Is DNA Fingerprinting the Gold Standard for Estimation of Adoption and Impacts of Improved Lentil Varieties?

Yigezu A. Yigezu[1], Jeffrey Alwang[2], M. Wakilur Rahman[3], M. Bazlur R. Mollah[3], Tamer El-Shater[1], Aden Aw-Hassan[1] and Ashutosh Sarker[1]*

[1] International Center for Agricultural Research in the Dry Areas (ICARDA), Amman, Jordan
[2] Virginia Tech, 250 Drillfield Drive, Blacksburg, VA 24061.
[3] Bangladesh Agricultural University (BAU)

*Corresponding author: Department of Agricultural and Applied Economics, Virginia Tech, 250 Drillfield Drive, Blacksburg, VA 24061, Tel. +1-540-231-6517, e-mail: alwangj@vt.edu

Abstract: In the early 1980s, disease susceptibility in short-season lentil landraces began to limit productivity in areas where relay cropping took place in Bangladesh. Since then, several improved high-yielding lentil varieties, which are resistant to rust and blight and suitable in the relay cropping system, have been released jointly by national and international research centers. This study used three methods, namely a panel of experts, a survey of 1,000 households where the respondents named the variety they used, and DNA fingerprinting of seed samples collected from all lentil plots cultivated by survey households to estimate adoption. Double hurdle and instrumental variables regression methods were applied to the household survey and DNA fingerprinting data to identify determinants of adoption and measure their impacts. Of particular interest was whether estimates of adoption, determinants of adoption and impacts varied by method of variety identification. Results showed that the expert panel overestimated the adoption of more recent varieties while about 89 percent of the farmer-reported varieties were accurate, as verified by DNA fingerprinting. DNA fingerprinting appears to have little advantage for estimating the level of adoption in this case, where few varieties of lentils are found, local variety names do not exist, and most seed is obtained through a formal system. However, even under these conditions, determinants of adoption vary by identification method, and use of farmer-reported information on the variety can lead to erroneous conclusions about determinants of adoption. Because recent breeding efforts have focused on taste and cooking considerations, yield impacts were not significantly different from zero.

Keywords: Adoption and impact estimation; DNA fingerprinting; lentils; household survey; expert panel; Bangladesh.

Funding: Funding was provided by the Standing Panel on Impact Assessment of the CGIAR under the project "Strengthening Impact Assessment in the CGIAR System", signed 1 July 2015.

1. Introduction

Agricultural research improves productivity, raises incomes among farmers and others and reduces global poverty by lowering prices of major food staples. Efforts to assess the impacts of agricultural research begin by focusing on the area planted to a particular variety or under a particular management practice (Alwang et al. 2017). Extent of diffusion, together with per-area changes in yields, cost of production, labor use and other outcomes help determine global impacts. Variety identification is a critical part of impact assessment; without knowledge of diffusion and evidence of the effect on per-area outcomes, the impact of variety-specific research cannot be measured. When farmer own assessments or expert opinions are used, substantial error in variety identification can occur (Walker and Alwang 2015).

Recent innovations have lowered the cost of DNA fingerprinting and made the method a feasible alternative to relying on farmer identification (Floro et al. 2017; Kosmowski et al. 2016; Maredia et al. 2016). Despite the attractiveness of using DNA fingerprinting to identify varieties, to date there have been few assessments of how fingerprinting affects estimates of diffusion, adoption and on-field impacts of new technologies. This paper presents such an assessment with a focus on improved lentil varieties in Bangladesh.

The paper contributes to the literature in three ways. First, we provide an assessment of the need to conduct DNA testing in a case where few varieties of a crop are planted, and farmer identification is likely to be good. Existing literature shows large gains to DNA testing in cases where many varieties are planted, variety names vary by locality, and informal seed systems predominate (Floro et al. 2017; Kosmowski et al. 2016; Maredia et al. 2016). In these cases, DNA testing is necessary to obtain accurate estimates of diffusion. The lentil case is likely to be different, because few varieties are available and lentil seed markets in Bangladesh are largely formal. Second, we examine how inferences about determinants of adoption and extent of adoption are affected by variety identification. Floro et al. (2017) is the only known study that examines differences in parameter estimates by type of identification, but they examine impacts on coefficients reflecting the determinants of adoption, not area planted, which is an important indicator of diffusion. Third, the paper analyzes yield impacts of new varieties in a case where the focus of the breeding research is on factors other than yield.

2. Background: Lentils in Bangladesh

Global efforts documenting the contribution of agricultural research to enhanced food security and other outcomes have focused on major cereal food crops such as wheat, rice and maize (Walker and Alwang 2015)[1]. Less attention has been paid to the quiet revolution associated with diffusion of modern varieties (MV) of food legume crops. Legumes are vital components in diversification of Bangladesh's rice-based cropping system, constituting the major source of dietary protein and providing several essential micronutrients (Ali et al. 2014; Datta et al. 2013; Singh et al. 1994). Due to their nitrogen-fixing ability, legumes can improve soil fertility and agricultural sustainability (Schmidtke et.al. 2004).

Prior to the 1980s, lentil was produced in Western Bangladesh during the dry season, as few alternatives existed to follow the rainy season (Aman) rice harvest. Lentil cultivators take advantage of residual moisture from the rice crop and plant the crop with few purchased inputs. In the 1980s, area expansion of irrigation infrastructure enabled irrigated dry season (Boro) rice

---

[1] Of nine food legume groups covered by the 2015 Diffusion and Impact of Improved Varieties in Africa (DIIVA) study of modern varieties in Africa (Walker and Alwang 2015), seven were new (not covered in the Evenson and Gollin, 2003 global study).

to displace lentil production (Sarker et al. 2004). In a drive for cereal self-sufficiency, the government promoted Boro rice production using price supports and outreach messages. Farmers, however, discovered that short-season lentil varieties can be planted between rice crops using a method known as relay cropping[2], and lentil production subsequently rebounded in transplanted rice areas (Sarker et al. 2004).

By the mid-1980s, susceptibility of short-season lentil landraces to two important diseases -Stemphylium blight (*Stemphylium botryosum*) and rust (*Uromyces fabae*) began to constrain production in relay-cropped regions. The Bangladesh Agricultural Research Institute (BARI) introduced improved varieties with disease resistance from India, but these did not fit into the short-season system. Suitability in the transplanted rice relay system is a paramount concern to producers and, without suitable varieties, the lentil-production components of the system were in danger of collapse.

A partnership between the International Center for Agricultural Research in the Dry Areas (ICARDA) and BARI began in the late 1980s in response to the lentil disease crisis. Since the 1990s, several improved varieties have been released with resistance to main diseases and yields that exceed those of landraces (table 1). All released varieties fit into the relay cropping system (Sarker et al. 2004) and area planted to lentils has grown steadily. After 1996, breeding objectives by ICARDA/BARI and the Bangladesh Institute of Nuclear Agriculture (BINA) pivoted toward a focus on consumption qualities and cooking time, and while 11 MVs have been released since 2001, experimental yields have not increased significantly (table 1).

---

[2] Relay cropping involves broadcast sowing lentil seeds directly into rice lands prior to the rice harvest. The lentil crop is established prior to rice harvest, taking advantage of residual soil moisture.

Table 1: Improved Lentil varieties released in Bangladesh, their characteristics and area share (based on DNA fingerprinting)

