Does Survey Mode Matter? Evidence from Phone and In-Person Agricultural Surveys in India

Ellen Anderson, Travis Lybbert, Ashish Shenoy, Rupika Singh, Daniel Stein

July 8, 2022

Abstract

Phone surveys have become more common in developing countries and are increasingly used for data collection in randomized control trials. However, the consequences of switching from in-person to phone surveys in the context of RCTs are not fully understood. Of particular concern is whether measurement error in outcomes collected from phone surveys could bias treatment effects. We compare responses from phone and in-person surveys conducted for an overlapping set of questions and households that were part of an RCT studying the effects of a pulses farming promotion program on crop production in Bihar, India. We find differences in properties of the outcomes distributions between the phone and in-person surveys. The differences are driven by both selection and mode effects. However, we find similar treatment effects of the program by survey mode for both intent to treat and local average treatment effects.

1 Introduction

Phone surveys are a potentially cheaper and faster alternative to in-person data collection. Public health risks with in-person data collection due to the Covid-19 pandemic necessitated a switch from in-person to phone surveys. However, research organizations and governments increasingly used phone surveys before Covid-19 as mobile phone ownership became more common. There are important concerns around the quality of data obtained from phone surveys and whether there is increased measurement error associated with data collected over the phone. Of particular concern is whether measurement error from phone surveys could bias treatment effects from randomized control trials. When designing RCTs, researchers may face a trade-off between sample size and measurement error based on survey mode. Phone surveys are generally cheaper, allowing researchers to reach more households, but they may increase measurement error when compared to more expensive in-person surveys. There is evidence that those who respond to phone surveys differ from in-person respondents on observable characteristics (Ferreira et al. 2011; Ballivian et al. 2015; Leo et al. 2015; Greenleaf et al. 2020). However, there is mixed evidence on measurement error resulting from being surveyed over the phone versus other modes, once accounting for selection (Ballivian et al. 2015; Mahfoud et al. 2015; Garlick, Orkin, and Quinn 2020; Greenleaf et al. 2020). In this paper, we aim to answer whether responses to the same questions for agricultural surveys administered via phone and in-person can be drawn from the same underlying distribution, how much of the differences between phone and in-person surveys can be attributed to selection and mode effects, and the extent to which RCT treatment effects differ by data collection mode.

We compare phone and in-person surveys conducted as part of a larger RCT measuring the effects of a pulses farming promotion program in Bihar, India on households' adoption of pulses farming and crop production. The phone and in-person surveys were administered a week apart on average, from May-June 2018 with an overlapping sample of farmers and an overlapping set of questions. The first year of the RCT (2017-2018) was designed jointly by research teams at IDinsight and the University of California, Davis (UC Davis). Results from the first year were intended to identify cost-effective interventions to increase adoption of pulses farming that could potentially be scaled up by governments. After conducting a joint in-person survey covering the 2017 Kharif (rainy) cropping season and beginning of the 2017-2018 Rabi (dry) cropping season, IDinsight and UC Davis conducted separate phone and in-person surveys, respectively, to follow up on Rabi production. The IDinsight survey covered a more limited set of questions but had a larger sample size. In this paper, we compare responses to questions that were identical in both surveys, both for the entire sample of each survey and for only the individuals that appear in both samples. We also separate selection and mode effects by comparing overall differences in responses to differences for individuals that were surveyed using both modes.

There is a growing literature on the use of computer-assisted telephone interviewing (CATI), Short Message Service (SMS), and Interactive Voice Response (IVR) surveys in developing countries. In particular, the literature focuses on selection and mode effects that result from using CATI compared to computer-assisted personal interviewing (CAPI). Selection effects result from differences in the type of respondent who answers in-person versus phone surveys. Greenleaf et al. (2020) find that respondents to a contraceptive use survey in Burkina Faso selected via random digit dialing are more likely to be younger, have a secondary education, and speak French than respondents to a nationally representative in-person survey. Similarly, Ballivian et al. (2015) conduct phone surveys in Peru and Honduras after initial in-person surveys and find that those who respond to the phone survey are younger, more educated, and more affluent. Although conducted via landline phones, Ferreira et al. (2011) find that phone respondents to a public health survey in Brazil were more likely to have white skin, were more educated, and had higher access to medical care but also more chronic conditions than in-person respondents. Leo et al. (2015) show that those who indicate owning a mobile phone in Demographic and Health Surveys tend to be richer and more urban than the overall population.

The literature on mode effects between phone and in-person surveys is mixed. Mode effects refer to differences in individual responses due to the survey format (CAPI, CATI, SMS, etc.). Garlick, Orkin, and Quinn (2020) do not find major differences between phone and in-person responses in a survey experiment with small enterprises in South Africa. However, they do find that phone respondents report lower resources taken from the enterprise by the owner or her family, higher use of written records, and fewer hours worked. Additionally, they find that responses were significantly lower on phone surveys for questions where the respondent needed to make an estimate such as stock/inventory and sales. Ballivian et al. (2015) find no statistically significant differences between phone and in-person responses for a consumption survey in Honduras and Peru. Similarly, Mahfoud et al. (2015) find strong agreement between phone and in-person responses for a public health survey in Lebanon, although with small differences between phone and in-person for reported level of physical activity and certain chronic conditions. Greenleaf et al. (2020) find that self-reported contraceptive use is 14 percentage points higher over the phone versus in-person among women in Burkina Faso. This difference remained large even after post-stratification weights were applied (12.7 percentage points). Kilic et al. (2021) conduct an experiment comparing different self-reported data collection methods for cassava production in Malawi. They find that daily diary keeping of production with phone-based follow-ups produces larger yearly production estimates than daily diary keeping with in-person follow-ups, 6-month recall surveys, and 12-month recall surveys. All self-reports were lower than crop cuts which they consider an upper bound for yearly production. The authors suggest that the diaries with phone follow-ups are the most accurate in their context. We also find larger average production over the phone in our context, however, we did not have respondents keep diaries and our recall period ranges from a few weeks to two months. [1]

High income countries incorporated CATI into national sample surveys over the last thirty years. Most studies find only small mode effects and produce similar estimates when controlling for observable characteristics. Nord and Hopwood (2007) compare food security prevalence

1. The Rabi harvest in our study region generally occurs from end of March through April.

between phone and in-person interviews as part of the US Current Population Survey Food Security Supplement. They find small differences between mode on individual items and higher prevalence of food insecurity in in-person interviews. However, most of these differences are accounted for by observable characteristics as those who responded in-person had lower socio-economic status. St-Pierre and Béland (2004) compare CAPI and CATI responses to a national health survey in Canada and do not find major differences overall, however, they find small differences in height for weight and level of physical activity. In a survey experiment in Switzerland, Scherpenzeel and Eichenberger (2001) find no differences between CAPI and CATI for a range of questions which could induce social desirability and memory effects. They also find no difference in validity and reliability of survey questions based on mode. Biemer (2001) estimate measurement error and non-response bias in the 1994 US National Health Interview Survey using repeated phone and in-person surveys. They find that both modes have significant biases but the biases are similar in magnitude between modes.

Additionally, the literature suggests social desirability bias and verifiability as two potential drivers of phone and in-person survey differences. For instance, Mahfoud et al. (2015) find that men report higher alcohol consumption over the phone as compared to in-person and women report lower levels of physical activity over the phone, potentially due to social desirability around health behaviors. However, St-Pierre and Béland (2004) find respondents are less likely to report being obese and physically inactive over the phone than in-person. Garlick, Orkin, and Quinn (2020) attribute higher reported use of written records for small enterprises in their phone survey to verifiability as the telephone interviewer cannot physically see if records are kept in the shop.

We contribute to the literature first by separating selection and mode effects for agricultural surveys in India. Consistent with the literature, we find that phone survey respondents are younger, more male, and more affluent than non-respondents. We also find suggestive evidence of mode effects. We find differences in important characteristics of the outcome distributions by mode including the mean, proportion of non-zero observations, and proportion of observations in the right tail. We also test whether the means and deciles observed in the phone survey data could be drawn from the in-person distribution by taking bootstrap resamples of the in-person data, computing the means and deciles of each outcome for each the resamples, and comparing them to the phone means and deciles. We find that it is rare to get means and deciles in the in-person resamples as large as the phone survey.

There is little evidence in the literature on whether survey mode can bias treatment effects in the context of a randomized control trial. Crawfurd et al. (2021) look at differences by survey mode within an RCT studying the effect of teacher tutoring calls on student learning during Covid-19 lock-downs in Sierra Leone. They assessed student learning through both in-person and phone-based tests, randomizing students into mode within each treatment arm. They find that both math and language test scores were much lower for in-person assessments. The authors suggest that this could be due to non-compliance as many students assigned to phone surveys were assessed in person and vice versa. When using random assignment to in-person assessments as an instrument for actually being surveyed in-person, the test score differences are no longer significant. Although the authors find differences in test scores by survey mode, they do not find heterogeneity in treatment effects of the intervention by survey mode. We find similar results in our context, treatment effects from phone and in-person survey data are similar in both sign and magnitude.

The paper proceeds first by describing the context of the surveys in Section 2. In Section 3, we provide descriptive statistics by mode. In Section 4, we test whether responses from different modes could be drawn from the same underlying distribution. We separate differences between the surveys into selection and mode effects in Section 5. Lastly, in Section 6, we compare treatment effects by survey mode and we conclude with recommendations for researchers in Section 7.

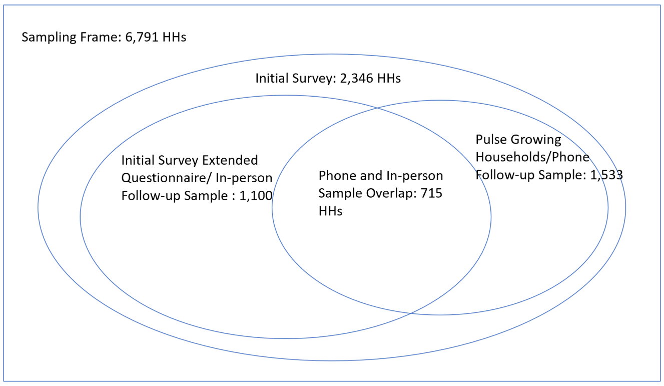
2 Context

The phone and in-person surveys we compare were administered in May and June 2018 as part of research on the India Grain Legume Cluster Development Program in Bihar, India. Researchers from IDinsight and UC Davis collaborated on the first year of a cluster randomized control trial measuring the effect of the program on adoption and production of pulses farming. The RCT had three treatments and a control arm. The initial phase of the program was focused on providing farmers with access to inputs and technical assistance in order to spur adoption of pulses farming. Pulses production had declined in the study areas in the years following the Green Revolution but there was potential to re-introduce pulses as a source of income and improved household nutrition. The three treatment arms were additive with a “low” intensity arm distributing certified pulse seed to farmers via a voucher system, a “medium” intensity arm distributing certified pulse seed directly to farmers and providing extension services, and a “high” intensity arm distributing seeds, providing extension services, and establishing demonstration plots to showcase best farming practices in each village. The first year of the evaluation was focused on identifying which of the three treatment arms was the most cost-effective at increasing pulse production and had potential to scale. In order to increase statistical power, we also pool the three treatment arms when looking at treatment effects by mode.

IDinsight and UC Davis collaborated on an initial in-person survey following the 2017 Kharif (rainy) season harvest, which covered cultivation and harvest outcomes for Kharif as well as crops planted for the 2017-2018 Rabi (dry) season. The sampling strategy for the initial survey set up the sampling strategy for the follow-up phone and in-person surveys. Households for the initial survey were drawn from a listing of households that attended a program kick-off meeting in May or June of 2017. Kick-off meetings were held in both treatment and control villages. Implementing partners were blinded to the treatment status of the village and the content of the meetings primarily consisted of a general discussion of agriculture in the village. Of the 6,971 households in the sampling frame, 2,346 were randomly selected for the initial survey, which was conducted from November 2017- January 2018. Of the 2,346 households, 1,100 randomly selected to receive an extended questionnaire pertaining only to UC Davis' research questions.

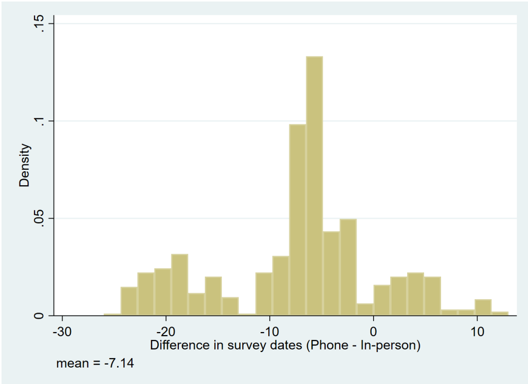
Following the 2018 Rabi harvest, IDinsight and UC Davis conducted phone and in-person surveys, respectively, on harvest outcomes. UCD surveyed 1,058 households of the original 1,100 who had completed extended questionnaires. As IDinsight's research focused on pulses production (and not production of other crops), IDinsight followed up over the phone with only the households who reported growing pulses in the previous survey. However, both surveys only asked about production of crops households had reported planting in the previous survey (i.e. if a household had not reported planting pulses then they were not asked about pulses production in either survey). IDinsight was able to reach 1,266 of the of the 1,533 households that reported growing pulses. Overall, there were 715 households that were in both the phone and in-person samples and 584 of those households were actually reached for both surveys. The sampling strategy is depicted in the figure below. For 429 of the 584, the respondent was the same household member across both surveys. We refer to the 489 households in which the same respondent completed both surveys as the overlapping sample. The two surveys covered an overlapping subset of outcomes. Both surveys asked farmers about quantity of crops harvested, quantity sold, quantity kept for home consumption, quantity saved for next year's seed, and reason for not harvesting (if crop was not harvested) for the most recent Rabi season. However, the phone survey only asked about pulses crops, which included pigeon pea, lentil, pea, fava bean, kidney bean, chickpea, urad, cowpea, and grasspea. The in-person survey covered all crops.

Figure 1: Survey Samples



The joint initial survey targeted the male head of household for agricultural questions and the female head of household for nutrition questions. The in-person survey kept the same targeting for the Rabi 2018 survey. The follow-up phone survey attempted to interview the same respondent from the agricultural portion of the previous survey. If the original respondent was unavailable, then someone in the household who knew about crop areas and harvest was surveyed instead. On average, the phone survey was conducted 7 days prior to the in-person survey for the overlapping sample. The difference in timing ranges from the phone survey being conduced 26 days before the in-person survey to 13 days after. The distribution of the lag between surveys is shown in the figure below.

Figure 2: Distribution of survey timing lag



3 Descriptive Statistics

Descriptive statistics for outcomes by survey mode are presented in Tables 1 and 2 below. Table 1 includes all observations and Table 2 only includes households that responded to both surveys and had the same member of the household respond both times. The difference in means and standard deviations (Phone - In-Person) are shown in the last two columns. Outcomes were winsorized at the top 5% for pulses crops which had over 100 observations (pigeon pea, lentil, fava bean, and pea). Each outcome is winsorized separately implying that the averages of production sold, home production, and seed may not add up to total production. The in-person survey provides information on whether peas were produced as vegetables or dried as a pulse. Therefore, we break up peas in the overlapping sample into green and dry. We cannot break up the full sample as we do not know the type of peas for households that are only in the phone survey. Generally, outcome means and standard deviations are larger in the phone survey for both the full and overlapping samples. For the full sample, the phone survey means are greater than in-person for 72% of outcomes and the standard deviations are greater for 64% of outcomes. For the overlapping sample, the phone survey is greater for 70% of means and 73% of standard deviations. We also present descriptive statistics for an inverse hyperbolic sine transformation of all outcomes in the appendix. We see the same pattern with bigger means and standard deviations in the phone data.

