



# Andhra Pradesh Rural Inclusive Growth Project: Findings from the Baseline Survey

Lucas Kitzmueller and PJ Paul<sup>1</sup>

23 March 2018

---

<sup>1</sup> Contact: [lucas.kitzmueller@idinsight.org](mailto:lucas.kitzmueller@idinsight.org), [pj.paul@idinsight.org](mailto:pj.paul@idinsight.org). We gratefully acknowledge help and guidance from Doug Johnson, Deeptha Umapathy, Jeff McManus, Ronald Abraham, Amy Chen, Nidhaanjit Jain, Alison Connor, Syed Maqbool, and Akib Khan. IDinsight also gratefully acknowledges the contributions to the design of this evaluation from SERP staff, particularly Evaluation and Learning Director Udayabhanu Prakash V and the World Bank team, particularly Abhishek Gupta and Giovanna Prennushi. All errors are our own.

## Table of Contents

<b>1</b>	<b>Introduction .....</b>	<b>4</b>
<b>2</b>	<b>Background .....</b>	<b>6</b>
<b>3</b>	<b>Assessment of Data Quality.....</b>	<b>8</b>
<b>4</b>	<b>Validity of Control Group .....</b>	<b>9</b>
<b>5</b>	<b>Household Profile.....</b>	<b>10</b>
<b>6</b>	<b>Income and Consumption .....</b>	<b>14</b>
<b>7</b>	<b>Crop Cultivation .....</b>	<b>16</b>
<b>8</b>	<b>Animal Husbandry.....</b>	<b>25</b>
<b>9</b>	<b>Drinking Water.....</b>	<b>25</b>
<b>10</b>	<b>Sanitation Facilities and Behavior.....</b>	<b>26</b>
<b>11</b>	<b>Handwashing .....</b>	<b>30</b>
<b>12</b>	<b>Anthropometric Summaries .....</b>	<b>30</b>
<b>13</b>	<b>Kirana Stores.....</b>	<b>31</b>
	<b>References .....</b>	<b>35</b>
	<b>Appendix.....</b>	<b>36</b>
	<b>Annex I Village and Household Selection .....</b>	<b>36</b>
	<b>Annex II Construction of Income Aggregate .....</b>	<b>44</b>
	<b>Annex III Construction of Consumption Aggregate .....</b>	<b>49</b>
	<b>Annex IV Field Observations and Comparisons with Back-Check Survey Data .....</b>	<b>53</b>

## Enclosures

- 1) Impact evaluation design document
- 2) Survey manual
- 3) Baseline questionnaire with SurveyCTO codes
- 4) Final datasets containing household and Kirana store data collected in the census and baseline surveys
- 5) Stata do-files of data cleaning, income and consumption aggregate construction and analysis, sampling weight construction, balance checks, graphs, and descriptive statistics

## Tables

Table 1. Balance Checks. ....	9
Table 2. Household Composition and Characteristics. ....	10
Table 3. Ownership Rates of Common Household Assets. ....	12
Table 4. NREGA. ....	13
Table 5. Household Participation in Government Schemes. ....	13
Table 6. Monthly Aggregated Household Income (in INR). ....	14
Table 7. Monthly Aggregate per Capita Consumption (in INR). ....	15
Table 8. Food Groups Consumed by $\geq 50\%$ of Households by Dietary Diversity Score. ....	15
Table 9. Crop Diversification. ....	17
Table 10. Agricultural Land Available for Crop Cultivation. ....	17
Table 11. Crop Intensity. ....	18
Table 12. Livestock Ownership. ....	25
Table 13. Determinants of Individual-Level Latrine Use. ....	29
Table 14. Characteristics of Hand Washing Facility. ....	30
Table 15. Kirana Store Characteristics. ....	31
Table 16. Other Occupations of Kirana Store Owners. ....	31

## Figures

Figure 1. Primary Occupation of Household Members. ....	12
Figure 2. Crop Cultivation. ....	16
Figure 3. Crop Productivity (in Tons per Hectare). ....	19
Figure 4. Farm Assets: Recommended vs. Actual Use. ....	20
Figure 5. Agricultural Operating Expenses. ....	21
Figure 6. Share of Crop Revenue from FVC Crops. ....	22
Figure 7. Reasons for Not Selling Crop. ....	23
Figure 8. Distribution of Buyer Types. ....	23
Figure 9. Reasons for Selling to Buyers. ....	24
Figure 10. Primary Sources of Drinking Water. ....	25
Figure 11. Household Latrine Access. ....	26
Figure 12. Private Latrine Types. ....	27
Figure 13. Well-Maintained Latrines by Type of Funding. ....	27
Figure 14. Household Latrine Use. ....	28
Figure 15. Anthropometric Summaries. ....	30
Figure 16. Major Suppliers of Kirana Stores. ....	32
Figure 17. Sanitation Products Sold by Kirana Stores. ....	33
Figure 18. Fruits and Vegetables Sold by Kirana Stores. ....	33
Figure 19. Nutritious Products Sold by Kirana Stores. ....	34

## Acronyms

APRIGP	Andhra Pradesh Rural Inclusive Growth Project
DWCRA	Development of Women and Children in Rural Areas
FPG	Farmer Producer Group
HH	Household
MNS	Mandal Nodal Store
NREGA	National Rural Employment Guarantee Act
PoP	Poorest of the Poor
SERP	Andhra Pradesh Society for the Elimination for Rural Poverty
SHG	Self Help Group

## Executive Summary

In 2015, the state of Andhra Pradesh, with support from the World Bank, launched the **Andhra Pradesh Rural Inclusive Growth Project (APRIGP)**. The project, which is implemented by SERP, an autonomous society of the Department of Rural Development, **seeks to increase agricultural incomes and secure access to human development services and social entitlements.**

This **report presents findings from a survey of 8943 households** conducted in APRIGP program areas and a set of similar control areas between July 2017 and October 2017. The primary purpose of this survey was to **serve as a baseline for an on-going impact evaluation of the overall APRIGP project.** We provide an assessment of the quality of the data collected and balance between treatment and control with the goal of allowing the organization which leads the end line survey and analysis to better determine what data to collect and what form the final analysis should take.

In addition, we **present descriptive statistics from the baseline data and highlight the policy implications of these data** for those involved in APRIGP and other agriculture or human development-focused programs. Key policy-relevant findings from the agriculture and sanitation module include:

- **There is tentative evidence suggesting that farmers' income could be increased by...**
  - **...diversifying towards high value crops:** Staple crops take up 69% of the cultivated area but contribute only 44% of the total revenue from crop sales while high value crops take up only 27% of the cultivation area but contribute 52% of the total revenue.
  - **...increasing crop intensity:** 70% of farm households cultivated their agricultural land in only one season, most of them in kharif.
  - **...increasing crop productivity:** Despite recent improvements, yield rates of all twelve most commonly grown crops are still far below global top yield rates.
  - **...securing access to productivity-enhancing farm assets:** Many farm households could improve their productivity by gaining access to hand or power sprayers, seed drills, threshers, and reapers.
- **Spot cash, trust, best prices, and proximity to home are the most important reasons for selling to middlemen, the major sales channel used by farm households.**
  - **Buyer types:** For all major crops, middlemen are by far the most relied on sales channel of households. Sales to middlemen make up 59% of all sale transactions. Commission agents and private markets are also frequently used.
  - **Reasons for choosing buyer types:** For the three most important buyer types, spot cash, trust, and best price are the most commonly cited reasons for selling to these buyers. Provision of loans by buyers plays only a minor role in choosing the sales channel.

- **Interventions must address both latrine access *and* use to achieve open defecation free communities.**
  - **Latrine access and maintenance:** Almost half of all households do not have access to a latrine. 10% of sample latrines were used for other purposes (e.g., storage).
  - **Latrine use:** In 31% of households with access to a latrine at least one member defecated in the open the last time they defecated. Women and educated individuals are more likely to use the latrines.
  - **Open defecation:** 58% of all individuals in the sample defecated in the open the last time they defecated.

The report further includes **findings on household characteristics, participation in government schemes, income and consumption, dietary diversity, animal husbandry, other WASH sectors, anthropometric indicators and Kirana stores.** It also provides background information on the APRIGP as well as the on-going impact evaluation. For detailed information on the evaluation design, sampling strategy, and data collection activities, please refer to the IDinsight Evaluation Design Document.

## 1 Introduction

In 2015, the state of Andhra Pradesh, with support from the World Bank, launched the Andhra Pradesh Rural Inclusive Growth Project (APRIGP). The project aims to combine innovative livelihoods enhancement and support activities with programs aimed at strengthening human development and access to social services. Taken together, these programs seek to reduce poverty across project areas.

This report presents findings from a survey of 8943 households conducted in APRIGP program areas and a set of similar control areas between July 2017 and October 2017. The primary purpose of this survey was to serve as a baseline for an on-going impact evaluation of the overall APRIGP project. We first provide an assessment of the quality of the data collected and balance between treatment and control with the goal of allowing the organization which leads the end line survey and analysis to better determine what data to collect and what form the final analysis should take.<sup>2</sup> In addition, we present descriptive statistics from the baseline data and highlight the policy implications of these data for those involved in APRIGP and other agriculture or human development-focused programs.

## 2 Background

### Andhra Pradesh Rural Inclusive Growth Project (APRIGP)

Despite strong economic growth in recent years, poverty in Andhra Pradesh remains high, in particular in rural areas and among Scheduled Castes (SC) and Scheduled Tribes (ST).<sup>3</sup> Further, while Andhra Pradesh performs better on human development indicators than many other major Indian states, it still lags behind other southern states, especially on health, nutrition, and education outcomes. For example, in 2011, 65% of rural households in erstwhile united Andhra Pradesh defecated in the open compared to only 5.6% households in Kerala (Ministry of Drinking Water and Sanitation, 2013).

Against this background, in 2015, the state of Andhra Pradesh launched the APRIGP. APRIGP is implemented by SERP, an autonomous society of the Department of Rural Development and seeks to *increase agricultural incomes and secure access to human development services and social entitlements*.

APRIGP aims to achieve this objective through interventions in three areas:

1. **Livelihoods** – Program activities under the livelihood component aim to increase the income in selected value chains of 250,000 small and marginal farmers by at least 50% through productivity enhancement, cost reduction, and improved market access. The livelihoods component consists of two programs:
  - The **Farm Value Chains (FVC)** program includes activities and investments aimed at organizing small and marginal farmers into producer organizations and providing them with high-quality support services like technology, credit, extension, marketing, as well as sustainable access to markets.

---

<sup>2</sup> IDinsight's scope of work is limited to supervision of baseline data collection and design of the overall impact evaluation. All other analyses will be conducted by another, yet to be identified, research team.

<sup>3</sup> As per the Socio-economic Survey, in 2011-12, the poverty rates for rural and urban areas of Andhra Pradesh were 10.96% and 5.81% respectively. However, there has been criticism that the number of poor are vastly underestimated due to unrealistic poverty lines.

- The **Rural Retail Chains (RRC)** program aims to connect rural producers and enterprises with rural consumers and enhance the quality and affordability of the consumption basket of poor households.
- 2. **Human Development** – The objective of the human development component is to improve WASH and nutrition indicators by strengthening demand for these services and increasing the supply of key nutrition, health, pre-school education, and water and sanitation services.
- 3. **Access to Services** – The access to services component seeks to improve the delivery of social protection entitlements to 500,000 poor households, mainly belonging to Scheduled Castes (SC) and Scheduled Tribes (ST) and Persons with Disabilities (PWD).

The project will be implemented over five years and in 147 mandals (about one third of the state). Project areas were selected on the basis of their share of SC and ST population, female illiteracy, and lack of irrigation.

### APRIGP Impact Evaluation

The primary purpose of the household survey analyzed in this report is to serve as a baseline for a larger impact evaluation of APRIGP. The impact evaluation focuses primarily on the Farm Value Chains (FVC) program of the APRIGP. The evaluation will assess whether the development objective of the program, namely an 50% increase in household income in the targeted value chain crop, has been achieved.

To identify the causal impact of the FVC program, the evaluation uses a matching design to pair villages in treatment mandals, which have been selected before the start of the program, to control villages outside of APRIGP mandals. For detailed information on the evaluation design and matching strategy please refer to the enclosed impact evaluation design document.

### Sampling

Treatment villages were sampled from a list of mandals in which SERP was planning to implement the FVC program in late-2017 and matched with control villages in the same district as similar as possible in observable characteristics. A census was then conducted in each selected treatment and control village to identify households which met the inclusion criteria and were deemed likely to participate in program. The final sample of 8943 households were selected from these eligible and interested households.<sup>4</sup> For detailed information on the sampling strategy and power calculations used to arrive at these sample sizes, please refer to the enclosed impact evaluation design document. Please note that, in the interest of simplicity, the sampling design was not taken into account in the analyses presented in this report (with the exception of the balance tests).

### Data Collection

Data collection for the baseline survey took place between July and October 2017 and was carried out by the independent survey firm AMS.

The baseline households survey used a multi-topic questionnaire, with a focus on crop cultivation and sales. The questionnaire also contained a household roster and sections on facilities, animal husbandry, income, consumption, access to social services, anthropometric

---

<sup>4</sup> Initially, 9900 households were sampled. Unfortunately, three treatment villages (135 households) had to be dropped from the sample. In 772 cases, households could not be located. In 50 cases, respondents were not present or did not give their consent when survey teams visited.



indicators, and water, sanitation, and hygiene (WASH). The average duration of the survey was around 3.5 hours.

For detailed information on the data collection exercise, please refer to the enclosed impact evaluation design document.

### 3 Assessment of Data Quality

The purpose of this section is to alert potential users of the data to concerns we have regarding the data quality and to provide guidance for the research team conducting the final impact evaluation. Unfortunately, field observations during spot checks, back check surveys, and plausibility checks indicate that the overall quality of the baseline data is low.

First, in-person observations by IDinsight staff revealed that, in many cases, the survey protocol was not followed closely by enumerators.<sup>5</sup> This is especially problematic for questions relating to household size and the occupation of household members. For example, IDinsight staff observed that instead of following the household definition provided during training, enumerators often asked respondents to list all living household members. Further, some enumerators selected the primary occupation on their own if the household member was doing more than one job/work or captured the same occupation for both questions on primary and secondary occupation. Given the way the survey was designed, wrong selection of household size and occupation of household members could result in large parts of the income section being skipped. For more information on the field observations collected by IDinsight staff, see Appendix IV.

Second, we found large discrepancies between original survey data and back check survey data.<sup>6</sup> For almost half of the key variables analyzed, the difference between back-check answers and the original survey answers is above 50% (measured as a proportion of the original answers). For a full list of these comparisons, see Appendix IV. While some difference between back-check and original survey answers is to be expected due to the different timings of the surveys or households having difficulties remembering the answers, differences this large point to low quality of survey data.

Third, some survey results appear implausible. Most concerning, a high proportion of household incomes from agriculture, animal husbandry, and self-employment are negative. Consequently, 28% of households have a negative aggregate income. Comparing this result with income data from other household surveys conducted in Andhra Pradesh, such as the India Human Development Survey 2011-12, this share seems unrealistically high. Unfortunately, extended data analysis and comparisons with back check data have not allowed us to pinpoint the reasons or isolate the variables that drive these results.<sup>7</sup> We

---

<sup>5</sup> IDinsight conducted field visits to observe enumerators in four districts (Prakasam, West Godavari, Krishna, and Nellore). All observed deviations from the survey protocol were immediately reported to SERP, AMS, and the World Bank team.

<sup>6</sup> Back-check surveys were conducted with 279 households in all 13 baseline districts.

<sup>7</sup> Some of the analyses we conducted to shed further light on the phenomenon of low household incomes were: 1) Increased the share of observations being trimmed to test if results are driven by outliers. 2) Analyzed key variables going into the aggregate individually for plausibility and outliers. 3) Motivated by field observations, checked if there have been double entries of expenses on yearly crops. 4) Analyzed the relationship between crop quantities harvested, sold, kept for domestic consumption, and paid as rent to check for irregularities. 5) Calculated income from crop cultivation using average market prices (retrieved from <http://agmarknet.gov.in/>) for six out of the seven most commonly grown crops as households may have a better sense of quantities sold

speculate that these low aggregate incomes are the result of overall low quality as indicated by back check data quality checks, and households generally over-reporting or -estimating expenses and under-reporting or -estimating revenues (a phenomenon frequently occurring in household surveys).

As a consequence, the following decisions have been made for the analysis presented in this report: First, we focus our analysis on parts of the survey where data quality appears sufficiently high to produce reliable, policy-relevant insights. Second, in cases where data quality seems low but still high enough to learn something about the lives of the people in Andhra Pradesh, we present results but mention our low confidence in the data in the text of the report. We urge readers to use caution when drawing any policy conclusion from these analyses. Third, some analyses originally intended were not performed altogether. For example, we did not assess the relative importance of different income sources of households or analyze self-employment activities in detail. Finally, all non-categorical variables have been trimmed at the 1<sup>st</sup> and 99<sup>th</sup> percentile.

Lastly, we urge the team tasked with conducting the final impact evaluation to take special care in collecting income data at end line and in using baseline income data. In particular, the evaluator should avoid estimating the impact of APRIGP on income using a fixed effect specification as originally planned as this is likely to generate highly misleading results. While measured income appears balanced between treatment and control at baseline (see section below), the low average baseline income values will likely result in overestimates of the impact of the program. Alternatively, the evaluator may consider using consumption rather than income as the key outcome variable.

## 4 Validity of Control Group

Table 1. Balance Checks.

	Control Villages	Treatment Villages	Difference (t-test)
Variable	Mean (SE)	Mean (SE)	p-value
Household size	3.887 (0.092)	3.906 (0.074)	0.869
Age	34.962 (0.631)	33.857 (0.867)	0.303
Whether HH head is literate	0.425 (0.023)	0.437 (0.017)	0.685
At least one member of a Velugu/SERP/DWRCA SHG	0.756 (0.019)	0.712 (0.021)	0.120
Aggregated monthly household income	2869.435 (532.874)	2647.347 (267.141)	0.709
Net income from crop cultivation	-1442.127 (539.665)	-1742.787 (387.163)	0.651
Net income from animal husbandry	-520.302 (120.753)	-375.169 (218.814)	0.561

than revenue received from crop sales. 6) Analyzed income (sub-)aggregates by district as well as consumption and interview time quartiles.