Name	Release year	Organizational origin	Characteristics	Maturity (days)	Yield (ton/ha)	Area share (%) from DNA	Rank of area share
BARI Masur 1	1991	BARI/ICARDA	High yield and rust resistance; white flower color	105-110	1.7-1.8	0.17	9
BARIMasur2	1993	BARI/ICARDA	High yield and rust resistance; tendril present at leaf.	105-110	1.5-1.7	0.29	8
BARI Masur 3	1996	BARI	High yield and rust resistance; seed coat is greyish and spotted, large seed	100-105	1.5-1.7	29.86	2
BARI Masur 4	1996	BARI/ICARDA	Resistance to blight and rust; high iron, high yield	110-115	1.6-1.7	22.37	3
BARI Masur 5	2006	BARI/ICARDA	Resistant to blight and rust; tolerant to foot rot; high yield	110-115	1.4-1.6	10.54	4
BARI Masur 6	2006	BARI/ICARDA	Resistant to SB and rust; tolerant to foot rot; high in iron and zinc; high yield.	105-110	2.2-2.3	31.24	1
BARI Masur 7	2011	BARI/ICARDA	Tolerance to SB and rust; red color; High yield. Good cooking quality; high crude protein (30-31%)	110-115	1.8-2.3	3.77	5
BARI Masur 8	2015	BARI/ICARDA	Tolerance to SB and rust; micronutrient-dense variety (Fe and Zn); can be planted late and high yield.	110-115	1.8-2.0	0.00	10
BINA Musur1	2001	BINA	Seed coat color is black, Grain is reddish yellow, the variety is tolerant to SB	125-130	Avg. 1.8	0.00	10
BINA Musur 2	2005	BINA	Early maturing, red color, good cooking quality, high protein (24-25%)	95-100	Max. 1.9 Avg. 1.8	0.00	10
BINA Musur 3	2005	BINA	Moderately resistant to rust, foot and root rot/wilt diseases, pod borer and tolerant to mild water stress, late sowing potential	95-100	Max. 2.4 Avg. 1.8	0.00	10
BINA Musur 4	2009	BINA/ICARDA	Moderately resistant to rust and diseases; good cooking quality	96-102	1.8	0.00	10
BINA Musur 5	2011	BINA/ICARDA	Tolerant to blight and rust diseases, red color with good cooking quality, high protein (29-30%)	99-104	Max- 2.2; Avg. 2.15	0.62	6
BINA Musur 6	2011	BINA/ICARDA	Tolerant to blight and rust diseases, red color with good cooking quality, high protein (30-31%),	105-110	Max. 2.0 Avg. 1.95	0.51	7
BINA Musur 7	2013	BINA	High yielding and tolerant to SB and rust	110-112	2.2 to 2.4	0.00	10

Sources:

1. Legumes varieties, Handbook of Agricultural Technology, Bangladesh Agriculture Research Institute, Available at: http://bari.portal.gov.bd/sites/default/files/files/bari.portal.gov.bd/page/1c204f38_4e6b_4ec6_a394_8af0a7f9088e/%E0%A6%AB%E0%A6%B8%E0%A6%B2%20%E0%A6%B8%E0%A6%AE%E0%A7%82%E0%A6%B9%20%281%29.pdf
2. Islam M N, MS Rahman, MS Alom and M. Akhteuzzaman (2015). Performance of different crops productivity enhancement thorough adaptation of crop varieties at Charland in Bangladesh. *Bangladesh Journal of Agricultural Research* 40 (4): 629-640.

The ICARDA/BARI partnership is seen as a major success as introduction of disease-resistant short-season varieties appears to have contributed to sustainability of the rice-lentil system (Sarker et al. 2004). This perception, however, has not found scientific support as little information is available on the spread and impacts of these new varieties. Lentil scientists in Bangladesh state that improved varieties have almost completely displaced landraces and have increased yield and farm income, but no verifiable evidence of these claims is available.

3. Methods

Release of MVs is a major mechanism by which agricultural research can improve productivity, enhance food security and reduce poverty. Variety identification is the first step in research evaluation as this information is used to measure aggregate diffusion and to classify observations (improved or not improved) for statistical analysis. Three methods are widely used for measuring diffusion: expert opinion, farmer reporting, and DNA fingerprinting. Expert opinion involves systematically gathering information from experts about their perceptions of variety spread, while farmer reporting is usually employed within the context of farm-household surveys. Estimates of adoption can be obtained by asking a random sample of farmers about which variety they plant and over how much area. Farmers often misidentify the variety(ies) they plant or are unaware of the name. Farmer-reported identification depends on the country and crop, but most studies show that farmer misidentification is quite common, with misclassification reaching rates of up to 71 percent (Maredia et al. 2016).

Misidentification of a variety as improved can lead to erroneous inferences. DNA fingerprinting of germplasm can verify the variety and ensure the accuracy of farmer reporting.

DNA analysis is considered to be a gold standard against which other variety identification strategies are compared[3] (Floro et al. 2017; Maredia et al. 2016).

Methods to measure diffusion imply different costs and, depending on the purpose of the analysis, different degrees of accuracy. As costs of DNA testing decline, the method is gaining in popularity (Kosmowski et al. 2016; Maredia et al. 2016; Labarta et al. 2015). Use of DNA fingerprinting is still nascent, however, and little evidence exists about how estimates of diffusion and evaluation of factors associated with adoption are affected by errors in variety identification. Floro et al. (2017) find that farmers in Colombia overestimate their use of modern cassava varieties and misidentification is associated with biased estimates of some determinants of adoption. In particular, the presence of more dependents in the farming family is positively associated with adoption of MVs when the variety is identified by the farmer, but negatively associated with it when DNA testing is used. Access to extension services has a strong positive association with adoption with farmer identification, while it is insignificant when DNA testing is used. CGIAR (2016) reports results from several studies and find that misclassification is ubiquitous and estimates of MV diffusion may vary substantially depending on the crop and country.

3.1 Expert panels and variety diffusion

While expert opinion estimates are subject to well-known biases, when elicited through a structured process they can be a good approximation of overall diffusion and are less expensive than survey sampling (Walker and Alwang 2015). Evenson and Gollin (2003) used

---

[3] The efficacy of DNA testing depends in part on the completeness of the reference library and the degree of variety homogeneity in the farmer's field (CGIAR 2016).

expert panels (and expert opinion) to measure diffusion world-wide, while Walker and Alwang (2015) assessed diffusion of 225 crop-country combinations in Africa using improved expert panel methods. Individual CGIAR centers use their own versions of panel procedures, and Walker (2015) summarizes lessons learned from these and other studies.

Expert panel estimates of diffusion of improved lentil varieties were derived during a meeting in Pubna district that included the research team from ICARDA and the Bangladesh Agricultural University (BAU), breeders from the Pulse Research Center (PRC) of BARI, BINA, and extension agents from neighboring districts. Participants were asked to consult one another and arrive at a consensus about the lentil area and share planted to local and improved varieties in each of the ten main lentil growing districts in Western Bangladesh (table 1).

3.2 Farmer identification

As an alternative to and a check of expert panel estimates, farmer own reporting and DNA fingerprinting are used. Careful enumeration of the farmer had been, until recently, the favored means of identifying varieties planted (Walker 2015). Farmer reporting, however, is fraught with difficulty when many varieties are available, substantial amounts of grain are recycled for use as seed, where local names that vary across locations are common, informal seed systems predominate, and when the plant morphology makes the MV difficult to distinguish from traditional varieties (CGIAR 2016; Maredia et al. 2016). For example, Larochelle et al. (2015) collected grain samples from Rwandan farmers to verify self-reported bean variety names. More than 400 variety names were claimed during a survey of 1,298 bean-producing households; experts were able to group these into 165 unique varieties, but neither farmers nor experts could identify about 15% of the bean samples.

Errors in farmer identification of the lentil variety were not expected to be a problem in Bangladesh, where few MVs have been released (table 1). In addition, the Bangladesh lentil seed system is characterized by a high degree of formality. Lentil seeds are produced by contract farmers overseen by the Bangladesh Agricultural Development Corporation (BADC), which distributes them widely through regional centers and affiliated seed dealers. Dealers are subject to quality checks from the national Seed Certification Agency of the Ministry of Agriculture (MoA). The result is a robust formal seed supply system with seed prices controlled by the MoA and few informal seed transactions occurring.

To verify that farmers are able to identify their varieties and to examine how misidentification might affect the empirical analysis, the study employed DNA fingerprinting of seed samples obtained from farmers during enumeration of the household survey.

3.3 DNA fingerprinting

We DNA fingerprinted lentils collected in each field where surveyed households planted lentils. The process began with collection of reference seed samples. Breeder seed samples were obtained for all released varieties from the BARI and BINA lentil breeding programs. During the household survey, samples of seeds were collected from the 1694 plots planted by 1000 sampled farmers. Prior to the analysis, a set of five Inter Simple Sequence Repeats (ISSR) and 41 Simple Sequence Repeats (SSR) markers (Andeden et al. 2015; Verma et al. 2015; Gupta et al. 2012; Hamwieh et al. 2009; and Pradeep et al. 2002) were identified from the lentil genome covering different linkage groups and individually tested for polymorphism in the breeder samples. Two ISSR and 20 SSR markers showed significant polymorphism across the Breeder seed samples and were used for varietal identification.

The allele sizes of the markers were determined by comparing the base pairs of DNA ladder and the SSR allele using the imageJ software (Schneider et al. 2012) and a data matrix was constructed. The data were subsequently analyzed using GenAlEx software package (Peakall and Smouse 2012) to calculate DNA Profile attributes.