Table 1: Outcome Descriptive Statistics Full Sample

	Phone			In-person				
	mean	sd	N	mean	sd	N	diff mean	diff sd
pigeon pea total production	35.08	54.73	554	17.12	32.00	237	17.96	22.74
pigeon pea production sold	5.23	17.25	367	2.10	8.93	115	3.14	8.32
pigeon pea home production	45.72	55.72	367	35.45	57.67	115	10.27	-1.96
pigeon pea seed	1.84	3.14	367	0.59	1.49	115	1.25	1.64
lentil total production	97.94	127.58	696	78.47	96.85	369	19.46	30.73
lentil production sold	19.60	51.17	655	9.10	27.01	344	10.50	24.16
lentil home production	57.40	57.17	657	43.76	41.52	344	13.64	15.65
lentil seed	8.92	13.43	661	5.59	10.18	344	3.33	3.24
fava bean total production	57.98	61.07	331	54.99	65.76	160	3.00	-4.69
fava bean production sold	11.75	28.29	306	3.54	12.23	151	8.21	16.06
fava bean home production	39.49	39.79	307	40.10	40.98	151	-0.61	-1.19
fava bean seed	5.81	6.39	309	3.50	5.47	151	2.31	0.93
pea total production	162.20	251.92	341	145.03	228.11	162	17.17	23.82
pea production sold	126.74	226.73	315	140.64	240.27	144	-13.90	-13.54
pea home production	28.80	27.88	322	18.86	23.15	143	9.94	4.73
pea seed	2.78	5.29	322	0.89	2.49	144	1.89	2.80
cowpea total production	72.00	96.54	5	352.50	260.18	4	-280.50	-163.64
cowpea production sold	100.00	100.00	3	350.00	264.58	4	-250.00	-164.58
cowpea home production	16.67	5.77	3	6.25	7.50	4	10.42	-1.73
cowpea seed	3.33	5.77	3	0.50	1.00	4	2.83	4.77
grasspea total production	152.73	395.90	43	135.92	231.13	25	16.81	164.77
grasspea production sold	51.72	129.44	32	37.14	73.29	21	14.58	56.15
grasspea home production	96.72	221.63	32	39.48	44.69	21	57.24	176.94
grasspea seed	50.20	126.40	32	16.48	44.68	21	33.73	81.72
kidney bean total production	55.93	86.10	59	59.81	83.77	27	-3.88	2.32
kidney bean production sold	21.21	51.81	52	21.30	64.97	23	-0.09	-13.16
kidney bean home production	30.51	49.69	52	35.35	28.83	23	-4.84	20.85
kidney bean seed	4.26	6.02	52	4.09	5.53	23	0.17	0.48
chickpea total production	44.42	63.97	66	27.48	36.39	29	16.94	27.58
chickpea production sold	4.31	19.39	58	0.00	0.00	21	4.31	19.39
chickpea home production	34.51	50.74	59	32.14	36.02	21	2.37	14.72
chickpea seed	8.29	15.69	59	5.86	21.67	21	2.43	-5.98
urad total production	23.08	22.06	12	31.83	31.57	6	-8.75	-9.51
urad production sold	0.00	0.00	9	0.00	0.00	5	0.00	0.00
urad home production	24.60	18.60	10	35.40	29.51	5	-10.80	-10.91
urad seed	1.60	1.35	10	1.10	2.19	5	0.50	-0.84

This table reports means, standard deviations, and observations for outcomes in both the phone and in-person surveys. All households included. Variables for pigeon pea, lentil, fava bean, and pea winsorized at top 5%.

Table 2: Outcome descriptive statistics overlapping sample

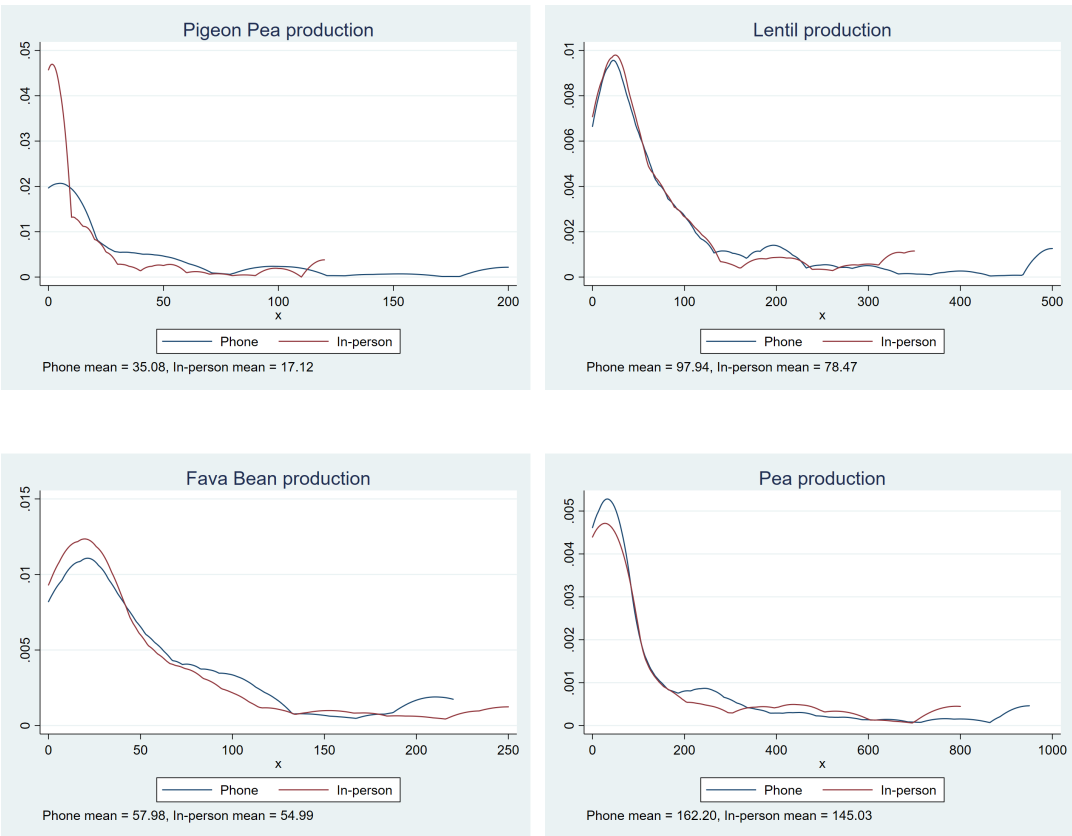
	Phone			In-person				
	mean	sd	N	mean	sd	N	diff mean	diff sd
pigeon pea total production	33.17	56.58	202	19.05	34.08	160	14.12	22.50
pigeon pea production sold	5.68	17.92	132	2.89	10.42	83	2.79	7.50
pigeon pea home production	43.16	57.28	132	39.42	64.94	83	3.75	-7.66
pigeon pea seed	1.41	2.80	132	0.52	1.37	83	0.89	1.42
lentil total production	101.48	127.92	224	83.06	96.31	222	18.42	31.61
lentil production sold	19.87	52.65	213	9.28	27.66	207	10.59	24.99
lentil home production	58.53	57.28	214	47.80	42.55	207	10.73	14.73
lentil seed	10.30	14.33	214	6.35	10.38	207	3.95	3.95
fava bean total production	47.20	46.83	100	49.17	55.26	97	-1.98	-8.44
fava bean production sold	8.15	22.23	92	4.20	13.47	94	3.95	8.77
fava bean home production	36.66	37.44	92	36.82	35.67	94	-0.17	1.77
fava bean seed	5.05	5.66	92	3.19	4.48	94	1.86	1.17
pea (overall) total production	179.86	267.31	120	169.20	244.23	116	10.66	23.08
pea (overall) production sold	138.58	238.19	109	162.61	255.62	106	-24.03	-17.43
pea (overall) home production	27.41	26.89	111	20.31	24.35	106	7.10	2.54
pea (overall) seed	2.88	5.44	111	0.94	2.64	106	1.94	2.79
pea (vegetable) total production	186.92	257.96	106	184.68	255.69	102	2.25	2.27
pea (vegetable) production sold	154.97	250.05	95	3.21	5.60	92	151.76	244.45
pea (vegetable) home production	27.35	26.51	97	18.78	23.90	92	8.56	2.61
pea (vegetable) seed	2.03	3.36	97	0.54	1.49	92	1.49	1.88
pea (dry) total production	62.07	70.75	14	56.43	64.29	14	5.64	6.47
pea (dry) production sold	27.39	62.86	14	1.75	5.85	14	25.64	57.01
pea (dry) home production	27.86	30.46	14	31.79	29.10	14	-3.93	1.36
pea (dry) seed	5.93	10.37	14	1.79	3.72	14	4.14	6.64
cowpea total production	10.00	14.14	2	305.00	417.19	2	-295.00	-403.05
cowpea production sold	0.00		1	300.00	424.26	2	-300.00	
cowpea home production	20.00		1	5.00	7.07	2	15.00	
cowpea seed	0.00		1	0.00	0.00	2	0.00	
grasspea total production	106.03	224.30	15	158.23	302.76	13	-52.20	-78.46
grasspea production sold	38.18	94.00	11	28.00	65.46	10	10.18	28.55
grasspea home production	77.45	152.00	11	35.80	45.21	10	41.65	106.79
grasspea seed	28.95	63.96	11	27.50	62.68	10	1.45	1.28
kidney bean total production	83.09	96.96	23	69.55	89.85	22	13.54	7.11
kidney bean production sold	27.05	71.37	21	24.50	69.32	20	2.55	2.05
kidney bean home production	41.95	48.63	21	36.90	30.41	20	5.05	18.22
kidney bean seed	5.81	7.76	21	4.70	5.69	20	1.11	2.06
chickpea total production	30.58	34.42	19	23.72	31.67	18	6.86	2.75
chickpea production sold	0.00	0.00	17	0.00	0.00	12	0.00	0.00
chickpea home production	27.06	28.72	17	25.42	27.12	12	1.64	1.60
chickpea seed	7.12	12.12	17	10.25	28.36	12	-3.13	-16.24
urad total production	32.50	26.30	4	32.75	32.53	4	-0.25	-6.23
urad production sold	0.00	0.00	4	0.00	0.00	4	0.00	0.00
urad home production	31.50	25.42	4	29.25	30.15	4	2.25	-4.73
urad seed	1.00	1.15	4	1.38	2.43	4	-0.38	-1.27

This table reports means, standard deviations, and observations for outcomes in both the IDinsight and UC Davis surveys. Only respondents tat were identical in both surveys are included. Variables for pigeon pea, lentil, fava bean, and pea winsorized at top 5%.

4 Distribution Tests

We first explore whether the in-person and phone survey data could be drawn from the same underlying distribution. The empirical distributions for total production of the most common pulses in both surveys, with the full sample winsorized at the top 5%, are shown below. In general, the IDinsight distributions have a fatter right tail (even with winsorization). The distributions for the other outcomes show similar patterns. Distributions for the inverse hyperbolic sine transformation of production for each crop are in the appendix. Although they are less right-skewed, we see a similar pattern with the phone distributions having more mass on the right.

Figure 3: Production densities by survey mode for most common pulses



We first test for equality of important properties of the in-person and phone survey distributions. We do not perform Kolmogorov-Smirnov tests for equality of distributions as the phone and in-person data are dependent and these tests would be too conservative (Ghanem 2017). We perform joint F-tests for equality of overall means, proportion of non-zero observations, proportion above the overall 75th percentile, and proportion above the 90h percentile by survey mode. We choose these properties as they are good characterizations of the outcome distributions in our setting. For the proportion above the 75th and 90th percentiles, we found the 75th and 90th percetiles for all of the data (using both phone and in-person) and then calculated the proportion above these overall percentiles in the phone and in-person data, respectively. We present descriptive statistics for each of these in our main crops for the full sample in Table 3 and the overlapping sample in Table 4. The phone survey tends to have a larger proportion of non-zero observations and more observations in the tails. We next perform joint tests of equality across crops within these outcomes. For example, for the overall sample means, we test if the means of total production are the same by survey mode for each crop jointly, if the means of production sold are equal for all crops, if the means of home production are equal for all crops, and if the means of production saved for seed are equal for all crops. This is the same as a joint F-test for balance by survey mode. P-values from the tests are reported in Tables 5 and 6 below. Table 5 uses the full sample and Table 6 restricts to the overlapping sample of households where the same individual responded to both surveys.

Table 3: Descriptive Statistics Proportion Non-Zero, Above Top Quartile, Above Top Decile Full Sample

	Proportion Non-Zero			Proportion Top Quartile			Proportion Top Decile		
	Phone	In-person	Diff	Phone	In-person	Diff	Phone	In-person	Diff
pigeon pea total production	0.15	0.05	0.11	0.15	0.05	0.11	0.10	0.03	0.07
pigeon pea production sold	0.02	0.00	0.01	0.02	0.00	0.01	0.02	0.00	0.01
pigeon pea home production	0.15	0.05	0.10	0.15	0.05	0.10	0.10	0.03	0.08
pigeon pea seed	0.06	0.01	0.05	0.06	0.01	0.05	0.06	0.01	0.05
lentil total production	0.28	0.15	0.13	0.21	0.10	0.10	0.07	0.03	0.04
lentil production sold	0.05	0.02	0.03	0.05	0.02	0.03	0.05	0.02	0.03
lentil home production	0.27	0.14	0.13	0.21	0.10	0.11	0.09	0.03	0.05
lentil seed	0.17	0.06	0.11	0.17	0.06	0.11	0.09	0.03	0.06
fava bean total production	0.13	0.06	0.07	0.13	0.06	0.07	0.09	0.04	0.05
fava bean production sold	0.03	0.01	0.02	0.03	0.01	0.02	0.03	0.01	0.02
fava bean home production	0.12	0.06	0.06	0.12	0.06	0.06	0.08	0.04	0.04
fava bean seed	0.09	0.03	0.06	0.09	0.03	0.06	0.09	0.03	0.06
pea total production	0.14	0.06	0.08	0.14	0.06	0.08	0.09	0.04	0.06
pea production sold	0.06	0.02	0.04	0.06	0.02	0.04	0.06	0.02	0.04
pea home production	0.13	0.04	0.08	0.13	0.04	0.08	0.09	0.02	0.06
pea seed	0.05	0.01	0.04	0.05	0.01	0.04	0.05	0.01	0.04

This table displays descriptive statistics for the proportion of non-zero observations, proportion of observations above the 75th percentile, and proportion above the 90th percentile by survey mode, as well as the difference by mode, using the full sample. The 75th and 90th percentiles are for all the data, not distinguishing by mode. The purpose is to determine whether more of the extremes are found in the phone or in-person data.