Net income from self-employment activities	-3458.145 (2349.490)	-252.089 (700.772)	0.192
Monthly Per capita consumption	4120.196 (819.783)	4611.389 (1366.395)	0.758
Dietary Diversity Score	5.900 (0.057)	5.441 (0.088)	0.000***
Land cultivated in kharif	1.269 (0.147)	1.162 (0.115)	0.566
Paddy yield rate (quintals/acre)	14.970 (0.820)	14.190 (0.946)	0.533
Whether HH sold all crops after harvest	0.578 (0.036)	0.530 (0.035)	0.333
Number of milch buffaloes owned	1.697 (0.055)	1.645 (0.052)	0.495
Whether household has access to private latrine	0.475 (0.044)	0.425 (0.030)	0.353
Whether at least one HH member has not used latrine	0.292 (0.035)	0.291 (0.024)	0.984

Note: Standard errors in parentheses are clustered at the village-level. Sampling weights have been applied. For each variable, only households were included for which the variable is relevant (e.g., net income from crop cultivation is only shown for farm households.) Four initially sampled villages could not be surveyed, out of which one was replaced (control group) and three have been dropped (treatment group). The impact evaluation design document provides more information on the construction of sampling weights and dropped villages.

The evaluation uses a matching design pairing treatment villages in APRIGP mandals to control villages outside of APRIGP mandals to identify the causal impact of the intervention. Data from the baseline survey shows that control and treatment sites are indeed similar with respect to a large number of observable characteristics, suggesting that the matching strategy was successful in ensuring balance on observables. Table 1 presents results from individual t-tests, accounting for weights and sampling strategy, between treatment and control for several key variables. Aside from dietary diversity, which shows large differences between treatment and control, all other variables appear balanced between treatment and control. Performing a logistic regression of the treatment variable on the above listed variables, the p-value for joint significance of the independent variables is 0.003. However, when dietary diversity is excluded from the regression analysis, the p-value for joint significance is 0.14.

## 5 Household Profile

### Household Composition and Characteristics

Table 2. Household Composition and Characteristics.

	N	Mean	SD
<b>Household composition</b>			
Household size	8943	4.0	1.9
Household with child (≤5 years)	8943	20.8%	
<b>Household member characteristics</b>			
Age	36132	31.5	19.7
Females	36114	49.4%	
Females among children (≤5 years)	2702	49.4%	

Years of schooling	36132	4.6	5.1
Literacy rate ( $\geq 7$ years)	32671	61.2%	
<b>Household head characteristics</b>			
Age	8825	48.2	13.7
Female household heads	8732	18.1%	
Years of schooling	8735	3.4	4.5
Literate household heads	8825	45.5%	
<b>Household religion</b>			
Hindu	8943	86.9%	
Islam	8943	8.0%	
Christianity	8943	4.7%	
Other	8943	0.3%	
<b>Household caste</b>			
SC	8943	28.9%	
ST	8943	8.2%	
BC	8943	43.0%	
OC	8943	19.5%	
Other	8943	0.2%	

The average household in the sample has four members, and 21% of households have at least one child under six. Most household heads are male (82%) and the majority of households are Hindu (87%).

### Missing Women

Reflecting state-wide trends,<sup>8</sup> there are fewer women than men in the sample, which may be the result of gender-based discrimination. Only 49.4% of individuals in the sample are female (sex ratio of 976), which is statistically significantly different from 50% at the 1%-level.

### Literacy Rates and Educational Attainment

Only 61% of individuals over the age of six in the sample are literate, which is below the state (67.4%) and national average (73.0%) (Government of Andhra Pradesh, 2014). The literacy rate is lower among women than among men (52% vs. 70%) and less than half of household heads are literate. Similarly, educational attainment is low. Only 39% of individuals over the age of 17 have completed primary school (5<sup>th</sup> grade) and 16% have completed secondary school (12<sup>th</sup> grade). Almost half of the individuals have never attended school.

### Characteristics of the Dwelling

Most households either use firewood (36%)<sup>9</sup> or LPG/PNG (62%) as a primary source of energy for cooking. Since firewood is both an inefficient source of energy and often the cause of harmful indoor pollution (Kankari et al., 2014), interventions improving access to other energy sources may have significant economic and health benefits for households and should be further explored.

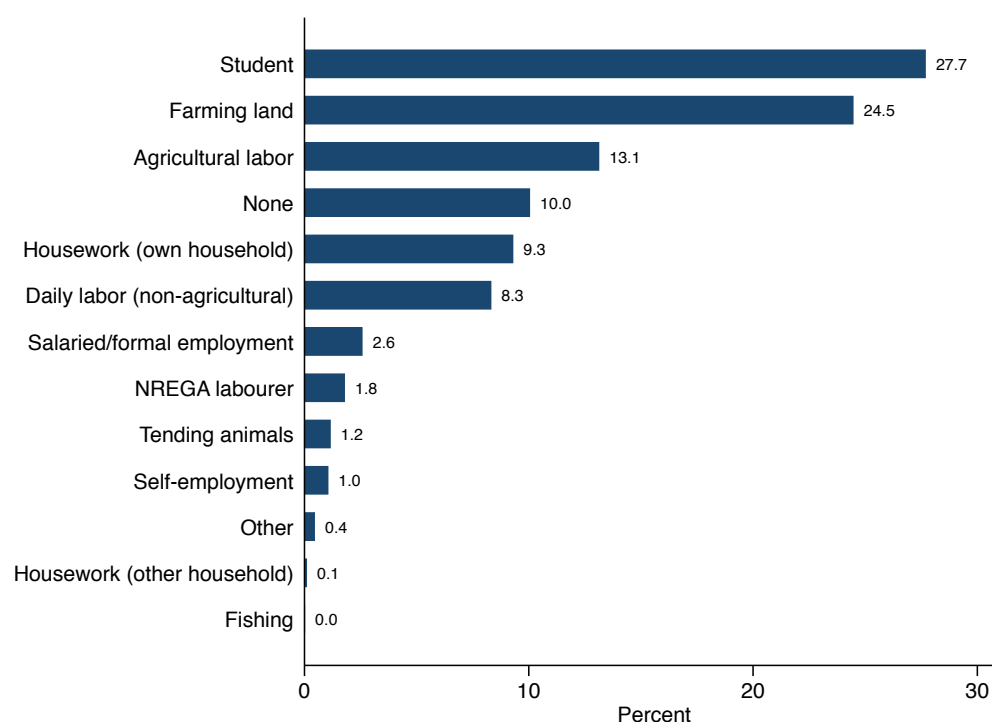
<sup>8</sup> According to a report by the Government of Andhra Pradesh (2014) using data from the 2011 census, the overall sex ratio was 943 and the sex ratio among children was 919 in Andhra Pradesh. The sex ratio is calculated as the number of females per 1000 males.

<sup>9</sup> While this figure is high, it compares favorably with data from the India Human Development Survey 2011-2012: Back then, only 28% of households reported *not* using firewood for cooking, lighting or heating.

Almost four out of five dwellings have a floor made from stone or cement (78%) and more than half a roof made out of concrete (56%). The most commonly used material for the wall is brick (35%), followed by stone packed with mortar (25%). Nearly all households own their house (93%), use electricity for lighting (99%), and have no kitchen garden (99%).

## Primary Occupation

Figure 1. Primary Occupation of Household Members.



As described in the data quality section of this report, we do not have high confidence in the data on occupation of household members. However, according to the data collected, more than one third of individuals in the sample are primarily employed in agriculture: 25% engage in farming, 13% work as agricultural laborers, and 1% tend animals. Every tenth household member was unemployed. Only 1.1% of individuals work primarily in a business operated by a household member and only three individuals reported fishing as their primary occupation.

## Household Assets

Table 3. Ownership Rates of Common Household Assets.

	N	Ownership rate (in %)
Cot	8942	92.3
Electric fan	8941	91.9
Cell phone	8942	88.2
Chair/table	8942	86.1
Television	8943	74.9
Cable/Dish TV	8943	66.3
Mixer/grinder	8942	41.2
Scooter	8943	29.8
Bicycle	8943	27.6
Pressure cooker	8943	20.5

Refrigerator	8942	10.7
Sewing machine	8942	8.0
Air cooler	8943	3.8
Credit card	8943	2.9
Washing machine	8943	1.1
Laptop	8943	0.9
Telephone	8943	0.6
Generator	8943	0.5
Car	8943	0.5
AC	8942	0.5
Chaff cutter	8943	0.4
Computer	8942	0.3
Microwave	8943	0.1

Note: Credit cards do not include Kisan credit cards.

Nearly nine out of ten households own a cell phone. This is remarkable from a policy perspective: cell phones can serve as a platform for various interventions and services. For example, they can be used for rapid data collection on issues of interest to policymakers via phone surveys or short messaging services (SMS). Further, they provide a direct channel of communication to the population, for example allowing agriculture extension agencies to send farmers weather and price information.

## Participation in Government Schemes

### NREGA

Table 4. NREGA.

	N	Mean	SD
Days worked under NREGA in last 12 months	7957	40.5	21.5
Days paid for work under NREGA in last 12 months	8068	32.6	22.9
Cash amount received per day of work under NREGA	7115	107.9	77.9

Note: All figures presented in the table are on the individual level.

71% of households have a NREGA job card, and 22% of individuals reported participating in NREGA in the last 12 months. On average, individuals reported working 8 days more under NREGA than they received payment for.

### Other Government Schemes

Table 5. Household Participation in Government Schemes.

	N	Participation Rate (in %)
Ration card	8942	97.5
BPL	8942	93.0
Antyodaya	8942	3.8
APL	8942	0.9
Annapurna	8942	0.1
At least one member of a Velugu/SERP/DWCRA SHG	8940	73.7 <sup>10</sup>
Direct Transfers such as for LPG cylinder	8943	47.8
Old Age Pension	8943	18.8

<sup>10</sup> Excluding households with no adult women younger than 60, the share raises to 76.4%.

Widow Pension	8943	10.8
Disability Pension	8943	3.5
Swachch Bharat Mission (SBM)	8943	2.5
Chandranna Bima	8943	1.5
Abhayastham	8943	1.4
Pregnant Womens' Assistance	8943	1.2
Pradhan Mantri Gramin Awaas Yojana/other housing scheme	8943	0.4

Nearly all households have a ration card (mostly BPL ration card). In three out of four households at least one household member is currently a member of a Velugu/SERP/DWCRA SHG. Half of all households received some form of direct transfer. Participation in all other government schemes asked about in the survey is low.<sup>11</sup>

## 6 Income and Consumption

### Income

Table 6. Monthly Aggregated Household Income (in INR).

	N	Mean	Median	SD	Min	Max
Net income from crop cultivation	66%	-2074	-950	7307	-107281	107458
Net income from animal husbandry	34%	-561	-417	2053	-9750	6475
Wage and salary income	74%	3636	2792	3502	2	65625
NREGA income	47%	533	417	402	2	3358
Net income from self-employment activities	4%	-49	204	9608	-60833	36500
Net income from transfers	19%	2551	417	5854	-16667	41667
Other income (rent, interest, etc.)	4%	5790	929	15172	~0	125000
<b>Aggregated monthly household income</b>	<b>8943</b>	<b>2114</b>	<b>1800</b>	<b>8689</b>	<b>-110610</b>	<b>128208</b>

Note: Percentages in the first column and mean values are calculated for the number of households reporting income for the sub-aggregate. The total aggregated monthly household income is calculated for all households.

Due to data issues mentioned in the data quality assessment of the report, we have low confidence in the aggregated income data and did not perform any further analysis on them. For detailed information on the construction of the income aggregate, see Appendix II.

As described in the data quality section, a high proportion of households have negative income from animal husbandry, agriculture, and self-employment and overall mean income in the sample from these activities is negative. We speculate that negative mean income from animal husbandry may be explained by recent increases in investments in livestock. On average households active in animal husbandry spent 495 INR per month on the purchase of livestock while generating only 152 INR per month from the sale of livestock. Excluding the purchase and sale of livestock and focusing on operating revenue and expenses alone results in a monthly household income from animal husbandry of -87 INR (compared to -561 INR).

<sup>11</sup> Statistics on the access to bank accounts are not reported due to data quality considerations.

## Consumption

### Aggregate Household Consumption

Table 7. Monthly Aggregate per Capita Consumption (in INR).

	N	Mean	SD	Min	Max
Total	8909	3682	25561	109	1505060

Real aggregate household consumption was deflated by dividing the aggregate consumption by the household size. For more information on the construction of the consumption aggregate, see Appendix III.

### Dietary Diversity

Table 8. Proportion of Households by Dietary Diversity Score.

Dietary Diversity Score Category	Proportion of Hs	Food groups consumed by ≥50% of households with given diversity score category
Lowest Dietary Diversity (≤4 food groups)	22%	Cereals Pulses
Medium Dietary Diversity (5 –7 food groups)	64%	Cereals Pulses Dairy products Food made with oil Sugar or honey Condiments, tea, coffee, pickles
High Dietary Diversity (≥8 food groups)	14%	Cereals Pulses Dairy products Food made with oil Sugar or honey Condiments, tea, coffee, pickles Food made from tubers Vegetables Fruits

Dietary diversity is measured using the FANTA project Household Dietary Diversity Scale (HDDS), i.e., as the count of the number of different food groups (out of 12) consumed by the household during the last 24 hours. On average, households had consumed products from 5.7 food groups. The most commonly consumed food groups are cereals (97% of households), other foods such as condiments, pickles, coffee, or tea (80%), foods made with oil, ghee, fat, or butter (75%) and dairy products (72%). Table 8 shows the proportion of households by dietary diversity score category and typical foods eaten by each type of household. The table clearly illustrates a correlation between dietary diversity and micronutrient access. Households with low levels of dietary diversity rely on cereals and pulses, while households with medium and high dietary diversity score consume nutrient-rich foods such as dairy products and vegetables.



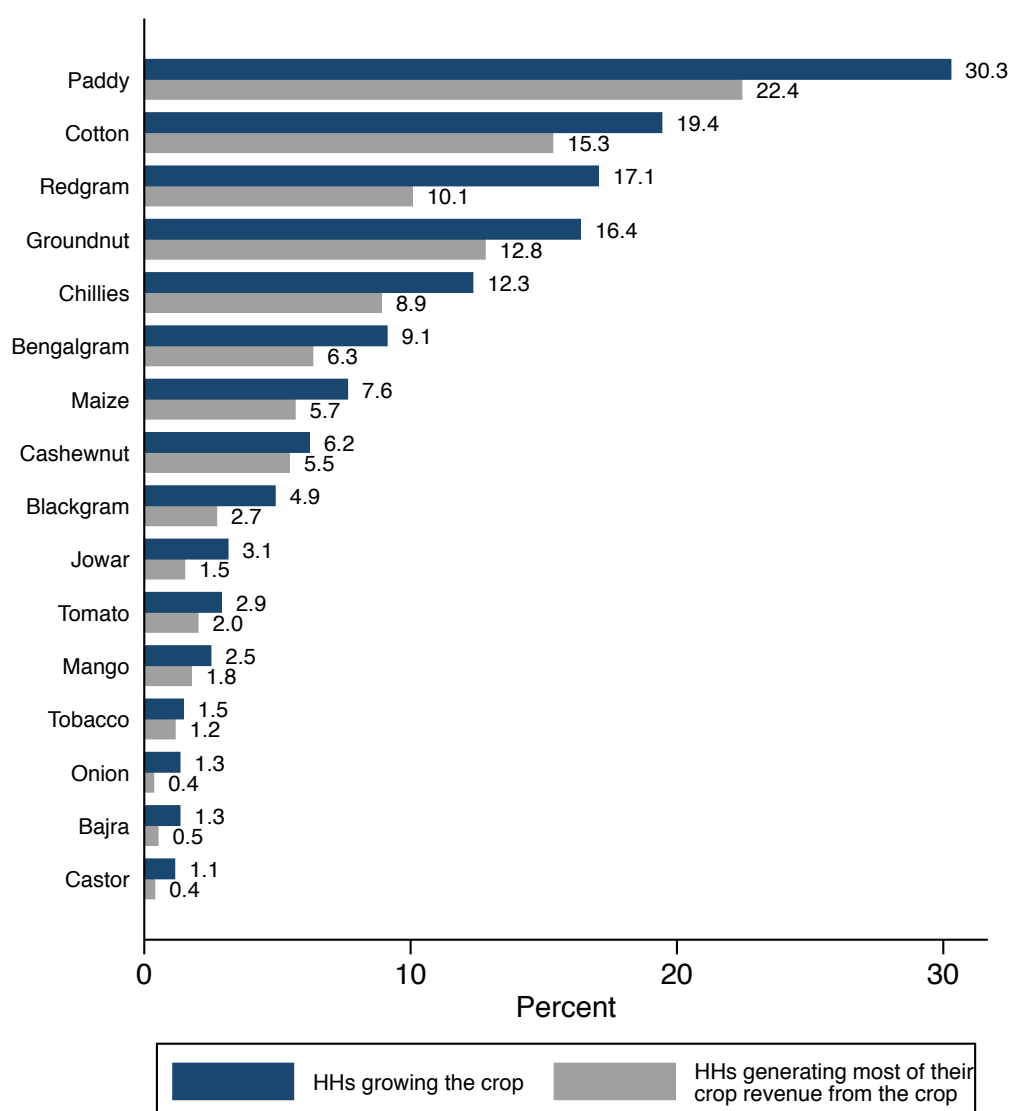
## 7 Crop Cultivation

### Crop Production

#### Crops Cultivated

Two thirds of households cultivated at least one crop in the last 12 months. These households are referred to as farm households in the remainder of the report. As figure 2 shows, the most commonly grown crops are paddy, red gram, ground nut, cotton, and chillies. 63% of farm households grew only one crop in the last 12 months. Hence, most crops also tend to be the main source of crop revenue when grown by a household. Comparing back check data and original survey data, we have high confidence in the data on crop production.

Figure 2. Crop Cultivation.



Note: Crops grown by less than 1% of households are not shown in the graph.

## Diversification Towards High Value Crops

Table 9. Crop Diversification.

	Revenue (in INR)	Area (in ha)	Productivity (in INR/ha)	Revenue (in %)	Area (in %)
Cereals	70,321,588	2,314	30,385	22.5	24.1
Pulses	38,264,612	2,602	14,706	12.3	27.1
Oilseeds	29,233,177	1,713	17,064	9.4	17.9
<b>Staple Crops Total</b>	<b>137,819,377</b>	<b>6,630</b>	<b>20,789</b>	<b>44.1</b>	<b>69.1</b>
Fruits	10,413,134	246	42,266	3.3	2.6
Vegetables	61,000,685	789	77,328	19.5	8.2
Other high value crops	92,002,579	1,508	61,012	29.5	15.7
<b>High Value Crops Total</b>	<b>163,416,398</b>	<b>2,543</b>	<b>64,257</b>	<b>52.3</b>	<b>26.5</b>
Other crops	11,048,418	416	26,567	3.5	4.3
<b>Total</b>	<b>312,284,193</b>	<b>9,589</b>	<b>32,568</b>	<b>100</b>	<b>100</b>

Note: Values are aggregated across all households. The 56 crops listed in the survey form were classified into staple and high value crops. Crops not listed in the survey form and entered by surveyors themselves are reported as “other” crops.