All distinct bands or fragments were identified according to size and scored visually based on their presence or absence. Scores obtained using primers were pooled to create a single data matrix. This matrix facilitates comparison of frequencies of all polymorphic ISSR markers among lentil varieties using Version 1.31 of the POPGENE software (Yeh et al. 1999). These data matrices were used to generate genetic information on each lentil variety. Only unambiguous bands were scored for the estimation of genetic similarity between the varieties using Jaccard’s similarity coefficient. Then, DNA obtained from field samples was compared with DNA extracted from pure seeds of the released variety (benchmark DNA) for verification.

While estimates from the expert panels were obtained for districts, estimates using farmer reporting and DNA fingerprinting were obtained by planted plot. Once estimates were obtained using each method, total lentil area in the respective units were aggregated from lower levels to higher levels (household, village, district and region). The survey data were aggregated to region and national levels using the survey weights derived from our sampling procedure (discussed below).

3.4 Defining “improved”

Standard practice in MV diffusion and impact assessment is to consider varieties released after a certain date as improved and those released before that date as unimproved. For example, the DIIVA studies (Walker and Alwang 2015) used 1998 as a cutoff because one of the objectives was to compare differences in MV diffusion in Africa since the last comprehensive global assessment (Evenson and Gollin 2003). Use of arbitrary cutoffs ignores the fact that crop improvement involves a steady flow of MV releases and adopters of newer varieties often replace older, but still improved, varieties. An arbitrary cut-off date implies that the counterfactual is what productivity (or another outcome measure) would have been had the farmer planted either unimproved or earlier-released improved releases. We employ different definitional cutoffs and examine the sensitivity of the findings to these cutoffs.

3.5 Adoption and intensity of adoption

Following Just and Pope (1978) and Feder (1980), we assume the producer choses to adopt an MV based on maximization of expected utility of income. Production of lentils is associated with risk and farmers, when faced with the possibility of adopting a new technology, are uncertain about profitability of earnings from new varieties. Attributes affecting risk aversion such as farmer age, wealth and family structure should all affect the decision to adopt. As a result of these considerations, the literature identifies a number of variables that enter into the adoption decision and we include as many of these as feasible.

Literature on agricultural technology adoption commonly uses two definitions as the outcome of technology choice. The most common is a binary measure, taking the value of zero if no land is devoted to an MV and one if any land is. An alternative indicator is known as the adoption intensity which is sometimes expressed as the share of land planted to the MV. This measure reflects the fact that adoption can be partial. The empirical models examine the determinants of whether lentil farmers adopt MVs and, if so, the area over which they adopt.

A double-hurdle (Cragg 1971) framework is employed; this framework incorporates the idea that the decision to adopt a new variety results from two sub-decisions: the first hurdle, determining whether the decision maker would ever adopt, and the second, determining the intensity (area) of adoption (Rickert-Gilbert et al. 2011; Amankwah et al. 2016). The basic idea is that part of the sample is comprised of farmers who would never adopt, while others might, and the latter group's decision to adopt might be affected by different variables and in different ways than their decision about intensity of adoption.

The decision to adopt the MV lentil is modeled as a binary function; the latent variable underlying household i's decision to use modern lentil varieties MVi* is specified as:

$$MV_i^* = x_{1i}'\beta_1 + \varepsilon_{1i} \qquad \textbf{(1)}$$

Where the vector x_{1i} reflects determinants of the adoption decision, β_1 are parameters, and ε_{1i} is a normally distributed error term with mean zero and constant variance. The corresponding probit is estimated on the observed outcome $MV_i=1$ if $MV_i^*>0$ and 0 otherwise.

The desired area planted to MVs is also an unobserved latent value that can be specified as:

$$A^*_i = x_{2i}'\beta_2 + \varepsilon_{2i} \qquad \textbf{(2)}$$

where x_{2i} are determinants of area, β_2 are parameters and ε_{2i} is a normally distributed error term. Since A^*_i is a latent variable, we work with observed area (A_i). Observed area = A^*_i if $MV_i^*>0$ and = 0 if $MV_i^* \leq 0$. Because we use observed area, the error term is a truncated normal distribution. The parameters β_1, β_2 can be estimated separately because the Cragg likelihood function is separable; the marginal effects, however, need special attention (Burke 2009).

Elimination of uncertainty about whether the variety is improved allows investigation of the effects of misreporting on parameter estimates. Two parameter estimates are available:

$\beta_j^{R,D}$, where j=1,2 (based on the equation estimated) and the super-script refers to farmer self-reported (R) and DNA tested (D). The D estimates are assumed to be the correct ones.

4. Data

Western Bangladesh is the main lentil-growing region and its ten-major lentil-growing districts constitute about 74 % of total national lentil area. As district-level data on the number of lentil growers was not available from the Bangladesh Bureau of Statistics (BBS), the survey team made telephone calls to the bureaus of agriculture and collected estimates of numbers of lentil growers in each of the ten districts. Using the estimated number of growers, power analysis was carried out and the minimum sample size required to ensure 95% confidence and 2% precision on the level of adoption was determined to be 864; this number was increased to 1000 households to gain additional power.

We subsequently learned that lentil is a minor crop in Bangladesh and unlike major crops such as rice, for which detailed, and statistically defendable area estimates are generated by BSS, area estimation for lentils uses a “subjective estimation method” (BBS 2017). This method involves subjective estimates by statistical office staff using the opinions of five farmers per upazila calibrated using information from the 2008 Agricultural Census. Use of this method casts serious doubt about the accuracy of the official area estimates. As a result, the sample was distributed across the ten districts proportional to the number of lentil growers (instead of lentil area) in each district. As there were no secondary data on number of growers at the sub-district level, the team took random samples of two sub-districts from each district and 20 farm households from each village. The number of villages to be randomly selected from each sub-district was then calculated as the ratio of the total sample size and the sample size per village.

The sample of 1,000 households was distributed among 20 sub-districts and 52 villages. After villages were randomly drawn, survey teams collected information on the number of lentil growers and lentil area from each village. Then, the survey team, in collaboration with village agricultural sub-officers and officers (where available), prepared a sampling frame with the list of farmers who planted lentil during the previous (2015) season. Between 16 and 23 households were randomly selected, with the village sample size based on the number of lentil growers. Sample weights were constructed to represent the probability of being surveyed and population estimates were obtained using these weights. Since the sample is representative at the 10-district level, the weighted estimates of total area under lentils and area under improved lentil varieties are comparable to the aggregate estimates from the expert panel.

The questionnaire covered household demographic and economic conditions, asset ownership and other relevant factors. Information on lentil farming was obtained by asking detailed questions on varieties planted, input use, management practices, yields and use of production for all lentil plots cultivated. Community-level information on access to infrastructure, farm services, extension, etc. was obtained from a separate village-level survey.

5. Results

5.1 Variety diffusion

As noted, any variety formally released could be called "improved" and we begin by presenting estimates of aggregate MV diffusion depending on the release date (table 2). Experts agreed that varieties released before 2006, even though called "improved" in the official varietal catalog, are too old to be considered improved. For the analysis, adoption is defined as the use of lentil varieties released in 2006 or after for at-least two years.

The experts believe that in addition to local varieties, 9 of 15 improved varieties released since 1991 are currently being planted – and 5 of 8 MV released on or after 2006 are also currently planted. Estimated adoption improved lentil varieties released on or after 2006 in Western Bangladesh was 67% (table 2). Covering 28% of lentil area, BARI-6 is estimated to be most widely-planted followed by BARI-7 and BARI- 3 which cover 20 and 19%, respectively. According to the experts, BARI-1, BARI-2 and BARI-8 varieties are not currently being planted.

Farmer reporting showed that five (BARI-3, BARI-4, BARI-5, BARI-6, and BARI-7) of 15 varieties released since 1991 are currently planted. All were developed by the ICARDA/BARI breeding partnership. Estimated adoption of varieties released on or after 2006 was 49%, where 17, 14 and 10% is covered by BARI-6, BARI-3 and BARI-4, respectively. Chuadanga and Jhenaidah districts have the highest (67%) and lowest (31%) adoption respectively. In districts with low adoption of latest varieties, BARI-3 and BARI-4 are still the dominant varieties.

DNA fingerprinting showed that the area under MVs released on or after 2006 was 45% and small areas (0.5%) were planted to unclassified varieties[4]. With area coverage of 31, 30, and 22% respectively, BARI-6, BARI-3 and BARI-4 dominate the landscape.

---

[4] Ungrouped varieties are those that did not match seeds in the reference library. All ungrouped were claimed by farmers to be local varieties.

Table 2: Estimates of aggregate diffusion (in June 2015) by method of estimation and time when varieties were released

Release dates	**Method of estimation**								
	Expert panel			**Farmer recall**			**DNA fingerprinting**		
	Total area (ha)[a]	**Area share (%)**	**Cumulative (%)**	**Total area (ha)[a]**	**Area share (%)**	**Cumulative (%)**	**Total area (ha)[a]**	**Area share (%)**	**Cumulative (%)**
1991-1995	0	0.00	0.00	0	0.00	0.00	540	0.50	0.50
1996-2005	34,049	31.66	31.66	53,248	49.51	49.51	57106	53.10	53.60
2006-2014	72,605	67.51	99.17	52,742	49.04	98.55	48831	45.40	99.00
Land races	895	0.83	100.00	1,559	1.45	100.00	DK[d]	DK[d]	DK[d]
Ungrouped[b]	0	0	100.00	0	0	100.00	775	0.72	99.72
Error[c]	0	0	100.00	0	0	100.00	297	0.28	100.00
Total	107,549	100.00	100.00	107,549	100.00	100.00	107,549	100.00	100.00
All improved varieties (regardless of release date)	106,654	99.17	97.17	105,990	98.55	98.55	106,477	99.00	99.00

Notes:

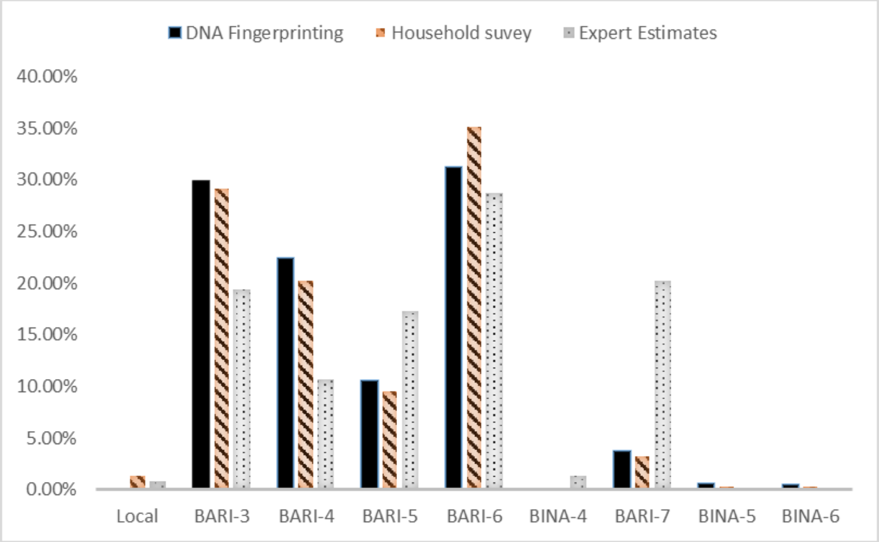
[a] Total area is based on the 2014 official statistics and is about 39% less than what we estimated using the survey data. As noted, the official data are generated using “subjective methods” and our survey is representative at the 10-district level.

[b] “Ungrouped” are varieties for which the DNA did not match any DNA extracted from the reference library (obtained from breeder seed samples).

[c] Error represents DNA samples which went bad (degenerated) and hence were not classified (i.e., unclassified samples due to technical problems).

[d] DK means “don’t know” as no pure seeds were available for landraces that could be used as benchmarks for DNA analysis

Figure 1: Variety-specific estimates of adoption (area share) by estimation method



Note: Annex table 1 provides exact details

5.2 Comparison of results from the three methods

While the farmer survey and expert opinion, respectively, identified 9 and 10 unique varieties as being grown in Western Bangladesh, DNA fingerprinting showed 11 varieties[5]. All three methods show BARI-6 to be the most widely adopted variety with area share estimates of 29%, 31% and 35% by experts, DNA testing, and farmer reporting, respectively. While experts list BARI-7 and BARI-3 as the second and third varieties with most area coverage, farmer reporting and DNA testing place BARI-3 and BARI-4 as the second and third most adopted varieties. Experts overestimate adoption of BARI-7 and BARI-5 by high margins relative to other methods. Experts also underestimate areas covered by BARI-3 and BARI-4, showing a bias toward later-released varieties. Estimates of adoption from the household survey and DNA fingerprinting are consistent (figure 1). Farmers in Western Bangladesh are knowledgeable

---

[5] Estimates are aggregated from the household survey, using the survey weights. They are representative at the 10-District area of Western Bangladesh and are comparable to the expert panel.

about the lentil varieties they plant; for the purposes of estimation of aggregate adoption, a farmer survey without the added expense of DNA fingerprinting would have sufficed.

The small difference between estimates from the household survey and DNA fingerprinting can be easily explained. First, even the oldest lentil variety released in Bangladesh is only 25 years old. Second, only 14 varieties have been officially released. Third, lentil varieties are known to farmers by their formal names, thereby reducing confusion. With the compressed time scale, unique names for each variety, and the limited number of releases, farmers are able to identify their varieties.

5.3 Determinants of adoption

Two binary dependent variables were used to analyze adoption: (i) DNA fingerprinted determination of adoption and (ii) farmer-reported assessment; the variables refer to whether the household has planted the MV (under each determination) on any of its plots. Variable descriptions and summary statistics are in table 3. For the DNA-tested outcome, participation in participatory variety selection (PVS) and in farmer field days, and receipt of assistance to plant improved lentils are statistically significant determinants of adoption (table 4). Farmers who engaged in PVS are about 38 percentage points more likely to adopt MVs compared to those who do not, and field day participants are 20 percentage points more likely. Those who received support to plant improved lentils are about 25 percentage points more likely to adopt compared to farmers not receiving support. Support consisted of seed, credit and fertilizer; most received a package with more than one form of support.

The presence of a seed dealer in a village raises the probability of adoption by about 7 percentage points, and villages where adoption appeared relatively early have about 2 percentage points more adoption for every year that the MVs are present. For example, for households living in villages where adoption occurred in the first year of MV availability (2006), adoption is about 18 percentage points higher than those where an improved variety emerged only in the past year. Farmers become more likely to adopt over time as experience in their own village with the variety increases. These findings indicate that increased awareness about the new varieties and access to seed are important determinants of adoption.

No farmer characteristic, including education or family structure is significantly associated with adoption. While variables such as education and family structure are often found to be significant determinants of adoption of MVs, their significance depends on the context. For example, Zeng et al. (2015) find household size is positively associated with adoption of improved maize varieties in Ethiopia, farmer education is not significant and farmer age, as in our case, is only marginally significant. Yigezu et al. (2018) showed that education and area planted are significant determinants of adoption of zero tillage (ZT) which makes sense as ZT is a knowledge and capital-intensive technology. El Shatter et al. (2016) also showed that farmer experience is an important determinant of adoption of ZT. On the other hand, Verkaart et al. (2017) do not find household demographic variables to be significant in improved chickpea adoption, also in Ethiopia. Farmer education is not significantly associated with adoption and no socio-demographic variable affected area planted to improved sweet potato in Uganda (Alwang et al. 2017).

Table 3: Variable descriptions and summary statistics[a]

		Non-adopters		Adopters[b]		Diff (mean or proportion test)[d] or row total
Variable name	**Description**	**Mean**	**SD**	**Mean**	**SD**	
age	Household head age (years)	46.27	11.68	45.59	11.45	NS
education[c]	Years education household head	5.78	4.60	6.08	4.53	NS
workers	# people of working age in household	3.30	1.45	3.34	1.74	NS
school_kids	# school-aged children in household	1.19	0.97	1.26	1.14	NS
owned_land	Owned land (hectares)	0.684	0.626	0.667	0.678	NS
swcons	DV (=1 if any plot has soil and water conservation structure)	0.04	0.19	0.03	0.18	NS
irrigation	DV (=1 if any plot has irrigation)	0.97	0.17	0.97	0.17	NS
pvs	DV (=1 if participated in participatory variety selection)	0.00	0.04	0.01	0.12	**
field_day	DV (=1 if participated in field day related to lentil growing)	0.02	0.15	0.06	0.24	***
improved_support	DV (=1 if household received "support" to try improved lentils)	0.08	0.27	0.19	0.39	***
support_date	Years since support was received	0.28	1.10	0.65	1.62	***
member_exec	DV (=1 if s member of executive committee of local group)	0.10	0.30	0.11	0.31	NS
seed_dealer	DV (=1 if village has improved lentil seed dealer)	0.40	0.49	0.48	0.50	***
village_seed_outreach	DV (=1 if extension activities in village promoted improved lentils)	0.70	0.46	0.78	0.42	***
years_village	Years since first improved variety was adopted in village	8.03	2.93	8.89	2.44	***
village_price_lentil	Village-level price of lentil	66.36	2.12	65.97	1.95	***
village_price_urea	Village-level price of urea	17.39	0.57	17.36	0.58	NS
village_price_tillage	Village-level price of tillage	397.09	113.55	414.12	103.04	**
yield	Lentil yield (kg/ha)	1451.50	16.39	1444.24	17.60	NS
lentilareahh[e]	Lentil area per household (ha)	0.37	0.01	0.34	0.01	**
plot_area[f]	Area of a lentil plot (ha)	0.25	0.01	0.245	0.01	NS
no_households	Number of households in the sample	501		499		1,000
no_plots	Number of plots in the sample	807		887		1,694

Source: Farm-household survey

[a] We do not show summary statistics for the asset index, computed using principal components, because it has no inherent meaning. The asset index includes: # of rooms in house, ownership of television, refrigerator, cellphone, motorcycle, bicycle, and tube-wells and access to electricity; dummy variables reflecting the wealth quintile of the household (e.g. wealth_q2) are entered in the regressions.