Overall, we reject equality between the phone and in-person surveys for most variables for

Table 4: Descriptive Statistics Proportion Non-Zero, Above Top Quartile, Above Top Decile Overlapping Sample

	Proportion Non-Zero			Proportion Top Quartile			Proportion Top Decile		
	Phone	In-person	Diff	Phone	In-person	Diff	Phone	In-person	Difference
pigeon pea total production	0.30	0.19	0.11	0.30	0.19	0.11	0.17	0.10	0.07
pigeon pea production sold	0.03	0.01	0.02	0.03	0.01	0.02	0.03	0.01	0.02
pigeon pea home production	0.29	0.19	0.10	0.29	0.19	0.10	0.18	0.10	0.08
pigeon pea seed	0.10	0.03	0.07	0.10	0.03	0.07	0.10	0.03	0.07
lentil total production	0.50	0.48	0.02	0.36	0.35	0.01	0.14	0.11	0.02
lentil production sold	0.09	0.06	0.03	0.09	0.06	0.03	0.09	0.06	0.03
lentil home production	0.48	0.47	0.01	0.36	0.35	0.02	0.17	0.14	0.04
lentil seed	0.32	0.23	0.09	0.32	0.23	0.09	0.19	0.14	0.05
fava bean total production	0.21	0.22	0.00	0.21	0.22	0.00	0.14	0.13	0.01
fava bean production sold	0.03	0.02	0.01	0.03	0.02	0.01	0.03	0.02	0.01
fava bean home production	0.20	0.21	-0.01	0.20	0.21	-0.01	0.13	0.15	-0.02
fava bean seed	0.14	0.11	0.03	0.14	0.11	0.03	0.14	0.11	0.03
pea total production	0.26	0.25	0.01	0.26	0.25	0.01	0.18	0.16	0.02
pea production sold	0.11	0.11	0.01	0.11	0.11	0.01	0.11	0.11	0.01
pea home production	0.24	0.18	0.06	0.24	0.18	0.06	0.16	0.10	0.06
pea seed	0.09	0.03	0.06	0.09	0.03	0.06	0.09	0.03	0.06

This table displays descriptive statistics for the proportion of non-zero observations, proportion of observations above the 75th percentile, and proportion above the 90th percentile by survey mode, as well as the difference by mode, using the overlapping sample. The 75th and 90th percentiles are for all the data, not distinguishing by mode. The purpose is to determine whether more of the extremes are found in the phone or in-person data.

both the full and overlapping samples. P-values are slightly larger but still significant when restricting to the overlapping sample. The one exception is production sold, where we cannot reject equality of phone and in-person across crops for proportion non-zero, proportion above the 75ht percentile, and proportion above the 90th percentile. These tests suggest that important properties of the distributions differ by survey mode.

Table 5: P-Values from Joint Tests of Significance Full Sample

	Overall	Proportion Zero	Proportion Top Quartile	Proportion Top Decile
Total Production	0.000	0.000	0.000	0.000
Production Sold	0.000	0.000	0.000	0.000
Home production	0.000	0.000	0.000	0.000
Seed	0.000	0.000	0.000	0.000

This table displays p-values from joint tests across crops for equality of means between phone and in-person using the full sample. The first column tests for equality of means for the outcomes of total production, production sold, home production, and seed. The second column tests for differences the proportion of non-zero observations for each outcome. For the third and fourth columns, we found the 75th and 90th percentiles of each outcome with all the data and then test if there's differences in the proportion of observations above the 75th and 90th percentiles,respcectively, by mode.

To further test whether the phone and in-person data come from the same underlying distri-

Table 6: P-Values from Joint Tests of Significance Overlapping Sample

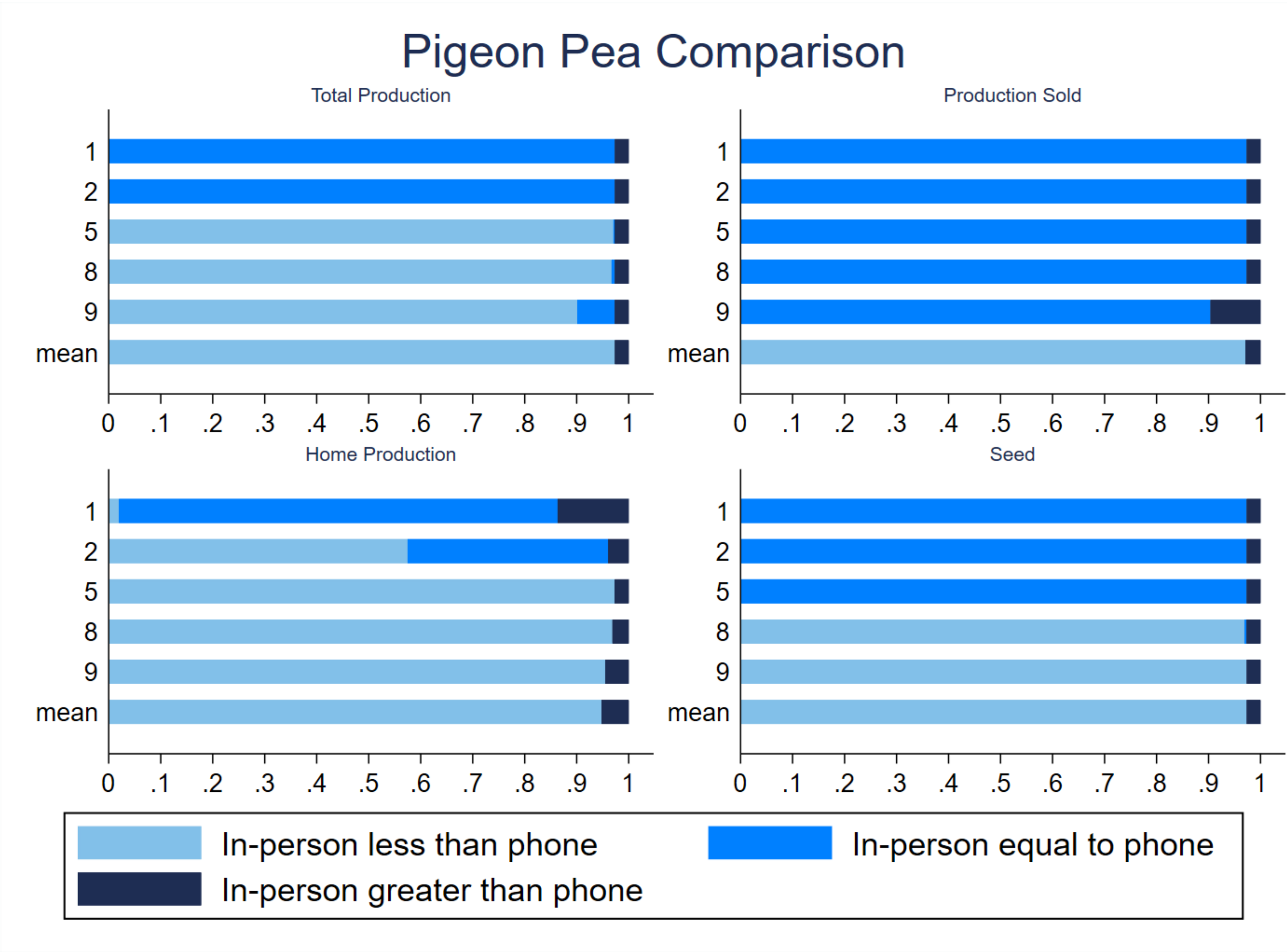
	Overall	Proportion Zero	Proportion Top Quartile	Proportion Top Decile
Total Production	0.004	0.004	0.004	0.030
Production Sold	0.022	0.199	0.199	0.199
Home production	0.014	0.001	0.000	0.000
Seed	0.000	0.000	0.000	0.000

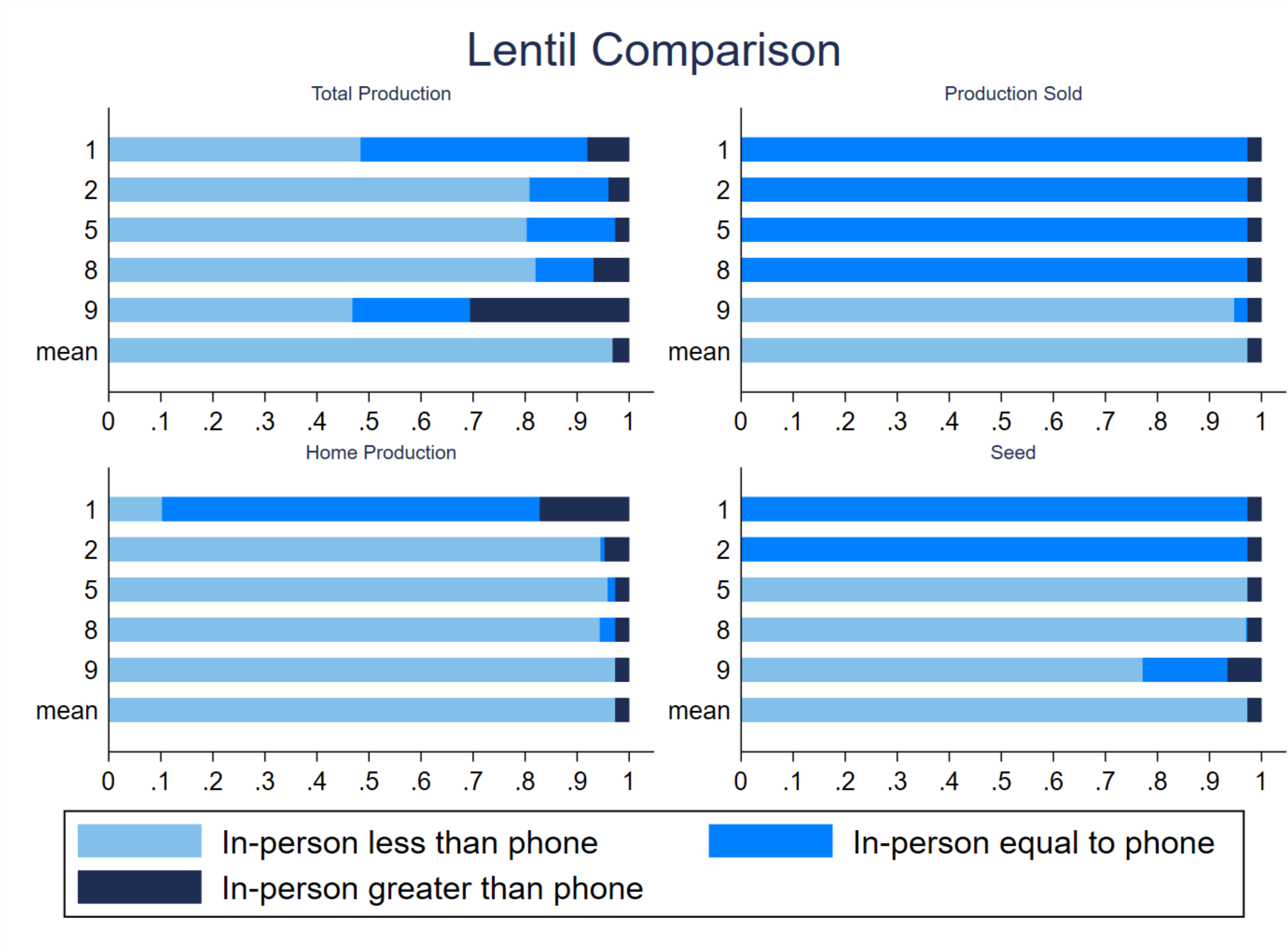
This table displays p-values from joint tests across crops for equality of means between phone and in-person using the overlapping sample. The first column tests for equality of means for the outcomes of total production, production sold, home production, and seed. The second column tests for differences the proportion of non-zero observations for each outcome. For the third and fourth columns, we found the 75th and 90th percentiles of each outcome with all the data and then test if there's differences in the proportion of observations above the 75th and 90th percentiles,respcectively, by mode.

bution, we run simulations to test how likely it is to get means and percentiles in draws of the in-person data as extreme as the phone means and percentiles. We first compute the means and 10th to 90th percentiles of outcomes for the full phone sample. We then take 10,000 bootstrap resamples from the in-person data and compute the means and 10th to 90th percentiles for each resample. Lastly, we calculate the proportion of in-person resamples which have means and percentiles less than the phone sample, equal to the phone sample, and greater than the phone sample. This allows us to assess how likely it is that the phone survey means and percentiles are a draw from the in-person survey distribution. The results for pigeon pea and lentil (the most common crops) with the full sample are shown in the Figure 4 below. Results for pea and fava bean are included in the appendix. The light blue represents the proportion of in-person resamples less than the phone means and deciles, the medium blue shows the proportion equal, and the dark blue shows the proportion of in-person resmaples greater then the phone sample. These proportions are shown for the 10th, 20th, 50th, 80th, and 90th percentiles as well as for the mean of each outcome. The dark blue is small for each outcome, meaning that it is rare to get means and deciles in the in-person resamples as extreme as the phone means and deciles.

We next run the same simulations using the overlapping sample. We compute means and deciles and resample from only the overlapping observations. The results for pigeon pea and lentil are included in Figure 5 below. Although the dark blue area is still small, it does greatly increase compared to the full sample, especially for total production and home production. This suggests either that households that responded on the phone but not in-person have larger outcome values, households that responded in-person but not over the phone have smaller outcomes, households that responded to both but had different household members