As has been shown for India as a whole (Chand, 2017), table 9 suggests that in Andhra Pradesh diversification towards high value crops offers great potential to increase farmers’ income. Staple crops (cereals, pulses, oilseeds) take up 69% of the cultivated area but contribute only 44% of the total revenue from crop sales while high value crops (fruits, vegetables, cotton, sugarcane, condiments, spices) take up only 27% of the cultivation area but contribute 52% of the total revenue. The productivity of high value crops was 64,000 INR per hectare as compared to 21,000 INR per hectare for staple crops. While this analysis does not take into account expenses and other considerations such as soil quality and irrigation, the size of the difference in productivity points to the high potential for income gains from shifting from staple crops to high value crops.

## Agricultural Land Use

### Agricultural Landholdings

Table 10. Agricultural Land Available for Crop Cultivation.

	Number of farm households	Share of farm households (in %)
Marginal (<1 ha)	2595	46.4
Small (1 - 2 ha)	1537	27.5
Semi-medium (2 - 4 ha)	1064	19.0
Medium (4 - 10 ha)	245	4.4
Large (>10 ha)	148	2.6
<b>Total</b>	<b>5589</b>	<b>100</b>

Note: Agricultural land available is calculated by taking the sum of agricultural land owned and rented in and subtracting agricultural land rented out. The table reports the mean across the three seasons (kharif, rabi, summer).

Almost half of all farm households fall in the “marginal” size category (<1.00 ha) and more than a quarter in the “small” size category (1.00 - 2.00 ha). On average, each farm household had 1.37 hectares of agricultural land available for crop cultivation in each season (kharif, rabi, and summer).

61% of households owned agricultural land in at least in one season and had not rented in any agricultural land. 25% of households owned land in at least one season, but also rented in land in at least in one season. 14% of households rented in and did not own any agricultural land.

## Crop Intensity

Table 11. Crop Intensity.

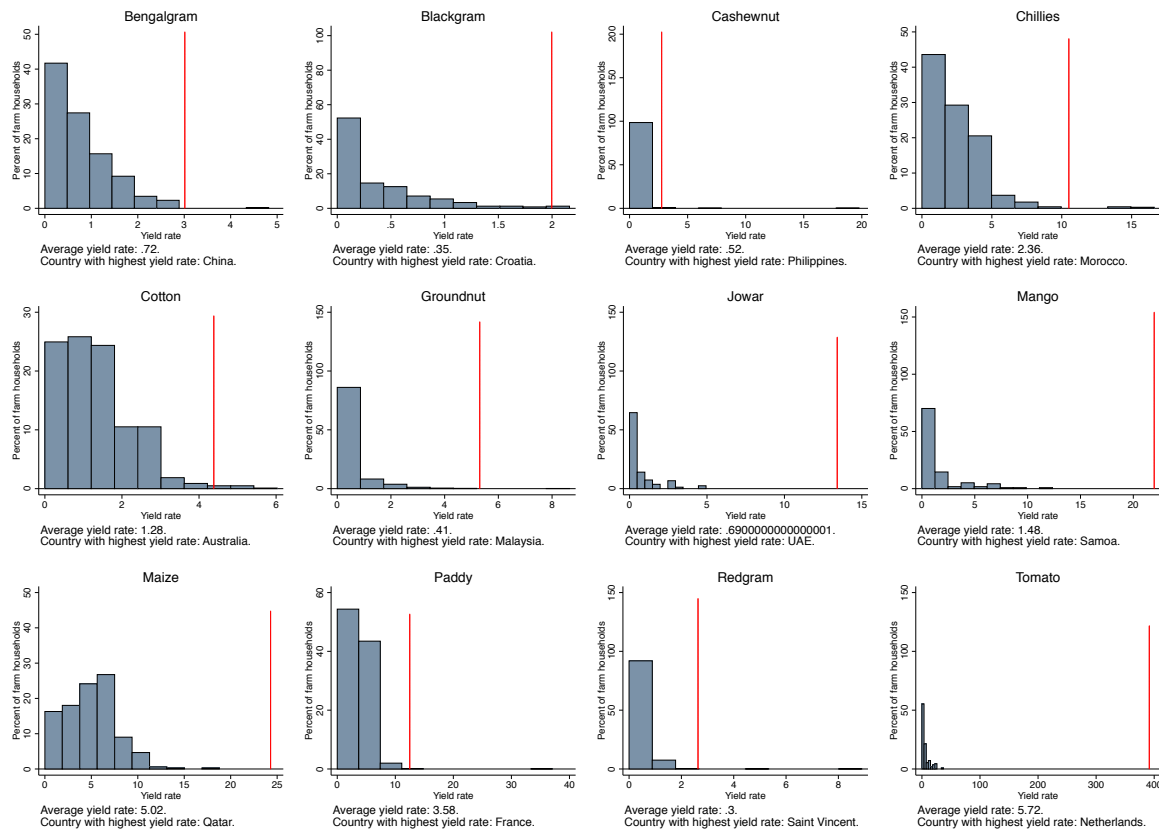
	Proportion of farm households who cultivated any of the available agricultural land	Proportion of land cultivated out of total agricultural land available to households
Kharif 2016	81.9%	0.75
Rabi 2016-17	34.7%	0.32
Summer 2017	16.4%	0.18
Total	43.5%	0.43

Increasing crop intensity may be another way to increase farm households' income. 70% of farm households cultivated their agricultural land in only one season, most of them in kharif. In rabi, 68% of agricultural land remained unused and in summer, 82% of agricultural land remained unused. Lack of access to water is often cited as reason for nation-wide low crop intensity (Chand 2017). Unfortunately, low quality of data on irrigated land<sup>12</sup> prevents us from further investigating this hypothesis.

<sup>12</sup> IDinsight staff observed that enumerators were confused about the distinction between irrigated land and non-irrigated land. Enumerators often treated rain-fed land as irrigated land.

## Crop Productivity

Figure 3. Crop Productivity (in Tons per Hectare).

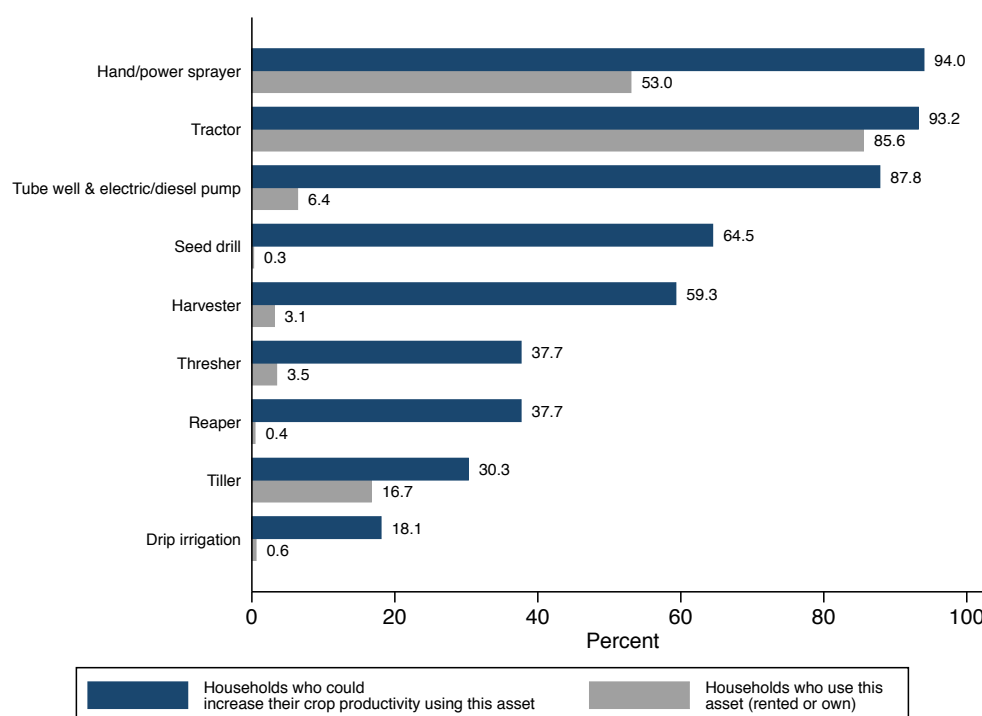


Note: The figure shows the histograms of households by crop and yield rate. For graphing, variables on cultivation area and crop production have been trimmed at the 5% and 95% level. The captions note the average yield rate in the sample and country with the highest yield rate globally. The red line indicates the average yield rate of the ten countries with the highest yield rates globally. Source: FAO, 2014.

Figure 3 shows histograms of farm households' crop yield rates (tons per acre) for each of the twelve most commonly cultivated crops. While there exists significant variation in yield rates across households, for all of the analyzed crops, yield rates for nearly all farm households in the sample are far below global top yield rates. Again, while these results do not take into account potential differences in irrigation, soil quality, or other factors between Andhra Pradesh and other countries, the size of the difference suggests there is still considerable scope to raise crop productivity and, consequently, farmers' income.

## Farm Assets

Figure 4. Farm Assets: Recommended vs. Actual Use.



Note: The share of households requiring an asset was calculated by analyzing the crop profile of every household and interviewing agricultural experts to make assumptions about which assets would increase the productivity of a given crop. Only the twelve most commonly grown crops were included in the calculation.

Data on farm asset ownership and rental suggests that crop productivity could be raised by ensuring that farmers have access to productivity enhancing assets. We first conducted interviews with agriculture experts to determine which assets would enhance productivity for farm households in our sample and then compared data on actual use of farm assets to this imputed data on recommended use.<sup>13</sup>

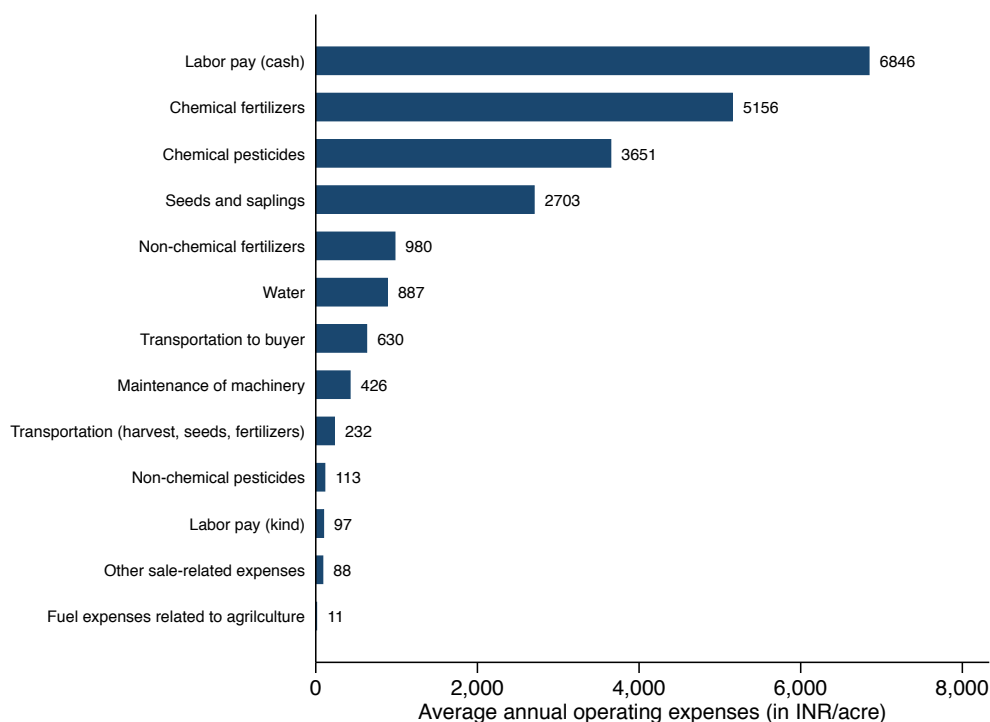
While nine out of ten farm households could increase their productivity by using a hand or power sprayer, only half of them used one during the last three seasons. Further, almost two thirds of households could increase their output by using a seed drill or harvester and one third of households by using a thresher and reaper. However, almost none of these households reported to have used these assets.

<sup>13</sup> The following assumptions about productivity enhancing farm assets were made for the calculation: (1) The productivity of bengal gram and black gram could be increased by access to a tractor, harvester, and sprayer. (2) The productivity of chillies and mango could be increased by access to a tractor, tube wells, and sprayer. (3) The productivity of cotton could be increase by access to a tractor, tube wells, sprayer, and seed drill. (4) The productivity of groundnut could be improved by access to a tractor, tube wells, sprayer, and drip irrigation. (5) The productivity of jowar could be increased by access to a tractor, reaper, thresher, tube wells, sprayer. (6) The productivity of maize could be increased by access to a tractor, reaper, thresher, harvester, tube wells, sprayer, and seed drill. (7) The productivity of paddy could be increased by access to a tractor, tiller, reaper, thresher, harvester, and seed drill. (8) The productivity of red gram could be increased by access to a tractor, harvester, tube wells, sprayer, and seed drill. (9) The productivity of tomato could be increased by access to tube wells, sprayer, and drip irrigation. (10) The productivity of cashew nut could be increased by access to tube wells and sprayer.

Almost all households rented the assets they used. Exception are hand/power sprayers, which were owned (individually or collectively) by roughly 15% of households, and sprinkler sets and drip irrigation, which were owned by almost half of households who used them.

## Expenses

Figure 5. Agricultural Operating Expenses.



Note: Graph shows the average annual operating expenses by farm households. The averages were calculated by first summing up the seasonal costs of each household and then calculating the average across households.

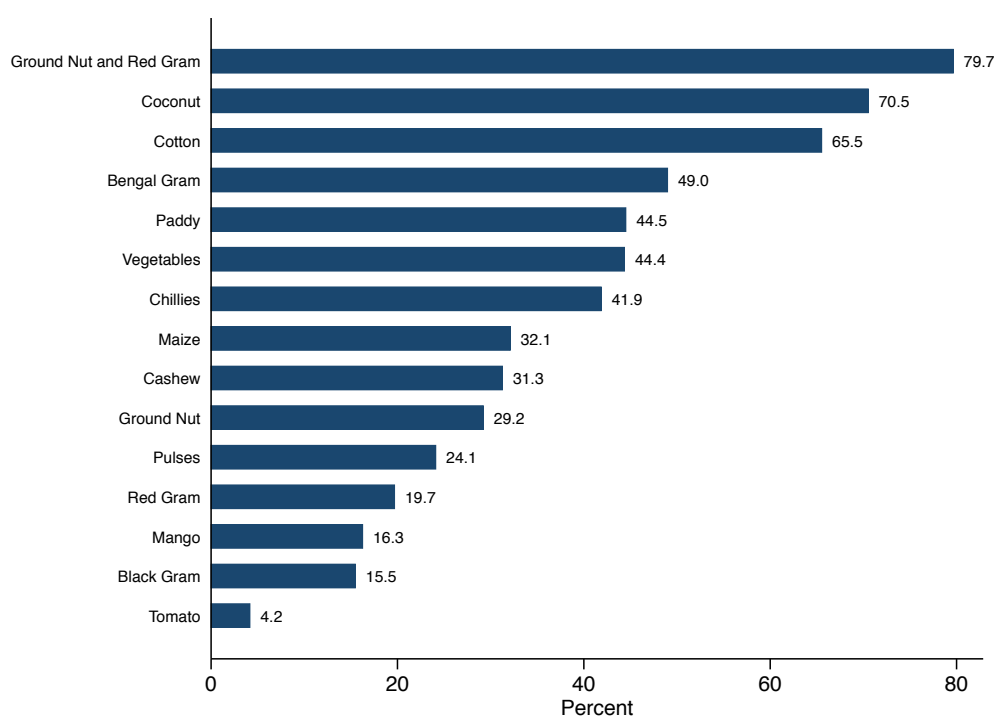
Analysis of agricultural expenses may shed light on potential ways in which farm income could be increased by lowering expenses. Labor compromises 32% of all annual operating expenses by farm households in the sample, chemical and non-chemical fertilizers make up 28% of annual expenses and chemical pesticides contribute 17% to annual expenses. Labor pay could potentially be reduced by facilitating mechanization of agricultural processes though such policies could have unintended effects on the poorest households. (Data from the second round of the India Human Development Survey reveal that a large portion of the poorest households in Andhra Pradesh rely on wage labor as their primary source of income (Desai et al., 2011-12).) In one third of all crop sales, farm households paid for transportation to sell the crop to the buyer. They paid, on average, 7% of the crop revenue.

## Crop Sales

### Share of Crop Revenue from FVC Crops

As Figure 6 shows, the share of revenue generated from the sale of the value chain crop varies substantially by crop. In villages assigned coconut, cotton, and both ground nut and red gram as the value chain crop, farm households derive a large portion of crop revenues from sale of the value chain crop. In villages assigned tomato, black gram, mango, and red gram as the value chain crop, revenues from the value chain crop make up a much smaller portion of overall crop revenue.

Figure 6. Share of Crop Revenue from FVC Crops.



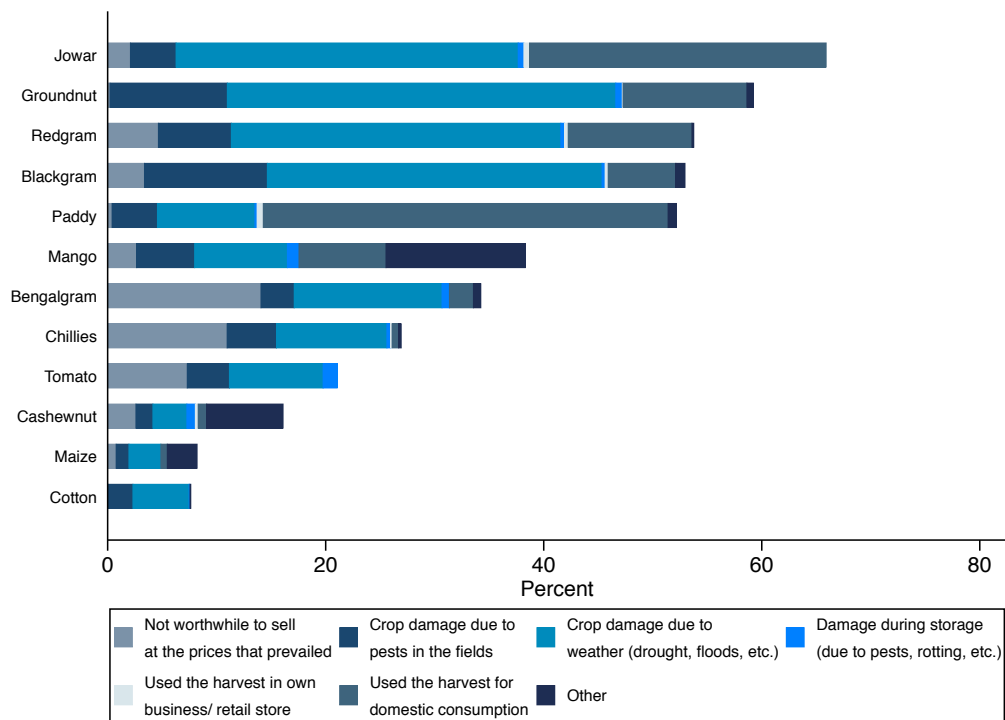
Note: The graph shows revenue from the assigned FVC crop as a share of total sales revenue from crop cultivation. Shares were calculated by first calculating the share of revenue from the assigned FVC crop for each household and then averaging across households. The most commonly assigned FVC crops are paddy (18 treatment villages), cashew (13), red gram (11), bengal gram (9) and red gram (9). 8 treatment villages were assigned both ground nut and red gram as FVC. Only 1 treatment village was assigned coconut as FVC crop.