[b] adopters are those verified by DNA testing and using post 2005 varieties.

[c] education is divided into 3 dummy variables: >0 and <6 years, >5 and < 11 years, >10 years. These appear as education_d in the following tables.

[d] NS=not significant; *** p<0.01, ** p<0.05, * p<0.1

[e] Lentilareahh is computed by dividing the total lentil area cultivated by all sample households by the total number of sample households (1,000).

[f] plot_area is computed by dividing the total lentil area cultivated by all sample households by the total number of their lentil plots (1,694).

Table 4: Determinants of adoption of MVs of lentils (marginal effects from probit estimation), farmer-reported and DNA-tested.

	Farmer-reported improved				**DNA-tested**			
Variables	**Marginal effects**	**Standard error**	**Marginal effects**	**Standard error**	**Marginal effects**	**Standard error**	**Marginal effects**	**Standard error**
age	-0.003*	0.002	-0.003*	0.002	-0.002	0.002	-0.002	0.002
education_d=1	-0.000	0.041	-0.004	0.0417	0.033	0.045	0.028	0.045
education_d=2	0.016	0.043	0.014	0.044	0.051	0.046	0.049	0.046
education_d=3	0.001	0.060	-0.005	0.062	0.045	0.060	0.035	0.061
workers	0.007	0.010	0.008	0.010	0.005	0.010	0.007	0.010
school_kids	0.005	0.016	0.005	0.016	0.005	0.017	0.006	0.016
wealth_q=2	0.018	0.060	0.022	0.061	-0.022	0.054	-0.020	0.055
wealth_q=3	0.027	0.062	0.030	0.064	-0.011	0.058	-0.008	0.060
wealth_q=4	0.054	0.066	0.055	0.069	-0.024	0.060	-0.025	0.063
wealth_q=5	0.135*	0.070	0.137*	0.072	0.037	0.068	0.038	0.070
owned_land	-0.096***	0.038	-0.097***	0.038	-0.053	0.038	-0.054	0.037
swcons	-0.076	0.086	-0.082	0.086	-0.059	0.085	-0.064	0.084
irrigation	0.020	0.109	0.026	0.108	-0.044	0.102	-0.031	0.105
pvs	0.269*	0.156	0.277*	0.152	0.378***	0.132	0.386***	0.128
field_day	0.147**	0.065	0.155**	0.064	0.204***	0.076	0.209***	0.072
improved_support	0.185*	0.096	0.174*	0.096	0.240***	0.088	0.227**	0.089
support_date	-0.012	0.020	-0.009	0.020	-0.018	0.020	-0.015	0.020
member_exec	0.033	0.043	0.039	0.041	0.011	0.045	0.018	0.044
village_seed_dealer	0.076	0.051	0.064	0.049	0.073*	0.044	0.058	0.042
village_seed_outreach	0.025	0.049	0.030	0.049	0.065	0.043	0.069	0.044
years_village	0.020**	0.010	0.021**	0.010	0.020**	0.008	0.021**	0.008
village_price_lentil	-0.006	0.014	-0.006	0.014	-0.010	0.010	-0.011	0.011
village_price_urea	0.002	0.043			-0.012	0.042		
village_price_tillage	0.000	0.000			0.000	0.000		
N	1,000		1,000		1,000		1,000	
Pseudo R^2	0.070		.067		0.057		0.053	

*** p<0.01, ** p<0.05, * p<0.1

Note: The education variable was broken into three dummy variables, and the wealth index (see note to last table) was divided into quintiles with dummy variables entered into this regression. The comparison groups are: no education and the lowest wealth quintile.

Two variables reflecting household economic status-land holding and asset quintile- are not significantly associated with adoption. Access to the improved lentil varieties is not biased toward well-off producers; this finding is consistent with other studies of adoption of staple crops (Alwang et al. 2017). However, it is not universally the case: many studies have shown that landholding size affects adoption of improved (Zeng et al. 2015; Feder, Just and Zilberman 1985). Our results that household conditions do not affect adoption of lentil MVs are consistent with some literature, but not all. Two factors drive the findings. First, area agro-ecological and socio-demographic conditions are relatively homogeneous in Western Bangladesh (table 3). Second, varieties released before and after 2005 are all disease-resistant and high-yielding. Both factors make it difficult to distinguish between adopters and non-adopters. The sample was adequately powered; 1,000 observations should be adequate to detect differences in adoption.

Comparison of results from farmer own-reported planting of lentil MVs (first two sets of results, table 4) and the DNA-verified results (second two sets, table 4) shows that relying on farmer-reported variety names will lead to errors in inference. In particular, estimates using farmer-reported adoption as the dependent variable indicate that owned land is a highly significant determinant of adoption, and wealth, although only marginally significant, also affects adoption. These findings reverse a key relationship of interest—while the DNA-tested varieties showed no significant relationship between land ownership and wealth, when relying on own-reporting, the analyst will conclude that land ownership and wealth are associated with adoption. Specifically, the owned land coefficient in the farmer-reported regressions indicate that an additional hectare of land owned is associated with an almost ten percentage point decrease in the probability of adoption of (farmer-reported) improved varieties. On the other hand, households in the highest wealth quintile are found, when using farmer reporting, to be about 14 percentage points more likely to adopt (a relatively large wealth effect).

Using the own-reported dependent variable also leads to a conclusion that participation in PVS, field days and receipt of support are less important than estimated using the DNA-verified results. While these three variables are significant and positive in the own-reported case, their magnitude and significance are much lower than with DNA-verified results.

5.4 Double-hurdle estimates

The Cragg (double-hurdle) model estimates adoption of MVs in Tier 1 and area planted (in hectares) to MVs in Tier 2 (table 5)[6]. These estimates are not directly of interest because the Cragg model combines a probit with a truncated normal distribution. Marginal effects of key variables, generated using techniques in Burke (2009) are presented in table 6. These show the change in area planted to MVs (in hectares) given a one-unit increase in the independent variable[7], accounting for the two-staged decision.

The three key variables affecting area planted—wealth status in the 5th wealth quintile, owned land and village=level price of tillage—have similar effects regardless of which method is used to identify the variety (table 6). Area planted to MVs is greater for wealthiest households (about 0.11 ha more) compared to those in the first wealth quintile, for those with larger landholdings (an increase of 1 ha in landholding is associated with a 0.1 hectare increase in area

---

[6] The double-hurdle was compared to a Tobit. The Tobit estimator can be used to examine the adoption/ area planted decision but is restrictive because it requires that decisions about adoption and area planted are determined by the same process (same variables and same coefficients; see Ricker-Gilbert et al. 2011).

[7] Marginal effects are average partial effects (Wooldridge 2002) of relevant independent variables on MV area.

planted to MVs), and for residents of villages where tillage is more expensive. Statistical significance is high.

Table 5: Double-hurdle estimates of probability of adoption (Tier 1) and improved lentil area planted by household (in hectares) (Tier 2), farmer-reported and DNA-tested MVs.