Figure 4: Simulation results full sample

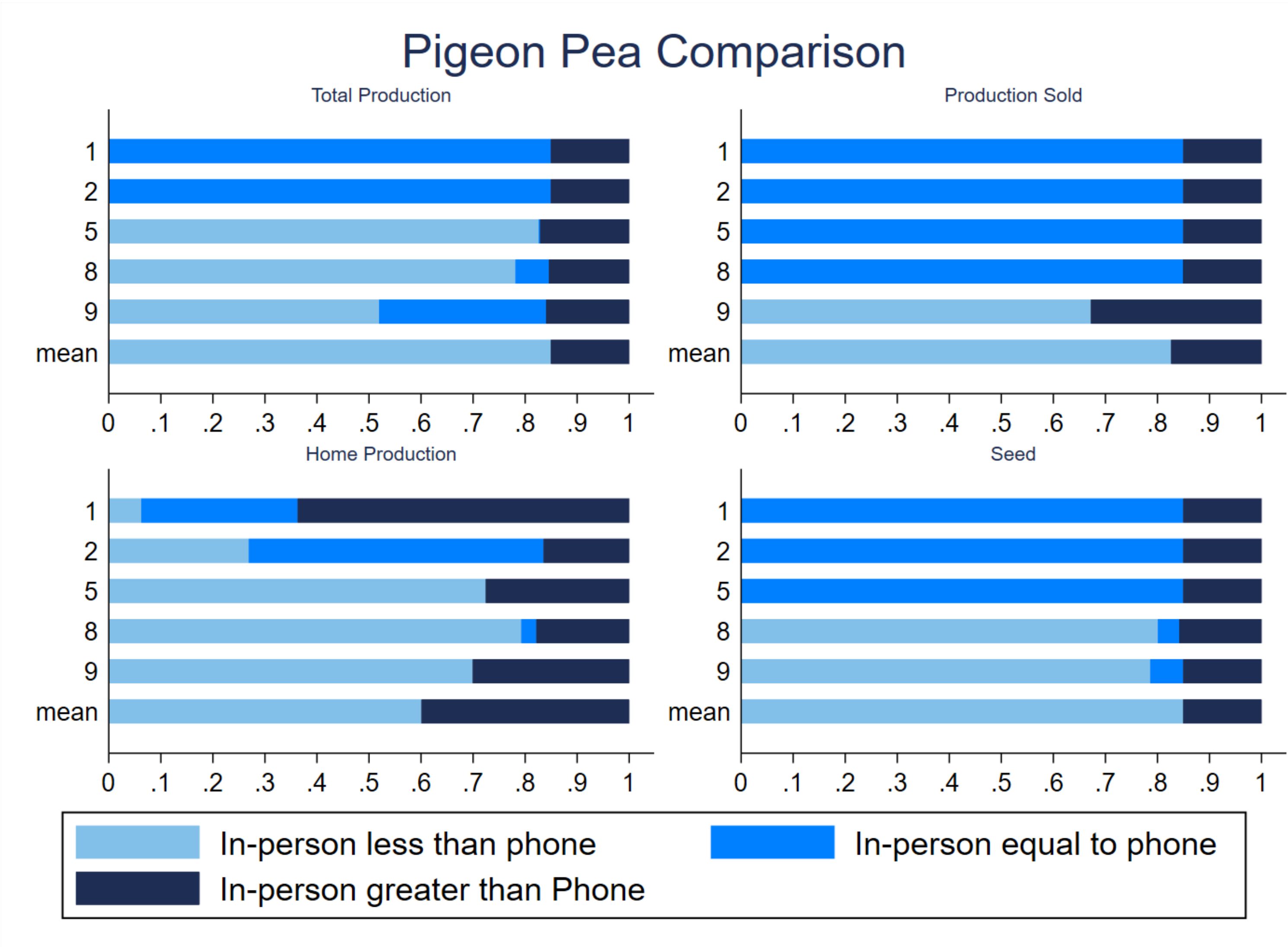




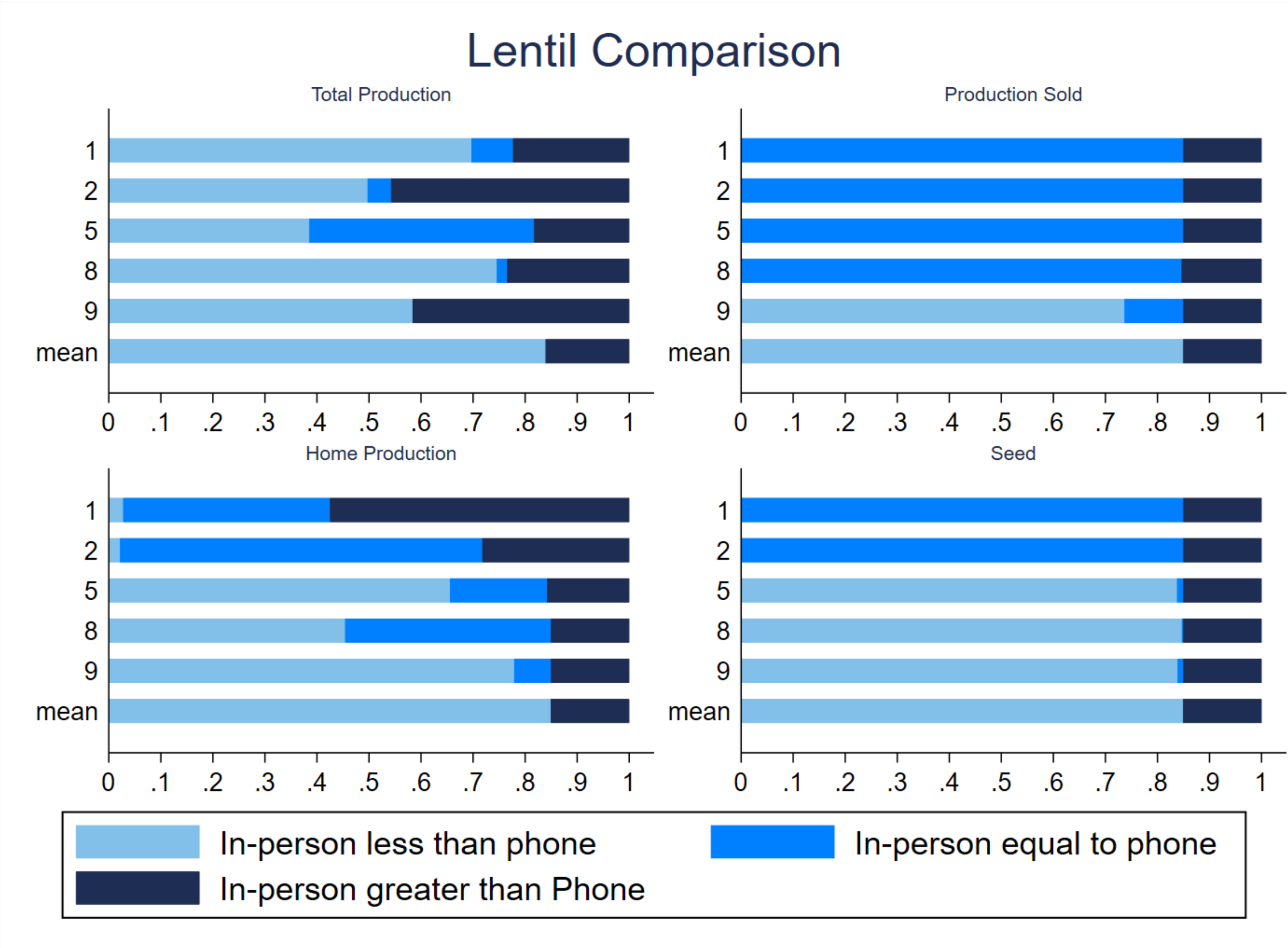
Proportion of UCD resamples less than, equal to, and greater than IDinsight for the 10th, 20th, 50th, 80th, and 90th percentiles, as well as the mean.

answer gave bigger responses over the phone, or some combination of these.

Figure 5: Simulation results overlapping sample



As the distributions are right-skewed and the phone survey has a longer right tail, it could be the case that the distributions look identical up to a certain percentile, at which point the phone survey starts to diverge. If this is the case, additional winsorizing in the phone survey could make the two distributions closer. Below, we plot the value at each decile by survey mode for production of each crop in the full sample. If additional winsorizing could get the phone survey distribution closer to that of the in-person survey, then we would expect the phone and in-person lines to match in the lower deciles and for the the phone survey to increase faster in the upper deciles. However, with the exception of pigeon pea, we see that the deciles



Proportion of UCD resamples less than, equal to, and greater than IDinsight for the 10th, 20th, 50th, 80th, and 90th percentiles, as well as the mean.

in each mode are fairly close with the phone survey increasing faster at some points and the in-person survey at others. We also see this for the overlapping sample. When we restrict to only the respondents who answered both surveys, the deciles get even closer and the phone and in-person lines track each other closely.

Figure 6: Production Values at Each Decile by Mode Full Sample

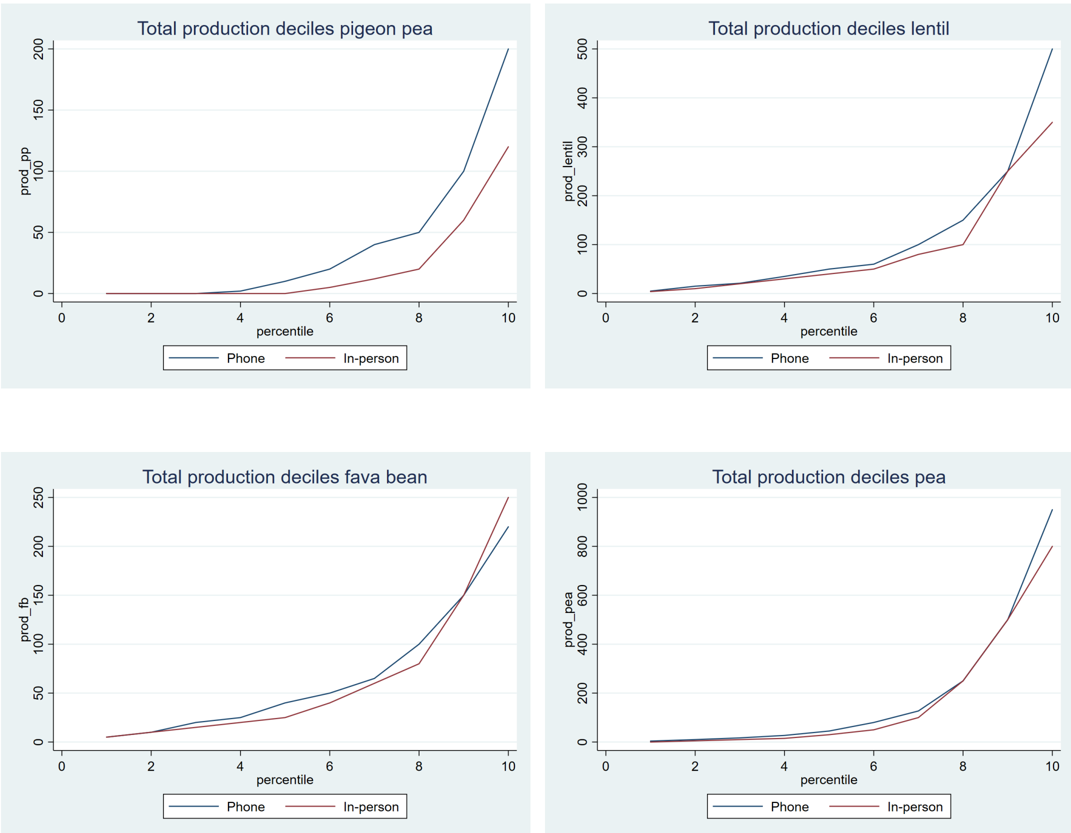
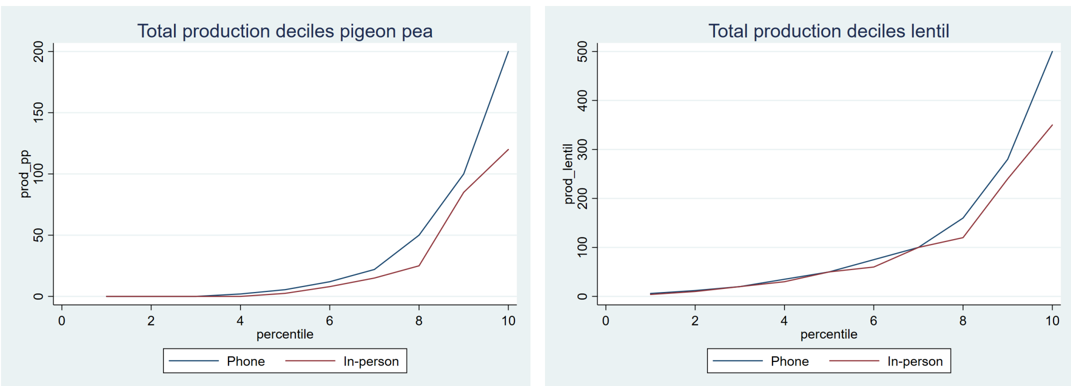
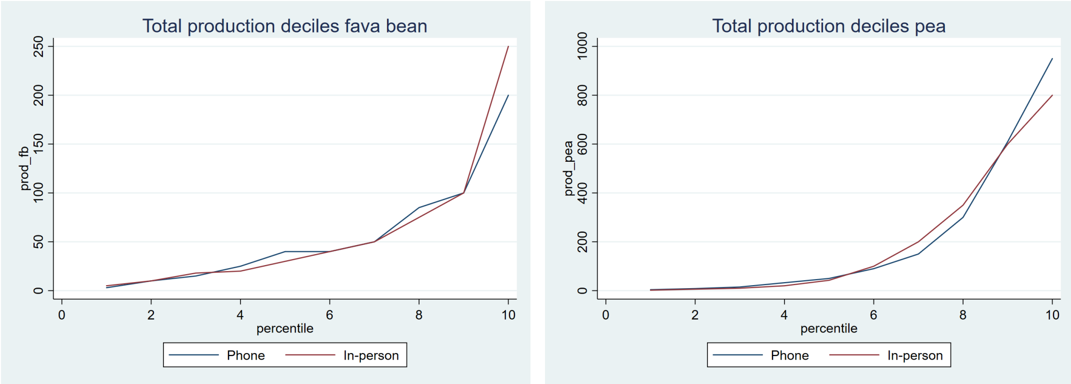


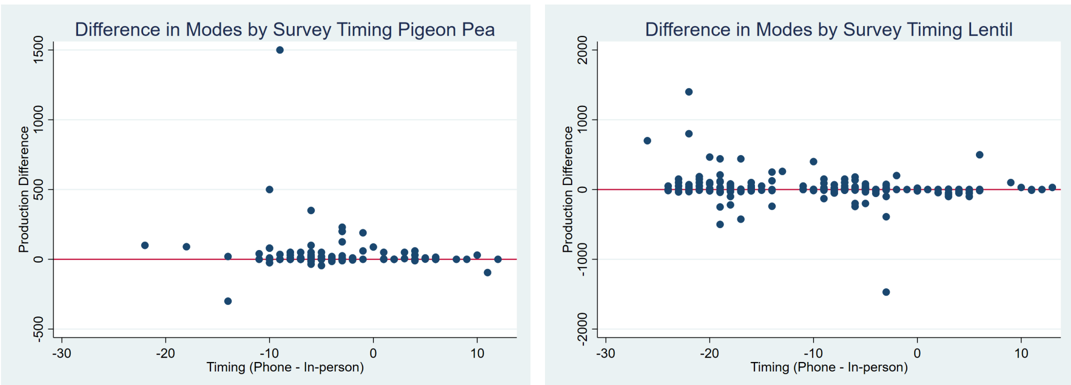
Figure 7: Production Values at Each Decile by Mode Overlapping Sample





The timing of the two surveys could be an important factor contributing to differences we see by mode. In general, the phone survey was conducted fist but there is some variation in timing. The graphs below plot the difference in response value for the same question and the same respondent against the difference in timing between the two surveys. Negative values on the horizontal axis indicate that the phone survey was conducted first and positive values indicate that the in-person survey was conducted first. We see the the differences are mostly around zero and do not differ based on which mode was conducted first.

Figure 8: Differences in Production by Mode Over Time



Additionally, Rounding may be an important driver of differences between distributions. Gourlay, Kilic, and Lobell (2019) find that overestimation of self-reported crop production, compared to crop cuts and remote sensing, is due in part to rounding. The graphs in Figure 9 below show the densities of the right-most digit (ones place) for production of each crop by mode. Although responses ending in 0 and 5 are more common than other numbers, the extent of rounding appears to be the same across modes.