### Crop Commercialization

For many crops, households often did not sell any of the produce from cultivating the crop. The proportion of households not selling any of their produce after harvest varies considerably by crop. More than half of the times farm households grew jowar, ground nut, black gram, red gram, and paddy, they did not sell any of their harvest. Similarly, the reasons for not selling any of the produce from cultivating a crop vary significantly by crop. For paddy the main reason for not selling any produce was domestic consumption. For jowar, ground nut, black gram, and red gram the main reason for not selling any produce was crop damage due to weather.

The high prevalence of crop damage points to the potential for crop insurance to increase farmer welfare. Yet less than 6% of farm households had insurance for any of their crops. However, among those who had crop insurance, 81% received some form of compensation for crop loss.

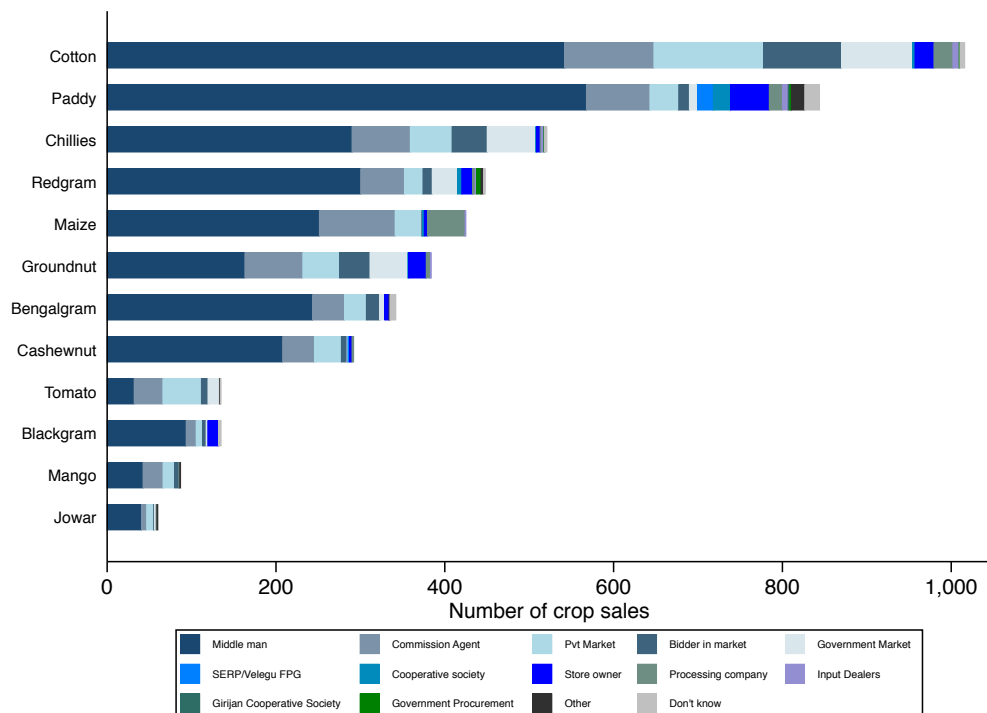
Figure 7. Reasons for Not Selling Crop.



Note: The graph shows for the twelve most commonly grown crops, the rate at which households cultivating this crop did not sell any of their produce after harvest. The graph also shows the share of the different reasons households cited for not selling the crop. Since household could cite more than one reason, stacked bars have been normalized.

## Buyer Types

Figure 8. Distribution of Buyer Types.



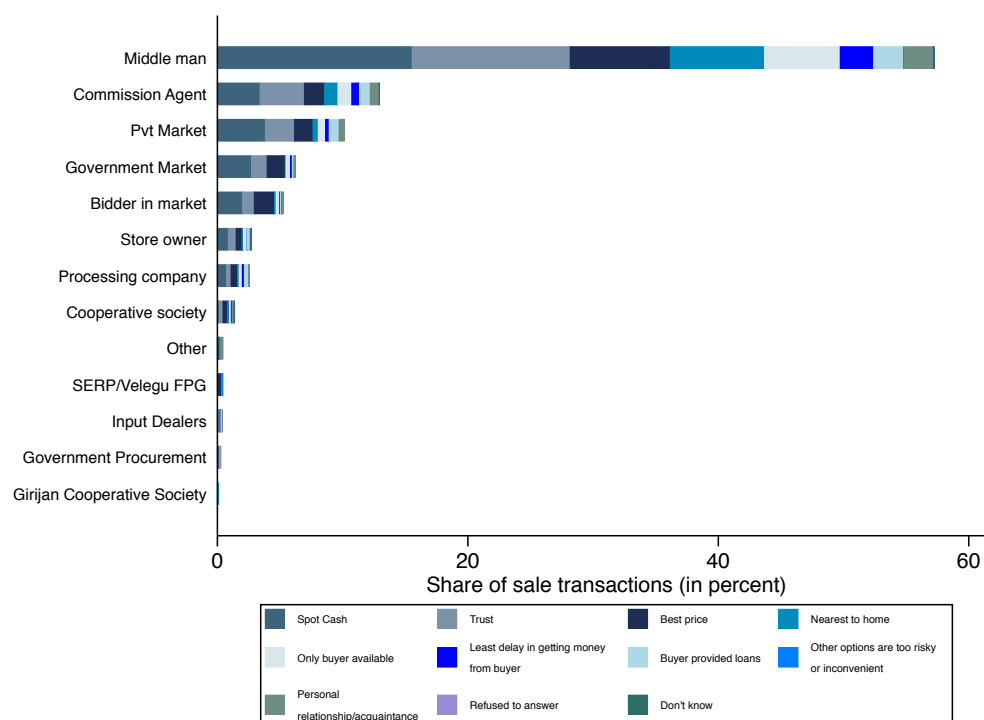
Note: Bar length indicates for the twelve most commonly grown crops, the number of times the crop has been sold to at least one buyer by a household after harvest. Additionally, the graph shows for each crop, the share in transactions different buyer types make up among all sale transactions for this crop.



For most crops (with the exception of tomatoes), middlemen are by far the most relied on sales channel for farm households. Sales to middlemen make up 59% of all sale transactions. Commission agents and private markets also play an important role, making up 13% and 10% of all transactions respectively.

## Reasons for Selling to Buyers

Figure 9. Reasons for Selling to Buyers.



Note: Bar length indicates the share of transactions the buyer type makes up among all crop sales. Additionally, the graph shows for each buyer type, which reasons the households cited for selling to this buyer type. Since households could mention more than reason, the share of the different seasons have been normalized. All crops were included in the analysis.

For the five most important buyer types, spot cash, trust, and best price are the most commonly mentioned reasons for selling to these buyers. When selling to middlemen, who make up more than half of all crop sales, proximity to home is also often cited as a reason for choosing this buyer type. Provision of loans by buyers plays only a minor role in choosing the sales channel. Only 10% of crop sales involved an advance from the buyer.

The importance of spot cash and trust in farmers' decision on sale channels suggest that interventions aiming at the adoption of new sales channels, such as the APRIGP, will have to gain farmers trust and ensure that payments for crop are made immediately to be successful. While 'best price' is a commonly cited reason for choosing a particular buyer type, the data provides suggestive evidence that awareness of prices among farm household is in fact rather low. Only 64% of households who said they chose their buyer for the best price have in fact received a price which was above the median price for the particular crop in the particular village in the particular season.<sup>14</sup>

<sup>14</sup> In this analysis, only crop-season-village combinations with more than four sales transactions are included. This tries to avoid that households, who sold their crop for a price just marginally below the top price, are not erroneously considered to have received a low price. Note, that this analysis remains only suggestive as it does

## 8 Animal Husbandry

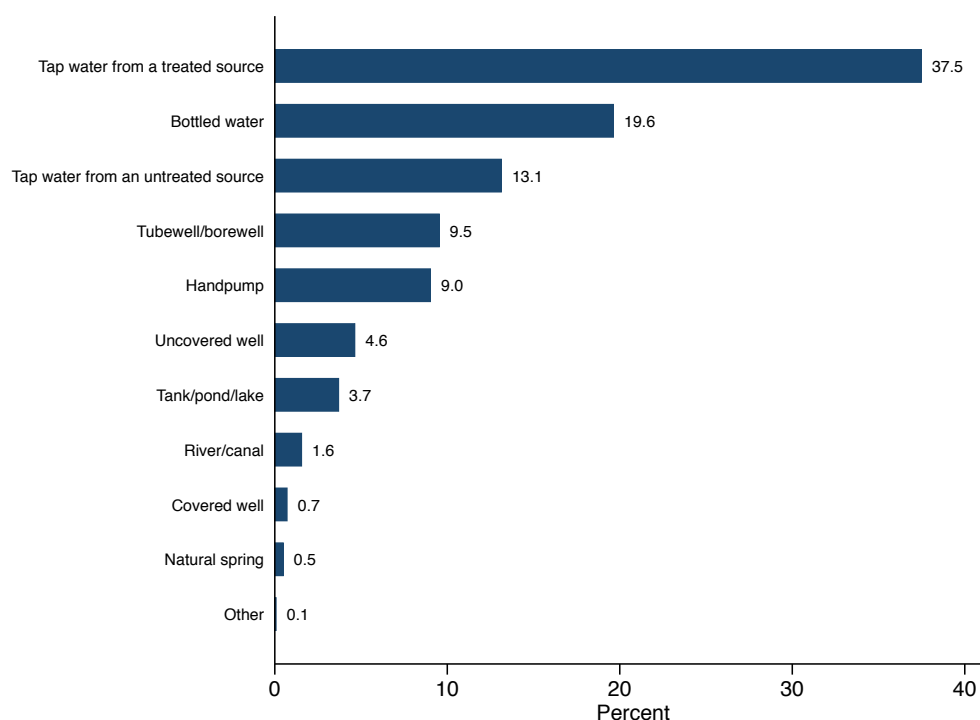
Table 12. Livestock Ownership

	% of households owning at least one animal	Average number of animals owned	SD
Milch buffaloes	17.3	1.8	1.0
Poultry, chicken, ducks	10.5	3.8	2.9
Milch cows	10.1	1.5	0.8
Draught animals (bullocks, buffaloes)	6.9	1.9	0.4
Goats	3.4	6.0	5.6
Sheep	2.7	10.9	11.2
Other animals	1.6	1.8	1.2

Overall, 39% of households own at least one animal. Milch buffaloes (17%), poultry, chicken, ducks (10%), and cows (10%) are the most commonly owned livestock.

## 9 Drinking Water

Figure 10. Primary Sources of Drinking Water.



More than one fifth of households use an unimproved water source (tap water from untreated source, uncovered well, tank/pond/lake, river/canal)<sup>15</sup> as their primary source for drinking water. However, we have low confidence in these data: IDinsight staff observed enumerators asking households generally about their water source rather than specifically about the source

not account for the quality of crops: Even if below the village median, household may have actually received the best price given the quality of their crops.

<sup>15</sup> Water sources as classified as unimproved water sources according to the WHO definition.

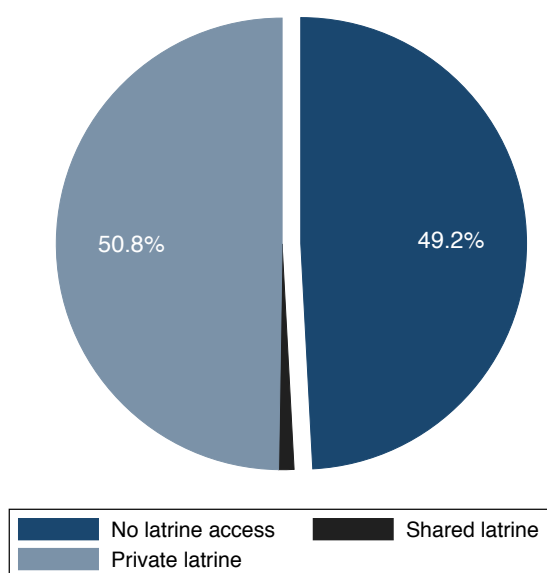
for drinking water. Hence, if households use unimproved water sources for other purposes than drinking, this figure may overestimate the share of unimproved sources in households' drinking water.

One out of four households have a water tap within the premises or get their water delivered to them. The average distance to the water source, if it is not located within the dwelling, is 130 meters. Roughly one third of households report boiling water before drinking it, one third use a cloth to filter the water, and one third of households do not treat their drinking water at all.

## 10 Sanitation Facilities and Behavior

### Latrine Access

Figure 11. Household Latrine Access.



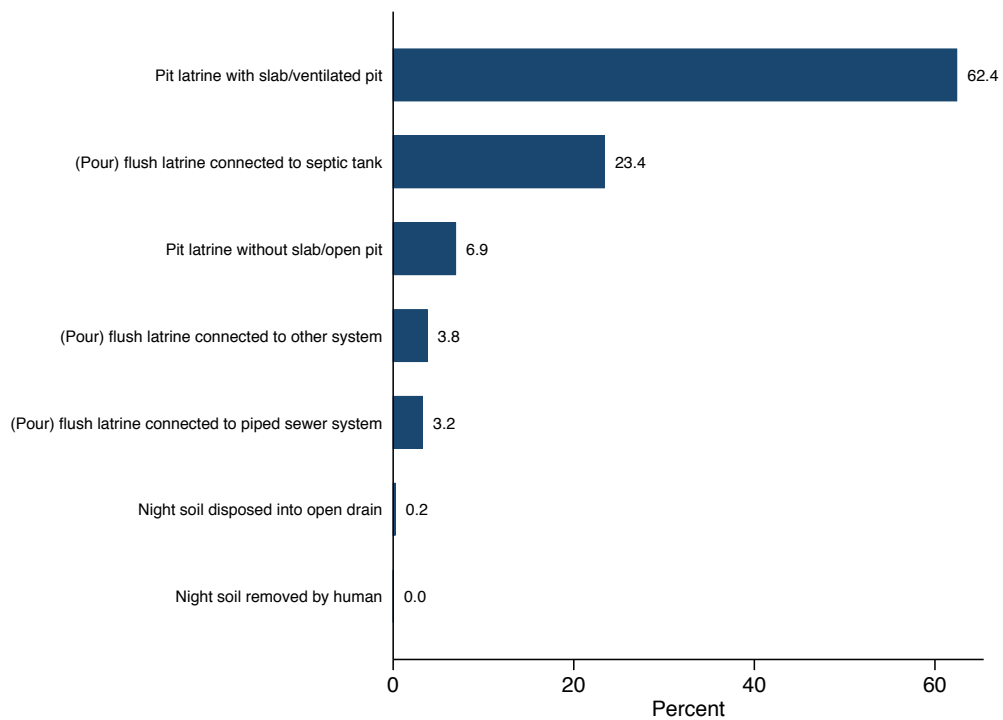
Almost half of all households do not have access to a latrine. Due to high consistency between original survey answers and back check survey answers we have high confidence in these data.

### Latrine Types

Among households owning a private latrine, roughly two thirds own a pit latrine with slab/ventilated improved pits and one fourth own a (pour) flush latrine connected to a septic tank. However, the share of latrines connected to a septic tank is likely to be overestimated as IDinsight staff repeatedly observed enumerators selecting the latrine type 'septic tank' in cases where no septic tank existed.

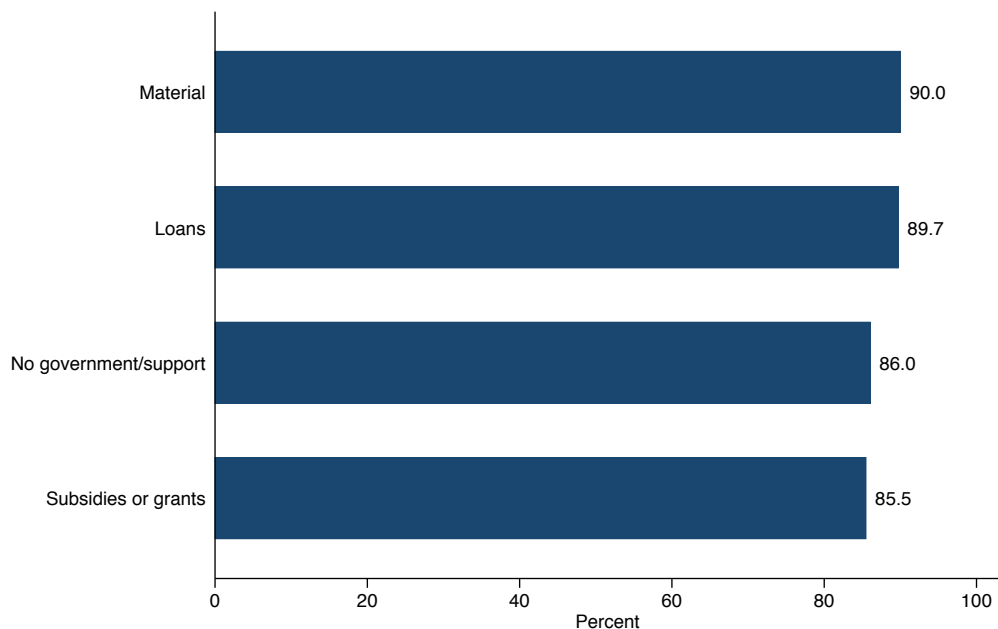
According to official government guidelines, all latrines should have two pits. Latrines which were built after the launch of APRIGP (November 2015) with the support of government/NGOs, are more likely to be two pit latrines (ventilated/with slab) or two pit pour flush latrines (31%) than those built before the start of APRIGP and/or not funded by the government/NGOs (22%). The difference is statistically significant at the 1% level.

Figure 12. Private Latrine Types.



## Maintenance

Figure 13. Whether Latrine is Well-Maintained by Type of Funding.



Note: Graph shows the share of latrines which are well-maintained by type of funding. Well maintained latrines are defined as not being used for other purposes nor having clogged squatting pans. 97% of households gave enumerators permission to inspect latrines. 10 households who reported more than one type of funding are excluded from the graph.

10% of latrines were used for other purposes (e.g., storage) and 5% of latrines had clogged squatting pans. There is no evidence that latrines for which households received subsidies,

grants, or construction material are less likely to be well maintained (not used for storage, not clogged) than latrines for which households received loans or no support at all.

