ARE WE MEASURING WHAT WE THINK WE ARE MEASURING? AND DOES IT MATTER?

Summary of Validation Studies Using DNA Fingerprinting

Mywish K. Maredia
maredia@msu.edu
Focus of this presentation

Two pilot studies conducted in 2013-14 in Africa under SIAC objective 1 (Methods)

- Cassava (*Manihot esculenta*) in Ghana
- Beans (*Phaseolus vulgaris*) in Zambia
Collaborators on these two studies

- Byron Reyes, Currently at CIAT

For Cassava study:
- Joe Manu, Ghana Crops Research Institute, and Awere
- Dankyi, Agriculture Innovations Consult (AIC)
- Peter Kulakow, Ismail Rabbi, Elizabeth Parkes,
- Tahirou Abdoulaye and Gezahegn Girma from IITA
- Ramu Puna (Cornell)

For Beans study:
- Petan Hamazakaza and Kennedy Mui Mui, Zambia Agricultural Research Institute
- Enid Katungi, Bodo Raatz, Clare Mukankusi and Allan Male, CIAT
Acknowledgement

- Funding support from the Standing Panel on Impact Assessment (SPIA) of the Independent Science and Partnership Council under the Strengthening Impact Assessment in the CGIAR (SIAC) program

- Partial funding support from the CRP on Roots and Tubers and PABRA, and technical support from IITA and CIAT

- ALL the enumerators, field staff and farmers / respondents who participated in the survey and shared their time and knowledge
Objectives

■ To test the effectiveness of different methods of varietal identification against the benchmark of DNA fingerprinting

■ To come up with ‘lessons learned’ and recommendations on methods / approaches that can be scaled up

These were PILOT studies - Not representative at the country level

■ Ghana study: representative of three regions – Brong Ahafo, Ashanti and Eastern (which account for 61% of cassava production)

■ Zambia study: representative of two provinces – Muchinga and Northern (which represent 70% of bean area in the country)
Methods of varietal identification tested in the pilot studies

<table>
<thead>
<tr>
<th>Methods</th>
<th></th>
<th>Ghana (cassava)</th>
<th>Zambia (beans)</th>
</tr>
</thead>
<tbody>
<tr>
<td>V</td>
<td>DNA fingerprinting (used as a benchmark to compare/validate other methods)</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>A</td>
<td>Farmer elicitation (A1-name and A2-type of variety) \a</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>B</td>
<td>Farmer elicitation based on series of photographs of plants and later identifying varieties based on morphological characteristics</td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>C</td>
<td>Farmer response on type of variety he/she had planted that match seed samples presented by the enumerators</td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>D</td>
<td>Trained enumerators/experts visiting the field and: 1. Recording observations on varietal characteristics (phenotyping); and 2. Identifying the variety based on observation (phenotyping)</td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>E</td>
<td>Taking photos of the plant in the field or seeds harvested by farmers for latter identification by experts (i.e., breeders, agronomists, etc.)</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>F</td>
<td>Collecting harvested seeds from farmers for latter identification by experts (i.e., breeders, agronomists)</td>
<td></td>
<td>X</td>
</tr>
</tbody>
</table>
DNA fingerprinting

- In both the pilots, DNA fingerprinting was used as a benchmark against which alternate approaches were evaluated / validated (method V).

- This involved first establishing a reference library of DNA fingerprints, and then applying the same or a sub-set of markers used to establish the reference library to genotype the samples (plant tissues or seeds) collected during the farm surveys.

- Sampling:
  - **Cassava:** Leaf tissues from one plant representing each variety on one plot per HH
  - **Beans:** 10-15 seeds representing each variety harvested in the previous season by a HH
MAIN RESULTS
Measures of effectiveness of different methods of varietal identification for Beans in Zambia: Comparison of outcomes against the benchmark of DNA fingerprinting

<table>
<thead>
<tr>
<th>Measures of effectiveness</th>
<th>Method A1 Farmer reported name</th>
<th>Method A2 Farmer reported type</th>
<th>Method C Farmer identification based on matching seed samples</th>
<th>Method E Expert group opinion based on photos</th>
<th>Method F Expert group opinion based on seed sample</th>
<th>Benchmark (DNA)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Varietal data points compared</td>
<td>N 855</td>
<td>802</td>
<td>783</td>
<td>736</td>
<td>847</td>
<td>855</td>
</tr>
<tr>
<td>Outcome: Number of data points classified as ‘released’ (or improved) varieties</td>
<td>N 34</td>
<td>106</td>
<td>553</td>
<td>135</td>
<td>123</td>
<td>141</td>
</tr>
<tr>
<td>% 4%</td>
<td>13%</td>
<td>71%</td>
<td>18%</td>
<td>15%</td>
<td>16%</td>
<td></td>
</tr>
<tr>
<td>Type I Error</td>
<td>A local variety incorrectly identified as IV</td>
<td>N 21</td>
<td>85</td>
<td>441</td>
<td>101</td>
<td>84</td>
</tr>
<tr>
<td>% 62%</td>
<td>80%</td>
<td>80%</td>
<td>75%</td>
<td>68%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Type II Error</td>
<td>An improved variety incorrectly identified as a local variety or by an incorrect IV name</td>
<td>N 129</td>
<td>108</td>
<td>59</td>
<td>87</td>
<td>97</td>
</tr>
<tr>
<td>% 91%</td>
<td>84%</td>
<td>48%</td>
<td>73%</td>
<td>70%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Accuracy of name: Data points correctly identified as IV by name</td>
<td>N 12</td>
<td>--</td>
<td>63</td>
<td>32</td>
<td>41</td>
<td></td>
</tr>
<tr>
<td>% 9%</td>
<td>52%</td>
<td>27%</td>
<td>29%</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Accuracy of category: Data points correctly identified as IV by Type</td>
<td>N --</td>
<td>21</td>
<td>--</td>
<td>34</td>
<td>44</td>
<td></td>
</tr>
<tr>
<td>% 16%</td>
<td>29%</td>
<td>32%</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Main findings from Zambia case study

- When simply asked for the name and type of bean variety a farmer planted, they underreported the adoption of improved varieties.