	Farmer-reported improved		**DNA-tested**	
Variables	**Coefficient**	**Standard error**	**Coefficient**	**Standard error**
Tier1 (adoption, N=1000)				
age	-0.007	0.004	-0.008*	0.004
education_d=1	0.029	0.111	0.005	0.107
education_d=2	0.047	0.114	0.013	0.110
education_d=3	-0.005	0.154	-0.041	0.156
workers	0.004	0.026	0.010	0.026
school_kids	0.012	0.041	0.017	0.041
wealth_q=2	0.039	0.166	0.026	0.162
wealth_q=3	0.072	0.168	0.033	0.169
wealth_q=4	0.066	0.184	0.069	0.182
wealth_q=5	0.317	0.199	0.297	0.197
owned_land	-0.043	0.066	-0.040	0.064
swcons	-0.212	0.233	-0.211	0.232
irrigation	0.070	0.283	0.041	0.284
pvs	1.224*	0.634	0.766	0.535
field_day	0.559***	0.194	0.441**	0.187
improved_support	0.743***	0.253	0.478*	0.268
support_date	-0.074	0.053	-0.029	0.054
member_exec	0.039	0.118	0.075	0.111
village_seed_dealer	0.226*	0.117	0.177	0.131
village_seed_outreach	0.108	0.125	0.080	0.130
years_village	0.068***	0.023	0.059**	0.030
village_price_lentils	-0.021	0.030	-0.021	0.039
Intercept	0.689	2.063	0.917	2.695
Tier2 (improved lentil area, N=503)				
age	-0.002	0.013	0.003	0.012
age-squared	0.000	0.000	0.000	0.000
workers	0.001	0.013	0.003	0.013
wealth_q2	-0.012	0.053	-0.016	0.055
wealth_q3	0.058	0.078	0.016	0.080
wealth_q4	0.040	0.044	0.034	0.045
wealth_q5	0.131*	0.065	0.105	0.065
owned_land	0.514***	0.104	0.487***	0.104
owned_land-squared	-0.078***	0.018	-0.072***	0.018
swcons	-0.006	0.095	-0.027	0.092
irrigation	0.128	0.088	0.094	0.087

	Farmer-reported improved		DNA-tested	
Variables	Coefficient	Standard error	Coefficient	Standard error
pvs	-0.050	0.171	0.026	0.151
field_day	0.070	0.093	0.128	0.089
improved_support	0.220***	0.071	0.198***	0.064
support_date	-0.008	0.015	-0.013	0.013
village_seed_dealer	0.136*	0.075	0.174**	0.072
village_price_lentil	-0.018	0.019	0.001	0.016
village_price_urea	0.007	0.045	0.016	0.044
village_price_tillage	0.002***	0.000	0.002***	0.000
Intercept	-0.079	1.632	-1.578	1.411
Sigma intercept	0.305***	0.031	0.293***	0.032
Chi-square (22 df)	56.16	P<0.0001	40.33	P<0.05

*** p<0.01, ** p<0.05, * p<0.1

Marginal effects of participation in PVS and field days and receipt of support for improved varieties are also in table 6. These show, that despite the probit model findings of a significant positive impact on adoption, participation in the PVS or field day promotional efforts is not significantly associated with area planted to MVs (adoption intensity). Receipt of support for improved lentil adoption is associated with a 0.08 hectare greater area planted compared to farmers who did not receive support; differences between estimates and their significant are negligible when comparing farmer-reported and DNA-verified indicators of adoption.

5.5 Impacts of adoption of improved lentil varieties

An attempt was made to measure impacts of adoption of improved lentil varieties on plot-level yield and whether the estimates of impacts varied depending on variety identification method. Because adoption is likely to be endogenous to observed yield (the outcomes are simultaneously determined), we used a two-stages least squares (2SLS) instrumental variables (IV) regression (Zeng et al. 2015), where the first stage was plot-level adoption and the second stage was plot-level yield.

Table 6: Marginal effects and bootstrapped standard errors, effect of independent variables on area planted to MV, DNA-tested versus farmer-reported

Variables	Coefficient	Standard error	z
DNA-tested			
wealth_q_5	0.047*	0.027	1.77
owned_land	0.219***	0.036	6.11
village_price_tillage	0.001***	0.000	13.11
pvs	0.012	0.123	0.10
field_day	0.570	0.350	1.62
improved_support	0.089**	0.043	2.05
Farmer-reported improved			
wealth_q_5	0.058*	0.031	1.88
owned_land	0.228***	0 .032	7.11
village_price_tillage	0.001***	0.000	10.87
pvs	-0.022	0.088	-0.25
field_day	0.031	0.036	0.85
improved_support	0.097**	0.042	2.31

*** p<0.01, ** p<0.05, * p<0.1
Note: 50 draws were taken to bootstrap the standard errors

Use of IV inevitably leads to questions about the appropriateness of the instruments; in this case, instruments are variables assumed to affect adoption but only affect yield through their effects on adoption. As yield is a function of inputs applied, several plausible instruments are available. The second-stage equation is a Cobb-Douglas production function, where log yield (in kg per hectare) is expressed as a function of log values of inputs (area planted, seeds, fertilizer, pesticides, machinery and labor)[8], access to irrigation and soil conservation structures, farmer education (a proxy for ability) and participation in field days (the extension method used in Western Bangladesh), possibly associated with enhanced management ability. Controlling for these, other factors affecting MV adoption (table 4) should only affect yield through effects

---

[8] Summary statistics for the input variables are available from the authors upon request.

on adoption. These variables were used as instruments[9]. We subjected these instruments to standard tests: relevance and exogeneity[10] (Wooldridge 2010).

Table 7: Two-stage least squares estimates of impacts of improved varieties on plot-level lentil yields in kg/hectare, DNA-tested versus farmer-reported (dependent variable is log(yield))

		Farmer-reported improved		**DNA-tested improved**	
VARIABLE	**Variable description**	**coef**	**se**	**coef**	**se**
Adoption=1[a]	Instrumented dummy variable	-0.044	0.087	-0.052	0.099
Lnplot_area	Log area of plot (hectares)	-0.175**	0.069	-0.171**	0.069
Lnseedkg	Log(seed in kg)	0.094**	0.047	0.092*	0.048
Lnureakg	Log(urea in kg)	0.012*	0.007	0.012*	0.007
Lntspkg	Log(TSP in kg)	-0.012	0.010	-0.012	0.010
Lnmopkg	Log(MOP in kg)	0.032***	0.009	0.032***	0.009
Lndapkg	Log(DAP in kg)	0.019*	0.010	0.019*	0.010
Lnfungicidegm	Log(fungicide in grams)	0.015***	0.003	0.015***	0.003
Lninsecticideml	Log(insecticide in grams)	0.017***	0.004	0.017***	0.004
Lvmech	Log(value of machinery)	0.037***	0.008	0.036***	0.008
lworkers	Log(number of workers)	0.008	0.022	0.007	0.022
irrigation	b	-0.197***	0.035	-0.197***	0.034
swcons	b	0.015	0.053	0.014	0.053
education_d=1	b	-0.053*	0.028	-0.052*	0.029
education_d=2	b	-0.049*	0.028	-0.048*	0.029
education_d=3	b	-0.177***	0.038	-0.178***	0.038
field_day	b	0.044	0.048	0.046	0.051
Intercept		6.643***	(0.210)	6.660***	(0.218)
Observations		1,692		1,692	
R-squared		0.114		0.113	
Relevance test (F-test)		9.89 (p=0.0000)		7.60 (p=0.0000)	
Test of over-identifying restrictions (Chi-squared)		35.464 (p=0.0002)			

*** p<0.01, ** p<0.05, * p<0.1. Standard errors are robust.

---

[9] Instruments include: age, age-squared, wealth_q, pvs, improved_support, village_seed_dealer and years_village. The other covariates in the probit and double-hurdle regressions were excluded entirely from the impact estimation because they failed to attain significance in any of these regressions.

[10] The relevance test is an F-test from a regression of adoption on the instruments; standard criteria are that the F-test value should be greater than 10. The exogeneity test, which tests overidentifying restrictions, is known as the J-test and is only valid if the system is identified. We use logic to establish this: participation in PVS, used to help select among alternative lentil varieties (for release) can logically only affect yield through its impact on adoption. Since large numbers of farmers are invited to participate in PVS, participation is not likely to be associated with yield (in contrast to participation in farmer field schools where participants are purposefully selected). Farmers become exposed to the new variety during the PVS and are, hence, more likely to adopt.