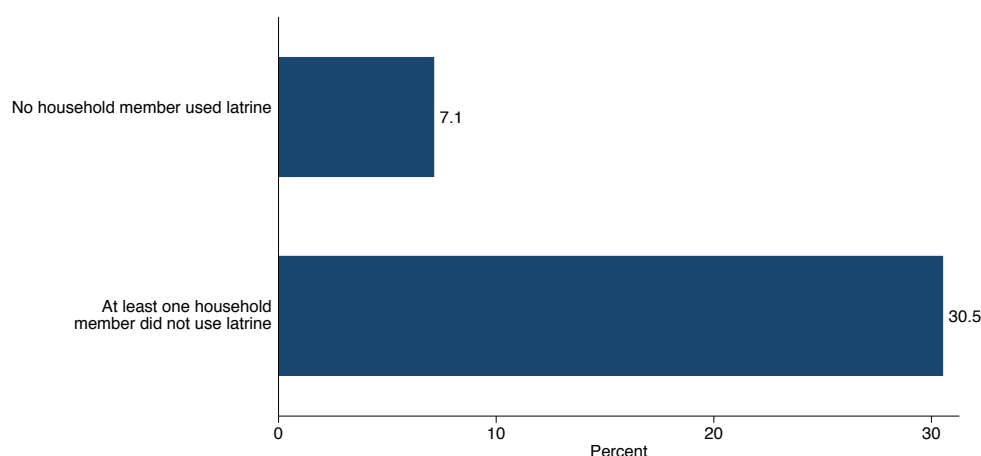
## Latrine Emptying

Among households with twin pit latrines, only 42 households reported that a pit has filled up in the past. Among these 42 households, 32 (76%) hired a tanker to empty the latrine pit, which indicates households are unaware of correct pit emptying procedures.<sup>16</sup> Overall, less than 2% of all households with latrines older than 10 years report that a pit of their latrine has filled up in the past. These findings are consistent with other recent research which suggests that Indian households seek to avoid pit emptying due to lack of information and cleanliness / religious concerns and thus build unnecessarily large and expensive latrines or limit latrine use (Coffey & Spears, 2017).

## Latrine Use

### Household Latrine Use

Figure 14. Household Latrine Use.



Note: Graph shows the share of households with access to a private latrine in which no member or at least one household member did not use the latrine the last time he or she defecated.

In 7% of households with access to a latrine all household members nevertheless defecated in the open the last time they defecated. In 31% of households with access to a latrine at least one member defecated in the open the last time they defecated. Overall, 18% of individuals in households with access to a latrine, and 58% of all individuals in the sample defecated in the open the last time they defecated. This suggests that there is still large scope to reduce the level of open defecation and that effective policies must address both access *and* use to achieve open defecation free communities.

<sup>16</sup> If one pit of a two pit latrine fills up, households can start using the second pit. After the first pit has stood for several years, the decomposed sludge can easily be removed, will not cause any health problems, and can even be used as a fertilizer.

## Determinants of Individual-Level Latrine Use.

Table 13. Determinants of Individual-Level Latrine Use.

	Whether household member used latrine	
	Coefficient	SE
<b>Household member-level characteristics</b>		
Age	0.000948 <sup>***</sup>	(0.000278)
Gender (female = 1)	0.174 <sup>***</sup>	(0.0121)
Years of schooling	0.00639 <sup>***</sup>	(0.00104)
Literacy	0.0413 <sup>***</sup>	(0.0117)
<b>Household religion (Hinduism is reference category)</b>		
Islam	0.0648 <sup>***</sup>	(0.0214)
Christianity	-0.0106	(0.0317)
Other	-0.171	(0.0836)
<b>Household cast (open category is reference category)</b>		
BC	-0.0299 <sup>**</sup>	(0.0160)
ST	0.0313	(0.0335)
SC	-0.0693 <sup>***</sup>	(0.0208)
Other	-0.00875	(0.0839)
<b>Government/NGO support for latrine construction (no support is reference category)</b>		
Received Loans	0.0490 <sup>***</sup>	(0.0133)
Subsidies	-0.0178	(0.0173)
Material	0.0305	(0.0473)
Constant	0.650 <sup>***</sup>	(0.0299)
Observations	16507	
$R^2$	0.065	

Note: \*  $p < 0.05$ , \*\*  $p < 0.01$ , \*\*\*  $p < 0.001$ . Coefficients are the result of fitting a linear probability model. Standard errors in parentheses are clustered at the village level. Only individuals living in households with access to a latrine are included in the regression analysis.

Individual latrine use is positively correlated with being female, years of schooling, and literacy ( $p < 0.001$ ). Women are 17 percentage points more likely to use latrines than men after controlling for other factors. Individuals in SC and BC households are 3 percentage points and 7 percentage points respectively less likely to use the latrine than individuals in OC households. Further, individuals in households who received loans for the latrine are 5 percentage points more likely to use the latrine than individuals in households who received no support. We find no evidence that individuals in households who received subsidies or construction materials were less likely to use the latrine.

## 11 Handwashing

Table 14. Characteristics of Hand Washing Facility.

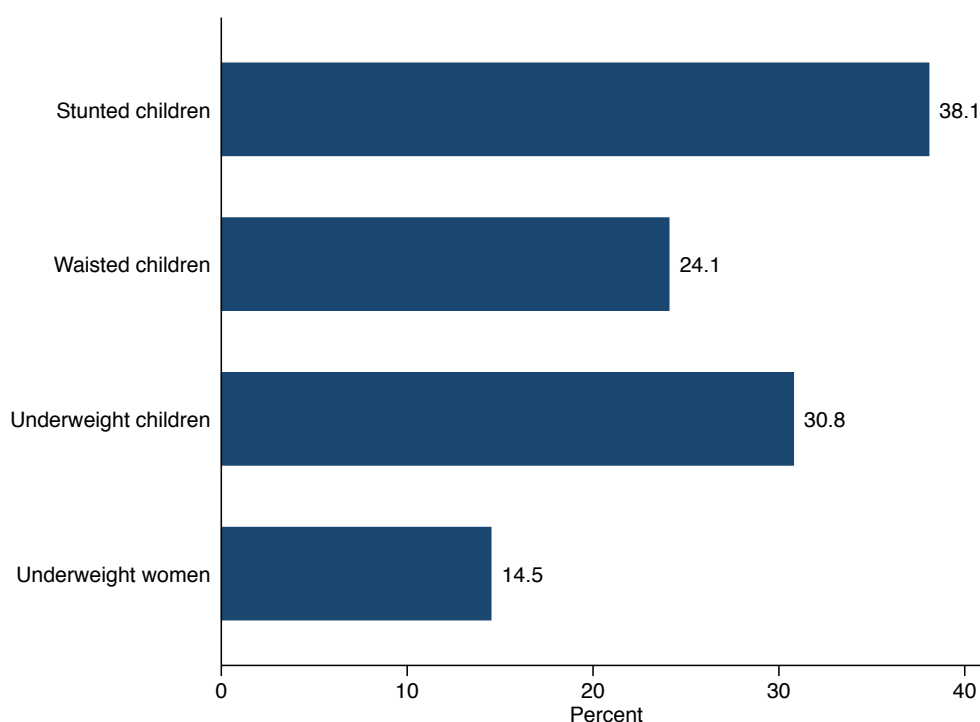
	N	Share of households (in %)
<b>Location of hand washing facility</b>		
Within the dwelling/plot	8943	74.2
Less than 100 meters from dwelling/plot	8943	24.0
More than 100 meters from dwelling/plot	8943	1.5
<b>Quality of hand washing facility</b>		
Water available at hand wash location	8668	85.6
Soap available at hand wash location	8668	68.6

Note: 97% of households gave enumerators permission to inspect hand washing facilities.

74% of households have a hand washing facility located within the plot. Water was available at 86% of the observed handwashing facilities and soap (not including ash, mud, or sand) was available at 68% of the facilities.

## 12 Anthropometric Summaries

Figure 15. Anthropometric Summaries.



Note: The first three bars show the share of children ( $\leq 5$  years) stunted, wasted, and underweight. Children are considered stunted, wasted, and underweight if they are two standard deviations below the WHO Child Growth Standards median. The last bar shows the share of underweight adult women. Women are considered underweight if their BMI is below 18.5.

Prevalence of malnutrition levels is high among children and women in the sample. 38% of children under the age of six are stunted (extremely low height for age), 20% are wasted (extremely low weight for height), and 29% are underweight (extremely low weight for age). Children are considered stunted, wasted, and underweight if they are two standard deviations

below the WHO child growth standard median. 15% of adult women are underweight (BMI < 18.5).

## 13 Kirana Stores

### Kirana Store Characteristics

Table 15. Kirana Store Characteristics.

	N	Mean	SD
<b>General characteristics</b>			
Years household has been running the store	717	9.6	8.2
Number of different customers last month	707	322.3	459.9
Number of different customers last week	713	86.5	118.7
Number of household members participating in running the store	720	1.8	0.8
Years of schooling of store owner	720	6.16	4.51
<b>Participation in schemes</b>			
At least one household member is member of a Velugu/SERP/DWCRA SHG	720	76%	
At least one household member is member of a SERP/Velugu FPG	720	5%	
Whether a household member is member of a MNS retail association	720	4%	
<b>Store Ownership Structure</b>			
Owned by male member of household	720	65%	
Owned by female member of household	720	34%	
Owned in partnership with members of the same household	720	1%	

The average Kirana stores reaches 87 different customers a week and 322 different customers a month. However, there is a high variation in the number of customers reached across stores. More than half of all Kirana stores reached less than 100 different customers a month. In three out of four Kirana store households, at least one household member is a member of a Velugu/SERP/DWCRA SHG, while less than 6% of Kirana store households have a member in a SERP/Velugu FPG or MNS retail association for Kirana stores. Around two thirds of Kirana stores are owned by male and one third by female household members.

Table 16. Other Occupations of Kirana Store Owners.

	N	Share of Kirana store owners (in %)
Farming land	141	33
Agricultural labor	141	27
NREGA labourer	141	18
Self-employment	141	13
Daily labor (non-agricultural)	141	12
Housework (own household)	141	8
Salaried/formal employment	141	2

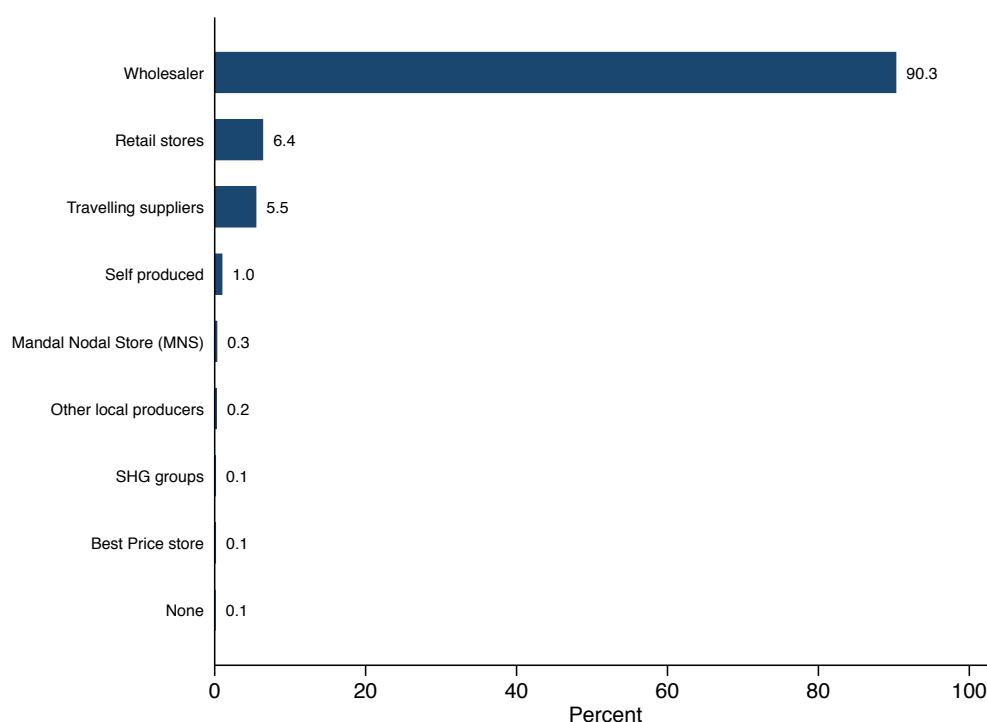


Other	141	2
Fishing	141	1
Housework (other household)	141	0
Student	141	0

20% of Kirana store owners have other occupations. Half of them work in agriculture: 33% farm land and 27% are agricultural laborers.

## Suppliers of Kirana Stores

Figure 16. Major Suppliers of Kirana Stores.



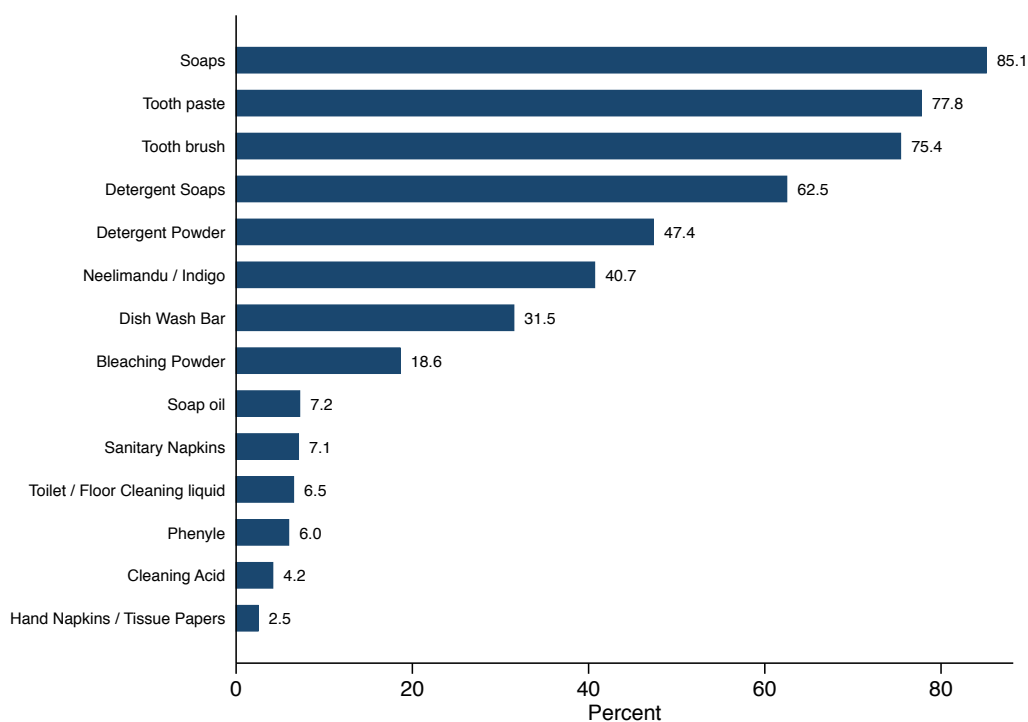
Note: Graph shows for each supplier how often it was cited as a major supplier by Kirana store owners. Respondent could place one supplier in more than one category.

88% of Kirana store owners cited wholesalers as a major supplier. Mandal Nodal Stores, SHGs and other local producers play only a marginal role. Only 2% of Kirana store owners purchased any products through SHGs in the last 30 days.

## Products Sold by Kirana Stores

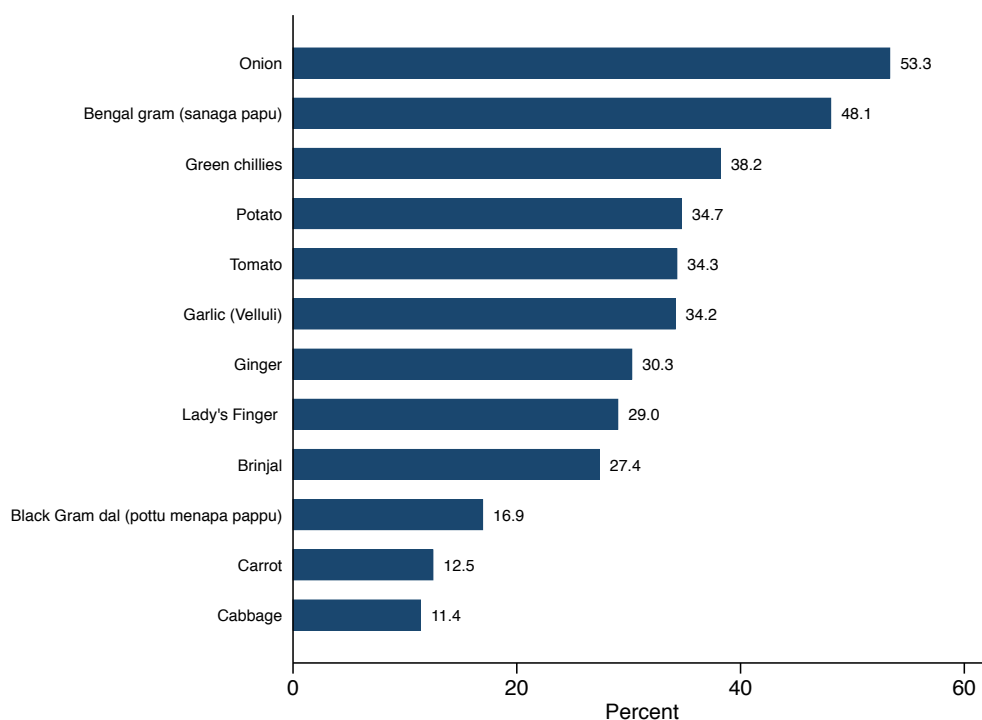
While soaps and products for dental hygiene are widely available, sanitary napkins were sold by less than 8% of Kirana stores. It is unclear from the data collected if this low availability of sanitary napkins stems from a lack of demand (e.g., if households cannot afford them) or lack of supply (i.e., Kirana stores not offering them). However, in either case this result points to an opportunity for an effective policy intervention, as lack of access to menstrual products has been shown to make adolescent girls miss school several days a month and more likely to drop out of school.

Figure 17. Sanitation Products Sold by Kirana Stores.



Note: Graph shows the share of Kirana stores which sold the item in the last 30 days.