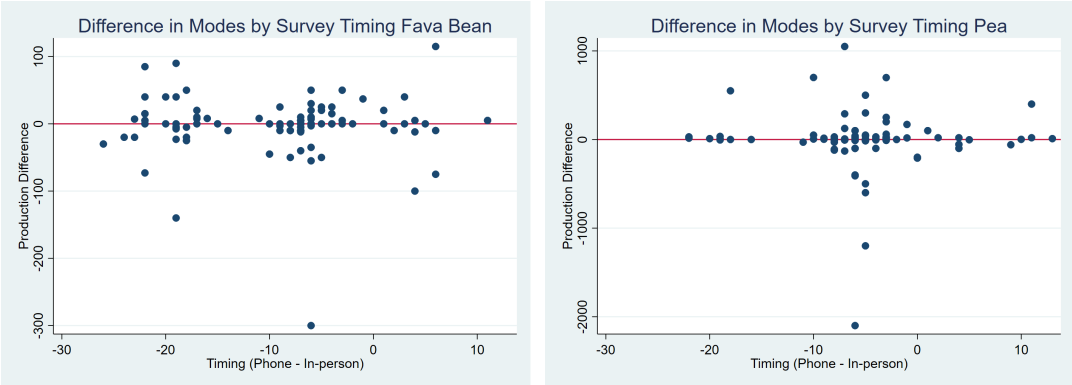
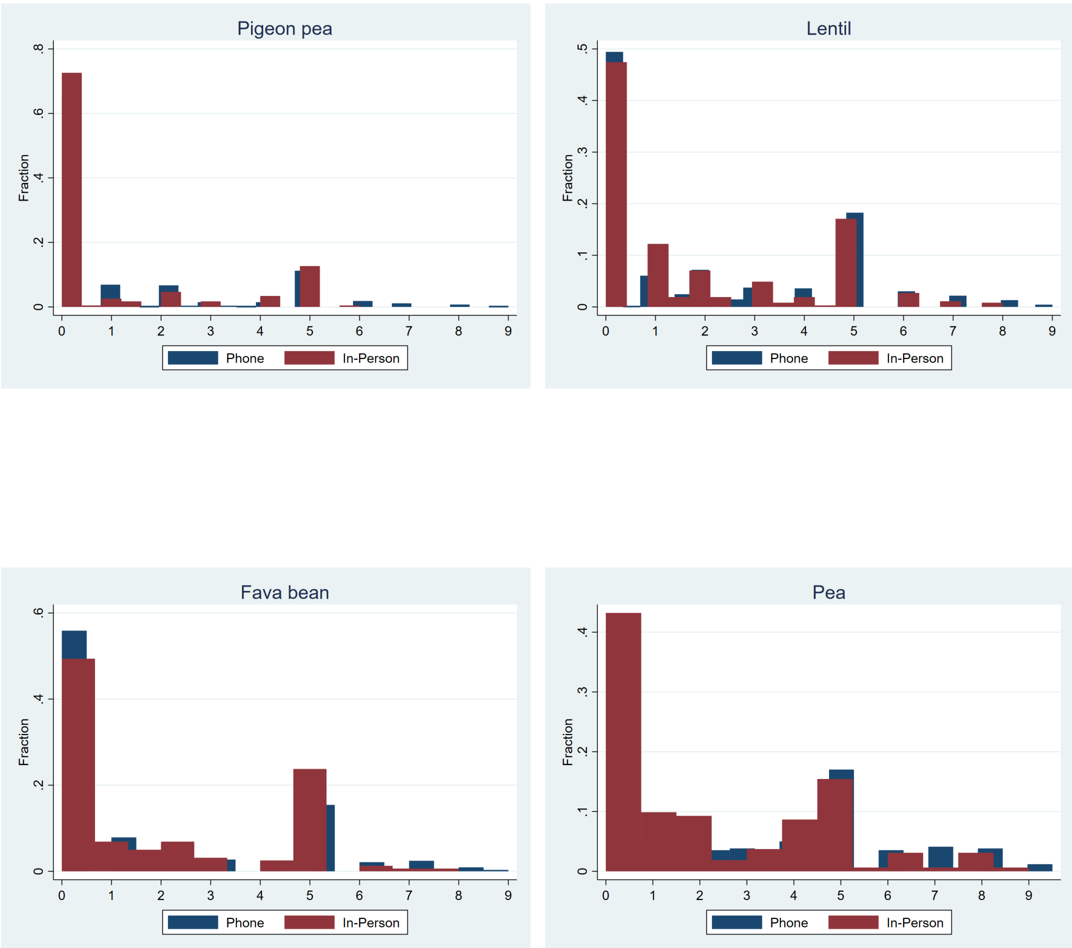


Figure 9: Density of right-most digit for pulses production



Overall, the joint tests and simulations suggest there are important differences between the in-person and phone distributions. These differences are less pronounced but still present when restricting to the overlapping sample. Additionally, we cannot attribute the differences to survey timing or rounding. In the next section, we tease out the extent to which these differences are due to mode and selection effects.

5 Selection versus Mode Effects

We next try to separate the differences between phone and in-person surveys into selection and mode effects. In terms of selection, we find evidence that the phone survey respondents are different from non-respondents on observables. Table 7 shows the mean characteristics of phone survey respondents and non-respondents. Column (1) includes all households the phone survey attempted to reach but did not and column (2) includes all households the phone survey was actually able to reach. Phone survey respondents tend to be younger, more male, and have more permanent housing structures than non-respondents. They are also are less likely to engage in sharecropping, less likely to receive Public Distribution System rations, and less likely to participate in MNREGA. The housing type, MNREGA and PDS questions were only asked to households that received the long questionnaires, therefore, the sample size is smaller.[2] We can also reject that coefficients are jointly equal to zero. This is consistent with the previous literature which finds that phone survey respondents are younger and more affluent than nationally representative samples.

Although there are differences in which individuals select into the phone and in-person surveys, it is possible that re-weighting on observables can produce similar estimates. We next conduct a weighting exercise where we first regress in-person outcomes on in-person covariates including age, caste, gender, number of plots, total hectares, whether the household sharecrops, and previous experience growing pulses. This is represented by Equation (1) below. Subscript i refers to the individual household and UCD refers to the in-person survey mode.

$$y_{i,UCD} = \beta X_{i,UCD} + \varepsilon_{i,UCD} \tag{1}$$

2. The long questionnaire households should look identical to short questionnaire households as the survey length was randomized.

Table 7: Balance Tests Phone Survey Respondents and Non-Respondents

Variable	(1) Overall Mean	(2) Respondent Mean	(3) Difference	(4) N
age	48.899	45.949	-2.949***	1,533
	(16.258)	(15.920)	(1.043)	
Male	0.772	0.886	0.115***	1,533
	(0.421)	(0.318)	(0.029)	
caste OBC	0.494	0.506	0.011	1,533
	(0.501)	(0.500)	(0.033)	
caste SC	0.131	0.122	-0.009	1,533
	(0.338)	(0.328)	(0.023)	
caste ST	0.049	0.042	-0.007	1,533
	(0.216)	(0.200)	(0.020)	
plots	6.629	7.261	0.632	1,533
	(4.790)	(4.575)	(0.424)	
sharecrop	0.371	0.295	-0.076**	1,533
	(0.484)	(0.456)	(0.038)	
total hectare	1.150	1.161	0.011	1,533
	(2.125)	(1.595)	(0.150)	
pakka	0.405	0.581	0.176***	715
	(0.493)	(0.494)	(0.054)	
mnrega	0.339	0.246	-0.093*	715
	(0.475)	(0.431)	(0.049)	
pds	0.744	0.626	-0.118***	715
	(0.438)	(0.484)	(0.043)	

F-Stat = 3.33

The above table tests for balance on baseline characteristics between phone survey respondents and non-respondent.

We then take the estimated coefficients from Equation (1) and use them to predict phone survey outcomes using phone survey data as shown in the equation below. The subscript IDI refers to the phone survey mode. Finally, we calculate average predicted outcomes for the phone survey and compare them to the actual outcomes. This is shown in Table 8.

$$\hat{y}_{i,IDI} = \hat{\beta} X_{i,IDI} \tag{2}$$

When using coefficients from the in-person regression, predicted phone means are close to but generally smaller than the actual in-person means. However, the predicted phone survey means are much smaller than the actual phone survey means. This implies that we cannot correct for selection bias on observables.

Table 8: Predicted Phone Survey Means using In-Person Coefficients

	In-person mean	Predicted phone mean	Actual phone mean
pigeon pea total production	17.12	13.63	35.08
pigeon pea production sold	2.10	1.65	5.23
pigeon pea home production	35.45	28.52	45.72
pigeon pea seed	0.59	0.42	1.84
lentil total production	78.47	67.49	97.94
lentil production sold	9.10	6.62	19.60
lentil home production	43.76	40.88	57.40
lentil seed	5.59	4.80	8.92
fava bean total production	145.03	129.16	162.20
fava bean production sold	140.64	124.74	126.74
fava bean home production	18.86	15.08	28.80
fava bean seed	0.89	0.60	2.78
pea total production	54.99	39.90	57.98
pea production sold	3.54	2.75	11.75
pea home production	40.10	30.45	39.49
pea seed	3.50	2.89	5.81

Phone survey means are predicted by regressing outcomes on covariates for in-person data and using the regression coefficients to predict outcomes with phone survey data.

In addition, we decompose the mean differences in outcomes to understand the direction of the selection and mode effects. To identify how much of the difference in the full sample is accounted for by the mode effect, we calculate the ratio of the difference in means in the overlapping sample to the difference in means in the entire sample. We attribute the residual (1-mode) to selection. This process is shown by Equations 3 and 4.

$$mode = \frac{\bar{y}_{overlapping,IDI} - \bar{y}_{overlapping,UCD}}{\bar{y}_{full,IDI} - \bar{y}_{full,UCD}} \tag{3}$$

$$selection = 1 - mode \tag{4}$$

The decomposition is presented in Table 9. The proportion of the difference in the full sample attributed to mode is greater than one for half of the outcomes. This suggests that the mode and selection effects work in opposite directions.

Table 9: Decomposition of Overall Differences by Mode and Selection

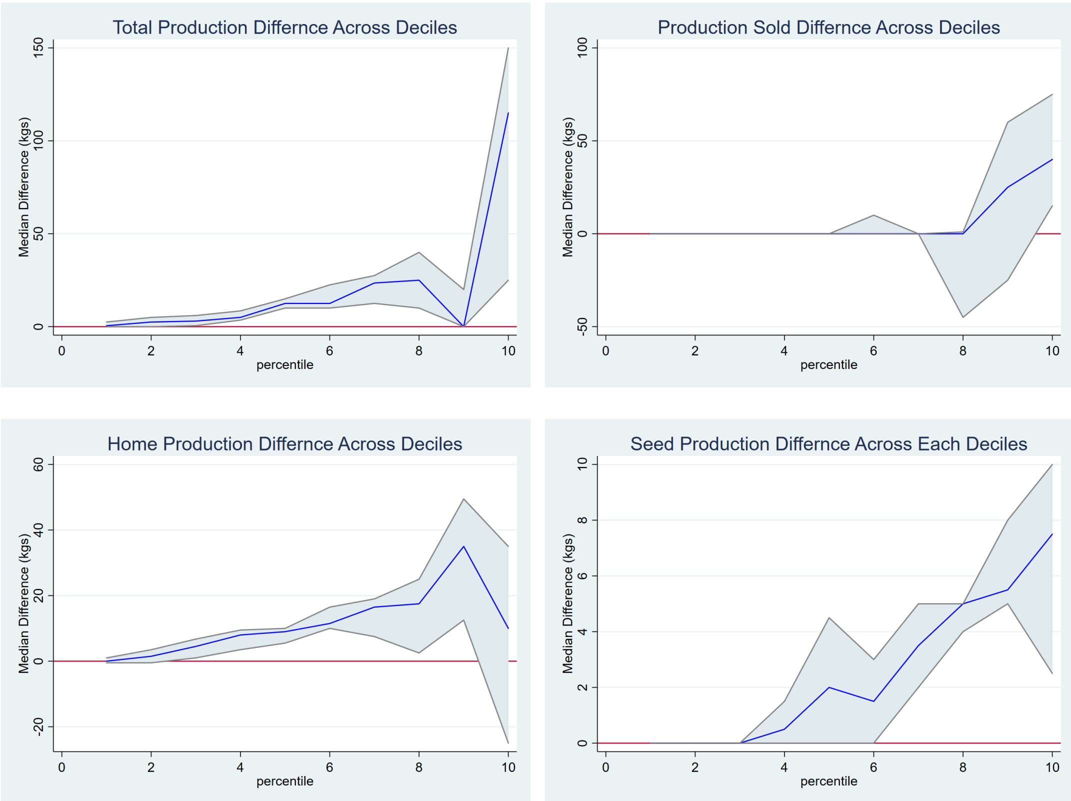
	Mode	Selection
pigeon pea total production	0.95	0.05
pigeon pea production sold	1.09	-0.09
pigeon pea home production	0.94	0.06
pigeon pea seed	0.77	0.23
lentil total production	1.04	-0.04
lentil production sold	1.01	-0.01
lentil home production	1.02	-0.02
lentil seed	1.15	-0.15
fava bean total production	0.81	0.19
fava bean production sold	0.69	0.31
fava bean home production	0.93	0.07
fava bean seed	0.87	0.13
pea total production	1.11	-0.11
pea production sold	1.09	-0.09
pea home production	0.95	0.05
pea seed	1.04	-0.04

This table reports the mode effect as the proportion of the overall difference in means accounted for by the difference in the overlapping sample and attributes the residual to selection.

In Table 9, we use differences in means to decompose mode and selection. However, as the data are right skewed, we also look at differences at each decile to better understand where in the distribution these differences are the biggest. We calculate the difference in values at each deciles for each crop and outcome in both the full and overlapping sample. Then, we calculate the median and interquartile range of differences within each outcome (total production, production sold, home production, and seed). Figure 10 below graphs the median difference at each decile in the full sample with a shaded region for the interquartile range. The median differences are bigger in the upper deciles, however, the upper deciles also have a larger interquartile range.

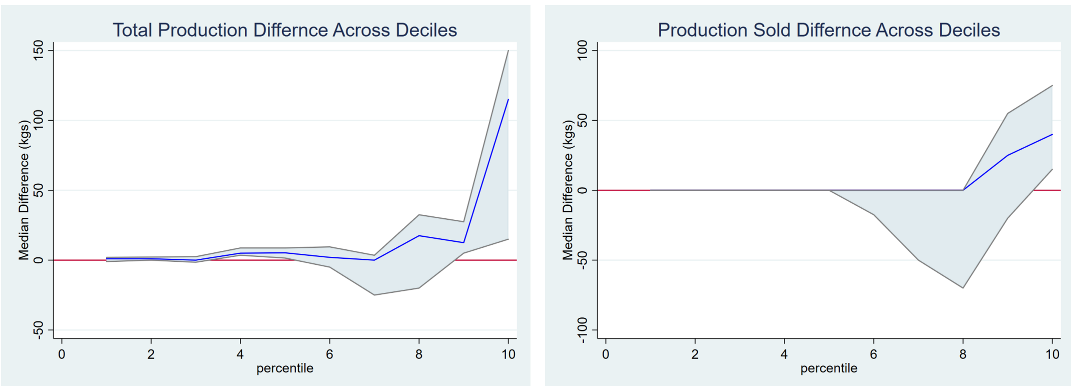
We also present the median difference at each decile for outcomes with the overlapping sample in Figure 11. We see the same pattern as the full sample where differences in the upper deciles are larger but also have a larger range. Together with the re-weighting and decomposition,

Figure 10: Median Difference at Each Decile Full Sample



this suggest that both mode and selection effects are driving differences in outcomes.

Figure 11: Median Difference at Each Decile Overlapping Sample



6 Treatment Effects

Lastly, we compare treatment effects from the RCT by survey mode. Table 10 presents estimated treatment effects for the three treatment arms: low, medium, and high intensity on production (kgs) of the most common pulses by mode. The odd columns use the phone survey data and the even columns use the in-person survey data. We use the following regression equation.

$$Y_{ijk} = \alpha_0 + \alpha_1 L_{jk} + \alpha_2 M_{jk} + \alpha_3 H_{jk} + \alpha_4 X_{ijk} + \gamma_k + \varepsilon_{ijk} \tag{5}$$

where Y_{ijk} is crop production for household i, in village j, and block k. L_{jk}, M_{jk}, and H_{jk} are binary variables indicating whether village j in block k was assigned to the "low", "medium", or "high intensity" treatment group respectively, with control as the reference group. X_{ijk} is a vector of additional household-level covariates which include: a binary variable for having cultivated pulses at least once in the 2 years prior to the 2017 Kharif season; binary variables for caste categories; respondent age; and respondent gender. γ_k is a block fixed effect and ε_{ijk} is the error term for household i in village j and block k. Standard errors are clustered at the village level (unit of randomization).

Overall, coefficients are similar by mode. We also conduct Wald post-estimation tests between the coefficients for each arm and find that we cannot reject equality of coefficients between the arms in either mode. To increase statistical power and better compare magnitudes, we pool treatment arms and present average treatment effects in Table 11. Again, coefficients are very similar across modes. We also present the percent increase in the control mean resulting from the treatment to better compare magnitudes across mode. With the exception of pigeon pea, the percent increase is similar across the phone and in-person surveys.