- However, when shown the seed samples of IV, they substantially overreported the adoption of varieties matching the seed samples (which in this case were all IVs).

- Methods based on expert elicitation ex post of the survey based on photos or seed samples collected from the farmers gave overall adoption estimates closest to the benchmark.

- But **ALL the methods** had high type I and type II errors or low accuracy rates when the outcome is compared for each data point.
Ghana Results: Variety names as reported by farmers

- About **180 variety names** reported across 914 accessions collected from farmers’ fields
- Names of collected accessions are shown on the right. The font size is proportional to the counts
- Most common variety names are Debor and Ankra
DNA results for cassava in Ghana

- Several interesting findings
  1. Some improved varieties are genetically identical (e.g., IFAD and UCC)
  2. Many released varieties are hybrids or admixtures
  3. Library accessions representing both ‘released varieties’ and ‘landraces’ fall under the same varietal cluster groups (e.g., variety group 2, 3, 4)

- The last bullet point (#3) especially poses a challenge for varietal identification

- The problem is: How to classify farmer samples that fall in these three variety cluster groups? Should they be classified as ‘improved/released’ varieties or local/landrace varieties?
DNA results for cassava in Ghana (cont’d)

- To address this dilemma, the analysis of effectiveness of different methods against the benchmark of DNA fingerprinting is done under two scenarios / assumptions
  - **Liberal scenario**: which assumes that all the farmer samples that fall in a variety cluster in which there is at least one released variety are essentially improved varieties
  - **Conservative scenario**: assumes the opposite (i.e., farmers samples that match the DNA results of a variety group in which there are both released varieties and landraces, the variety group is considered not-improved).
Main findings from the cassava case study

- Individual methods studied provided different estimates of adoption of improved varieties, ranging from 2% to 16%.
- Like Zambia study, no one method stands out to be most effective on all measures.
- On aggregate level adoption of IV, methods based on farmer elicitation and field observations by an expert provide closest estimates under the conservative scenario, but with high error rate.
- Identifying cassava varieties accurately by NAME in a setting where hundreds of variety names exist is a challenge across all the methods tested.
- Adoption estimates by the experts (based on photos) are substantially higher than other methods and has much higher type I error (false positives).
Implications and Key Insights from this Study...
Key methodological implications for tracking varietal adoption

1. No one method stood out to be most effective in identifying varieties accurately on all measures of effectiveness. 

   Thus, we are not able to recommend any method tested in these two pilots as an identification method for tracking varietal adoption

2. Results for both beans and cassava indicate that when there is a diversity of names by which farmers call their varieties and the informal seed system is dominant (or the seed system is ineffective in maintaining purity), the traditional method of farmer elicitation will give an underestimate of adoption of improved varieties by names.

   Thus the current gold standard of eliciting varietal adoption from farmer surveys may not be an accurate method for measuring varietal turnover and assessments of type II benefits of plant breeding research (i.e., benefits from varietal replacement)
3. Methods to measure adoption of ‘improved varieties’

Data from farmers’ surveys and experts are better able to give an aggregate assessment of the adoption of improved varieties as a category than variety specific adoption. However, all the methods evaluated are prone to both type I and type II errors which has implications on the accuracy of analyses that use adoption estimates derived from such methods as an outcome variable (as is done in the determinant of adoption studies) or as one of the treatment variables (as done in impact studies that try to estimate the causal link between varietal adoption and welfare outcomes).

How significant are these potential errors in such analyses needs to be investigated and a topic of further research.
Implications on Seed System R&D

4. The results of these two studies raise critical questions on the reliability and effectiveness of the seed system in developing countries.

More attention needs to be devoted in developing a functional and effective seed system that can:

1. Ensure ‘Distinctiveness’ of released varieties (as determined by DNA fingerprinting)

2. Maintain the purity of seeds throughout the varietal release, multiplication of early and later generation seeds, and during the dissemination/sale of varieties as per the identity claims.
Lessons learned on DNA fingerprinting as a method of varietal identification, its potential, and limitations....

...See Maredia et al. (2016) – MSU IDWP Paper 149 for this discussion.
Why don’t farmers know which varieties they are growing?

- Could be many reasons....
- But not a focus of this presentation
What are the implications on the broader impact analysis METHOD that farmers don’t know which varieties they are growing?

1. Is this ‘knowledge gap’ one of the reasons for the **Yield gap**?
2. Isn’t the welfare ‘impact’ of a varietal technology a result of three things?
   - Its genetic quality (i.e., traits embedded in seeds)
   - Its seed quality in the planting season; and
   - The behavioral response of farmers (i.e., decisions on other inputs and practices)?
So what are the implications on the broader impact analysis METHOD that farmers don’t know which varieties they are growing? (cont’d)

Points to ponder...

a. If farmers don’t know or know incorrectly which varieties they are growing, will finding out through DNA fingerprinting lead to any more accurate or credible estimates of impacts of MVs if farmers’ behavior is already influenced/altered by what the farmer believes he/she is growing?

b. Doesn’t type I and type II errors completely change the counterfactual and thus the accuracy of impact estimates even with DNA fingerprinting?

c. Thus, beyond tracking varietal adoption and varietal turnover, does DNA fingerprinting for the purpose of ‘determinants of adoption’ and ‘ex post IA’ really matter?
THANKS