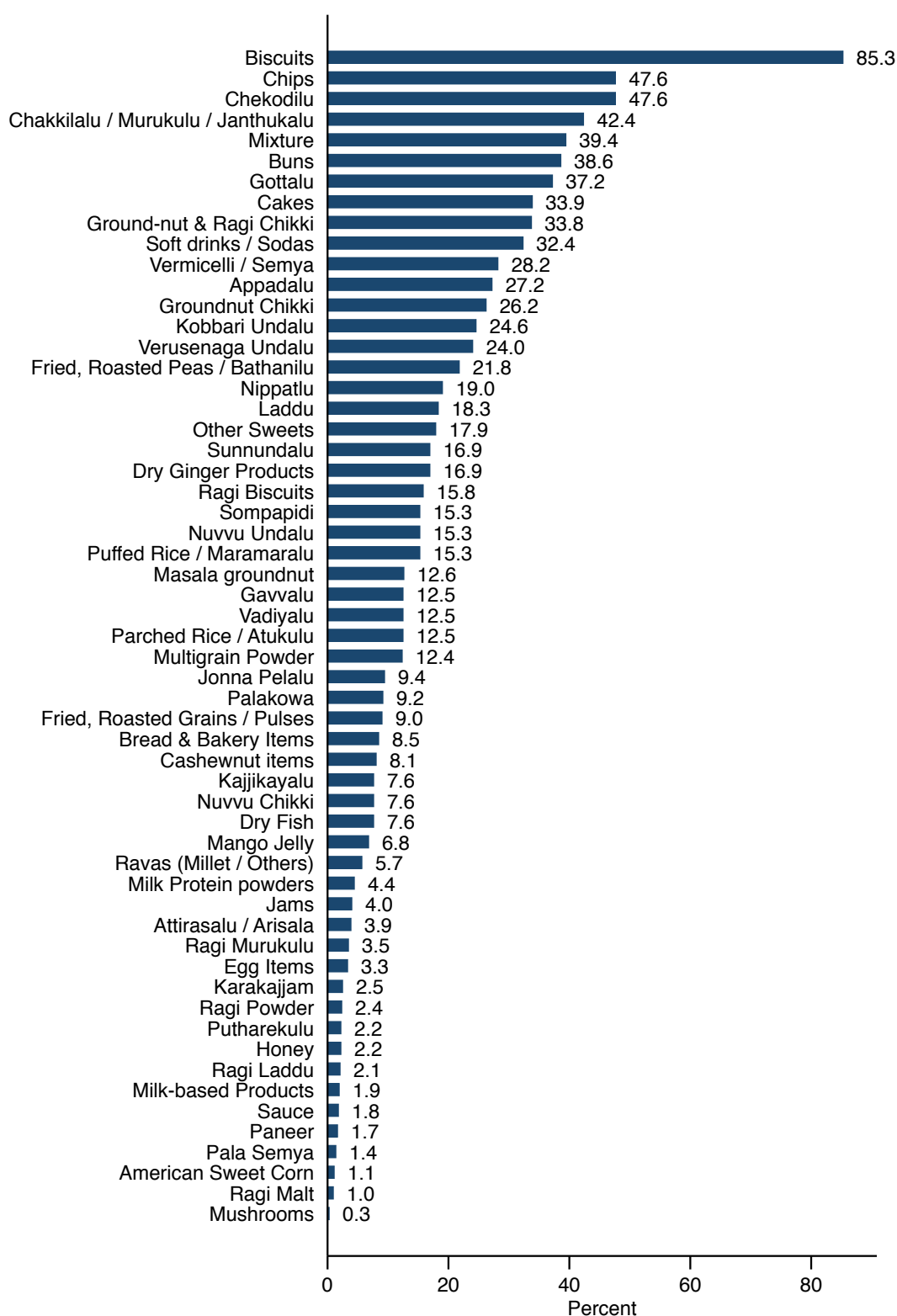
Figure 18. Fruits and Vegetables Sold by Kirana Stores.



Note: Graph shows the share of Kirana store which sold the item in the last 30 days.

Overall, few Kirana stores sell fruits and vegetables. One third of Kirana stores had not sold any fruits or vegetables in the last 30 days.

Figure 19. Nutritious Products Sold by Kirana Stores.



Note: Graph shows the share of Kirana store which sold the item in the last 30 days.

With the exception of biscuits, no nutritious product is sold by more than half of all Kirana Stores. 8% of Kirana stores sold no nutritious product and one third of Kirana stores sold fewer than 6 different nutritious products in the last 30 days. This indicates that there is scope to increase the diversity and nutritious quality of products offered by Kirana stores.

## References

- Chand, R. (2017). Doubling Farmers' Income. Rationale, Strategy, Prospects, and Action Plan. NITI Policy Paper No. 1/2017.
- Coffey, D., & Spears, D. (2017). Where India Goes: Abandoned Toilets, Stunted Development and the Costs of Caste. Harper Collins.
- Deaton, A., & Zaidi, S. (2002). Guidelines for Constructing Consumption Aggregates for Welfare Analysis (Vol. 135). World Bank Publications.
- Desai, S., Vanneman, R., and National Council of Applied Economic Research, New Delhi (2011-12). India Human Development Survey-II (IHDS-II). ICPSR36151-v2. Ann Arbor, MI: Inter-university Consortium for Political and Social Research [distributor], 2015-07-31. <http://doi.org/10.3886/ICPSR36151.v2>
- Food and Agriculture Organization of the United Nations (2014). FAOSTAT – Crops. <http://www.fao.org/faostat/en/#data/QC> (accessed 10 October 2017)
- Government of Andhra Pradesh (2014). Andhra Pradesh State Statistical Abstract. <http://www.ap.gov.in/andhra-pradesh-state-statistical-abstract/> (accessed on 5 December 2017).
- Government of Andhra Pradesh (2017). Socio Economic Survey 2016-17. <http://www.ap.gov.in/wp-content/uploads/2017/03/SOCIO-ECONOMIC-SURVEY-2016-17.pdf> (accessed on 5 December 2017).
- Kankaria, A., Nongkynrih, B., & Gupta, S. K. (2014). Indoor Air Pollution in India: Implications on Health and its Control. *Indian Journal of Community Medicine: Official Publication of Indian Association of Preventive & Social Medicine*, 39(4), 203–207. <http://doi.org/10.4103/0970-0218.143019>
- Ministry of Drinking Water and Sanitation (2013). Construction of Toilets. <http://pib.nic.in/newsite/PrintRelease.aspx?relid=101749> (accessed on 5 December 2017).

# Appendix

## Annex I Village and Household Selection

### Introduction

This annex describes the sampling and matching procedures that were used to select and identify the sample of study households for the APRIGP Impact Evaluation.

Households were selected through several stages of sampling:

1. Treatment village selection
2. Control village selection
3. Habitation selection
4. Household selection

Individual sections of this document describe each stage in detail.

After outlining the sampling procedure, details are also provided on calculating final sample weights as well as a note on what population the final sample (adjusting for weighting) is representative of.

### Treatment Village Selection

The following was the procedure followed to select the pool of potential treatment villages<sup>17</sup>:

- From the list of all 5665 Census villages (as per 2011 Census), in the 147 APRIGP mandals<sup>18</sup> in Andhra Pradesh, the following categories of villages were dropped:
  - Villages where Farmer Producer Groups (FPGs) had already opened bank accounts (230)
  - Villages for which SERP did not provide implementation status data (460)
  - Villages where SERP did not plan to implement (1743)
- This left behind 3239 villages as the pool of potential treatment villages available for selection.

From the pool of potential treatment villages, an initial set of 130 treatment villages were selected using the following procedure:

- Drop villages with fewer than 45 households as per Census data. This threshold was chosen because the minimum sample to be collected from any surveyed village was 45.
- In the list of remaining Census villages, form crop-based strata (16 such strata formed) based on the value-chain crop of the village. This mapping of villages to their value chain crops was provided by SERP.
- Within each stratum select treatment villages using Probability Proportional to Size (PPS) Sampling with the number of households in the village serving as the 'size' in PPS sampling. The number of villages to be selected from each stratum was calculated as  $130 * (\text{Population of Stratum} / \text{Total Population summed across all Stratums})$ .

---

<sup>17</sup> Villages here and in the rest of the document refers to Census villages as per the 2011 Census of India.

<sup>18</sup> There are 150 APRIGP mandals in which only 147 mandals received Farm Value Chain related interventions; the remaining 3 mandals received interventions related to animal husbandry alone.

From the selected initial set of 130 villages, 20 villages were later dropped since SERP had opened bank accounts in them after agreeing to freeze the implementation in those villages. The remaining 110 villages formed the final list of treatment villages.

### Control Village Selection

The 110 control villages were selected by matching the above mentioned treatment villages (110 villages) along socio-economic and agricultural characteristics to other Census villages in non-APRIGP mandals within the same district as the treatment village. The socio-economic variable data was obtained from Census 2011, and the village-level data relating to agricultural characteristics was obtained from the Andhra Pradesh government's CORE Dashboard.<sup>19</sup> The CORE Dashboard contains village-level information on the area and number of farmers cultivating different agricultural and horticultural crops. The data populated in the CORE Dashboard is collected bottom-up from village- and mandal-level agricultural department offices for each agricultural season.

The overall matching strategy was to stratify villages based on crop cultivation patterns, and then conduct matching within these strata based on socio-economic variables.

Strata were formed based on which crops were the "largest" crops cultivated in a village, as defined by area under cultivation for each crop in the village, using the following procedure:

- *Discarding data for less relevant crops:* Using the information on the area and number of farmers cultivating different crops from the CORE Dashboard, the "largest" crop in each district which is not a value chain crop found in the selected treatment villages of that district is identified, and data for that crop in all eligible control villages in the district are ignored for stratification purposes. For example, we ignore Paddy data for villages in Kadapa because there are no treatment villages with Paddy as the value chain in Kadapa.
- *Forming strata:* Villages were grouped into strata formed using the top two<sup>20</sup> remaining crops (in terms of area under cultivation) grown in that village, taking into consideration the ordering of the top two crops. For example, Village A growing Paddy and Maize in that order of area would be grouped into the same strata as Village B growing Paddy and Maize, but not Village C growing the crops in the order Maize and Paddy.
- *Strata reassignment for cases where value chain crops for treatment villages were missing from top two crops of original strata:* In some cases, the value chain crop of a treatment village was not part of the top two crops of the strata the village was assigned to. This could introduce problems if control villages in the same strata did not actually grow the value chain crop, hence not being able to serve as a good match for villages growing that value chain crop. These cases were manually resolved by reassigning such villages to a new strata according to district and just the value chain crop (rather than the top two crops). For example, treatment Village P in Guntur with value chain crop Chillies falling in a strata with top-2 crops Paddy-Cotton was manually assigned the strata Chillies-Guntur.

---

<sup>19</sup> The CORE Dashboard is accessible online at <http://45.114.143.92/CMWebService/#/>.

<sup>20</sup> Initial attempts to use top-3 crops to generate strata resulted in too large a number of strata to be useful.

- Control villages were allotted to these new district – value chain crop strata (and taken out of their original strata as defined by top two crops) if the area under cultivation for the specified value chain crop was above a certain threshold and they belonged to the specified district.
- The threshold was determined using the number of treatment villages that needed to be matched. For example, to generate a set of control villages (from a pool of all control villages within a district) to match with one treatment village that needed to be reassigned to a new strata, the 90th percentile of cultivated area under the unmatched value chain crop across all villages in the district was used as the threshold. However, to generate a set of control villages to match 10 treatment villages against, the 80th percentile of cultivated area was selected. This discretion was to ensure that the number of control villages reassigned to new strata was kept to a minimum.
- In order to keep such strata reassignments to the minimum, we considered only those treatment villages where the value chain crop did not appear in the top three crops of that strata: for example, if Village T with value chain crop Chillies falls in a strata with top three crops Paddy-Cotton-Maize, we would reassign village T to a new strata.
- *Strata reassignment for too few control villages in strata:* Strata reassignment was also done in the same manner as above for cases where strata had fewer control villages than treatment villages for matching.

Finally, within each strata and within districts, control villages were matched to treatment villages based on Principal Component Analysis (PCA) scores generated using the following socio-economic variables:

- Proportion of SC/ST population in the village,
- Proportion of literate women in the village,
- Total population of the village, and
- Proportion of irrigated area (at the mandal level, hence constant within mandal).

Thus, the villages in each final pair of matched treatment and control villages fall within the same district, have similar crop-cultivation patterns, and have similar socio-economic characteristics.

#### Habitation Selection (for both Treatment and Control Villages)

For all treatment and control villages, SERP staff provided a list of habitations in each village along with the following information for each habitation:

- No of households
- Whether or not the habitation is “FPG-suitable”<sup>21</sup>
- # of households cultivating the value chain crop
- whether the habitation is majority SC/ST

Villages were then split into three groups based on the habitation information:

---

<sup>21</sup> The factors considered by SERP to assess “FPG suitability” includes number and proportion of households cultivating the value-chain crop, ease of mobilizing them, and their past participation in similar programs.

- **Group 1:** Villages in which there is at least 1 habitation suitable for FPG which has at least 30 households cultivating the value chain crop
- **Group 2:** Villages in which there is at least one FPG-suitable habitation, but no habitations that are both FPG-suitable and have at least 30 households cultivating the value chain crop
- **Group 3:** Villages which have no habitations that are FPG-suitable

Habitations were randomly selected based on the following set of rules which varied according to the groups defined above:

- For **group 1** villages:
  - First, remove all habitations that are not FPG-suitable or which don't have at least 30 value chain cultivating households.
  - If at least one remaining habitation is SC/ST majority, remove all habitations that are not SC/ST majority.
  - From remaining habitations, select using PPS sampling, taking size as the total number of households in the habitation from SERP data.
- For **group 2** villages:
  - First, remove all habitations that are not FPG-suitable.
  - For the remaining habitations, sort habitations first by whether habitation is SC/ST majority and then by descending number of households which cultivate the value chain crop.
  - Select all habitations until the total number of households cultivating the value chain crop in selected habitations is at least 30. (Note that these habitations should not be segmented.)
- For **group 3** villages:
  - Sort habitations first by whether habitation is SC/ST majority and then by descending number of households which cultivate the value chain crop.
  - Select all habitations until total number of households is at least 200. (Note that these habitations should not be segmented unless only one habitation was selected and this habitation has a lot more than 200 households.)

Note that whether or not the final selected habitation should be segmented or not thus depends not just on size but also on the way in which it was selected.

**Dropped Villages:** During the Habitation selection stage we dropped the following four villages:

- In Anantapur district, Chennavaram village (Treatment village) in Singanamala mandal was dropped since it is in the drainage area of a dam and families were evacuated from there a couple of years back.
- In Visakhapatnam district, Gudem Kotha Veedi village (Treatment village), was selected under the assumption that the value chain grown in the village was turmeric. However, later SERP revised the value chain to turmeric. There are no matching turmeric control villages in Visakhapatnam.
- In West Godavari district, Tekuru village (Treatment village) in Polavaram mandal was dropped since this village comes under the Polavaram irrigation project. As a result the families have been relocated from the village.



- In Srikakulam district, the village Ranastalam (Control village) in Ranastalam mandal was selected as a control village under the assumption that it grew paddy. However, there is no paddy cultivation in the village, only maize is cultivated.

**Replacement Villages:** After consultations with SERP officials including the DPM of Srikakulam district, we selected Korada village in Hiramandalam mandal as a replacement for Ranastalam village in the same district. The remaining three treatment villages could not be replaced, as the intervention had reached an advanced stage already in previously unsampled treatment villages.

**Significant deviation in population estimates provided by SERP, and on-the-ground estimates:** In a few habitations, it was observed that the actual population of the habitation (as noted by the enumerators surveying the habitation) differed significantly from the estimates provided by SERP. The list of such habitations along with the difference in estimate, and the course of action taken to address the differences is provided below:

District	Mandal	Census Village	Habitation	Expected HH count	Actual HH count
Chittoor	Mulakalacheruvu	Mulakalacheruvu	Patha Mulakalacheruvu	82	150-160
East Godavari	Rowthulapudi	Srungavaram	Srungavaram	200	1100-1150
East Godavari	Jaggampeta	Balabhadrapuram	Balabhadrapuram	535	140-150
East Godavari	Gangavaram	Nellipudi	Nellipudi	86	950-1000
East Godavari	Rajavommangi	Rajavommangi	Rajavommangi	190	1000-1100
East Godavari	Rajavommangi	Singampalle	Singampalli	85	130-140
Guntur	Pedanandipadu	Varagani	Varagani	2500	1100-1150
Guntur	Piduguralla	Peda Agraharam	Peda Agraharam	600	200-220
Krishna	Musunur	Chintalavalli	Chintalavalli	700	1050-1100
Krishna	Musunuru	Chintalavalli	Chintalavalli	700	1100-1150
Krishna	Kanchikacherla	Pendyala	Pendyala	1700	2000-2100
Krishna	Chandarlpadu	Veladi	Kothapalem	305	700-730
Krishna	Kanchikacherla	Kanchikacherla	Kanchikacherla	6251	5000-5100
Krishna	Veerullapadu	Jujjuru	Jujjuru	1920	5700-5800
Kurnool	Peapally	Mettupalle	Mettupalle	1024	430-450
Nellore	Podalakur	Biradavolu	Biradavolu	385	295-310
Nellore	Naidupeta	Karumanchivari Khandrika	Karumanchivari Khandrika	120	90-92
Nellore	Bogole	Thalluru	Chaitanyanagar	110	60-70
Nellore	Marripadu	Allampadu	Pathalampadu	98	60-70
Prakasam	Cumbum	Hazarat Gudem	Kagithala Gudem	220	70-80
Prakasam	Marripudi	Marripudi	Gangaalem	150	90-100
Prakasam	Tripuranthakam	Ramasamudram	Ramasamudram	600	350-400
Prakasam	Komarolu	Reddicherla	Mulapalli	1304	120-130
Prakasam	Ballikurava	Konidena	Konidena	350	830-850
Prakasam	Naguluppala_Padu	Mattigunta	Mattigunta	350	190-200
Prakasam	Naguluppala_Padu	Raparla	Nettevaripalem	8	100-110
Prakasam	Inkollu	Inkollu	Inkollu	900	4500-4600

Prakasam	Tangutur	Marlapadu	Marlapadu	375	630-650
Prakasam	Yerragondapalem	Gollavidipi	Gollavidipi	2000	700-750
Srikakulam	Pathapatnam	Routhupuram	Routhu Laxmi Puram	192	380-400
Srikakulam	Amadalavalasa	Kalivaram	Kalivaram	289	150-160
Visakhapatnam	Madugula	Sankaram	Krishnam Palem	82	150-160
West Godavari	Jeelugu Milli	Mulagalampalle	Mulagalampalle	150	950-1000
West Godavari	Dwarakaturumala	Gunnampalle	Gunnampalli	100	830-850

### Household Selection

Within each treatment and control habitation, 30 households were randomly selected from among all eligible farmer households and 15 households were randomly selected from among all remaining households.<sup>22</sup> Treatment and control households were selected independently of each other, i.e. no matching procedures were used to select the control households. Farmer households were eligible for sampling (as farmer households) if they cultivated the value chain crop in the past two years.

In case the number of farmer households fell below the required number in a village, we topped up the sample by drawing first from the pool of farmer households which did not cultivate the value chain crop, and then from the remaining general households in that same village.

However, in 40 villages the number of farmer households (irrespective of their value chain crop cultivation status) was significantly lower than the required sample size as per power calculation. The survey agency performed a review of these villages to assess the feasibility of additional census survey and short-listed 17 villages for a second round of census survey. This second round of census survey was performed either in a different habitation within the same village as the original survey, or if no other surveyable habitations were present in the village, then in a different segment of the habitation originally surveyed. The following table lists all these 40 ‘problem village’ along with their resurvey status.