We next compare local average treatment effects across modes, shown in Table 12. We use assignment to treatment as an instrument for cultivation of each pulse to look at the effect of adopting cultivation on production. The empirical specification is below, where $A\hat{dopt}_{ijk}$ are predicted values of crop adoption. X_{ijk} are the same household controls as equation (5) and γ_k is a block fixed effect. Similar to the average treatment effects, we find that the phone and in-person coefficients are similar in both sign and magnitude. [**Have also included alternative table 12a where instead using treatment to instrument for crop adoption, I use treatment as an instrument for program participation (which I define as household having received seed, participated in/observed extension, or was aware of demo plots).**]

$$Y_{ijk} = \beta_0 + \beta_1 A\hat{dopt}_{ijk} + \beta_2 X_{ijk} + \ \gamma_k + u_{ijk} \tag{6}$$

Interestingly, the phone and in-person regression coefficients have similar standard errors, with the exception of pigeon pea. However, the phone survey had a larger sample size and larger outcome standard deviations as shown in Tables 1 and 2. This provides useful information for researchers as phone surveys may be noisier, yet they may still be more cost-effective. If phone surveys are substantially cheaper, researchers may be able to increase the sample size and achieve the same or greater precision as an in-person survey. To estimate how much greater the phone survey sample size needs to be to get standard errors as precise as the in-person survey, we divide the phone outcome variance by the in-person outcome variance. We derive this below.

$$\frac{\sigma_{IDI}}{\sqrt{bN}} = \frac{\sigma_{UCD}}{\sqrt{N}} \tag{7}$$

where b is how much the sample size needs to be scaled in the phone survey to make the standard errors equal.

$$b = \frac{\sigma^2_{IDI}}{\sigma^2_{UCD}} \tag{8}$$

We estimate b for all outcomes using the sample variances. The median $\hat{b}$ is 1.59 with an inter-quartile range of 0.87 to 3.11 in the full sample. That means the phone survey sample size needs to be 59% larger than in-person to get as precise standard errors. In the overlapping sample, the median $\hat{b}$ is 1.22 with an inter-quartile range of 1.03 to 2.74. Using the upper bound of this range, researchers can expand the sample size and use phone surveys as opposed to in-person without losing precision if the phone survey is three times cheaper.

An additional concern with using phone surveys in RCTs, especially in a panel setting, is whether those who attrit in in-person surveys are different than those who attrit in phone surveys. To better understand this, we compare initial characteristics for individuals that both the in-person and phone surveys attempted (both successfully and unsuccessfully) to follow-up with to those who attrit in each mode. Mean characteristics are presented in Table 13. Column 2 presents characteristics of those who attrit in the phone survey only, column 3 shows those who attrit in-person only, and column 4 show those who attrit in both surveys. Those who attrit in the phone survey look fairly similar to the overall population. However, there are only a small number of observations for those who attrit in-person and in both modes, therefore, we cannot draw conclusions about their characteristics.

The results of the RCT were intended to identify which program model was the most cost-effective at increasing adoption and production of pulses farming in the short-term. In the initial phone survey analysis, IDinsight aggregated production across different types of pulses (Anderson et al. 2021) as the goal of the program was to increase pulses production as a whole. In Table 14, we present average treatment effects of the three treatment arms on aggregate production by mode. As above, odd columns use phone survey data and even columns use in-person survey data. We present coefficients for the treatments separately as well as pooled. Coefficients are similar for each arm and there are no significant differences between treatments in either mode. The Rabi estimates for the pooled treatment are similar in sign and magnitude across modes. Although the pooled treatment effects look similar for Kharif, the in-person survey estimate is bigger in magnitude and is also statistically significant.

7 Discussion and Conclusion

When choosing a data collection mode, it is important to keep in mind the estimates the data will be used for. Although we do not know the true outcomes of households in our study, we do find that outcomes measured by phone surveys have larger means and standard deviations than in-person. We also find that the phone survey distributions tend to have a longer right tail. These differences arise from both mode and selection effects. Therefore, phone surveys may not be the best mode for estimating overall population means (at least for agricultural

Table 10: Average Treatment Effects on Pulse Production All Treatments

	(1) pigeon pea production IDi	(2) pigeon pea production UCD	(3) lentil production IDi	(5) lentil production UCD
low	0.700	0.762	2.393	1.967
	(1.160)	(0.526)	(4.961)	(6.342)
medium	0.944	0.659	2.504	1.436
	(1.317)	(0.479)	(3.953)	(5.040)
high	0.601	1.041*	-0.774	4.908
	(0.954)	(0.625)	(4.225)	(5.370)
low_med	0.875	0.873	0.983	0.932
low_high	0.937	0.720	0.554	0.647
med_high	0.806	0.602	0.456	0.530
control mean	4.683	0.890	27.923	25.161
ar2	0.095	0.055	0.173	0.107
N	2066	1055	2079	1055

	(5) fava bean production IDi	(6) fava bean production UCD	(7) pea production IDi	(8) pea production UCD
low	0.864	0.015	8.228**	4.883*
	(1.205)	(1.081)	(3.345)	(2.862)
medium	1.307	1.765	7.506**	6.075**
	(1.185)	(1.098)	(2.929)	(2.391)
high	1.143	1.022	9.198***	9.360***
	(1.091)	(1.196)	(2.318)	(2.420)
low_med	0.875	0.873	0.983	0.932
low_high	0.937	0.720	0.554	0.647
med_high	0.898	0.549	0.618	0.298
control mean	5.321	4.134	4.457	2.482
ar2	0.077	0.043	0.139	0.094
N	2079	1055	2079	1055

This table displays estimated average treatment effects of the three program models: low, medium, and high intensity on production of the most common pulses.The odd columns use data from the phone survey and the even columns use data from the in-person survey. We also present p-values for Wald test of equality of coefficients for each arm.

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 11: Average Treatment Effects on Pulse Production Pooled Treatment

	(1) pigeon pea production IDi	(2) pigeon pea production UCD	(3) lentil production IDi	(5) lentil production UCD
treat	0.751	0.817**	1.372	2.743
	(0.825)	(0.355)	(3.334)	(4.359)
control mean	4.683	0.890	27.923	25.161
percent control mean	0.160	0.918	0.049	0.109
ar2	0.096	0.056	0.173	0.109
N	2066	1055	2079	1055
	(5) fava bean production IDi	(6) fava bean production UCD	(7) pea production IDi	(8) pea production UCD
treat	1.112	0.969	8.300***	6.777***
	(0.875)	(0.894)	(1.917)	(1.650)
control mean	5.321	4.134	4.457	2.482
percent control mean	0.209	0.234	1.862	2.730
ar2	0.078	0.042	0.140	0.092
N	2079	1055	2079	1055

This table presents average treatment effects for the pooled treatment on production of major pulses crops. Odd columns use phone survey data and even columns use in person survey data. The percent increase in the control mean is displayed to compare magnitudes.

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 12: Local Average Treatment Effects on Pulse Production

	(1) pigeon pea production IDi	(2) pigeon pea production UCD	(3) lentil production IDi	(4) lentil production UCD
pigeon pea adoption	6.134	5.303***		
	(5.919)	(1.977)		
lentil adoption			13.646	21.184
			(30.861)	(30.133)
control mean	4.683	0.890	27.923	25.161
percent control mean	1.310	5.959	0.489	0.842
F first stage	27.665	17.798	16.412	13.664
N	2066	1055	2079	1055
	(5) fava bean production IDi	(6) fava bean production UCD	(7) pea production IDi	(8) pea production UCD
fava bean adoption	32.460**	26.368**		
	(14.639)	(13.335)		
pea adoption			68.265***	53.855***
			(9.952)	(9.207)
control mean	5.321	4.134	4.457	2.482
percent control mean	6.100	6.378	15.315	21.697
F first stage	2.505	1.867	30.277	23.999
N	2079	1055	2079	1055

This table presents local average treatment effects of adoption of cultivation on production of major pulses crops. Assignment to treatment is used as an instrument for cultivating each crop. Odd columns use phone survey data and even columns use in person survey data. The percent increase in the control mean is displayed to compare magnitudes.

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 12a: Local Average Treatment Effects on Pulse Production

	(1) pigeon pea production IDi	(2) pigeon pea production UCD	(3) lentil production IDi	(4) lentil production UCD
exposure	3.255	3.798**	5.919	12.748
	(3.457)	(1.750)	(14.370)	(20.260)
control mean	4.683	0.890	27.923	25.161
percent control mean	0.695	4.267	0.212	0.507
F first stage	57.860	23.801	58.721	23.801
N	2066	1055	2079	1055
	(5) fava bean production IDi	(6) fava bean production UCD	(7) pea production IDi	(8) pea production UCD
exposure	4.799	4.503	35.808***	31.498***
	(3.758)	(4.225)	(9.089)	(10.291)
control mean	5.321	4.134	4.457	2.482
percent control mean	0.902	1.089	8.033	12.690
F first stage	58.721	23.801	58.721	23.801
N	2079	1055	2079	1055

This table presents local average treatment effects of program exposure on production of major pulses crops. Assignment to treatment is used as an instrument for exposure to the program. We define exposure as having received pulse seed from an implementing NGO, being part of a farmers group or attending a farmer group meeting, being aware of demonstration plots nearby, or having participated in/observed one of the following extension activities: training, street play, night meeting, video screening, puppet show, wall painting, billboard, mic announcement, public song, or literature distribution. Odd columns use phone survey data and even columns use in person survey data. The percent increase in the control mean is displayed to compare magnitudes.

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

production). In our case, we do not find large differences in treatment effects between the phone and in-person data. Even though phone responses are larger and noisier they do not bias treatment effects, suggesting measurement error is classical in our context. When choosing a data collection mode, researchers should consider ways in which measurement error from mode may be correlated with treatment.

Our RCT results are especially relevant for researchers performing power calculations for future studies. As phone surveys are cheaper, researchers can reach a larger sample size, potentially increasing power. However, phone survey estimates may be less precise. We find similar standard errors on treatment effects using the phone and in-person survey data, but the phone survey had a larger sample size. Using the upper bound of our estimates, researchers can increase sample size and get similar precision if the phone survey is three times cheaper than in-person.

Table 13: Baseline Characteristics for Attrition by Mode

Variable	(1) Overall Mean	(2) Phone	(3) In-Person	(4) Phone and In-Person
age	46.650	47.864	50.500	51.000
	(16.418)	(16.827)	(17.577)	(24.759)
gender	0.857	0.780	0.900	1.000
	(0.350)	(0.416)	(0.316)	(0.000)
caste_obc	0.555	0.593	0.200	0.000
	(0.497)	(0.493)	(0.422)	(0.000)
caste_sc	0.103	0.076	0.100	0.333
	(0.305)	(0.267)	(0.316)	(0.577)
caste_st	0.024	0.051	0.100	0.000
	(0.152)	(0.221)	(0.316)	(0.000)
plot_roster_count	7.283	6.237	8.100	17.333
	(4.753)	(3.743)	(3.725)	(13.279)
sharecrop_yn	0.324	0.390	0.000	0.000
	(0.469)	(0.490)	(0.000)	(0.000)
total_hectare	1.127	1.165	1.050	2.563
	(1.345)	(1.882)	(1.032)	(2.922)
pakka	0.551	0.398	0.700	0.667
	(0.498)	(0.492)	(0.483)	(0.577)
mnrega	0.262	0.331	0.400	0.667
	(0.440)	(0.472)	(0.516)	(0.577)
pds	0.646	0.754	0.800	0.333
	(0.478)	(0.432)	(0.422)	(0.577)
Observations	715	118	10	3

This table displays baseline characteristics for those that both surveys attempted to reach in column (1), those who attrit only in the phone survey in column (2), those who attrit only in the in-person survey in column (3), and those who attrit in both survey in column (4).

Table 14: Average Treatment Effects on Aggregate Pulse Production

	(1) Kharif production IDi	(2) Kharif production UCD	(3) Kahrif production IDi	(4) Kharif production UCD
low	0.721	0.700		
	(1.156)	(0.546)		
medium	0.909	0.654		
	(1.324)	(0.484)		
high	0.761	0.974		
	(0.943)	(0.630)		
treat			0.800	0.773**
			(0.826)	(0.360)
low_med	0.904	0.945		
low_high	0.974	0.731		
med_high	0.915	0.666		
control_mean	4.718	1.020	4.718	1.020
percent_control_mean			0.170	0.758
ar2	0.095	0.060	0.096	0.062
N	2066	1055	2066	1055

	(5) Rabi production IDi	(6) Rabi production UCD	(7) Rabi production IDi	(8) Rabi production UCD
low	19.903	16.931		
	(12.997)	(14.685)		
medium	19.370**	18.063		
	(9.583)	(11.622)		
high	17.893**	29.549***		
	(7.922)	(10.773)		
treat			19.044**	21.474**
			(7.398)	(9.178)
low_med	0.970	0.943		
low_high	0.878	0.398		
med_high	0.882	0.365		
control_mean	53.088	45.820	53.088	45.820
percent_control_mean			0.359	0.469
ar2	0.147	0.066	0.148	0.065
N	2079	1055	2079	1055

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

References

Anderson, Ellen, Rupika Singh, Daniel Stein, and Kate Sturla. 2021. "What are the barriers to pulse cultivation in India? Evidence from a randomized controlled trial." *Unpublished Manuscript.*

Ballivian, Amparo, João Pedro Azevedo, Will Durbin, J Rios, J Godoy, and C Borisova. 2015. "Using mobile phones for high-frequency data collection." *Mobile Research Methods* 21.

Biemer, Paul P. 2001. "Nonresponse bias and measurement bias in a comparison of face to face and telephone interviewing." *Journal of Official Statistics* 17 (2): 295.

Crawfurd, Lee, David K Evans, Susannah Hares, Justin Sandefur, et al. 2021. "Teaching and testing by phone in a pandemic."

Ferreira, Aline Dayrell, Cibele Comini César, Deborah Carvalho Malta, Amanda Cristina de Souza Andrade, Cynthia Graciane Carvalho Ramos, Fernando Augusto Proietti, Regina Tomie Ivata Bernal, and Waleska Teixeira Caiaffa. 2011. "Validity of data collected by telephone survey: a comparison of VIGITEL 2008 and'Saude em Beaga'survey." *Revista Brasileira de Epidemiologia* 14:16–30.

Garlick, Robert, Kate Orkin, and Simon Quinn. 2020. "Call me maybe: Experimental evidence on frequency and medium effects in microenterprise surveys." *The World Bank Economic Review* 34 (2): 418–443.

Ghanem, Dalia. 2017. "Testing identifying assumptions in nonseparable panel data models." *Journal of Econometrics* 197 (2): 202–217.

Gourlay, Sydney, Talip Kilic, and David B Lobell. 2019. "A new spin on an old debate: Errors in farmer-reported production and their implications for inverse scale-Productivity relationship in Uganda." *Journal of Development Economics* 141:102376.

Greenleaf, Abigail R, Aliou Gadiaga, Georges Guiella, Shani Turke, Noelle Battle, Saifuddin Ahmed, and Caroline Moreau. 2020. "Comparability of modern contraceptive use estimates between a face-to-face survey and a cellphone survey among women in Burkina Faso." *PloS one* 15 (5): e0231819.

Kilic, Talip, Heather Moylan, John Ilukor, Clement Mtengula, and Innocent Pangapanga-Phiri. 2021. "Root for the tubers: Extended-harvest crop production and productivity measurement in surveys." *Food Policy* 102:102033.

Leo, Benjamin, Robert Morello, Jonathan Mellon, Tiago Peixoto, and Stephen T Davenport. 2015. “Do mobile phone surveys work in poor countries?” *Center for Global Development Working Paper,* no. 398.

Mahfoud, Ziyad, Lilian Ghandour, Blanche Ghandour, Ali H Mokdad, and Abla M Sibai. 2015. “Cell phone and face-to-face interview responses in population-based surveys: how do they compare?” *Field methods* 27 (1): 39–54.

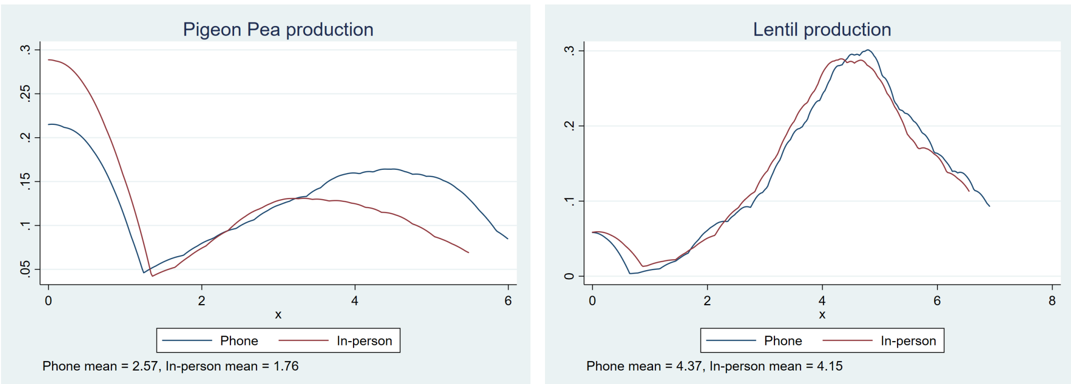
Nord, Mark, and Heather Hopwood. 2007. “Does interview mode matter for food security measurement? Telephone versus in-person interviews in the Current Population Survey Food Security Supplement.” *Public Health Nutrition* 10 (12): 1474–1480.

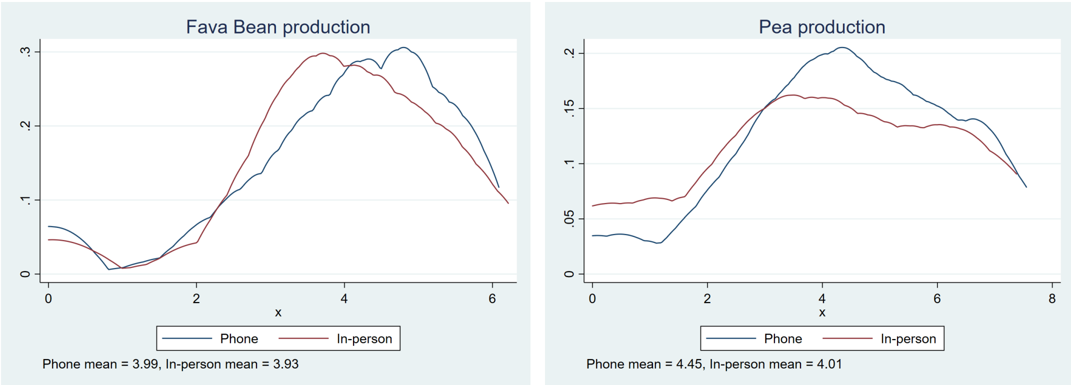
St-Pierre, Martin, and Yves Béland. 2004. “Mode effects in the Canadian Community Health Survey: A comparison of CAPI and CATI.” In *Proceedings of the Annual Meeting of the American Statistical Association, Survey Research Methods Section, August 2004.*

Scherpenzeel, Annette, and Philippe Eichenberger. 2001. *Mode effects in panel surveys: A comparison of CAPI and CATI.* Bundesamt für Statistik.

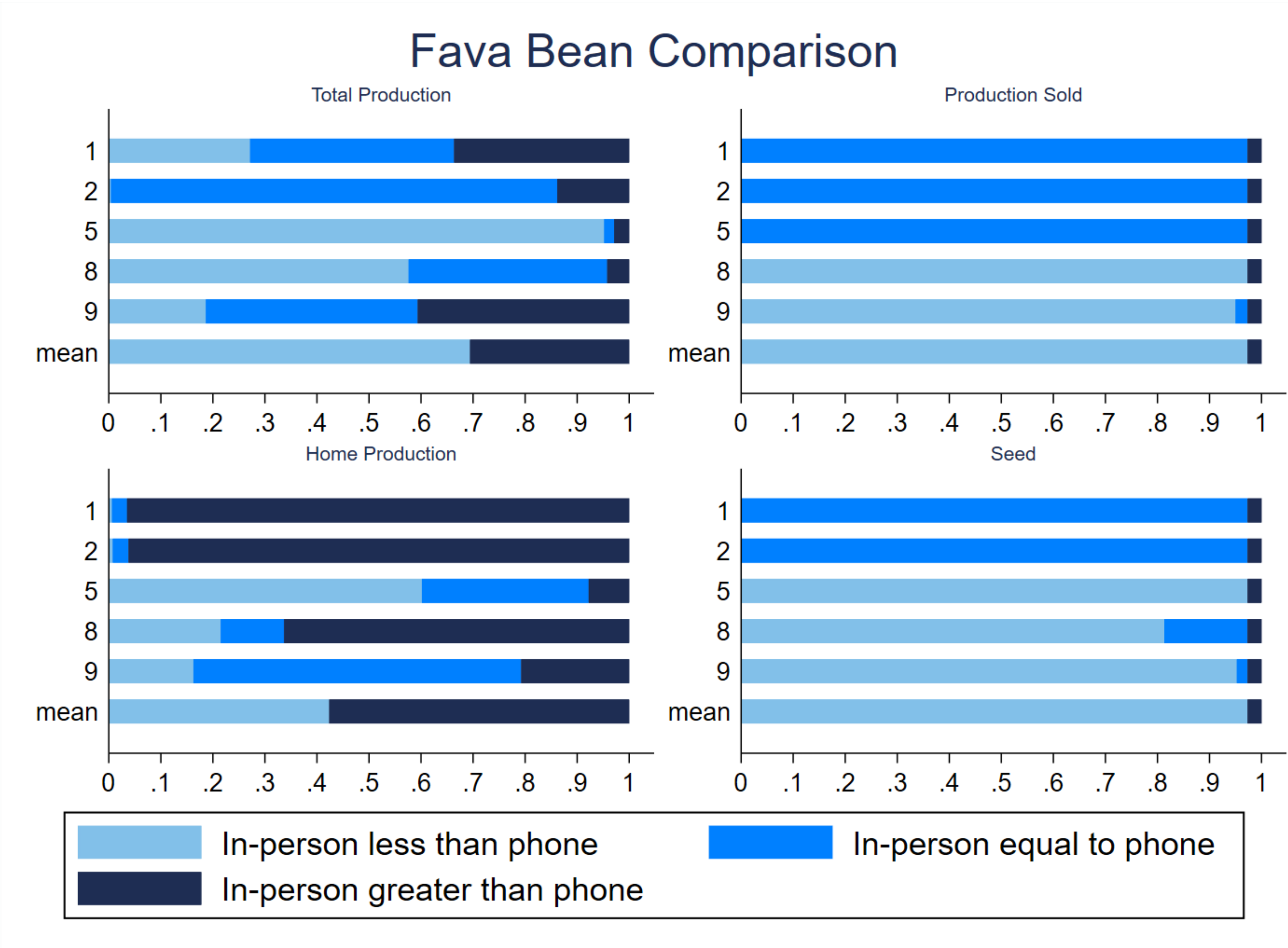
8 Appendix

Appendix Figure 1: Production densities by survey mode for most common pulses inverse hyperbolic sine transformation





Appendix Figure 2: Simulation results other crops



Appendix Table 1: Outcome Descriptive Statistics Full Sample Inverse Hyperbolic Sine Transformation

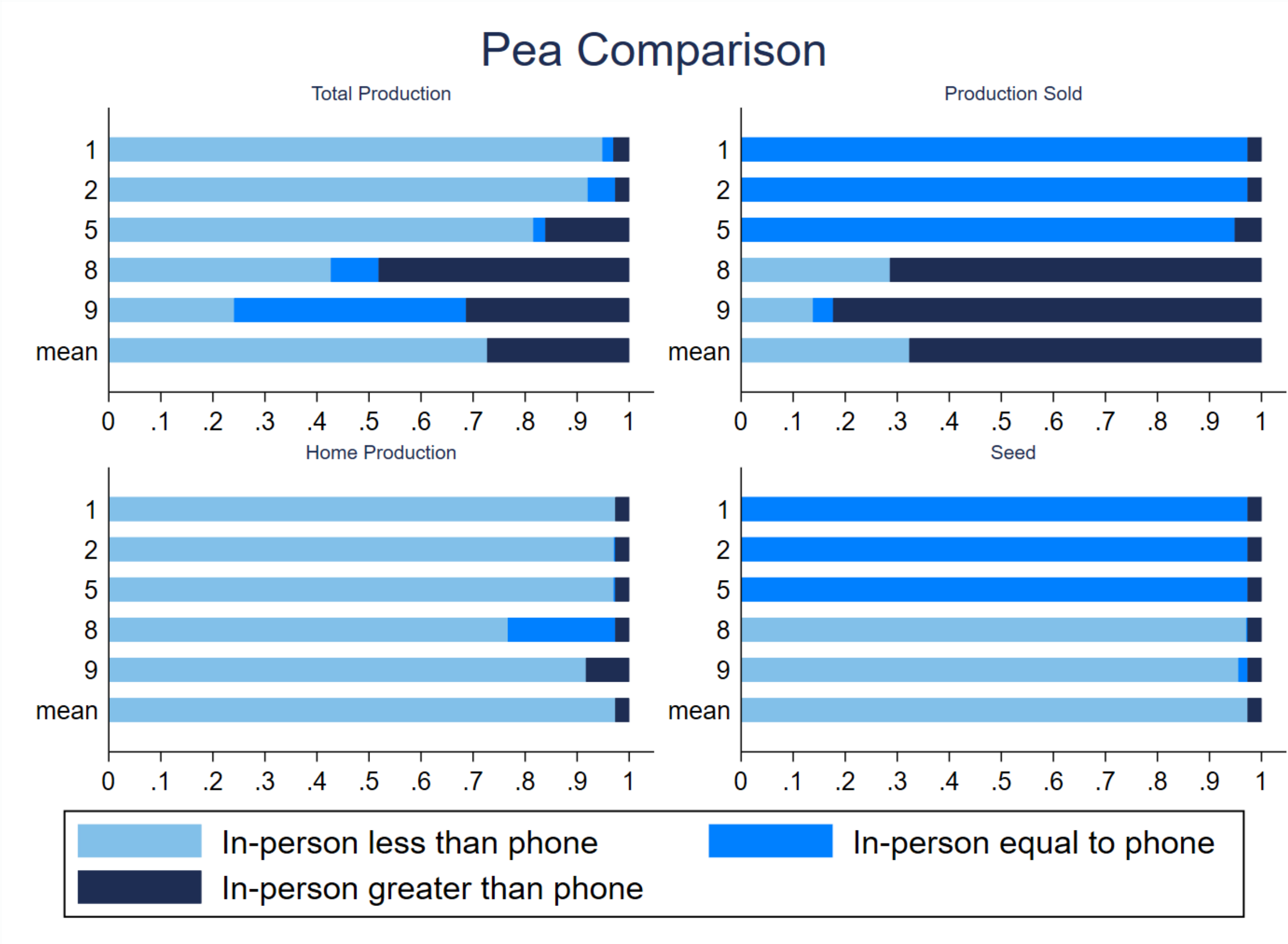
	Phone			In-person				
	mean	sd	N	mean	sd	N	diff mean	diff sd
pigeon pea total production	2.57	2.18	554	1.76	2.00	237	0.81	0.18
pigeon pea production sold	0.44	1.36	367	0.24	0.98	115	0.21	0.38
pigeon pea home production	3.67	1.52	367	3.33	1.43	115	0.34	0.09
pigeon pea seed	0.78	1.09	367	0.31	0.74	115	0.47	0.35
lentil total production	4.37	1.61	696	4.15	1.65	369	0.22	-0.04
lentil production sold	0.94	1.99	655	0.60	1.61	344	0.34	0.39
lentil home production	4.10	1.37	657	3.89	1.30	344	0.22	0.07
lentil seed	1.77	1.62	661	1.23	1.55	344	0.55	0.08
fava bean total production	3.99	1.55	331	3.93	1.46	160	0.06	0.08
fava bean production sold	0.91	1.84	306	0.37	1.23	151	0.53	0.62
fava bean home production	3.71	1.39	307	3.82	1.20	151	-0.11	0.19
fava bean seed	1.73	1.39	309	1.14	1.32	151	0.59	0.07
pea total production	4.45	1.91	341	4.01	2.22	162	0.45	-0.30
pea production sold	2.56	2.98	315	2.39	3.11	144	0.18	-0.12
pea home production	3.39	1.43	322	2.49	1.82	143	0.90	-0.39
pea seed	0.86	1.27	322	0.34	0.86	144	0.52	0.41
cowpea total production	3.05	2.92	5	5.85	1.92	4	-2.80	1.00
cowpea production sold	3.76	3.28	3	5.10	3.41	4	-1.34	-0.13
cowpea home production	3.46	0.40	3	1.60	1.86	4	1.86	-1.46
cowpea seed	1.00	1.73	3	0.36	0.72	4	0.64	1.01
grasspea total production	3.29	2.48	43	4.00	2.34	25	-0.71	0.14
grasspea production sold	1.22	2.42	32	1.35	2.48	21	-0.13	-0.06
grasspea home production	3.67	1.84	32	3.31	1.92	21	0.36	-0.07
grasspea seed	2.34	2.11	32	1.34	1.98	21	1.00	0.13
kidney bean total production	3.38	1.96	59	3.64	1.95	27	-0.25	0.02
kidney bean production sold	1.41	2.14	52	1.02	2.05	23	0.39	0.09
kidney bean home production	3.01	1.68	52	3.75	1.30	23	-0.74	0.38
kidney bean seed	1.38	1.32	52	1.27	1.44	23	0.11	-0.11
chickpea total production	3.48	1.71	66	2.81	1.96	29	0.67	-0.24
chickpea production sold	0.26	1.13	58	0.00	0.00	21	0.26	1.13
chickpea home production	3.52	1.34	59	3.63	1.13	21	-0.11	0.21
chickpea seed	1.68	1.61	59	0.78	1.41	21	0.90	0.20
urad total production	3.14	1.62	12	3.18	1.96	6	-0.04	-0.35
urad production sold	0.00	0.00	9	0.00	0.00	5	0.00	0.00
urad home production	3.67	0.71	10	3.36	2.09	5	0.30	-1.38
urad seed	1.06	0.79	10	0.56	1.00	5	0.50	-0.21