[a] Endogenous variable

[b] See table 3 for variable description

Results (table 7) show no significant impact of adoption on yield (kg/hectare) and no difference in findings when DNA-verified or farmer-reported adoption is used. Those input variables with significant coefficients had the expected signs and the fit of the regression was relatively good. We attribute the non-significance of the instrumented adoption variable to reflect the pivot in breeding objectives away from yield and toward cooking, nutrient and taste traits. The team conducted a workshop that included ICARDA and BARI lentil breeders and discussions at the workshop indicated that the shift toward these objectives strengthened after 2006. Evidence from this discussion is contained in table 1. Of the eight listed varieties released after 2005, six mention cooking quality, taste or nutrient content as a primary breeding objective, and only BARI Masur 5 and BINA Musur 7 (not found on farmer fields) do not include these characteristics as breeding objectives (table 1). As noted, no significant yield gains were observed after 2005 compared to varieties released between 1996 and 2005. Thus, the finding that adoption of post-2005 varieties has no impact on yields is logical.

We ran a second test to examine if improved cooking and taste attributes led to higher prices for sales of post-2005 varieties[11]. This 2SLS regression had sales prices as the dependent variable in the second-stage of estimation, with household size and composition, education of head, wealth, land ownership and access to irrigation as covariates. Excluded instruments from the first stage included participation in pvs and field days, access to improved support, years

---

[11] We thank an anonymous reviewer for suggesting these regressions. Unfortunately, despite their passing the same tests presented above, they did not provide evidence of a price premium for the post-2005 varieties.

that the improved varieties were found in the village and the presence of an improved seed dealer. All of these variables could legitimately be expected to affect adoption but have no logical pathway to affect sales prices except through their impact on adoption. In the price regressions, as in the yield regressions in table 7, adoption of post-2005 lentil varieties was not significantly associated with higher sales prices (results available from authors upon request). A plausible explanation for these findings (lack of significance) is that lentil sales were not separated by variety in the questionnaire. Ideally, we would know sales prices by variety, but this information was not available in the data.

Absence of a significant yield effect was expected given the attributes of new lentil varieties in table 1 and the discussions with breeders during the workshop. Detection of taste preferences and how changes in breeding objectives affect adoption of new varieties required better information on stated and revealed preferences; our survey was inadequate for a detailed analysis of these factors.

6. Conclusions

The main objectives of this study are to assess the extent of diffusion of improved lentil varieties in Western Bangladesh, to understand how estimates of diffusion vary by measurement method, and to understand factors affecting adoption of new varieties. The household survey shows that lentil area in Western Bangladesh during 2014-15 was approximately 149,000 hectares, about 39% higher than official statistics show. Official statistics use a “subjective method”-not a sample survey or census- and do not accurately account for lentils which are often “hidden” between crops (as relay crops or intercropped).

Expert elicitation did not produce reliable estimates of lentil diffusion. The ranking of variety diffusion by experts was not consistent with the ranking from the representative sample. In particular, experts failed to capture the existence of old improved varieties. This finding is consistent with the general message from the Walker and Alwang (2015) study that the validity of elicitation measures depends on the context. Special care in elicitation methods is needed when varieties have similar attributes and morphological differences are minor.

Farmer own-reporting is relatively accurate, mainly because there are only few improved lentil varieties with relatively short history in Bangladesh. In contrast to other experience, DNA fingerprinting was not necessary to estimate aggregate diffusion. DNA fingerprinting is the most reliable identification method, but the high correspondence between the DNA and own-reported results means that for the purpose of estimates of diffusion the expense could have been avoided.

The literature shows large gains to DNA testing in cases where many varieties are planted, variety names vary by locality, and seeds are mainly transferred to farmers through informal mechanisms (Floro et al. 2017; Kosmowski et al. 2016; Maredia et al. 2016). When few varieties are present and formal seed systems predominate, farmer-identification appears to be good. The presence of few lentil varieties, the absence of local names and the predominance of formal seed transfer means that farmer identification is sufficiently accurate to estimate aggregate diffusion. These findings are similar to results from a study of adoption of potato varieties in Yunnan China. In Yunnan, farmer identification was about 95% accurate, and China is also characterized by the presence of a few potato varieties and a highly formalized seed system (Myrick 2016). We suggest that researchers conduct an assessment of country-specific conditions, perhaps being followed by small-scale pilot studies prior to implementation of DNA testing at scale. Costs of DNA testing need to be balanced against its benefits; one under-appreciated cost is disruption in timing of field operations (e.g. in some cases, leaf samples may be the preferred sampling method).

While DNA-verified and own-reported estimates of diffusion (overall adoption) matched well, regressions using the farmer-reported outcome led to minor (erroneous) conclusions about determinants of adoption. The farmer-reported results indicated that owned land and wealth are significantly associated with adoption. These findings would have led to the conclusion that poorer farmers or those with larger landholdings face obstacles to adoption, indicating that lentil improved varieties are not pro-poor. The correct (DNA-verified) findings show no such bias.

However, the adoption/area results from the double-hurdle model show that area planted to MVs of lentils is greater for wealthier households compared to households in the first wealth quintile, for households with larger landholdings, and for households residing in villages where tillage is more expensive. Thus, while the possibility of adopting MVs is independent of wealth and farm size, wealthier farmers and those with larger holdings are able to plant more area to MVs. Benefits of MV research will flow disproportionately to better-off farmers. These findings are independent of variety identification strategy.

The double-hurdle results, which jointly reflect adoption and intensity show no real difference in parameter estimates by variety verification method. The signs, magnitudes and significance of the key variables were virtually the same whether DNA verification or farmer reporting was used to identify and measure adoption. Minor differences emerge when the outcome is the probability of adoption, but not when the outcome is area planted, conditional on adoption. Since both outcomes are of interest to policy makers, the regression results suggest that DNA verification should be used, but gains to DNA verification are relatively small.

An analysis of yield differences between post- and pre-2006 lentil varieties showed no significant differences in yields. Yield impact measurement was not sensitive to the method of variety identification.

Several lessons can be drawn. First, lentil research produced disease-resistant varieties that allowed relay-cropped lentil production to continue in Western Bangladesh in the face of extreme disease pressure. Second, in cases where few varieties exist, and seed systems are formal, DNA fingerprinting is not likely to substantially improve estimates of diffusion. Estimates of determinants of adoption were altered when own-reporting is used, but these differences may not, on their own, justify the additional expense of DNA testing. When few varieties are found, local names are not common, seeds are procured through a formal system, if only adoption and area estimates are desired, own-reporting may suffice. For different purposes, DNA fingerprinting may be essential. For example, DNA fingerprinting opens the door to analysis of factors such as genetic diversity (in addition to variety diversity).

Third, when minor crops such as lentils emerge in a farming system, official statistics may fail to reveal their importance. As lentils employ under-used land between important rice and vegetable seasons, their prominence has been masked by official use of “subjective” measurement methods. This study reveals the degree to which lentils are missed in official statistics and the importance of a relatively minor crop to household-level outcomes.

References

Alwang, J., E. Gotor, G. Thiele, G. Hareau, M. Jaleta and J. Chamberlin. 2017. Pathways from research on improved staple crop germplasm to poverty reduction for smallholder famers. Agricultural Systems, https://doi.org/10.1016/j.agsy.2017.10.005

Ali, M.O., M. I. Zuberi. and A. Sarker, 2014. Management practices for lentil as relay crop in the rice-based cropping system, Journal of Agricultural Science and Technology B 4: 126-134.

Amankwah, A., K. K. Quagrainie and P. V. Preckel, 2016. Demand for improved fish feed in the presence of a subsidy: a double-hurdle application in Kenya. Agricultural Economics, 47: 633–643.

Andeden, E.E., Baloch, F.S., Çakır, E., Toklu, F. and H. Özkan, 2015. Development, characterization and mapping of microsatellite markers for lentil (Lens culinaris Medik.). Plant Breeding, 134(5): 589-598.

BBS (2017) Yearbook of Agricultural Statistics-2016, Bangladesh Bureau of Statistics (BBS) Statistics and Informatics Division, Ministry of Planning, Government of the People's Republic of Bangladesh.

Burke W.J., 2009. Fitting and interpreting Cragg's Tobit alternative using Stata. Stata Journal, 9(4): 584–592.

CGIAR, 2016. Innovative Methods for Adoption of Agricultural Technology. Workshop proceedings, Boston, 3-4 August 2016. https://ispc.cgiar.org/meetings-and-events/adoption-boston-2016.

Cragg J.G., 1971. Some statistical models for limited dependent variables with application to the demand for durable goods. Econometrica, 39(5): 829–844.

Datta, S.K., M.A.R. Sarkar and F.M.J. Uddin, 2013.Effect of variety and level of phosphorus on the yield and yield components of lentil, International Journal of Agricultural Research Innovation and Technology, 3 (1): 78-82.