District	Mandal	Census Village	Additional Survey Status
Anantapur	guntakal	obulapuram	
Chittoor	baireddipalle	kambhampalle	
Chittoor	chinnagottigallu	bhakarapet	Additional Survey
Chittoor	kurabalakota	mudivedu	
Chittoor	mulakalacheruvu	burakayalakota	
Chittoor	punganur	chandramakula palle	
East Godavari	seethanagaram	nallagonda	
Guntur	bollapalle	bollapalle	
Guntur	gurazala	gottimukkala	
Kadapa	galiveedu	pandikunta	
Kadapa	lakkireddipalle	brahmana yerragudi	
Kadapa	sambepalle	guttapalle	

<sup>22</sup> Households were not stratified for by PoP status.

Kadapa	sambepalle	motakatla	Additional Survey
Krishna	kanchikacherla	kanchikacherla	Additional Survey
Krishna	kanchikacherla	pendyala	
Krishna	kankipadu	uppaluru	Additional Survey
Krishna	mylavaram	mylavaram	
Krishna	vatsavai	kakaravai	Additional Survey
Krishna	vissannapet	narasapuram	Additional Survey
Kurnool	kolimigundla	kolimigundla	Additional Survey
Kurnool	nandavaram	nandavaram	
Nellore	marripadu	allampadu	
Prakasam	lingasamudram	lingasamudram	Additional Survey
Prakasam	maddipadu	edugundlapadu	Additional Survey
Prakasam	marripudi	kuchipudi	
Prakasam	santhanuthala padu	matti padu	
Prakasam	singarayakonda	somarajupalle	
Prakasam	tangutur	marlapadu	Additional Survey
Prakasam	voletivaripalem	voletivaripalem	
Prakasam	yerragondapalem	gollavidipi	
Prakasam	zarugumilli	narasimhanayani khandrika	
Prakasam	zarugumilli	zarugumilli	Additional Survey
Srikakulam	amadalavalasa	kalivaram	Additional Survey
Srikakulam	hiramandalam	korada	Additional Survey
Srikakulam	pathapatnam	korasavada	Additional Survey
Srikakulam	pathapatnam	pathapatnam	Additional Survey
Srikakulam	saravakota	agadala	
Visakhapatnam	madugula	vommali jagannadhapuram	
Vizianagaram	parvathipuram	sudigam	Additional Survey
Vizianagaram	vepada	kondagangupudi	Additional Survey

### Sampling Weight Calculation

Estimation of the final regression equations requires that each observation be weighted with the inverse probability of its occurrence. Since the observations are derived via sampling, these weights will be calculated as sampling weights. Sampling weights are calculated as the inverse of the product of the probability of selection at each stage of sampling.

As detailed in the previous sections, the selection of the sample for the survey involved three different stages. In the first stage, the list of treatment villages was stratified based on the value-chain crop cultivated in the village. Each such stratum was allotted a sample size of villages based on relative population size of that strata. In the second stage, the allotted numbers of villages were randomly sampled from the strata using Probability Proportional to

Size sampling method with the population of the village serving as the size parameter. In the final stage farmer and non-farmer households were sampled from each habitation randomly. Accordingly, the sample weight for each observation of the sample may be approximated using the following formula:

Sampling Weight =  $1 / P(\text{Village}) * P(\text{Household})$ , where  $P()$  denoted the probability of selection at each stage of sampling.

$P(\text{Village}) = (\# \text{ Villages selected in the stratum}) * (\text{Population of the village as per census}) / \text{Total population of all villages within the stratum as per census}$

$P(\text{Household}) = \text{No of selected household from the sub-group} / \text{Total number of households in the sub-group}$ , where the sub-groups are farmer households and non-farm households. Household counts are based on survey mapping rather than the census or SERP estimates.

Note that we do not account for the probability of selecting habitations due to the complicated way in which habitations were selected.

#### Representativeness of the Sample (after-weighting)

After weighting, our final treatment sample is approximately representative of all households in all villages in the treatment mandals not excluded from the sampling process. Due to the various corner cases described above and that habitation selection process, this does not hold true exactly but may be considered a rough approximation.

## Annex II Construction of Income Aggregate

This annex provides an overview of the statistical and analytical procedures that will be used to construct the nominal household income aggregate from the APRIGP Baseline survey data. The sections below describe the construction of categories of income, which comprise the final income aggregate and the variables, which are included in these categories.

**Note on typographic convention:** Through out this document, text in fixed-width font refers to the names of variables from the survey questionnaire.

### Income Aggregate Construction

- It is most transparent to construct the income aggregate as the sum of distinctive sub-aggregates of total income. The sub-aggregates we use to split income are: agriculture and animal husbandry related income; income from wage and salary; income from self-employment activities; income from fishing and income from transfers.
- Detailed procedures to construct these aggregates are described below.

### Agriculture and Animal Husbandry Income aggregate

#### Agricultural Income Aggregate

- The total income from agriculture will be calculated in the following manner:
  - $\text{Net income from agriculture} = \text{Gross income from the sale of harvested crops} + \text{Gross income from the sale of byproducts} - \text{Net expenditure on leasing land} - \text{Operating expenses related to cultivations} - \text{Net capital expenses related to cultivation} - \text{Total expenses incurred during the sale to buyers of the harvest}$
- The gross income from the sale of harvested crops will be calculated as the total of the variable `buyer_money` aggregated over crops and seasons.
- Gross income from the sale of byproducts will be calculated as the total of the variable `harvest_byprod_value` aggregated over crops and seasons.
- The net expenditure on leasing land will be calculated as the sum of lease rent received and paid on land owned or leased. Thus net expenditure = value of rent paid in cash or kind + value of harvest paid in the form of rent for share-cropping - value of rent received in cash or kind - value of harvest received in the form of rent for share-cropping.
  - In terms of variables from the survey, net expenditure on leasing land =  $\text{value\_rent\_out} + \text{harvest\_crop\_lord}$  (valued at price paid by the buyer) -  $\text{value\_cash\_kind} - \text{value\_crop\_rent}$
- Operating expenses related to cultivation will be calculated as the total of the variable `ag_exp_item` over several different assets and seasons, along with the following variables `cost_water` and `labour_pay_cash` aggregated over all seasons.
- Net capital expenses related to cultivation will be calculated as the total of the rent paid for different capital assets recorded in `farm_asset_rental` after adjusting the rent received for different capital assets recorded in `farm_asset_lease`. In the case of assets owned by the household, the value of rent will be found via imputation as described in the section below.
- Total expenses incurred during sale to buyers of the harvest will be calculated by totaling the expenses incurred on such transportation for each crop in each season. These expenses are recorded in the variable `buyer_transport_cost`.

### Animal Husbandry Income Aggregate

- The net income from animal husbandry activities will be calculated using the following formula:
  - Net income from animal husbandry = Gross income from sale of animal produce - Gross expenses on animal husbandry + Gross income from sale of livestock - Gross expenses on purchase of livestock
- In terms of variables, the following formula will be used:
  - $$\text{money\_milk} - \text{money\_produce} - \text{value\_fodder\_self} - \text{value\_fodder} - \text{cost\_vet\_animal} - \text{cost\_labour\_animal} - \text{cost\_other\_animal} - \text{money\_livestock} - \text{money\_buy}$$
- The following table presents a summary of the different variables used in the above formula. Items of expense are colored in red.

Variable Name	Variable Bucket	Description of the variable
<b>money_milk</b>	Gross income from sale of animal produce	Indicates the income from selling milk and milk products in the last 12 months
<b>money_produce</b>	Gross income from sale of animal produce	Indicates income from selling poultry, eggs, meat, and wool in the last 12 months
<b>value_fodder_self</b>	Gross expenses on animal husbandry	Indicates the value in rupees of the feed and fodder the household produced/ collected in the last 12 months (excluding open grazing)
<b>value_fodder</b>	Gross expenses on animal husbandry	Indicates the value in rupees of the feed and fodder the household purchased in the last 12 months
<b>cost_vet_animal</b>	Gross expenses on animal husbandry	Indicates the cost incurred by the household for veterinary services in the last 12 months
<b>cost_labour_animal</b>	Gross expenses on animal husbandry	Indicates the cost incurred on labor the household hired to take care of animals in the last 12 months
<b>cost_other_animal</b>	Gross expenses on animal husbandry	Indicates the other costs incurred by the household, such as expenses on ropes, chain, transport, related to animal husbandry
<b>money_livestock</b>	Gross income from sale of livestock	Indicates the total amount of money the household earned selling livestock in the last 12 months
<b>money_buy</b>	Gross expenses on purchase of livestock	Indicates the total amount of money the household spent on buying livestock in the last 12 months

### Wage and Salary Income Aggregate

- The wage and salary income aggregate for a household will be constructed by totaling the total wage or salary earnings along with the value of bonus in cash or kind received in the last year over each member of the household.
- In terms of variables, the following formula will be used:
  - Gross wage and salary income aggregate = `ws_worked_rate` totaled over each industry in which each member of the household worked in the last year.
- Income from NREGA, totaling `nrega_payment` over each member of the household, will be reported separately.

## Self-Employment Income Aggregate

- The net income from self-employment activity will be calculated using the following formula for each self-employment activity engaged in by the household, and then totaled over all such separate employment activities.
- Net income from self-employment activity = Value of sales - Value of purchases - Value of business related expenses.
- The formula is expressed in terms of the variables collected via the survey in the following manner:
  - Net income from self-employment activity =  $\text{busin\_sales} + \text{busin\_value\_others} - \text{busin\_raw} - \text{busin\_interest} - \text{busin\_rent} - \text{busin\_utilities} - \text{busin\_transport} - \text{busin\_labour} - \text{busin\_tax} - \text{busin\_operating}$
- The following table presents a summary of the different variables used in the above formula. Items of expense are colored in red.

Variable Name	Variable Bucket	Description of the variable
<b>busin_sales</b>	Value of sales	Indicates the total value of sales over the last 12 months
<b>busin_value_others</b>	Value of sales	Indicates value of receipts from other business services offered by the enterprise
<b>busin_raw</b>	Value of purchases	Indicates the value of raw materials purchased over the last 12 months
<b>busin_interest</b>	Value of business related expenses	Indicates the value of interest paid on capital by the enterprise over the last 12 months
<b>busin_rent</b>	Value of business related expenses	Indicates the value of rent paid on building/ machinery over the last 12 months
<b>busin_utilities</b>	Value of business related expenses	Indicates the value of expenses incurred on electricity, water, gas, fuel etc over the last 12 months
<b>busin_transport</b>	Value of business related expenses	Indicates the value of expenses incurred on transportation and marketing over the last 12 months
<b>busin_labour</b>	Value of business related expenses	Indicates the value of expenses incurred on hired laborers over the last 12 months
<b>busin_tax</b>	Value of business related expenses	Indicates the total tax paid by this enterprise over the last 12 months
<b>busin_operating</b>	Value of business related expenses	Indicates the total of all other operating expenses over the last 12 months.

## Income from Fishing

- The variable `fishing_months` records all the months in the last 13 months from July 2016-July 2017 during which the household cultivated or captured fish. The remainder of the fishing income module contains questions relating to the last-3 months.

- Only 6 out of 8943 households reported to be active in fishing. Hence, income from fishing was not included in the income aggregate reported in the baseline report.
- The total income from fishing for the household will be calculated in the following manner:
  - $\text{Net income from fishing} = \text{Annualized gross income from fishing} - \text{Annualized net revenue expenses on fishing} - \text{Annualized net capital expenditure on fishing}.$
- Annualized gross income from fishing will be calculated as the annualized sum of the value of the variable `fish_sold_value` over each different category of fish listed in `fish_category`.
- The total of the operating expenses incurred for fishing the household over the last three months of fishing will be calculated as the total of the variable `spend_opex_item` across different items. This total will then be annualized to obtain the net revenue expenses on fishing.
- Another category of expenses incurred for fishing by the households is the expense relating to capital assets used in fishing such as fishing ponds and boats. The annual expenses on these items will be calculated in the following manner:
  - In the case of assets owned by the households, we will find the median expenditure on account of leasing each such capital item within each district i.e. `fish_asset_rental`. This median expenditure will be imputed to all households who own these capital assets. Such imputed expenditure will be scaled-up to the fishing activity in the last 12 months.
  - The total of capital expenses will be reduced by adjusting for any income received from leasing out of these assets.
  - The total of such capital expenses will be annualized as appropriate.

#### Net-income from Transfers

- The net income from transfers will be calculated in the following manner:
  - $\text{Net income from transfers} = \text{value of transfer received in cash} - \text{value of transfer given in cash}.$
  - The exact variables used will be: `transfer_received_cash` - `transfer_given_cash`
- Since the reference period for variables relating to transfers is 12 months, annualization is not necessary.

#### Technical Summary: Annualization Formula

- The survey collects information on variables for a given period of time known as the reference period.
- The reference period varies from variable to variable-- for example, in the consumption module the reference period for most items of everyday use (vegetables, food grains etc.) is 'the last 30 days' while the reference period for items such as education, vacation etc. is 'the last 365 days'.
- However, for the final analysis, we require the value of these variables corresponding to the same time period, e.g., 30 or 365 days.
- The annualization formula given below converts values for the reference period to annualized values.



- Annualized value = Reference period value x Length of a year in reference period units / Length of reference period.
- The following examples will make the above formula clearer:
  - Annualized value of food-grain consumption = Value of food-grain consumption in the last 30 days \* 365 days/ 30 days.
  - Annualized value of last month's consumption of food-grains = Value of food-grain consumption in the last month \* 12 months/1 month.

#### Technical Summary: Rent Imputation for Owned Assets

- There are several capital assets which are used in agriculture and fishing which can either be leased-in or owned by the household.
- In the case of leased-in assets, the expenditure associated with the asset for a given period is the pro-rated lease rental paid for the asset. However, in the case of assets owned by the households, such expenditure figures are not easily available.
- The following procedure will be used to impute lease rental for the assets owned by the household:
  - Within each district, we will find the median expenditure on account of leasing each such capital asset, i.e. we will find the pro-rated median rent paid for that capital asset by households who lease the asset. This median expenditure will be imputed to all households who own these capital assets. Such imputed expenditure will also be annualized if necessary.

## Annex III Construction of Consumption Aggregate

The consumption module of the APRIGP baseline survey consists of several questions about the consumption expenditure of the household on various categories of items for different reference periods. These questions collect information on consumption expenditure on food, non-food, and spending on durable assets. This document details the procedure that will be followed to arrive at the final household level monthly consumption aggregate.

The consumption module of the survey consists of 49 questions with two different reference periods. Some questions, such as those relating to consumption of everyday food items such as cereals, grains, and vegetables have a reference period of 30 days, while questions relating to consumption expenditure on items such as education, durables etc. have a reference period of 365 days.

The final reporting of the consumption aggregate will be as a monthly figure, i.e. items of consumption pertaining to an annual reference period will be pro-rated to monthly figures.

### Consumption Aggregate for 30-day Reference Period Items

In the consumption module of the questionnaire, questions 7.1 to 7.28 have a reference period of 30 days, and relate to routine consumption on food items and other household expenses. For all these questions, the value of consumption in the last 30 days is collected via the survey. The 30-day reference period consumption aggregate will be constructed by totaling the responses to these questions about the value of consumption of different items.

The 30-day consumption aggregate is further divided into two sub-buckets. Questions 7.1 to 7.17 relate to food items, some of which are supplied through PDS stores as well. Questions 7.18 to 7.28 relate to non-food items of expenditure, which are still routine in nature. These include spending on fuel, telephone charges, water charges etc.

### Consumption Aggregate for 365-day Reference Period Items

In the consumption module of the questionnaire, questions 7.30 to 7.48 have a reference period of 365 days. For all these questions as well, the value of consumption in the last 365 days is collected via the survey. The 365-day reference period consumption aggregate will be constructed by totaling the responses to these questions about the value of consumption of different items. This aggregate will then be pro-rated down to a monthly figure.

However, from the larger list of all items collected with a 365-day reference period, we will drop some variables that are lumpy and infrequent expenditure items as per Deaton and Zaidi (2002). For example, the variable relating to purchase of furniture and fixtures and transportation equipment are to be dropped since they are large expenses that are incurred only sporadically. The table below lists all variables, which will be similarly excluded.

Table 17. Variables Excluded from Consumption Aggregate.

Variable	Description of the variable	Variable bucket	Reason for exclusion
y_furniture_fixtures	Purchase of furniture and fixtures (including bedstead, almirah, suitcase, carpet, etc.)	365 day consumption of household items	Lumpy, infrequent expenditure. Excluded as per Deaton and Zaidi
y_cooking_appliances	purchase of cooking and household appliances (includes electric fan, AC, sewing machine, washing machine, stove, pressure cooker, refrigerator, water purifier, iron, oven, toaster, etc.)	365 day consumption of household items	Lumpy, infrequent expenditure. Excluded as per Deaton and Zaidi

y_recreation	purchase of durables for recreation (TV, radio, tape recorder, musical instruments, CD/DVD players, camera, etc.)	365 day consumption of household items	Lumpy, infrequent expenditure. Excluded as per Deaton and Zaidi
y_transport	personal transport equipment (includes bicycle, scooter, car, tyres, etc.)	365 day consumption of household items	Lumpy, infrequent expenditure. Excluded as per Deaton and Zaidi
y_therapy	therapeutic appliances (includes glasses, contact lenses, hearing aids, etc.)	365 day consumption of household items	Lumpy, infrequent expenditure. Excluded as per Deaton and Zaidi
y_personal_good	other personal goods (includes clock, watch, computer, mobile, landline phone etc.)	365 day consumption of household items	Lumpy, infrequent expenditure. Excluded as per Deaton and Zaidi
y_durables_other	other durables	365 day consumption of household items	Lumpy, infrequent expenditure. Excluded as per Deaton and Zaidi
y_ornaments	jewelry and ornaments	365 day consumption of household items	Lumpy, infrequent expenditure. Excluded as per Deaton and Zaidi
y_functions	social functions (marriage, funerals, gifts, etc.)	365 day consumption of household items	Lumpy, infrequent expenditure. Excluded as per Deaton and Zaidi

Finally, the above 30-day reference period aggregate, the pro-rated 365-day reference period aggregate, and the imputed rental value of self-consumed housing will be added together to derive the total annual consumption expenditure of the household.

### Imputed Rental Value for Housing Services

Deaton and Zaidi (2002) identify housing services as constituting an integral component of the consumption basket of households, and recommends that in the event of self-owned housing, the rental value of housing services be calculated using hedonic regressions and other methods. However, in our sample, only 5% of the total sample lives in rented houses which makes estimation of hedonic regressions difficult.