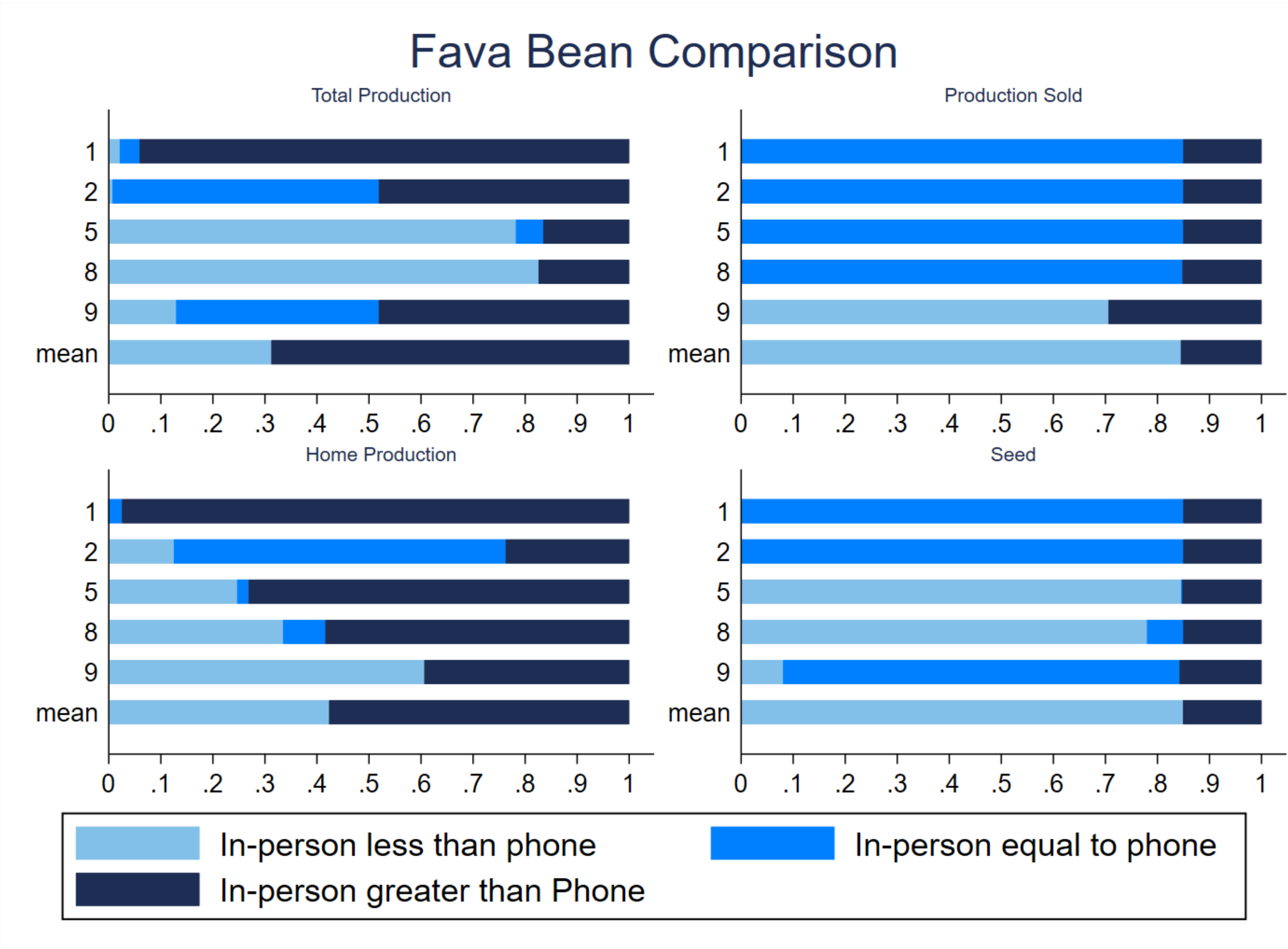
This table reports means, standard deviations, and observations for outcomes after an inverse hyperbolic sine transformation in both the phone and in-person surveys. All households included.

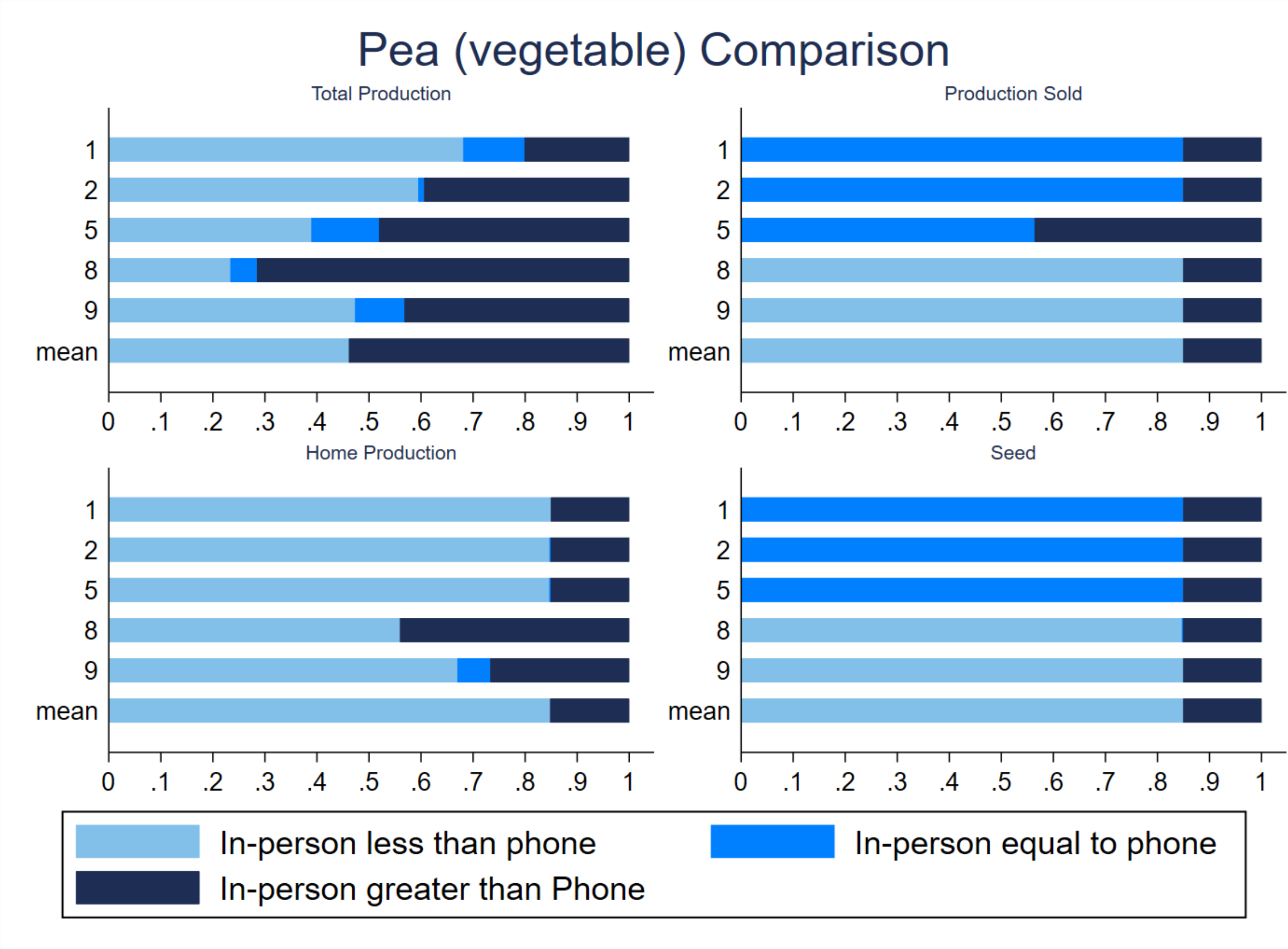
Appendix Table 2: Outcome Descriptive Statistics Overlapping Sample Inverse Hyperbolic Sine Transformation

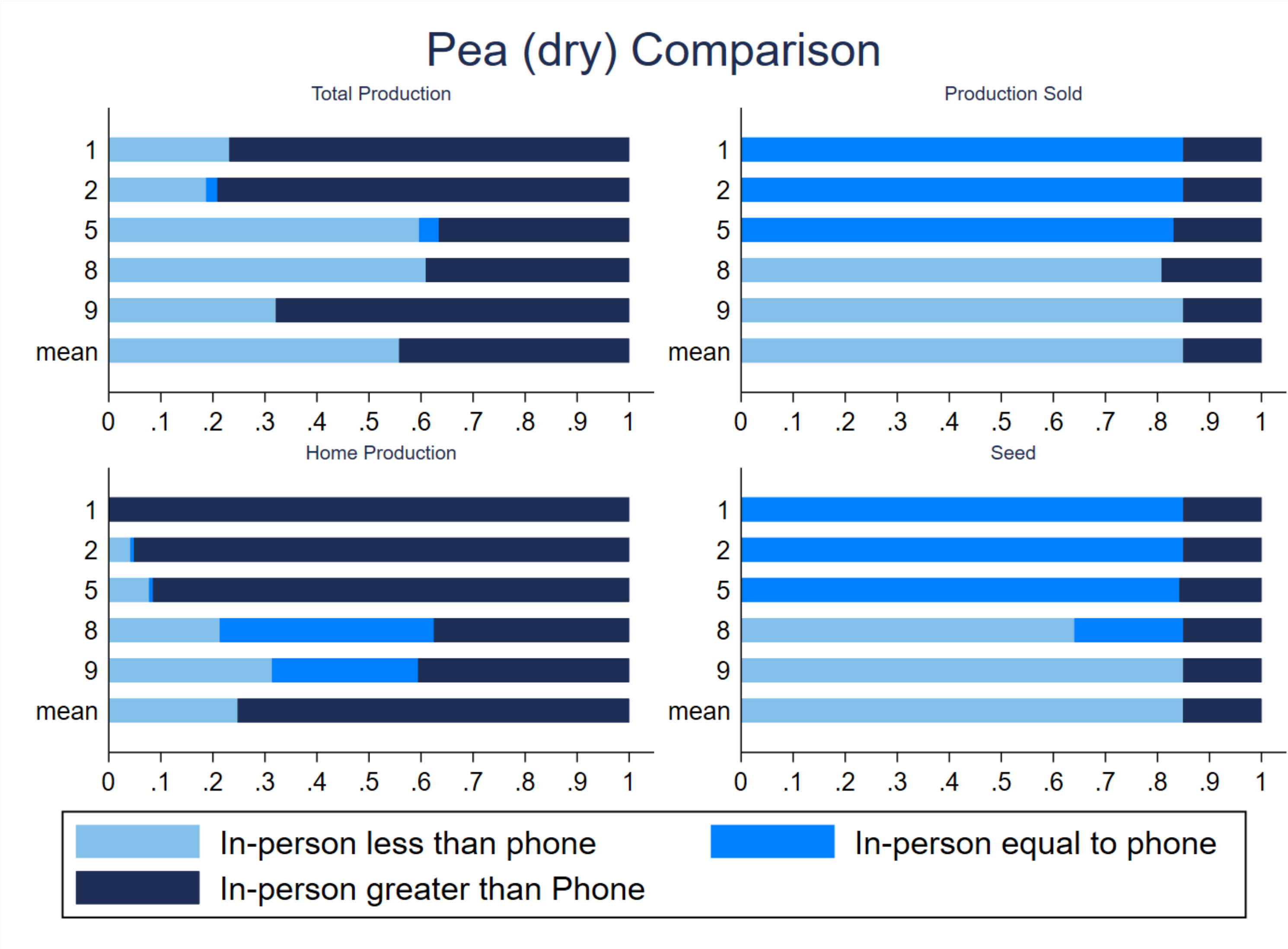
	Phone			In-person				
	mean	sd	N	mean	sd	N	diff mean	diff sd
pigeon pea total production	2.43	2.15	202	1.89	2.03	160	0.54	0.12
pigeon pea production sold	0.48	1.42	132	0.32	1.14	83	0.16	0.27
pigeon pea home production	3.50	1.57	132	3.36	1.47	83	0.14	0.10
pigeon pea seed	0.61	1.00	132	0.28	0.70	83	0.33	0.31
lentil total production	4.43	1.60	224	4.24	1.66	222	0.19	-0.06
lentil production sold	0.91	1.99	213	0.59	1.60	207	0.32	0.39
lentil home production	4.09	1.45	214	4.01	1.28	207	0.08	0.17
lentil seed	1.94	1.66	214	1.41	1.58	207	0.53	0.07
fava bean total production	3.87	1.49	100	3.98	1.29	97	-0.10	0.20
fava bean production sold	0.71	1.65	92	0.42	1.31	94	0.29	0.34
fava bean home production	3.64	1.40	92	3.80	1.15	94	-0.16	0.25
fava bean seed	1.64	1.33	92	1.18	1.26	94	0.47	0.07
pea (overall) total production	4.47	2.06	120	4.33	2.13	116	0.14	-0.07
pea (overall) production sold	2.63	3.05	109	2.72	3.18	106	-0.09	-0.13
pea (overall) home production	3.34	1.43	111	2.59	1.83	106	0.76	-0.40
pea (overall) seed	0.88	1.29	111	0.35	0.89	106	0.53	0.41
pea (vegetable) total production	4.51	2.11	106	4.35	2.24	102	0.15	-0.12
pea (vegetable) production sold	2.80	3.12	95	1.04	1.28	92	1.76	1.84
pea (vegetable) home production	3.35	1.42	97	2.42	1.86	92	0.94	-0.45
pea (vegetable) seed	0.80	1.16	97	0.27	0.72	92	0.53	0.44
pea (dry) total production	4.14	1.41	14	4.20	1.08	14	-0.07	0.34
pea (dry) production sold	27.39	62.86	14	1.75	5.85	14	25.64	57.01
pea (dry) home production	27.86	30.46	14	31.79	29.10	14	-3.93	1.36
pea (dry) seed	5.93	10.37	14	1.79	3.72	14	4.14	6.64
cowpea total production	1.84	2.61	2	5.04	2.89	2	-3.20	-0.28
cowpea production sold	0.00		1	300.00	424.26	2	-300.00	
cowpea home production	20.00		1	5.00	7.07	2	15.00	
cowpea seed	0.00		1	0.00	0.00	2	0.00	
grasspea total production	2.96	2.53	15	3.70	2.66	13	-0.74	-0.13
grasspea production sold	38.18	94.00	11	28.00	65.46	10	10.18	28.55
grasspea home production	77.45	152.00	11	35.80	45.21	10	41.65	106.79
grasspea seed	28.95	63.96	11	27.50	62.68	10	1.45	1.28
kidney bean total production	4.07	1.88	23	3.93	1.81	22	0.14	0.07
kidney bean production sold	27.05	71.37	21	24.50	69.32	20	2.55	2.05
kidney bean home production	41.95	48.63	21	36.90	30.41	20	5.05	18.22
kidney bean seed	5.81	7.76	21	4.70	5.69	20	1.11	2.06
chickpea total production	3.20	1.70	19	2.56	2.04	18	0.63	-0.33
chickpea production sold	0.00	0.00	17	0.00	0.00	12	0.00	0.00
chickpea home production	27.06	28.72	17	25.42	27.12	12	1.64	1.60
chickpea seed	7.12	12.12	17	10.25	28.36	12	-3.13	-16.24
urad total production	3.93	0.81	4	3.57	1.41	4	0.36	-0.60
urad production sold	0.00	0.00	4	0.00	0.00	4	0.00	0.00
urad home production	31.50	25.42	4	29.25	30.15	4	2.25	-4.73
urad seed	1.00	1.15	4	1.38	2.43	4	-0.38	-1.27

This table reports means, standard deviations, and observations for outcomes after an inverse hyperbolic sine transformation in both the phone and in-person surveys. Only respondents that were identical in both surveys are included.


Pea Comparison
Total Production
Production Sold
Home Production
Seed
1
2
5
8
9
mean
0
.1
.2
.3
.4
.5
.6
.7
.8
.9
1
In-person less than phone
In-person equal to phone
In-person greater than phone


Fava Bean Comparison
Total Production
Production Sold
Home Production
Seed
1
2
5
8
9
mean
0 .1 .2 .3 .4 .5 .6 .7 .8 .9 1
In-person less than phone
In-person equal to phone
In-person greater than Phone


Pea (vegetable) Comparison
Total Production
Production Sold
Home Production
Seed
1
2
5
8
9
mean
0
.1
.2
.3
.4
.5
.6
.7
.8
.9
1
In-person less than phone
In-person equal to phone
In-person greater than Phone


Pea (dry) Comparison
Total Production
Production Sold
Home Production
Seed
1
2
5
8
9
mean
0
.1
.2
.3
.4
.5
.6
.7
.8
.9
1
In-person less than phone
In-person equal to phone
In-person greater than Phone