El-Shater, T., Y. A. Yigezu, A. Mugera, C. Piggin, A. Haddad, Y. Khalil, S. Loss and A. Aw-Hassan, 2016. Does Zero Tillage improve the Livelihoods of Smallholder Cropping Farmers? Journal of Agricultural Economics, 67(1):154-172

Evenson, R., and D. Gollin (eds.), 2003. Crop Variety Improvement and Its Effects on Productivity. Wallingford, UK: CAB International.

Feder, G., 1980. Farm size, risk aversion and the adoption of new technology under uncertainty. Oxford Economic Papers, 32: 263-283.

Floro, V. O., Labarta, R. A., Becerra López-Lavalle, L. A., Martinez, J. M. and T.M. Ovalle, 2017. 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. doi:10.1111/1477-9552.12247

Gupta, M., Verma, B., Kumar, N., Chahota, R.K., Rathour, R., Sharma, S.K., Bhatia, S. and T.R. Sharma, 2012. Construction of intersubspecific molecular genetic map of lentil based on ISSR, RAPD and SSR markers. Journal of genetics, 91(3): 279-287.

Hamwieh, A., Udupa, S.M., Sarker, A., Jung, C. and M. Baum, 2009. Development of new microsatellite markers and their application in the analysis of genetic diversity in lentils. Breeding Science, 59(1): 77-86.

Just R.E. and R. Pope, 1978. Stochastic specification of production functions and economic implications. Journal of Econometrics, 7: 67-86.

Kosmowski, F., Aragaw, A., Kilian, A., Ambel, A., Ilukor, J., Yigezu, B. and J. Stevenson, 2016. Varietal Identification in Household Surveys: Results from an Experiment Using DNA Fingerprinting of Sweet Potato Leaves in Southern Ethiopia, Policy Research Working Paper No. 7812, Washington, DC: The World Bank.

Labarta, R., Martinez, J., Lopera, D., Gonzalez, C., Quintero, C., Gallego, G., Viruez, J. and R. Taboada, 2015. Assessing Impacts of the Adoption of Modern Rice Varieties using DNA Fingerprinting to Identify Varieties in Farmer Fields: A Case Study in Bolivia Palmira, Colombia: CIAT.

Larochelle C., J. Alwang, G. W. Norton, E. Katungi and R. A. Labarta, 2015. Impacts of Improved Bean Varieties on Poverty and Food Security in Uganda and Rwanda, Chapter 16 in Thomas S. Walker and Jeffrey Alwang (Eds.). Improved varieties of food crops in sub-saharan Africa.

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. and B. Raatz, 2016. Testing Alternative Methods of Varietal Identification Using DNA Fingerprinting: Results of Pilot Studies in Ghana and Zambia. International Development Working Paper No. 149. East Lansing Michigan: Michigan State University.

Myrick, S. N. B. 2016 An Economic Impact Assessment of Cooperation-88 Potato Variety in the Yunnan Province of China. Unpublished M.S. Thesis, Blacksburg, VA: Virginia Polytechnic Institute and State University.

Peakall, R. and P. E. Smouse, 2012. GenAlEx 6.5: genetic analysis in Excel. Population genetic software for teaching and research—an update. Bioinformatics, 28(19): 2537-2539.

Pradeep R.M., Sarla, N. and E.A. Siddiq, 2002. Inter simple sequence repeat (ISSR) polymorphism and its application in plant breeding. Euphytica, 128(1): 9-17.

Ricker-Gilbert, J., T. S. Jayne and E. Chirwa, 2011. Subsidies and crowding out: A double-hurdle model of fertilizer demand in Malawi. American Journal of Agricultural Economics, 93(1): 26-42.

Sarker, A., M. A. Bakr, M. A. Afzal, W. Erskine, M. M. Rahman and M.C. Saxena, 2004. Lentil Improvement in Bangladesh. Bangkok, Thailand: Asia-Pacific Association of Agricultural Research Institutions.

Schmidtke, K., A. Neumann, C. Hof and R. Rauber, 2004. Soil and atmospheric nitrogen uptake by lentil (Lens culinaris Medik.) and barley (Hordeum vulgare ssp. nudum L.), as monocrops and intercrops, Field Crops Research. 87: 245–256.

Schneider, C. A., W.S. Rasband and K.W. Eliceiri, 2012. NIH Image to ImageJ: 25 years of image analysis. Nature methods, 9(7): 671-675.

Singh, G., R.H Mehta and O.P. Singh, 1994. Effect of seed rate and method of sowing of lentil.Indian J. Pulses Res.7 (2): 132-136.

Verkaart, S., B.G. Munyua, K. Mausch and J.D. Michler, 2017. Welfare impacts of improved chickpea adoption: A pathway for rural development in Ethiopia? Food Policy 66: 50–61.

Verma, P., Goyal, R., Chahota, R.K., Sharma, T.R., Abdin, M.Z. and S. Bhatia, 2015. Construction of a genetic linkage map and identification of QTLs for seed weight and seed size traits in lentil (Lens culinaris Medik.). PloS one, 10(10), p.e0139666.

Walker, T. S., 2015. Validating Adoption Estimates Generated by Expert Opinion and Assessing the Reliability of Adoption Estimates with Different Methods, ch. 20 in Walker, T. S. and Alwang, J. (eds.) Crop Improvement, Adoption and Impact of Improved Varieties in Food Crops in Sub-Saharan Africa. Boston, MA: CGIAR and CABI.

Walker, T. S. and Alwang, J. (eds.), 2015 Crop Improvement, Adoption and Impact of Improved Varieties in Food Crops in Sub-Saharan Africa. Boston, MA: CGIAR and CABI.

Wooldridge, J. M., 2002. Econometric Analysis of Cross Section and Panel Data. Cambridge, MA: MIT Press.

Yeh F.C., R.C. Yang and T. Boyle, 1999. POPGENE VERSION 1.31: Microsoft Windows- based Free ware for population genetic Analysis. ftp://ftp. Microsoft.com/Softlib/HPGL.exe.

Yigezu, Y.A., T. El-Shater, A. Mugera, A. Aw-Hassan, C. Piggin, A. Haddad, Y. Khalil and S. Loss, 2018. Enhancing adoption of agricultural technologies requiring high initial investment among smallholders. Technological Forecasting & Social Change, 134:199-206.

Zeng, D., J. Alwang, G.W. Norton, B. Shiferaw, M. Jaleta and C. Yirga, 2015. Ex-post impacts of improved maize varieties on poverty in rural Ethiopia. Agricultural Economics, 46(4), 515-526.

Annex I: Estimates of aggregate diffusion (in June 2015) by variety name and method of estimation

Variety name	Year of Release	DNA Fingerprinting		Household survey		Expert Estimation	
		Area share (%)	Estimated total area (ha)	Area share (%)	Estimated total area (ha)	Area share (%)	Estimated total area (ha)
Local	1900	DK[a]	DK[a]	1%	1555	1%	895
BARI-1	1991	0%	187	0%	0	0%	0
BARI-2	1993	0%	310	0%	0	0%	0
BARI-3	1996	30%	32005	29%	31477	19%	20892
BARI-4	1996	22%	23984	20%	21809	11%	11432
BINA-1	2002	0%	0	0%	0	0%	238
BINA-2	2005	0%	0	0%	0	1%	719
BINA-3	2005	0%	0	0%	0	1%	767
BARI-5	2006	11%	11334	10%	10400	17%	18612
BARI-6	2006	31%	33482	35%	37871	29%	30853
BINA-4	2009	0%	0	0%	50	1%	1435
BARI-7	2011	4%	4052	3%	3599	20%	21705
BINA-5	2011	1%	666	0%	386	0%	0
BINA-6	2011	1%	553	0%	402	0%	0
BINA-7	2011	0%	0	0%	0	0%	0
BARI-8	2012	0%	0	0%	0	0%	0
NA	NA	0%	156	0%	0	0%	0
Ungrouped[b]		0%	523	0%	0	0%	0
Error[c]		0%	297	0%	0	0%	0
Total		100%	107549[d]	100%	107549[d]	100%	107549[d]

Notes:

[a] DK means "don't know" as there were no pure seeds for landraces that could be used as benchmarks for DNA analysis

[b] Ungrouped represent varieties the DNA of which did not match any of the DNA extracted from the reference library (reference DNA of the varieties obtained from the breeder seed samples).

[c] Error represents DNA samples which went bad (degenerated) and hence were not classified (i.e., unclassified samples due to technical problem).

[d] Total area is based on the 2014 official statistics which is about 30% less than estimated using the survey data.