaves (potential for early maturity) Fewer days are better for food security, while more days is better for more grains Influenced by: disease, pest or moisture stresses	Determines days to 50% flowering from the planting date Indicates the possibility of a variety to escape drought and some pests and diseases
	Pods	Assess the number of pods per plant and seeds per pod (potential for high yield) Many pods and many grains per pod: high yielders	Determine the number of pods per plant and seeds per pod
	Grain yield	Assess the amount of grain produced by rating or measuring using tins or sacks More grains means fewer hunger days or more income Influenced by: quality of grain, preferences (taste, seed size, color.)	Measure the weight of yield produced per unit area Also, measure the weight of 100 seeds

3.2. Correspondence and trade-off assessment between individual observations

Grain yield

Results of correlation analysis showed that there were significant relationships between farmer and researcher variables (Fig. 1). There was a strong curve-linear relationship between the researcher grain yield variable and farmer volumetric grain yield variable (Fig. 1A). The variables were positively correlated with $r = 0.97$ and $p\text{-value} < .001$. However, Fig. 1A also shows five observations (marked 644, 645, 693, 700, and 701) with high error of prediction values. The simulated values of these observations were different from the actual values.

When correlation was performed without the five values, the coefficient value remained the same ($r = 0.97$). This means the values did not influence the relationship. The observations with the high error of prediction could be due to biases or errors from either methodology. The researcher's method of collecting grain yield data may be exposed to biases due to weighing problems, such as the use of inappropriate weighing scales (large scale intervals), faulty weighing scales, and taring container weights. These individual errors may be small, but the combination of errors can be significant (*Murphy, 1991*). Although farmers used standard tins to measure grain yield produced (section 2.2.2), the measurements could be exposed to errors. The errors may be due to the estimation of different levels, which may vary by farmers; for example, some farmers would say a full tin to level, while others it is heaped (*Fermont and Benson, 2011*).

Average correspondence was observed between grain yield scores by farmer and yield relative to KK8 as measured by researchers, with a coefficient of - 0.60 at p-value < 0.001 (Fig. 1B). This means that the lower the rating value, the higher the grain yield, which is in agreement with sections 2.2.1 and 2.2.2. The trend line in Fig. 1B had a steep slope from $y = 4$ to $y = 2$, and gentle slopes between $y = 5$ and $y = 4$, and from $y = 2$ to $y = 1$, which shows that farmers had difficulty in identifying the differences between the extreme ratings. It is also evident that for yield difference of - 0.5 to 0.5 Mg ha⁻¹, the farmer ratings ranged from 1 (much better than KK8) to 5 (much worse than KK8). The ratings were, however, slightly influenced by biases to varieties at the extreme scores (Fig. 1C). Focused group discussions with Nandi farmers revealed that they might consciously over- or under-rate the performance of a variety due to their preferences. For example, yields of two varieties may be the same. Still, a farmer may rate a variety with higher market value better than the one with lower market value because of the expected income. In addition, farmers adjust yields when rating to cater for the quality of the grain harvested, and losses that may occur during the cleaning, threshing, winnowing, and drying are taken into account. Similar findings were reported by Fermont and Benson (2011) and were referred to as Economic yield.

Diseases tolerance

Farmers rated disease severity irrespective of being single or multiple, while the researcher rated the severity of each disease. Researcher observations were averaged by calculating the

disease index of the bean varieties across four foliar diseases (equation 1). The relationship between the farmers' ratings of disease tolerance and the researcher's foliar disease index was significant, with a positive correlation coefficient of 0.60 (Fig. 1D). Farmers pointed out that disease rating may be influenced by other physical conditions, such as moisture stress, occurring at different growth stages. Disease index did not take into consideration the different severity levels of the foliar diseases, thus makes it difficult to express the effect on susceptible varieties. Different bean varieties have varying susceptibility levels to different diseases e.g., KK15 showed relatively more susceptibility to leaf rust than the rest of the varieties (Ochieng, 2018).

Emergence

The ability of a seed of a given variety to emerge from the soil when planted is an important factor to both farmers and researchers (Table 1). Farmers are also keen on how long the seed takes to emerge because, to them, it is an indicator of the maturity period (Table 1). Correlation analysis of the emergence data collected by farmers and that collected by researchers showed a strong significant relationship with $r = -0.80$ at $p\text{-value} < .001$ (Fig. 1E). However, it can also be observed that when the researcher recorded no differences between the other varieties and KK8 (x-axis = 0, Fig. 1E), the ratings ranged from 1 to 5. This indicates that there were some factors that may have influenced farmers while making observations. Therefore, they rated in favor of or against those factors.

Maturity

A lower correlation coefficient was obtained between maturity data collected by the researcher and that collected by the farmers with a coefficient of $r = 0.39$ (Fig. 1F). Farmers rated some early maturing varieties worse than variety KK8, probably due to their production objectives. A farmer producing for sale would prefer late-maturing varieties (varieties KK8, KK16, and KK194), while the ones producing for food security would prefer early maturing varieties (varieties KK15, Alulu, and KK72). FGDs revealed that farmer rating for maturity might be influenced by biophysical factors affecting the crop at that stage, leading to low convergence with researcher data. For example, a variety would be rated worse by a farmer if its maturity coincided with moisture, pest, or disease stresses leading to low yields. These factors confound information received from farmer maturity rating, making it more subjective.

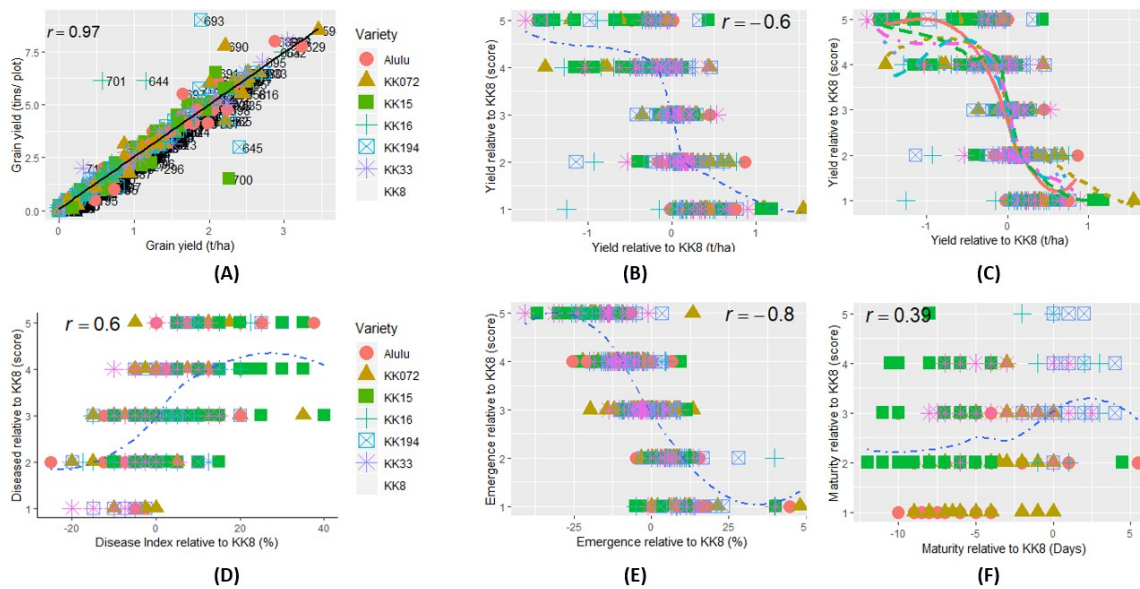


Fig. 1. Correlation between (A) grain yield (farmer volumetric vs. researcher metric), (B) grain yield (farmer scores vs. researcher metric), (C) grain yield (farmer scores vs. researcher metric) for each variety, (D) disease (farmer scores vs. researcher disease index), (E) emergence, and (F) maturity variables of the tested bean varieties in Nandi County.

Pest tolerance

Aphids were the main pests in bean production during the trial period. The results of the correlation analysis between pest data collected by farmers and that collected by the researcher indicated that the variables were significantly related, with a coefficient of 0.54 at p -value < 0.001 . An accuracy of 79.4% was determined when the pest data (collected by the farmers and those collected by the researcher) was cross-tabulated. The corresponding findings can be observed along the diagonal of the tabulation displayed on a balloon plot (Fig. 2).

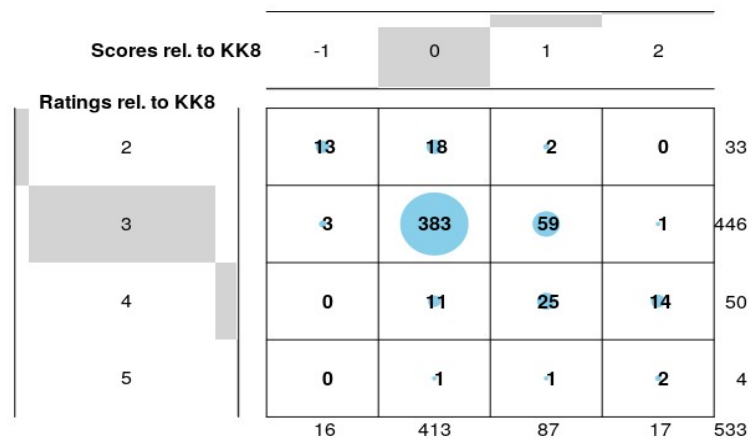


Fig. 2. Relationships between pest scored by farmers and that scored by the researcher on bean varieties tested in Nandi County. Researcher scores are on the horizontal axis (-1 to 2), and the farmer ratings are on the vertical axis (2 to 5).

3.3. Correspondence, trade-off assessment, and conclusions based on aggregated information

Grain yield

Based on the aggregated information (all farms and seasons), correlation analysis between mean researcher grain yield (Mg ha^{-1}) and mean volumetric yield (tins/plot) were significantly correlated, $r = 0.98$ and $p\text{-value} < 0.001$ (Fig. 3A). The results indicated that the conclusions that may be drawn from volumetric data collected by farmers would be similar to those that may be drawn from the metric data collected by the researcher. Similar trends were also observed between the percentages of farmers who rated each variety better than variety KK8 (from farmer grain ratings) and mean grain yield for each variety relative to KK8 (researcher grain yield). The outputs had strong positive linear relationship with $r = 0.85$ at $p\text{-value} = 0.03$ (Fig. 3B). The correlation coefficient of the outputs ($r = 0.85$) was relatively higher than that for the individual observations ($r = 0.60$). The output of aggregated farmer ratings is closely related to the researcher's findings. On assessing the validity of farmer grain yield versus the researcher observations, the output of the logistic regression was plotted (Fig. 4A). It was determined that farmers could hardly give accurate assessments for observations with differences of about ± 0.44 kg of the measured harvests, which is about $\pm 0.5 \text{ Mg ha}^{-1}$. The model also gave a prediction accuracy of 84%.

Diseases tolerance

The outputs of farmer disease rating and that of disease index had a strong negative correlation coefficient of -0.85 at $p\text{-value} = 0.03$ (Fig. 3C). This means that the higher the disease index of a variety relative to KK8, the lower the percentage of farmers rating the variety better than KK8. Logistic regression further confirmed the above findings, and it was also established that farmers could hardly make observations corresponding to that of the researchers when the severity differences were up to $\pm 10\%$. Figs. 4 E, F, G, H, and I show how farmer ratings corresponded with the severities of BCMV, leaf rust, CBB, anthracnose, and root rot, respectively.

Pest tolerance

Analysis of the variable outputs (percentage of farmers who rated the varieties better than KK8 and that of farms where the tested varieties were better than KK8) showed a stronger correlation with a coefficient of 0.96 at $p\text{-value} < 0.003$ (Fig. 3D). The percentages were relatively low (less than 10%) for both variable outputs, meaning the varieties were not superior to KK8 against aphids. Similarly, higher percentages of farms showed the tested varieties were the same as KK8 (output from researcher methodology), and higher percentages of farmers rated the varieties to have performed alike (output from farmer methodology) with $r = 0.84$ at $p = 0.04$.

Emergence

Similarly, correlation analysis of the outputs of emergence variables also showed a strong significant negative relationship with a coefficient of 0.86 , at $p = 0.03$ (Fig. 3E). This means that the higher the mean emergence of a variety relative to KK8, the higher the percentage of farmers rating the variety better than KK8. From the logistic regression, it was noted that most farmers could accurately assess the emergence of different varieties. However, a small percentage could not give an accurate assessment for differences of up to $\pm 7\%$ of plants (Fig. 4B).

Maturity

Similar to grain yield and emergence, the output of maturity drawn from the researcher methodology was related to that from farmer methodology (Fig. 3F). The y-axis of Fig. 3F shows the output from the researcher methodology (maturity periods of other plots minus that of KK8 plot). In contrast, the x-axis shows the output from farmer methodology (percentage of farmers who rated other varieties better than KK8). Most varieties were earlier than KK8,

except KK16 and KK194, on the y-axis. The late varieties were rated by relatively lower percentages of farmers to be better than KK8, as opposed to those that were relatively earlier than KK8, which were rated by more farmers to be better than KK8 (Fig. 3F). From the regression analysis, the output from farmer methodology explained about 66% of the output from the researcher methodology at p -value $< .05$. Analysis using logistic regression showed the accuracy of the ratings were very low (Fig. 4C). This is because farmer observations and ratings of maturity are influenced by many other factors, such as the production objective (i.e., for food security or market). Therefore, the comparison between farmer rating of maturity and the researcher counts of days to 50% flowering would not give the exact interests for both teams. Discussions with farmers after analysis revealed that farmer ratings are mainly influenced by their interests on the crop and the challenges experienced during the evaluation period; thus, it would be problematic to assume that the farmer and researcher experience should converge when assessing maturity. This is in agreement with the findings of Steinke et al. (2017) when working with the Honduras farmers, where they tested whether farmer-generated data in agricultural citizen science is good enough to generate valid statements about the research topic.

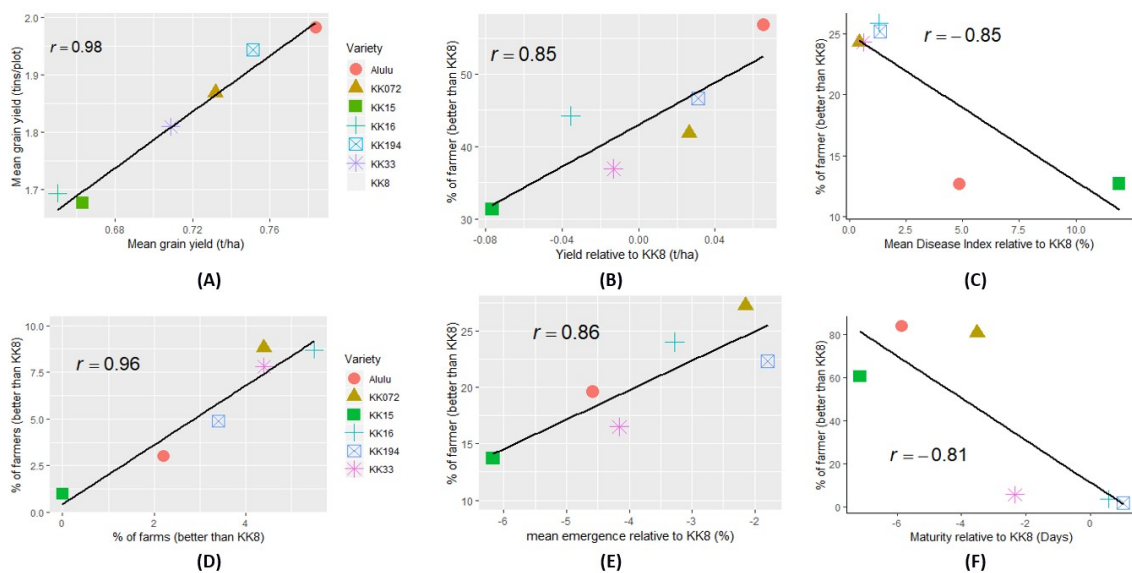


Fig. 3. Correspondence between (A) grain yield (farmer volumetric vs. researcher metric), (B) grain yield (farmer scores vs. researcher metric), (C) disease (farmer scores vs. researcher disease index), (D) pest (farmer vs. researcher scores), (E) emergence, and (F) maturity variable outputs (based on means) of the tested bean varieties in Nandi County.

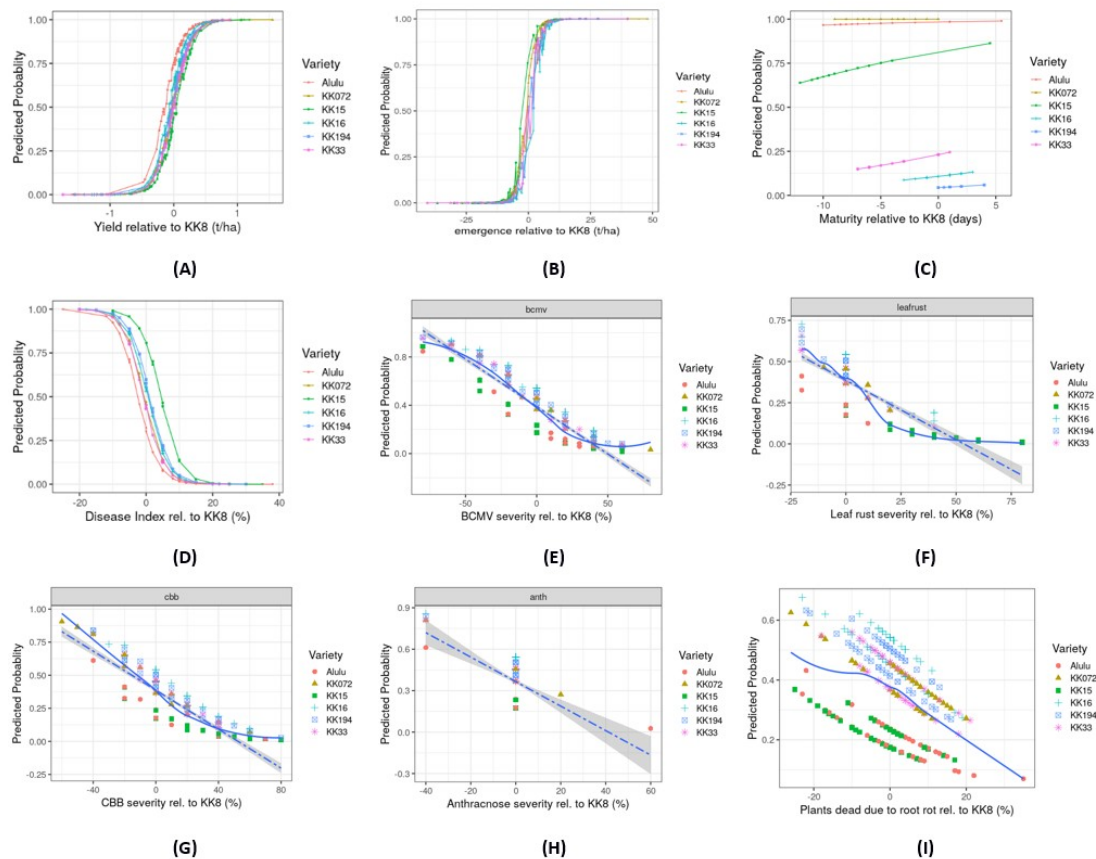


Fig. 4. Probability of farmer rating a variety better than KK8 in terms of (A) grain yield (Mg ha⁻¹), (B) emergence (%), (C) maturity (days), (D) disease index (%), (E) bcmv severity (%), (F) leaf rust severity (%), (G) cbb severity (%), (H) anthracnose severity (%), and (I) root rot severity (%).

4. CONCUSSIONS

The results of this study show that farmers conduct their research, and they have variables they use when evaluating technologies such as new bean varieties. Some of these variables have high levels of subjectivity, especially the ones collected at the post-harvesting stage of the crop, such as taste and grain color. Even when variables assessed were the same as those assessed by researchers, differences exist in how farmers assessed them. Farmers generally use ranking and comparative ratings when evaluating technologies. The results also show that there was a high convergence between farmer and researcher variables such as between farmer volumetric and researcher metric yields and emergence variables.

The low correspondence between individual farmer observations and that of the researcher was not only due to incorrect representations, e.g., challenges in rating maturity because of disease or drought stress, but also due to socioeconomic conditions experienced during data collection. These conditions make the process more subjective even for the relatively more

objective variables, thus divergent views in varietal indicators of quality among the observers. Therefore, for someone seeking to understand variations (for example, sources of differences between farmers), then researcher and farmer data may not be equivalent. This is because of (i) lack of precision in the subjective assessments of farmer relative to the objective measurements of the researcher, and (ii) of the fact that farmer and researcher measure different components (even though they call them the same).

However, researcher and farmer data seem to be equivalent if someone is seeking a conclusion about average results, based on means. This is evident even in cases where variable convergence is low; there is an agreement between the outputs of the two data sets, and the same conclusions are reached - clear evidence of the application of the Wisdom of Crowds (*Surowiecki, 2005*). Increased farmer participation can improve the convergence of farmer and research assessment results. This offers researchers' avenue for conducting large N trials with reduced costs. It also allows researchers to capture what farmers consider to be most informative and relevant, and may influence their decisions when evaluating the performance of technologies, e.g., growth habit. These can then be integrated with the researcher's knowledge to increase the levels of technology dissemination and uptake. Therefore, the use of farmer methodology instead of researcher methodology is suitable only when the researcher would like to know conclusions about average results rather than understanding the variation between farmers.

Acknowledgments: We thank the Director General, Kenya Agricultural & Livestock Research Organization for additional facilitation and support. We sincerely thank the smallholder farmers of Nandi County, western Kenya, for hosting the trials, and for their active participation in the research process. The support and valuable insights given by Ric Coe in data analysis is highly acknowledged.

Funding: The McKnight Foundation funded this work through the Collaborative Crop Research Program (CCRP) grant number 14-306. We are grateful for this support.

Conflict of interest: The authors declare no conflict of interest.

REFERENCES

- Adesina, A. A. and Baidu-Forson, J. (1995) Farmers' perceptions and adoption of new agricultural technology: evidence from analysis in Burkina Faso and Guinea, West Africa. *Agricultural economics*, 13(1), 1-9.
- Bandiera, O. and Rasul, I. (2002) Social Networks and Technology Adoption in Northern Mozambique. Discussion Paper Series. London, UK, Centre for Economic Policy Research CEPR. April 2002.
- CIAT, (1985) Annual Report 1984, Tropical Pastures Program Working Document No. 6, CIAT, Cali, Colombia. P. 23
- Coe, R. (2010) Analyzing Ranking and Rating Data from Participatory On-Farm Trials 1, (2002), 44–64.
- Deressa, T. T., Hassan, R. M., Ringler, C., Alemu, T. and Yesuf, M. (2009) Determinants of farmers' choice of adaptation methods to climate change in the Nile Basin of Ethiopia. *Global environmental change*, 19(2), 248-255.
- Doss, C. R. (2001) Designing agricultural technology for African women farmers: Lessons from 25 years of experience. *World development*, 29(12), 2075-2092.
- Fermont, A. and Benson, T. (2011) Estimating yield of food crops grown by smallholder farmers. International Food Policy Research Institute, Washington DC, 1-68.
- Hand, E. (2010) Citizen science: people power. *Nature* 466:685–687.
- Hellin, J., Bellon, M. R., Badstue, L., Dixon, J. and La Rovere, R. (2008) Increasing the Impacts of Participatory Research. *Experimental Agriculture*, 44(01), 81–95. <https://doi.org/10.1017/s0014479707005935>
- Ochieng, H. (2018) Farmer and Researcher methodologies trade-offs in participatory large N research: Common bean adaptation trials in Western Kenya. Maseno University, Kisumu, Kenya.
- Louv, R., and Fitzpatrick, J. W. (2012). *Citizen science: Public participation in environmental research*. Cornell University Press.
- Morris, M. L. and Bellon, M. R. (2004) Participatory plant breeding research: opportunities and challenges for the international crop improvement system. *Euphytica*, 136(1), 21-35.
- Murphy, J., Casley, D. J. and Curry, J. J. (1991) Farmers' estimations as a source of production data: methodological guidelines for cereals in Africa. The World Bank.

- Nkonya, E., Schroeder, T. and Norman, D. (1997) Factors affecting adoption of improved maize seed and fertiliser in northern Tanzania. *Journal of Agricultural Economics*, 48(1-3), 1-12.
- Simtowe, F., Kassie, M., Diagne, A., Asfaw, S., Shiferaw, B., Silim, S. and Muange, E. (2011) Determinants of agricultural technology adoption: The case of improved pigeonpea varieties in Tanzania. *Quarterly Journal of International Agriculture*, 50(892-2016-65202), pp.325-345.
- Steinke, J., van Etten, J. and Zelan, P. M. (2017) The accuracy of farmer-generated data in an agricultural citizen science methodology. *Agronomy for Sustainable Development*, 37(4), 32. <https://doi.org/10.1007/s13593-017-0441-y>
- Surowiecki J (2005) *The Wisdom of Crowds*. Anchor Books, New York
- Cleveland, D. A. and Soleri, D. (2007) Extending Darwin's analogy: bridging differences in concepts of selection between farmers, biologists, and plant breeders. *Economic botany*, 61(2), 121.
- Van Etten, J., Beza, E., Calderer, L., Van Duijvendijk, K., Fadda, C., Fantahun, B. and Kiambi, D. (2016) First experiences with a novel farmer citizen science approach: crowdsourcing participatory variety selection through on-farm triadic comparisons of technologies (TRICOT). *Experimental Agriculture*, 1-22.
- Yamamoto, T., Iketani, H., Ieki, H., Nishizawa, Y., Notsuka, K., Hibi, T. and Matsuta, N. (2000) Transgenic grapevine plants expressing a rice chitinase with enhanced resistance to fungal pathogens. *Plant Cell Reports*, 19(7), 639-6.