### Price Indices

Our data reflects significant variation in price of different food items between districts. To adjust for the inter-district differences in price, we will create a Paasche index as per Deaton and Zaidi (2002, pp. 43). The following items will be used to construct the price index: Rice, Sugar, Pulses, Chicken, eggs, and milk. Some food items have been left out due to sparse data on those items in some districts. The detail of the selection of food items included in the index is provided in the appendix.

### Equivalence Scales

The consumption aggregate constructed as per the above procedure, should be adjusted to account for the differences in composition of the households. This adjustment will be performed using an equivalence scale.

Although, Deaton and Zaidi (2002) recommend several different approaches to constructing equivalent scales, we have opted for the simple method of using the household size as the equivalence scale. Accordingly, the real aggregate household consumption will be deflated by dividing the aggregate by the household size.

## Construction of the Price Index based on Food Prices

This section describes the choice of food items whose prices will be used to construct the Paasche price index that will be used to deflate nominal consumption aggregate.

Our questionnaire contains several variables, which relate to expenditure on food items. However only a few of them have associated information on quantity of consumption which can be used to construct average prices of food items.

The general procedure we use to select prices for the Paasche index is the following: use the price of every food item for which average price is available, provided that the item is consumed by more than 40% of the households in each district in our sample. Accordingly, the following items have been omitted from the price index: Wheat, Palm oil, Edible oils and vanaspati, meat, and fish. Thus the price index will be a Paasche index constructed using the prices of rice, sugar, pulses, chicken, eggs, and milk.

The tables below provide the district-wise proportions of households consuming each different food item. Highlighted cells provide examples of instances where our selection criterion were not satisfied.

Table 18. Proportion of Households Consuming Core Food Items by District (in INR per kg/liter).

District	Rice	Wheat	Sugar	Palm Oil	Pulses
Anantapur	0.94	0.95	0.97	0.89	0.97
Chittoor	0.76	0.41	0.89	0.45	0.96
East_Godavari	0.75	<b>0.29</b>	0.94	0.92	0.95
Guntur	0.90	0.48	0.80	0.78	0.87
Kadapa	0.73	0.70	0.86	0.91	0.91
Krishna	0.84	0.42	0.94	0.72	0.95
Kurnool	0.92	0.58	0.87	0.82	0.88
Nellore	0.88	0.46	0.80	0.82	0.94
Prakasam	0.92	0.63	0.72	0.83	0.90
Srikakulam	0.49	0.60	0.97	<b>0.33</b>	0.96
Visakhapatnam	0.72	<b>0.35</b>	0.92	0.85	0.91
Vizianagaram	0.55	0.61	0.90	0.66	0.77
West_Godavari	0.79	<b>0.28</b>	0.93	<b>0.36</b>	0.97

Note: The above table summarizes the district-level proportion of households consuming specified food products.

Table 19. Proportion of Households Consuming Core Food Items by District (in INR per kg/liter).

District	Edible Oils and Vanaspti	Meat	Chicken	Fish	Eggs	Milk
Anantapur	<b>0.21</b>	0.57	0.73	<b>0.08</b>	0.58	0.77
Chittoor	0.61	<b>0.36</b>	0.78	<b>0.14</b>	0.60	0.46
East_Godavari	<b>0.13</b>	0.47	0.73	0.43	0.86	0.75
Guntur	<b>0.35</b>	0.53	0.57	<b>0.14</b>	0.67	0.78
Kadapa	<b>0.16</b>	0.55	0.69	<b>0.12</b>	0.59	0.56
Krishna	0.41	0.48	0.71	<b>0.20</b>	0.81	0.74

<b>Kurnool</b>	<b>0.38</b>	0.59	0.65	<b>0.08</b>	0.54	0.61
<b>Nellore</b>	<b>0.27</b>	<b>0.26</b>	0.71	<b>0.35</b>	0.70	0.46
<b>Prakasam</b>	<b>0.25</b>	0.45	0.63	<b>0.16</b>	0.69	0.66
<b>Srikakulam</b>	0.69	0.42	0.85	0.67	0.75	0.56
<b>Visakhapatnam</b>	<b>0.35</b>	0.58	0.66	0.54	0.80	0.56
<b>Vizianagaram</b>	0.62	0.78	0.79	0.51	0.85	0.52
<b>West_Godavari</b>	0.79	0.45	0.69	0.49	0.75	0.61

Note: The above table summarizes the district-level proportion of households consuming specified food products.

## Annex IV Field Observations and Comparisons with Back-Check Survey Data

This annex lists observations from field visits and audio audits by IDinsight field staff and comparisons between original survey data and back check survey data for key variables. Results are organized by questionnaire sections.

The purpose of this annex is twofold: First, we seek to alert potential users of the data to concerns we have about specific variables. Second, we want to highlight the challenges we faced in data collection so that the agency responsible for the endline data collection is aware of these challenges and can overcome them through extended enumerator training or changes to the questionnaire.

### Field Visits and Audio Audits

Overall, four field visits have been conducted to observe enumerators in the districts Prakasham, West Godavari, Krishna, and Nellore. All deviations from the survey protocol described in this annex were immediately reported to AMS, SERP, and the World Bank team and followed-up on by IDinsight staff. In many cases we found that they have not been resolved with the urgency they required. In addition to field visits, IDinsight also conducted audio audit checks of household surveys.

It should be noted that findings from field visits and audio audits are limited to a small sample of enumerators and we cannot assess how prevalent these deviations from the survey protocol were among all enumerators.

### Back-Check Survey

Back-check surveys were conducted by AMS with 279 households in all 13 baseline districts. The tables in this annex reports the following statistical information to assess the quality of the collected data:

- **Deviation Range:** The difference between the back-check answers and the original survey answers represented as a proportion of the original answers and back-check answers.
- **Proportion of variation explained within household/ member:** The within-group intra-cluster correlation for the variable across the original survey and the back-check survey. The groups could either be the household or a member of the household as appropriate. The value can be calculated both by using the trimmed set of data, or the untrimmed data.

### Section 1: Metadata, Consent and Introduction

Several enumerators did not ask for the household's consent before starting a survey, and failed to provide an appropriate introduction and explanation to the survey.

### Section 2: Household facilities survey and roster

#### *Field Observations on Data Quality*

**Household Roaster:** In some cases, enumerators have not followed the definition of the household which was provided to them during training. They casually asked about the members living in the house of the respondent and entered the names the respondent provided.

**Occupation:** Some of the observed enumerators did not follow the definition of primary occupation which was provided to them during training. They asked in general about the occupation of a household member or the work done by a household member. Sometimes, the enumerators selected the primary occupation on their own if the household

member was doing more than one job/work. Some enumerators captured the same occupation both for primary and secondary occupation.

### Section 3: Agriculture and Animal Husbandry Module

#### *Field Observations on Data Quality*

**Land Module:** In some cases, enumerators did not ask these questions per season. Instead, they asked how much land the household had in total and how much they have cultivated.

**Irrigated Land:** Enumerators were sometimes unclear about the distinction between irrigated land and non-irrigated land and treated rain-fed land as irrigated land.

**Permanent and Current Fallows:** In several instances, enumerators were unclear about the distinction between permanent fallows and current fallows.

**Wrong selection of seasons:** Some of the enumerators observed did not differentiate between seasonal crops and yearly crops. Although, IDinsight had provided information on yearly and seasonal crops (for example, red gram, cotton, chilies and sugarcane are yearly crops), enumerators were not sure about this, leading them to select these crops in both kharif and rabi season. This has may have led to double entry of data for these crops.

**Converting bags into quintals:** In some cases, enumerators were unclear about the conversion of harvest quantities from 'bags' to 'quintals'. The rough calculations they performed were often inaccurate.

#### *Comparison of Back-Check Data and Original Survey Data*

Table 20. Deviations in Key Reported Variables Relating to Agriculture.

Variable Name	Description of Variable	Median Deviation Range	Proportion of variation explained within household/ member (20% trimmed)	Proportion of variation explained within household/ member (untrimmed)
cultivate_kharif	Land cultivated during kharif	12 - 20%	0.63	0.75
cultivate_rabi	Land cultivated during rabi	40 - 100%	0.44	0.45
cultivate_summer	Amount of land cultivated during summer	100%	0.32	0.02
vc_crops_cult_1-vc_crops_cult_995	Dummy variable indicating the different crops cultivated	0 (Indicating that on average there was a median difference of 0 between the number of crops reported as per the original survey and the number reported as per the back-check survey)	-	-
exp_season_1 - exp_season_4	Dummy variable indicating the	0	-	-

	different seasons in which agriculture related expenses were reported
--	---

Table 21. Deviations in Reported Values of Agricultural Expenses.

Variable Name	Description of Variable	Median Deviation Range	Proportion of variation explained within household/ member (20% trimmed)	Proportion of variation explained within household/ member (untrimmed)
ag_exp_index_1	Expenditure on chemical fertilizers	50 - 53%	0.39	0.66
ag_exp_index_2	Expenditure on non-chemical fertilizers	50%	0.02	0.15
ag_exp_index_3	Expenditure on chemical pesticides	42% - 50%	0.54	0.51
ag_exp_index_5	Expenditure on seeds and saplings	45% - 50%	0.63	0.25
ag_exp_index_6	Expenditure on transportation	56% - 67%	0.81	0.80
ag_exp_index_8	Expenditure on maintenance of machinery	50% - 86%	0.00	0.00

Table 22. Deviations in Reported Values of Agricultural Harvest.

Variable Name	Description of Variable	Median Deviation Range	Proportion of variation explained within household/ member (20% trimmed)	Proportion of variation explained within household/ member (untrimmed)
area_crop	Area under a particular crop in a given season	9% - 11%	0.52	0.00
harvest_crop	Quantity of harvest from the area under cultivation	43% - 50%	0.70	0.99
crop_self_consume	Quantity of harvest consumed by the cultivator's household	83% - 90%	0.65	0.14
sell_crop	Dummy variable indicating whether the harvest was sold to an external buyer	18%	-	0.64



Table 23. Deviations in Reported Values of Animal Husbandry Related Variables.

Variable Name	Description of Variable	Median Deviation Range	Proportion of variation explained within household/ member (20% trimmed)	Proportion of variation explained within household/ member (untrimmed)
number_milch_cow	Number of milch cows owned by the household	0 units (Median)	0.72	0.08
number_milch_buff	Number of milch buffaloes owned by the household	0 units (Median)	0.84	0.93
number_poultry	Number of poultry, chicken, ducks owned by the household	0 units (Median)	0.57	0.20
money_milk	Revenue from selling milk and milk products	67% - 68%	0.59	0.64

## Section 4: Income

### Field Observations on Data Quality

**Selecting the Industry of occupation:** In several instances, enumerators did not know which industry was to be selected for construction work and selected agricultural labor. For other occupations as well, they have sometimes erroneously selected agriculture as the industry.

**Calculating the total amount received from this occupation:** Some enumerators were unclear about the reference period for the number of days worked and the amount of wages received from the above-mentioned industry. They have sometimes entered the daily wage instead of total amount received from each industry.

**Entering hand loans as remittance:** Some enumerators have captured hand loans (informal inter-personal loans) as remittances.

**Most of the Income section skipped due to the wrong selection of occupation:** In several cases, large parts of the income section were entirely skipped as a result of wrong or incomplete entry of occupation related data in the section on household roster.

### Comparison of Back-Check Data and Original Survey Data

Table 24. Deviations in Reported values of Wage and Salary Related Variables.

Variable Name	Description of Variable	Median Deviation Range	Proportion of variation explained within household/ member (20% trimmed)	Proportion of variation explained within household/ member (untrimmed)
ws_worked_days	Number of days worked by the member for wage or salary	33%	0.26	0.27

ws_worked_hours	Number of hours per day worked by the member for wage or salary	11 - 16%	0.15	0.17
ws_worked_rate	Total annual receipts of the member from wage or salary employment	40% - 50%	0.61	0.72
nrega_work_days	Number of days worked by the member under NREGA scheme	17% - 20%	0.46	0.45
nrega_paid_days	Number of days of payment received by the member under NREGA scheme	20% - 25%	0.53	0.52
nrega_payment	Total payment received by the worker under NREGA scheme	32% - 41%	0.67	0.48

## Section-6: WASH

### Field Observations on Data Quality

**Selection of proper respondent:** In many cases, the observed enumerators asked the questions in this section to the male respondent instead of a female respondent. Even when female respondents were present at the time of survey, some of the observed enumerators asked this section to the male respondent.

**Type of latrine in the household:** Some of the observed enumerators selected 'septic tank' for the type of latrine even though households did not have a septic tank toilet. Later on during data collection, this issue has been resolved by explaining the difference between septic tank and other types of toilets. IDinsight had originally provided pictures of the types of toilets during training, but enumerators had not followed the definition provided.

**Primary drinking water source:** In some instances, enumerators were not clear on the definition of the primary drinking water source. They asked generally about the water source used by households rather than about the source used for drinking water specifically.

**Cleaning of garbage:** Some of the observed enumerators assumed that the cleaning of garbage related to the cleaning of the house and premises rather than collection of garbage from the home by the panchayath, municipality or any of the household members.

### Comparison of Back-Check Data and Original Survey Data

Table 25. Deviations in Reported Values of WASH Related Variables.

Variable Name	Description of Variable	Mean Deviation Range	Proportion of variation explained within household
hh_pvt_latrine	Whether the household has a private latrine	8%	0.83

lat_storage	Whether the latrine is used for storing other material	15%	0.00
lat_water	Whether the latrine has access to water	25%	0.18
lat_slippers	Whether there are slippers outside the latrine	35%	0.24
lat_supplies	Whether the latrine has sanitation supplies like soap etc.	34%	0.12
lat_used	Whether, in the opinion of the enumerator, the latrine is being used	8%	0.28

Note: Trimmed values are not reported since the above variables are dummy variables taking values 1 or 0.

## Section 7: Consumption

### Field Observations on Data Quality

**Selection of proper respondent:** In many of the observed cases, enumerators asked the questions on consumption to the male respondent instead of female respondents. They had been informed during training that this section should be done with female respondents.

**Asking the questions in a wrong way:** Some of the observed enumerators asked consumption questions in a general way, not specifically asking for the consumption *in the last 30 days* and the price *of consumed products*.

### Comparison of Back-Check Data and Original Survey Data

Table 26. Deviations in Reported Values of Consumption Related Variables.

Variable Name	Description of Variable	Median Deviation Range	Proportion of variation explained within household (untrimmed)	Proportion of variation explained within household (20% trimmed)
palm oil	Value of palm oil purchased from the market in the last 30 days	41-46%	0.17	0.50
pulses and pulse products	Value of pulses and pulse products purchased from the market in the last 30 days	56%-60%	0.00	0.27
rice	Value of rice purchased from the market in the last 30 days	37-50%	0.16	0.58
sugar	Value of sugar purchased from the market in the last 30 days	50-53%	0.27	0.39

wheat/atta	Value of wheat/atta purchased from the market in the last 30 days	33%	0.07	0.56
chicken (including other birds)	Value of chicken purchased from the market in the last 30 days	65%	0.09	0.21
edible oils and vanaspati (kgs)	Value of edible oil purchased from the market in the last 30 days	100%	0.07	0.07
eggs (units)	Value of eggs purchased from the market in the last 30 days	100%	0.23	0.32
fish (including prawns etc.) (kgs)	Value of fish purchased from the market in the last 30 days	100%	0.15	0.23
meat(goat/ mutton, beef/buffalo et.c)	Value of meat purchased from the market in the last 30 days	50%	0.53	0.57
bajra and its products	Value of bajra and its products purchased from the market in the last 30 days	80% - 86%	0.009	0.66
jowar and its products	Value of jowar and its products purchased from the market in the last 30 days	100%	0.08	0.37
maize and its products	Value of maize and its products purchased from the market in the last 30 days	100%	0.00	-
other rice products (ruce flour etc.)	Value of other rice products purchased from the market in the last 30 days	100%	0.05	0.00
other wheat products (maida etc.)	Value of other wheat products purchased from the market in the last 30 days	100%	0.046	0.13

Table 27. Deviations in Reported Values of Consumption Related Variables (30 Day Reference Period).

Variable Name	Description of Variable	Median Deviation (as percentage of baseline value)	Proportion of variation explained within household (untrimmed)	Proportion of variation explained within household (20% trimmed)
Dietary diversity index	Index created by aggregating the responses to consumption of food group	37.5%		0.24
m_cons_service_value	Expenditure on consumption of services in the last 30 days	74%	0.03	0.55
m_consumables_other	Expenditure on other consumables in the last 30 days	100%	0.11	0.26
m_conveyance_value	Expenditure on conveyance and transport in the last 30 days	92%	0.38	0.25
m_entertain_value	Expenditure on entertainment in the last 30 days	20%	0.00	0.57
m_fuels_value	Expenditure on fuel in the last 30 days	69%	0.04	0.45
m_medical_value	Expenditure on medical services in the last 30 days	81%	0.04	0.35
m_phone_value	Expenditure on telephone in the last 30 days	50%	0.42	0.34
m_rent_value	Expenditure on rent in the last 30 days	50%	0.23	-
m_semidurable_value	Expenditure on semi-durables in the last 30 days	50%	0.34	0.37
m_toiletries_value	Expenditure on toiletries in the last 30 days	39%	0.49	0.38
m_water	Expenditure on water in the last 30 days	100%	0.02	0.37

Table 28. Deviations in Reported Values of Consumption Related Variables with Yearly Reference Period.

Variable Name	Description of Variable	Median deviation (as percentage of baseline value)	Proportion of variation explained within household (untrimmed)	Proportion of variation explained within household (trimmed 10-90 percentiles)
y_bedding	Annual expenditure on bedding by the household	100%	0.61	0.26
y_clothing	Annual expenditure on clothing by the household	50%	0.15	0.48
y_cooking_appliances	Annual expenditure on cooking appliances by the household	100%	0.54	0.00
y_durables_other	Annual expenditure on other durables by the household	100%	0.00	0.00
y_edu_materials	Annual expenditure on education materials by the household	60%	0.67	0.59
y_education	Annual expenditure on education by the household	55%	0.42	0.61
y_functions	Annual expenditure on functions by the household	66%	0.00	0.32
y_furniture_fixtures	Annual expenditure on furniture fixtures by the household	100%	0.04	-
y_insurance	Annual expenditure on insurance by the household	33%	0.86	0.54
y_medical	Annual expenditure on medical by the household	95%	0.72	0.41
y_ornaments	Annual expenditure on ornaments by the household	100%	0.71	-
y_personal_good	Annual expenditure on personal goods by the household	100%	0.00	0.00
y_recreation	Annual expenditure on recreation by the household	100%	0.44	-

y_repair	Annual expenditure on repair by the household	100%	0.46	0.15
y_shoes	Annual expenditure on shoes by the household	60%	0.45	0.35
y_therapy	Annual expenditure on therapy by the household	100%	0.00	-
y_transport	Annual expenditure on transport by the household	100%	0.02	0.19
y_utensils	Annual expenditure on utensils by the household	100%	0.00	-
y_vacation	Annual expenditure on vacation by the household	100%	0.00	-