Sampling strategy and data collection

Summary:
The data was collected in collaboration with Smart Power India (SPI), a Rockefeller Foundation Initiative, as part of a study designed to measure rural electricity demand and understand the customer perspective towards mini-grid electricity and grid electricity (served by public distribution companies or private distribution franchises (DF)).

For this purpose, research was designed to collect data from villages with interventions such as mini-grid (MG) or private distribution franchises (DF), and structurally similar non-intervention villages, with only public distribution companies. In order to sample 200 villages spread across four states, we employed purposive sampling. We started by randomly sampling 50 villages with mini-grid intervention (MG) and 50 villages being served by private distribution franchises (DF) from their respective sampling frames. Thereafter, we randomly sampled 50 villages without mini-grids, but similar to MG villages and 50 villages without DF, but similar to DF, from their respective sample frames.

Within each revenue village, we randomly sampled 50 households, proportionally distributed across the biggest hamlets in the village.

Research was designed to ensure that all villages in the sample frame were similar to mini-grid villages along two variables: village population and distance to the nearest town. Only villages within one standard deviation of the mean of these covariates for mini-grid villages were included in the sample frame. This was done to ensure that all villages are structurally comparable.

Detailed strategy:

1. Sampling strategy for villages

Overall, we sampled 200 villages spread across four energy deficit states in India, namely Bihar, Rajasthan, Odisha and Uttar Pradesh.

Our sampling frame has four categories of villages:

- *Mini-Grid Intervention (MG)* villages- These are villages where SPRD mini-grids are operational. There are 91 such villages spread across 19 districts in Uttar Pradesh and Bihar.
- Mini-grid Non-intervention (non-MG) villages – These are villages, which could offer prospective sites for SPRD mini-grids, and belong to same 19 districts as SPRD villages.
- Distribution Franchise Intervention (DF) villages – These are villages in the eight districts of Odisha, which are being served by distribution franchises.

- Distribution Franchise Non-Intervention (non-DF) villages- These are villages in select 20 districts across Bihar, Rajasthan, Odisha and Uttar Pradesh, which could be prospective sites for distribution franchises.

We sample 50 villages from each of the four categories and choose the number of villages within each of the four states to reflect the distribution of MG, non-MG, DF and non-DF villages in the sampling frame. Thus, our sample has 100 intervention and 100 non-intervention villages.[1] Detailed sampling strategy is discussed below:

A. Sampling MG villages:

From the list of 98 MG villages, we first dropped those villages, where the number of mini-grid customers is less than 10% of maximum customers across all MG sites. This left us with 91 MG villages, spread across 19 districts in UP and Bihar. From this list, we sampled 50 villages in proportion to the number of MG villages in every district, but blocking the sampling by district. This ensured that every district has at least one MG village selected in the sample.

B. Sampling non-MG villages:

For sampling non-MG villages, our sampling frame was limited to villages, which share some structural similarities with sampled MG villages. As the selection of MG villages for mini-grid installation is determined by availability of potential customers and unmet electricity demand, this criterion ensures that our sampled villages could serve as prospective sites for similar interventions in future. In order to ensure that non-MGs share structural similarities with MG villages, we use three key variables:

- Total village population, which directly determines the number of customers and electricity demand at village level.
- Distance to town, which has been found to be negatively co-related with the rural living standards (Asher, 2017).
- Presence of clustered shops (market), which influences electricity demand for productive purposes.

These variables can influence customers' demand for and attitude towards electricity, and are also exogenous variables. We used Census 2011 data for the 50 selected MG villages and performed a matching exercise to select 50 non-MG villages, which are also spread across districts having MG villages.

We use the first two variables to create a sampling frame of villages for selecting non-MGs. From a list of total villages in our sample districts, we first excluded all villages that do not

[1] