# Groundnut variety identification among smallholder farmers in northern Nigeria: DNA fingerprinting and farmers self-reported adoption status

by Geoffrey Muricho, Collins Odhiambo, and Chris Ojiewo

*Copyright 2021 by Geoffrey Muricho, Collins Odhiambo, and Chris Ojiewo. All rights reserved. Readers may make verbatimcopies of this document for non-commercial purposes by any means, provided that this copyright notice appears on all such copies.*

# Groundnut variety identification among smallholder farmers in northern Nigeria: DNA fingerprinting and farmers self-reported adoption status

**Geoffrey Muricho[a]*, Collins Odhiambo[b] and Chris Ojiewo[a],**

[a] International Crops Research Institute for the Semi-Arid Tropics (ICRISAT), Nairobi, Kenya

[b] Consultant, International Crops Research Institute for the Semi-Arid Tropics (ICRISAT), Nairobi, Kenya

***Corresponding author**: g.muricho@cgiar.org

## Abstract

Low agricultural productivity in sub-Saharan Africa has been associated with poor uptake of modern farming technologies like improved crop varieties. However, actual adoption of improved crop varieties is unclear because past empirical studies have relied on subjective self-reported adoption status that could be inaccurate. Weak seed systems have not helped the situation as rampant seed adulteration is reported. Fortunately, more accurate DNA fingerprinting (DNA-FP) analysis is fast gaining momentum as a gold standard to assess adoption. Using household survey data and DNA-FP grain samples collected from smallholder groundnut farmers in Nigeria, we estimate adoption levels and analyze farmers' ability to correctly identify varieties. Confusion matrix and multinomial logistic regression model are adopted. DNA-FP analysis showed that sampled households grew 35 distinct groundnut varieties of which 27 were improved. Further, though DNA-FP results showed an adoption spread of about 59%, farmers' self-reported adoption was 44%. This 15% underreporting of adoption by farmers points to a non-trivial variety knowledge gap. Econometric analysis showed that access to groundnut seed and variety information from formal sources were critical in minimizing type I and type II errors of variety identification.

**Keywords:** Nigeria, DNA Fingerprinting, Groundnut; Adoption

## Introduction

About 56% of people in sub-Saharan Africa (SSA) live in poverty and 80% of them are engaged in smallholder subsistence agriculture that is characterized by low yields (OECD, 2016). This unfortunate situation has been attributed to several factors including overreliance on rain fed agriculture, poor infrastructure, poor government policies and low usage of modern agricultural technologies like improved crop varieties among others (Muema, 2018). Low adoption of improved crop varieties has been cited as one of the main contributors to low yields which in turn has exacerbated poverty among rural smallholder farming households. The widely grown local crop varieties are highly susceptible to biotic and abiotic stresses whose frequency and intensity have increased in recent times due to climate change (Ojiewo et al., 2018). Fortunately, recent international, regional and national investments in agricultural research has led to development and release of many improved high yielding and well adapted modern crop varieties in many countries of SSA. Further investments have been put in place to promote and popularize these improved crop varieties among the targeted smallholder farmers in the region with the ultimate goal of increasing adoption and improving their welfare outcomes (Dreisigacker et al., 2019).

However, adoption levels of these promoted modern crop varieties is not clear despite literature being replete with empirical adoption studies. Most of these past studies on adoption of improved crop varieties typically rely on breeders' opinions, extension agent reports, seed producers and suppliers, and a large proportion of them are dependent on eliciting responses from sampled respondents in household surveys (Pannell & Zilberman, 2020). This latter and most popular source of data used in adoption studies (self-reported adoption) has made findings on adoption rates vary significantly. One main cause of this variation is recall bias caused by decay in farmers' ability to correctly recall varieties planted over long periods of time (Kosmowski et al., 2019;

Wineman et al., 2020). Besides, weak and poorly regulated seed systems in SSA, especially for legumes, do not have sufficient oversight and regulatory policies that has led to rampant adulteration of seed even those procured from formal sources (Bold et al., 2017). This adulteration problem is more pronounced for quality-declared seed (QDS) that is produced under less stringent certification process (Andrade-Piedra et al., 2016). Compounding this adulteration problem is the weak and poorly resourced extension system that makes access to proper information about agricultural technologies difficult (Wossen et al., 2019). In addition, lack of clear morphological differences between different crop varieties makes it very difficult for farmers to correctly identify the varieties they have grown.

These myriads of challenges lead to information asymmetry that causes misidentification of variety seed type that potentially skews self-reported adoption rates of crop varieties. The implications of these misidentifications are serious because inferential statistics drawn from such research may lead to flawed policy designs and suboptimal application of agronomic practices in farming systems where good agronomic practices are applied to improved crop varieties (Wineman et al., 2020). The imprecise estimates also have far reaching implications for agricultural research and development investments that desire to change farmers' livelihoods through promoting good quality and high yielding seeds. International, regional and local development practitioners may also have wrong impressions on project impacts/outcomes if adoption and impact assessments are based on wrong identification of the adopted varieties. This can easily lead to misallocation of resources which could result into minimal or no impacts at all.

To correct for these misreported adoption rates, a small but growing number of mainstream adoption literature is recognizing and concentrating on DNA fingerprinted (DNA-FP) data as a benchmark for analyzing household and or plot level variety adoption (Floro et al., 2018; Wineman

et al., 2020; Wossen et al., 2019).The emerging trends in the few studies conducted so far using DNA-FP highlight two main findings: - under-reporting and over-reporting of adoption rates. Maredia et al. (2016) compared farmers' self-reported adoption rates of common bean in Zambia and cassava in Ghana. In Zambia, farmers' self-reported adoption of improved common bean was 13% while the DNA-FP results showed 16%. On the other hand, in Ghana, farmers' self-reported adoption was 6% compared to 29% that was confirmed with DNA-FP results. This underestimation of adoption by farmers was attributed to confusion and misunderstanding of what constitutes an 'improved' or 'modern' variety. Similarly, Dreisigacker et al. (2019) in Afghanistan and Tizale et al., (2015) in Ethiopia found farmers underestimating their adoption of improved wheat varieties by about 4% and 36%, respectively. Also, Wossen et al. (2019), found that farmers underestimated adoption of improved cassava varieties in Nigeria by about 7%. On the other hand, Floro et al. (2018)found that farmers in Colombia overestimated adoption of improved cassava varieties by 8% when self-reported results were compared with DNA-FP results.

In these cases of misidentification, the classical Type I errors (false positives) and Type II errors (false negatives) often emerge. False positives are cases where farmers wrongly identify local varieties as improved varieties. On the other hand, false negatives are cases where improved varieties are wrongly reported as local varieties. However, with the advent of affordable DNA-FP technology, more precise crop variety identification is now possible upon collection of crop samples from the survey participating households. Therefore, using DNA-FP results generated from grain samples collected from smallholder groundnut farmers in northern Nigeria, we analyze variety identification accuracy among the sampled households and contribute to this limited but growing literature in two ways: - First, we use the emerging DNA-FP approach to resolve uncertainties around identification of groundnut varieties grown by sampled households as a

results of recall bias. Therefore, we generate more accurate adoption levels among the sampled producers. This method has not been widely used in past empirical adoption studies due to several reasons ranging from lack of data and cost implications that goes with collecting and analyzing samples. Second, to the best of our knowledge, no study has analyzed improved groundnut variety adoption using DNA-FP within the context of Nigeria. Yet groundnut production in Nigeria is very critical for reduction of food insecurity and poverty. According to 2019 statistics, the country is the leading producer of groundnut in west Africa, accounting for almost 44% of total production in the region, 23% of Africa's production and 13% of global production (FAOSTAT, 2021). Further, it is estimated that about 34% of cultivated land in Nigeria is under groundnut and this crop contributes about 23% of total household cash income in the country (Ajeigbe et al., 2015).

## 1. Data and methods

### 1.1 Data

DNA-FP is fast gaining acceptance among researchers and development practitioners, not only in determining the adoption rates of different varieties, but also as a standard tool for building a comprehensive crop variety reference library. In Nigeria, the International Crops Research Institute for the Semi-Arid Tropics (ICRISAT) and Institute for Agricultural Research (IAR) have been in the forefront of developing and releasing improved groundnut cultivars from early 1990s. Prior to 1992, IAR undertook most of the groundnut breeding work and about 20 groundnut varieties had been developed and officially released in Nigeria before ICRISAT came on board (Ntare et al., 2013). With Bill and Melinda Gates Foundation (BMGF) investments in TL I (2007-2011), TL II (2012-2014) and TL III (2015-2018) projects under the leadership of ICRISAT, a

total of six improved groundnut varieties were developed, tested, adapted and released in Nigeria[1]. Besides, the last two phases of this BMGF investment (TL II and TL III) involved very robust variety awareness creation and promotional activities that aimed at increasing adoption and eventual impacts of improved groundnut varieties among the targeted communities.

To estimate the adoption of promoted and other improved groundnut varieties among the target communities after almost 12 years of BMG investments, a household survey was commissioned in 2017 to collect data in five States of northern Nigeria (Bauchi, Jigawa, Kano, Katsina and Kebbi). Multi-stage sampling design was used to select survey units (households). First, the five States were purposively selected because these were the sites that TL III project activities were implemented. The second stage involved stratified sampling of Local Government Areas (LGAs) in each of the selected State where three LGAs were randomly selected in each State except for Kebbi that had two. The third stage involved random sampling of 42 villages within the selected LGAs. Finally, 1,470 households were randomly selected based on the list that was provided by area leadership with the help of extension officers. The actual data collection was conducted by trained and experienced enumerators who used a semi-structured questionnaire programmed in tablets.

During the household interviews, groundnut grain samples for DNA extraction were collected from each of the surveyed household. Each sample had a unique identification number that linked it to specific household that it was collected from. In total, 1470 samples were collected – one

[1] The six varieties are: SUMNUT24 (ICIAR19BT) released in 2011; SAMNUT25 (ICGX-SM-00020/5/P10) released in 2013; SAMNUT26 (ICGX-SM-00018/5/P15/P2) released in 2013; SAMNUT27 (ICGV-IS 07999) released in 2018; SAMNUT28 (ICGV-IS 09926) released in 2018; SAMNUT29 (ICGV 01276) released in 2018

sample from each of the sampled households. The DNA samples were shipped to Australia-based Diversity Arrays Technology (DArT) for DNA extraction and analysis. Fingerprinting was conducted to determine the degree of similarity between samples from the surveyed households against varieties in the reference library. DArTseq techniques that combines DArT complexity reduction and next generation sequencing methods were used for genotyping (Jaleta et al., 2020).

### 1.2 Methods

Descriptive and econometric approaches were used to analyze variety identification and adoption among sampled households. To understand farmers' knowledge of different groundnut varieties that they had grown, confusion matrix (CM) model was used to categorize and compare farmers' self-reported adoption rates and the more accurate DNA-FP results. Further, in order to understand the determinants of variety identification among sampled households (type I or type II error), multinomial logit regression model (MNL) was used.

#### 1.2.1 Confusion Matrix (CM)

Following Chicco & Jurman (2020), we used confusion matrix (CM) model to compare farmers' self-reported adoption rates and DNA-FP results. The CM is used to evaluate and describe the accuracy of classification or identification predictions based on confirmed true values of those predictions. In this case, it is known that farmers had identified/predicted groundnut varieties based on their own knowledge while a DNA-FP analysis was used to confirm/verify farmers self-reported identification. Therefore, there are two sources of binary predictions – from farmers' self-reported adoption and DNA-FP results based on genetic analysis of the grain samples collected from the same farmers. In this case, CM is an important approach that can be used to classify these binary predictions (identifications/classifications) based on expected true values from DNA-FP results.

In binary responses like the ones under this study, the CM has four main components (Table 1). First, true positives (TP) are values of improved groundnut varieties that were correctly identified by farmers when they self-reported their adoption status ($\lambda_{11}$ in Table 1). Second, false positive (FP) are local varieties that were incorrectly self-reported by farmers as improved groundnut varieties ($\lambda_{01}$ in Table 1). This $\lambda_{01}$ constitute what is called Type I error (Wineman et al., 2020). Third, false negative (FN) are confirmed improved groundnut varieties that farmers incorrectly self-reported as local varieties ($\lambda_{10}$ in Table 1). This erroneous identification of improved varieties as local varieties ($\lambda_{01}$) is referred to as Type II error (Wineman et al., 2020). Fourth, true negative (TN) are local varieties that were correctly self-reported as local groundnut varieties by the sampled households ($\lambda_{00}$ in Table 1). Therefore, using farmers self-reported adoption outcomes as predictions and DNA-FP as actual or true variety identification, we find that the total number of true improved groundnut variety adopters in the sample is $\varphi_{11}$ (Eqn 1) and the true total number of non-adopters (local groundnut variety growers) is $\varphi_{00}$ (Eqn 2). Therefore, the total sample size is $\Phi$ (Eqn 3).

$$\varphi_{11} = \lambda_{11} + \lambda_{10} \quad \text{Eqn (1)}$$

$$\varphi_{00} = \lambda_{01} + \lambda_{00} \quad \text{Eqn (2)}$$

$$\Phi = \varphi_{11} + \varphi_{00} = \lambda_{11} + \lambda_{10} + \lambda_{01} + \lambda_{00} \quad \text{Eqn (3)}$$

Table 1. The Confusion Matrix (CM)

| Actual: DNA-FP analysis | Predicted: Farmers' self-reported identification | | |
|---|---|---|---|
| | P=Positive | N=Negative | Row total |
| P=Positive | TP=$\lambda_{11}$ | FN=$\lambda_{10}$ | $\varphi_{11} = \lambda_{11} + \lambda_{10}$ |

| | | | |
|---|---|---|---|
| N=Negative | FP=$\lambda_{01}$ | TN=$\lambda_{00}$ | $\varphi_{00} = \lambda_{01} + \lambda_{00}$ |
| Column total | $\tau_1 = \lambda_{11} + \lambda_{01}$ | $\tau_0 = \lambda_{10} + \lambda_{00}$ | $\Phi = \varphi_{11} + \varphi_{00} = \lambda_{11} + \lambda_{10} + \lambda_{01} + \lambda_{00}$ |

From Eqn. 1 – Eqn. 3, we find that the total number of confirmed households that had adopted improved varieties based on DNA-FP analysis was $\varphi_{11}$ out of the total sampled households, $\Phi$. Therefore, from CM model (Table 1), seven key descriptive metrics can be derived to evaluate the accuracy of self-reported variety identification among the sampled farmers when the DNA-FP results are used as the true identification benchmark (Eqn. 4 – Eqn. 10). Eqn. 4 gives the accuracy rate with which the sampled households can correctly identify the groundnut variety they are growing regardless of whether it is improved or local. From Eqn. 5, we get the probability of the household misidentifying the variety that it is growing (the converse of Eqn. 4). This is also called the error rate or misidentification rate and it can also be derived as 1 – Eqn. 4. Further, the probability of the sampled households to correctly identify an improved variety if indeed they are growing improved groundnut variety is given by Eqn. 6. This is the true positivity rate and it is also called sensitivity or recall rate in the literature. On the other hand, the probability of the sampled households wrongly identifying a local variety as an improved variety is given by Eqn. 7. This Eqn. 7 is the rate of committing Type I error in the sample. The probability of households correctly identifying local varieties if indeed they are growing local groundnut varieties is given by Eqn. 8 (also called specificity rate). Besides these statistics, the precision rate is computed using Eqn. 9 and it is the probability that what was identified as improved variety by the sampled households is indeed an improved variety – it is a variant of the accuracy rate with the former targeting the true positives only. Finally, the prevalence rate that is derived from Eqn. 10 measures the true adoption rate among the sampled households based on the DNA-FP results.

$$Accuracy\ rate = \frac{(\lambda_{11}+\lambda_{00})}{\Phi} \quad \text{(Eqn. 4)}$$

$$Misidentification\ rate = \frac{(\lambda_{10}+\lambda_{01})}{\Phi} \quad \text{(Eqn. 5)}$$

$$True\ positivity\ rate = \frac{\lambda_{11}}{\varphi_{11}} \quad \text{(Eqn. 6)}$$

$$False\ positivity\ rate = \frac{\lambda_{01}}{\varphi_{00}} \quad \text{(Eqn. 7)}$$

$$True\ negativity\ rate = \frac{\lambda_{00}}{\varphi_{00}} \quad \text{(Eqn. 8)}$$

$$Precision\ rate = \frac{\lambda_{11}}{\tau_1} \quad \text{(Eqn. 9)}$$

$$Prevalence\ rate = \frac{\varphi_{11}}{\Phi} \quad \text{(Eqn. 10)}$$

### 1.2.2 Determinants of groundnut variety identification

From Table 1, we find four mutually exclusive possible variety identification outcomes (TP; FP; FN and TN) that could form a set of dependent variables ($Y$). We analyze the probability of a household falling into one of the four mutually exclusive variety identification outcomes (Table 1) using the multinomial logistic (MNL) regression model. This model is an extension of the binary logistic regression model where the dependent variable is nominal (dependent variable has more than two outcomes or categories that are not ordered). Therefore, the MNL approach is used to analyze the probability of a case (household) falling in any of the more than two dependent variable categories conditional on a set of predicators. These categories are mutually exclusive, and each is defined by a binary outcome (1=yes; 0=otherwise). If the dependent variable has $j$ categories, where $j$ is a positive integer and $j \geq 3$, then MNL model is used with one category designated as the base or reference group/category. In this study, there are four mutually exclusive possible

variety identification outcomes (TP; FP; FN and TN in Table 1) that form the dependent variables set ($Y$) and each of these outcomes is measured on a binary scale (1=if yes; 0=otherwise). On the other hand, the independent variables set ($X$) are household characteristics that have been hypothesized to condition the probability of the household identifying groundnut variety correctly (improved or local) or incorrectly (type I error or type II error). With the four mutually exclusive unordered dependent variables, the MNL model is specified as follows: -

$$P_{ij} = \beta_i X_i + e_i \quad \text{(Eqn. 11)}$$

Where $P_{ij}$ is the probability of household $i$ correctly (improved or local) or incorrectly (type I error or type II error) identifying the variety of the groundnut variety grown. Subscript $j$ in $P_{ij}$ denotes the four variety identification categories (TP, FP, TN and FN). The $X_i$ are household characteristics that determine the ability of the household to identify the groundnut variety correctly (TP and TN), make type I error (FP) or make type II error (FN). On the other hand, $e_i$ are the error terms that are assumed to be normally distributed with a mean of zero and constant variance. Therefore, the probability of a household identifying the groundnut variety that it grew is specified as follows: -

$$Prob\ (P_i = j) = \frac{e_{z_j}}{\sum_{k=0}^{j} e_{z_k}} \quad \text{(Eqn. 12)}$$

Where $z_j$ is the self-reported groundnut variety that the farmer made and $z_k$ are the other alternative variety reporting that the farmer could have reported (3 options). Given that there are alternative reporting options for the farmer, the log odds ratio is computed as follows: -

$$\ln\left(\frac{P_{ij}}{P_{ik}}\right) = \alpha + \beta_i X_i + e_i \quad \text{(Eqn. 13)}$$

Where $P_{ij}$ is the probability of the farmer identifying the variety as self-reported and $P_{ik}$ are probabilities for alternative variety identification reporting. $\ln\left(\frac{P_{ij}}{P_{ik}}\right)$ is the natural log of probability of reporting the variety as $j$ relative to the probability of other variety reporting, $k$. The constant in Eqn. 13 is $\alpha$ while $\beta$ is a vector of parameters to be estimated in the model. Since the coefficients of the MNL model only gives the direction of influence of the independent variable on the dependent variable (correlates), marginal effects of each independent variable are usually computed to give the actual magnitude of the effect of the independent variable on the dependent variable. These marginal effects from the MNL model are computed as follows: -

$$\delta = \frac{\partial P_i}{\partial X_i} = p_i\left(\beta_j - \sum_{k=0}^{j} P_i\beta_k\right) = P_i\left(\beta_j - \beta\right) \quad \text{(Eqn. 14)}$$

## 2. Results and Discussions

### 2.1 Adoption rates of groundnut varieties

After extensive further data cleaning, a total of 1428 samples out of 1470 were fit for the final analysis in this study. The average percentage purity for the extracted material was 99% with a 0.93 classification confidence. The DNA-FP sample analysis revealed that about 35 distinct groundnut varieties were grown by sampled households (Table 2). Out of these 35 varieties, about 27 (77%) were improved. The overall household level adoption rate of improved groundnut varieties based on DNA-FP results was about 59% compared to 44% that was reported by farmers in their own self-identification of the varieties they had grown. The 15% underestimation of adoption means that there are farmers growing improved groundnut varieties without knowing that they are indeed improved (type II error), and this is a clear indication that there exists a non-trivial variety knowledge gap among sampled households.

Table 2. DNA-FP and Farmer Self-reported Variety Classification (% households)

| Variety name | DNA Fingerprinted (N=1428) | | | Farmers self-reported (N=1428) | | |
|---|---|---|---|---|---|---|
| | Local | Improved | Total | Local | Improved | Total |
| 55-437(Yardakar) | 0.00 | 5.11 | 5.11 | 1.96 | 3.15 | 5.11 |
| BAHI DISTRICT NGOGO | 0.28 | 0.00 | 0.28 | 0.14 | 0.14 | 0.28 |
| DODOMA | 0.14 | 0.00 | 0.14 | 0.00 | 0.14 | 0.14 |
| Fleur 11 | 0.00 | 0.28 | 0.28 | 0.07 | 0.21 | 0.28 |
| ICG 12991 | 0.00 | 3.36 | 3.36 | 0.84 | 2.52 | 3.36 |
| ICG 8326 | 0.00 | 1.47 | 1.47 | 1.26 | 0.21 | 1.47 |
| ICGV 01276 | 0.00 | 0.56 | 0.56 | 0.21 | 0.35 | 0.56 |
| ICGV-IS-96894 | 0.00 | 0.21 | 0.21 | 0.07 | 0.14 | 0.21 |
| ICGV-SM 01711 | 0.00 | 0.91 | 0.91 | 0.42 | 0.49 | 0.91 |
| ICGV-SM 01731 | 0.00 | 1.26 | 1.26 | 0.77 | 0.49 | 1.26 |
| ICGV-SM 93530 | 0.00 | 1.05 | 1.05 | 0.84 | 0.21 | 1.05 |
| ICGV-SM-08501 | 0.00 | 0.28 | 0.28 | 0.21 | 0.07 | 0.28 |
| ICGX-SM-00018/5/P15/P | 0.00 | 0.21 | 0.21 | 0.21 | 0.00 | 0.21 |
| ICGX-SM-00020/5/P10 | 0.00 | 1.26 | 1.26 | 0.42 | 0.84 | 1.26 |
| ICIAR 19BT | 0.00 | 12.61 | 12.61 | 3.15 | 9.45 | 12.61 |
| JL 24 | 0.00 | 0.07 | 0.07 | 0.07 | 0.00 | 0.07 |
| JOHARI | 0.00 | 0.21 | 0.21 | 0.21 | 0.00 | 0.21 |
| Jargyada(Local name) | 19.40 | 0.00 | 19.40 | 12.54 | 6.86 | 19.40 |
| LOCAL 1 | 12.54 | 0.00 | 12.54 | 11.13 | 1.40 | 12.54 |

| | | | | | | |
|---|---|---|---|---|---|---|
| M-572.801 | 0.00 | 2.52 | 2.52 | 0.91 | 1.61 | 2.52 |
| M576-79 | 0.00 | 0.07 | 0.07 | 0.07 | 0.00 | 0.07 |
| MAIBARGO (LOCAL NAME) | 7.56 | 0.00 | 7.56 | 5.46 | 2.10 | 7.56 |
| MAMBOLEO UYUI | 0.14 | 0.00 | 0.14 | 0.00 | 0.14 | 0.14 |
| MANIPINTAR | 0.00 | 0.91 | 0.91 | 0.49 | 0.42 | 0.91 |
| MWARABU | 0.49 | 0.00 | 0.49 | 0.07 | 0.42 | 0.49 |
| Manipintar | 0.00 | 0.07 | 0.07 | 0.07 | 0.00 | 0.07 |
| NYOTA | 0.56 | 0.00 | 0.56 | 0.07 | 0.49 | 0.56 |
| Nkatiesari | 0.00 | 0.42 | 0.42 | 0.35 | 0.07 | 0.42 |
| QH 243C | 0.00 | 4.55 | 4.55 | 1.61 | 2.94 | 4.55 |
| RED KAHAMA | 0.00 | 0.84 | 0.84 | 0.35 | 0.49 | 0.84 |
| RMP 91 | 0.00 | 17.44 | 17.44 | 10.43 | 7.00 | 17.44 |
| RMP-91 | 0.00 | 1.54 | 1.54 | 0.98 | 0.56 | 1.54 |
| RRB (resistant red bull) | 0.00 | 0.21 | 0.21 | 0.00 | 0.21 | 0.21 |
| SGV 99031 | 0.00 | 0.35 | 0.35 | 0.14 | 0.21 | 0.35 |
| SH 470P | 0.00 | 0.07 | 0.07 | 0.00 | 0.07 | 0.07 |
| Shitaochi | 0.00 | 0.35 | 0.35 | 0.28 | 0.07 | 0.35 |
| TS 32-1 | 0.00 | 0.35 | 0.35 | 0.14 | 0.21 | 0.35 |
| UGA-2 | 0.00 | 0.35 | 0.35 | 0.21 | 0.14 | 0.35 |
| Total | 41.11 | 58.89 | 100.00 | 56.16 | 43.84 | 100.00 |

At individual variety level, the DNA-FP results showed that the most widely adopted improved groundnut varieties were RMP 91 (17%) and ACIAR 19BT (13%). The 17% adopters of RMP 91

constituted 10% who correctly identified the variety as improved and 7% who erroneously identified this improved variety as local (type II error). On the other hand, the 13% confirmed adoption of ACIAR 19BT constituted 3% who correctly identified it as improved and almost 10% who misidentified this improved variety as local (type II error). Similar analysis of DNA-FP results revealed that the most widely grown local groundnut varieties were Jargyada (19%) and Local 2 (13%). The 19% adoption of Jargyada constituted 12% who correctly identified this variety as local and 7% who made type I error of misidentifying this local variety as improved (Table 2). We also find that the 13% who adopted Local 1 variety composed of 11% who accurately identified it as local with about 2% making type I error of misclassifying this variety as improved. Therefore, these results showed a mismatch between farmers' identification of groundnut varieties that they had grown and the confirmed DNA-FP results. However, given that some of the grown varieties had very low observations to support further individual variety specific analysis (Table 2), we merged improved varieties into one group and local varieties into another for subsequent analysis.

Farmers' self-reported adoption of improved groundnut varieties and DNA-FP results were further analyzed using the confusion matrix model (Table 3). While farmers self-reported results showed an adoption rate of about 44%, the DNA-FP analysis showed a higher adoption rate of almost 59% (Table 2; Table 3). This underreporting of adoption rates in our current study is consistent with past empirical studies (Wossen et al., 2019); though it is also in contrary to other studies that found over reporting of adoption levels (Floro et al., 2018; Wineman et al., 2020). Besides, Wossen et al. (2019) demonstrated that this self-misreporting of adoption status (type I or type II errors) leads to significantly biased estimation of research impacts which may lead to misdiagnosis and ineffective policy options. The low self-reported adoption rates in this current study were due to higher proportion of sampled households making type II error i.e. mistaking improved varieties

for local varieties compared to type I error of incorrectly identifying local as improved varieties. The results showed that about 45% of the surveyed farmers made type II error while 28% made type I error (Table 3). Therefore, only 55% of the 59% DNA-FP confirmed adopters had correctly identified the varieties they were growing as improved compared to about 72% of the 41% DNA-FP confirmed non-adopters that were able to correctly identify the local varieties that they were growing (Table 3). The former (55%) is what is called true positivity rate or recall rate (Eq. 6) while the latter (72%) is called true negativity rate (Eqn. 8).

These results show that there is a significant proportion of smallholder groundnut growers in northern Nigeria who are growing improved groundnut varieties that they mistakenly consider as local. Such wrong perception of the type of groundnut variety grown (lack of information) is likely to have implications on how accompanying agronomic practices are adopted on plots where these varieties are grown. If sub-optimal agronomic practices are adopted on an improved variety plot because the farmer wrongly thinks that it is a local variety, then the productivity potential of that improved variety is likely to be compromised, thereby undermining the envisaged genetic gain from improved varieties. The overall improved variety identification precision rate among the sampled households was about 73% (Eqn. 9). Therefore, to improve precision rate, the true positivity (recall) rate should be as high as possible.

Table 3. Comparison of farmers' self-reported and DNA-FP adoption results (confusion matrix)

| Actual: DNA-FP analysis | Predicted: Farmers self-reported identification | | |
|---|---|---|---|
| | Improved (N=626) | Local (N=802) | Total (N=1428) |
| Improved (N=841) | TP=54.58 | FN=45.42 | 58.89 |
| Local (N=587) | FP=28.45 | TN=71.55 | 41.11 |

| | | | |
|---|---|---|---|
| Total (N=1428) | 43.84 | 56.16 | 100 |

Further, analysis of farmers self-reported adoption and DNA-FP results showed that the variety identification accuracy rate (Eqn. 4) among the sampled households is about 62%. This means that the probability of a household to correctly identify the variety that it is growing (whether improved or local) is about 62% with the remaining 38% being the probability of misidentification or error rate (Eqn. 5). If indeed farmers significantly adopt good agronomic practices like fertilization, spacing, weeding etc. based on their own perceived crop variety that they are growing, then this 38% misidentification rate is not trivial and calls for concerted efforts to address it. This misidentification is likely to compromise the genetic potential of improved varieties that normally need better agronomic practices to achieve their yield potentials. It could also lead to misallocation of resources if high input resources are used on plots with local varieties that are genetically low yielding – leading to low returns per unit of inputs used. A combination of these two potential misidentification outcomes (compromised yield and resource misallocation) could dissuade farmers from adopting improved varieties. Therefore, analyzing determinants of variety misidentification among smallholder farmers is critical for designing interventions that are likely to induce increased adoption of improved varieties which will lead to increased productivity and better welfare impacts of agricultural research.

### 2.2 Determinants of groundnut variety identification

The MNL model coefficients are presented in Annex A while marginal effects are in Table 4. We present, interpret and discuss marginal effects because coefficients from MNL model only indicate the direction of the effect but not the magnitude by which the independent variables affect the dependent variable – something that marginal effects address (Dung, 2020). The marginal effects results show that gender of the household head has a negative and significant effect on the

probability of the household making type I error. Compared to female headed households, male headed households were almost 7% less likely to commit type I error (Table 4). This means male headed households were less likely to mistake a local groundnut variety for an improved variety compared to their female counterparts. The finding could be associated with the possibility that it is easier for male household heads to access information about improved groundnut varieties through different forums like trainings and participation in other technology promotional activities - field days, seed fairs etc. (Ankrah et al., 2020; Lamontagne-Godwin et al., 2018). Further, education of the household head was negatively associated with type II errors. An increase in formal education of the household head by one year has a potential of reducing type II error by about 1%. Therefore, households with more educated heads are more likely to correctly identify improved groundnut varieties that they grew (not to mistake an improved variety for a local variety).

Table 4. Determinants of groundnut variety identification (MNL model marginal effects)

| Variables | True Positive | False Positive (Type I error) | False Negative (Type II error) | True Negative |
|---|---|---|---|---|
| Household head sex (1=male; 0=female) | -0.036 (0.039) | -0.069** (0.029) | 0.077 (0.048) | 0.027 (0.045) |
| Household head age (years) | -0.001 (0.001) | 0.001 (0.001) | 0.000 (0.001) | 0.001 (0.001) |
| Household head formal education level (years) | 0.002 (0.002) | -0.001 (0.002) | -0.008*** (0.002) | 0.007 (0.002) |

| | | | | |
|---|---|---|---|---|
| Improved seed available in the village (1=yes; 0=no) | 0.255***<br>(0.025) | 0.006<br>(0.024) | -0.085**<br>(0.035) | -0.175<br>(0.035) |
| Number of groundnut plots | -0.005<br>(0.005) | 0.003<br>(0.003) | -0.001<br>(0.004) | 0.003<br>(0.004) |
| Household head farm labour participation (1=full time; 0=otherwise) | -0.023<br>(0.026) | -0.013<br>(0.020) | 0.029<br>(0.032) | 0.007<br>(0.031) |
| Source of groundnut seed (1=formal; 0=informal) | 0.131***<br>(0.022) | 0.088***<br>(0.019) | -0.126***<br>(0.033) | -0.093<br>(0.033) |
| Household members engaged in agricultural technology training (1=yes; 0=no) | 0.044*<br>(0.026) | 0.035*<br>(0.021) | -0.198***<br>(0.035) | 0.119<br>(0.033) |
| Sources of groundnut variety information (1=formal; 0=informal) | 0.015<br>(0.031) | 0.036<br>(0.027) | 0.005<br>(0.040) | -0.057<br>(0.042) |
| Membership to farmer groups (1=yes; 0=no) | 0.014<br>(0.022) | 0.014<br>(0.018) | 0.017<br>(0.026) | -0.046<br>(0.026) |
| Visited by agricultural extension staff (1=yes; 0=no) | 0.075***<br>(0.026) | -0.056***<br>(0.021) | 0.052**<br>(0.025) | -0.071<br>(0.025) |
| TL III project target village (1=yes; 0=no) | 0.080***<br>(0.025) | 0.004<br>(0.021) | 0.023<br>(0.026) | -0.107<br>(0.026) |
| Jigawa (1=yes; 0=otherwise) | -0.045<br>(0.043) | -0.067***<br>(0.024) | 0.089*<br>(0.046) | 0.023<br>(0.047) |

| | | | | |
|---|---|---|---|---|
| Kano (1=yes; 0=otherwise) | -0.098** | 0.155*** | 0.002 | -0.058 |
| | (0.042) | (0.038) | (0.035) | (0.035) |
| Katsina (1=yes; 0=otherwise) | -0.097** | 0.066** | 0.020 | 0.012 |
| | (0.038) | (0.031) | (0.038) | (0.039) |
| Kebbi (1=yes; 0=otherwise) | -0.241*** | -0.056** | 0.218*** | 0.080 |
| | (0.039) | (0.026) | (0.044) | (0.042) |

*** p<0.01, ** p<0.05, * p<0.1; values in parenthesis show standard errors

On the other hand, we find that households found in villages where there is an improved seed seller were more likely to correctly identify improved variety and unambiguously less likely to commit type II in variety identification (mistake an improved variety for a local variety). Households in villages that have improved groundnut seed sellers were about 26% more likely to correctly identify improved groundnut varieties that they have grown compared to those found in villages without improved groundnut seed sellers (Table 4). Consistently, households in villages with improved groundnut seed sellers were almost 9% less likely to commit type II error in identifying the groundnut variety that they grew. This confirms the importance of formal seed dealers in disseminating information about improved agricultural technologies like seed. In fact, closely related to this finding on formal seed dealers is the positive and significant correlation between formal sources of groundnut variety information and the likelihood of households correctly identifying improved groundnut varieties that they grew. Households that accessed information on groundnut varieties from formal sources were about 13% more likely to correctly identify improved varieties that they grew and similarly, 13% less likely to commit type II error of variety identification (Table 4). However, they were also almost 9% more likely to commit type I error of mistaking a local variety for an improved variety. This latter finding shows that sampled

households are more certain about local varieties than improved varieties. Similar trends were also observed on the effect of engaging households in agricultural technology training where it was found that those engaged in these trainings were more likely to correctly identify improved groundnut varieties and less likely to commit type II error but more likely to commit type I error, just like households who received groundnut variety information from formal sources. Compared to otherwise, those engaged in agricultural technology training were 4% more likely to correctly identify improved varieties that they grew and 20% less likely to make type II error in groundnut variety identification though they were about 4% more likely to commit type I error (Table 4).

Also, visits of agricultural extension staff to households had a positive and significant association with the likely of a household correctly identifying improved groundnut variety and unambiguous negative association with the likelihood of making type I error. Households that were visited by an agricultural extension staff were almost 8% more likely to correctly identify the improved varieties that they grew compared with those that were not visited. Similarly, unlike households that were not visited by agricultural extension staff, those visited were about 6% less likely to make type I error of variety identification (Table 4). However, we also find a positive and significant association between visits by agricultural extension staff and the likelihood of making type II error. Those households that were visited by agricultural extension staff were about 5% more likely to make type II error of mistaking an improved groundnut variety for local variety. We also find a positive and significant association between TL II project intervention villages and the likelihood of a household correctly identifying improved groundnut variety. This latter finding points to the importance of development projects in disseminating agricultural technology information to target rural farming households.

## 3. Summary and conclusions

Majority of the poor in sub-Saharan Africa (SSA) live in rural areas and mainly derive their livelihood from agriculture that is characterized by low productivity. This low agricultural productivity has been associated with poor uptake of available modern farming technologies especially improved crop varieties among many other factors. However, the actual adoption levels of improved crop varieties is not clear since most of the past empirical investigations have relied on farmers' self-reported adoption status that is likely to be inaccurate due to recall bias caused by decay in farmers' ability to remember. The seed systems in SSA where quality assurance and control are not strong has not helped the situation because seed adulteration has been reported in several countries in the region. Therefore, dependence on farmers self-reported adoption status is likely to lead to erroneous conclusions on adoption levels of improved varieties. However, in recent past, there is an emerging trend in approaches used to estimate crop variety adoption levels where crop samples are collected from farmers and subjected to DNA fingerprinting (DNA-FP) analysis to confirm the true variety identity. We adopted this innovative DNA-FP approach in our current study to estimate the adoption levels of improved groundnut varieties in northern Nigeria and analyzed the determinants of farmers' ability to correctly identify the groundnut varieties that they have grown on their farms.

In order to understand the level of variety knowledge among sampled households, confusion matrix (CM) was used to compare farmers self-reported adoption status and the confirmed DNA-FP analysis results. Further, multinomial logistic (MNL) regression model was adopted to analyze the determinants of households' ability to correctly identify the variety that had been grown. Descriptive statistics from DNA-FP analysis showed that sampled households grew a total of 35 distinct groundnut varieties of which 27 were improved varieties. However, in terms of adoption, DNA-FP results revealed that the adoption rate of improved groundnut varieties in northern

Nigeria was about 59% though farmers' self-reported adoption rate was 44%. This 15% underreporting of adoption by farmers points to a non-trivial variety knowledge gap that need to be addressed. Further, econometric analysis showed that availability of groundnut seed dealers in villages, formal sources of groundnut seed and groundnut variety information, engaging farmers in technology trainings and increased visits by agricultural extension staff are critical in ensuring that farmers who grew improved groundnut varieties correctly identifies them. Type I errors of variety identification could be minimized with increased visits of agricultural extension staff while type II errors will be minimized if formal education of the household is creased, more seed dealers are established in the villages, farmers obtain variety information from formal sources and they are also engaged in agricultural technology trainings.

**References**

Ajeigbe, H., Waliyar, F., Echekwu, C., Ayuba, K., Motagi, B., D, E., & Inuwa, A. (2015). *A Farmer's Guide to Groundnut Production in Nigeria.*

Andrade-Piedra, J., Bentley, J., Almekinders, C., Jacobsen, K., Walsh, S., & Thiele, G. (2016). *Case studies of Roots, Tubers and Bananas seed systems* (No. 2016–3). https://doi.org/10.4160/23096586RTBWP20163

Ankrah, D. A., Freeman, C. Y., & Afful, A. (2020). Gendered access to productive resources – evidence from small holder farmers in Awutu Senya West District of Ghana. *Scientific African*, *10*, e00604. https://doi.org/https://doi.org/10.1016/j.sciaf.2020.e00604

Bold, T., Kaizzi, K. C., Svensson, J., & Yanagizawa-Drott, D. (2017). Lemon technologies and adop- tion: Measurement, theory, and evidence from agricultural markets in Uganda. *Quarterly Journal of Economics*, *132*, 1055–1100.

Chicco, D., & Jurman, G. (2020). The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation. *BMC Genomics*, *21*(1), 1–13. https://doi.org/10.1186/s12864-019-6413-7

Dreisigacker, S., Sharma, R. K., Huttner, E., Karimov, A., Obaidi, M. Q., Singh, P. K., Sansaloni, C., Shrestha, R., Sonder, K., & Braun, H. J. (2019). Tracking the adoption of bread wheat varieties in Afghanistan using DNA fingerprinting. *BMC Genomics*, *20*(1), 1–13. https://doi.org/10.1186/s12864-019-6015-4

Dung, L. T. (2020). A Multinomial Logit Model Analysis of Farmers' Participation in Agricultural Cooperatives: Evidence from Vietnam. *Applied Economics Journal*, *27*(1), 1–22. https://www.researchgate.net/publication/341235030

FAOSTAT. (2021). *Crops*. http://www.fao.org/faostat/en/#data/QC

Floro, V. O., Labarta, R. A., Becerra López-Lavalle, L. A., Martinez, J. M., & Ovalle, T. M. (2018). Household Determinants of the Adoption of Improved Cassava Varieties using DNA Fingerprinting to Identify Varieties in Farmer Fields: A Case Study in Colombia. *Journal of Agricultural Economics*, *69*(2), 518–536. https://doi.org/https://doi.org/10.1111/1477-9552.12247

Jaleta, M., Tesfaye, K., Kilian, A., Yirga, C., Habte, E., Beyene, H., Abeyo, B., Badebo, A., & Erenstein, O. (2020). Misidentification by farmers of the crop varieties they grow: Lessons from DNA fingerprinting of wheat in Ethiopia. *PLoS ONE*, *15*(7 July), 1–16. https://doi.org/10.1371/journal.pone.0235484

Kosmowski, F., Aragaw, A., Kilian, A., Ambel, A., Ilukor, J., Yigezu, B., & Stevenson, J. (2019). Varietal identification in household surveys: Results from three household-based

methods against the benchmark of DNA fingerprinting in southern Ethiopia. *Experimental Agriculture*, *55*(3), 371–385. https://doi.org/10.1017/S0014479718000030

Lamontagne-Godwin, J., Williams, F. E., Aslam, N., Cardey, S., Dorward, P., & Almas, M. (2018). Gender differences in use and preferences of agricultural information sources in Pakistan. *The Journal of Agricultural Education and Extension*, *24*(5), 419–434. https://doi.org/10.1080/1389224X.2018.1491870

Maredia, M. K., Reyes, B. A., Manu-Aduening, J. Dankyi, A., Hamazakaza, P. Muimui, K., Rabbi, I., Kulakow, P., Parkes, E., Abdoulaye, T., Katungi, E., & Raatz, B. (2016). *Testing alterna- tive methods of varietal identification using DNA fingerprinting: Results of pilot studies in Ghana and Zambia* (No. 149; International Development Working).

Muema, E. (2018). *Determinants of access and utilisation of seasonal climate information services among smallholder farmers in Makueni County, Kenya* (Issue 11) [University of Nairobi]. https://doi.org/10.1016/j.heliyon.2018.e00889

Ntare, B. R., Ajeigbe, H., Echekwu, C. A., Ibro, A., & Amadou, A. (2013). Adoption and Impacts of Modern Groundnut Varieties in Nigeria. *ICRISAT Open Access Repository*, 1–44.

OECD. (2016). Agriculture in Sub-Saharan Africa: Prospects and challenges for the next decade. In *OECD-FAO Agricultural Outlook 2016-2025* (pp. 59–95). OECD Publishing. https://doi.org/10.1787/agr_outlook-2016-5-en

Ojiewo, C., Jean Claude, R., John, W., Mathew, A., & Solomon, G. (2018). *Mainstreaming Efficient Legume Seed Systems in Eastern Africa*. Food and Agriculture Organization of the United Nations.

Pannell, D., & Zilberman, D. (2020). Understanding Adoption of Innovations and Behavior Change to Improve Agricultural Policy. *Applied Economic Perspectives and Policy*, *42*(1), 3–7. https://doi.org/https://doi.org/10.1002/aepp.13013

Wineman, A., Njagi, T., Anderson, C. L., Reynolds, T. W., Alia, D. Y., Wainaina, P., Njue, E., Biscaye, P., & Ayieko, M. W. (2020). A Case of Mistaken Identity? Measuring Rates of Improved Seed Adoption in Tanzania Using DNA Fingerprinting. *Journal of Agricultural Economics*, *71*(3), 719–741. https://doi.org/10.1111/1477-9552.12368

Wossen, T., Alene, A., Abdoulaye, T., Feleke, S., & Manyong, V. (2019). Agricultural technology adoption and household welfare: Measurement and evidence. *Food Policy*, *87*(June 2019), 101742. https://doi.org/10.1016/j.foodpol.2019.101742

**Annex**

Annex A. Determinants of groundnut variety identification (MNL model coefficients): True Positive as reference group

| Variable | False Positive Model (Type I error) | False Negative Model (Type II error) | True Negative Model |
|---|---|---|---|
| Household head sex (1=male; 0=female) | -0.426 | 0.630* | 0.416 |
| | (0.351) | (0.362) | (0.337) |
| Household head age (years) | 0.013 | 0.005 | 0.011 |
| | (0.008) | (0.007) | (0.007) |
| Household head formal education level (years) | -0.017 | -0.041** | 0.015 |
| | (0.020) | (0.019) | (0.017) |
| Improved seed available in the village (1=yes; 0=no) | -1.201*** | -1.932*** | -2.284*** |
| | (0.292) | (0.255) | (0.257) |
| Number of groundnut plots | 0.049 | 0.020 | 0.037 |
| | (0.041) | (0.038) | (0.037) |
| Household head farm labour participation (1=full time; 0=otherwise) | -0.002 | 0.277 | 0.190 |
| | (0.246) | (0.238) | (0.227) |
| Source of groundnut seed (1=formal; 0=informal) | 0.125 | -1.442*** | -1.290*** |
| | (0.225) | (0.220) | (0.222) |
| Household members engaged in agricultural technology training (1=yes; 0=no) | 0.145 | -1.093*** | 0.140 |
| | (0.252) | (0.258) | (0.235) |

| | | | |
|---|---|---|---|
| Sources of groundnut variety information (1=formal; 0=informal) | 0.227<br>(0.311) | -0.137<br>(0.273) | -0.362<br>(0.287) |
| Membership to farmer groups (1=yes; 0=no) | 0.042<br>(0.217) | -0.046<br>(0.193) | -0.284<br>(0.196) |
| Visited by agricultural extension staff (1=yes; 0=no) | -0.859***<br>(0.257) | -0.179<br>(0.213) | -0.676***<br>(0.213) |
| TL III project target village (1=yes; 0=no) | -0.375<br>(0.254) | -0.419**<br>(0.212) | -0.919***<br>(0.209) |
| Jigawa (1=yes; 0=otherwise) | -0.909**<br>(0.456) | 0.717**<br>(0.346) | 0.435<br>(0.350) |
| Kano (1=yes; 0=otherwise) | 1.255***<br>(0.346) | 0.308<br>(0.321) | 0.077<br>(0.303) |
| Katsina (1=yes; 0=otherwise) | 0.866**<br>(0.341) | 0.522*<br>(0.316) | 0.495<br>(0.301) |
| Kebbi (1=yes; 0=otherwise) | 0.371<br>(0.495) | 2.445***<br>(0.363) | 2.032***<br>(0.356) |
| Constant | -0.333<br>(0.506) | 0.165<br>(0.484) | 0.819*<br>(0.450) |
| Model description: | | | |
| Number of observations | 1,428 | | |
| LR chi2(48) | 833.310 | | |
| Prob > chi2 | 0.0000 | | |
| Pseudo R2 | 0.220 | | |

| | |
|---|---|
| Log likelihood | -1480.386 |

*** p<0.01, ** p<0.05, * p<0.1; values in parenthesis show standard errors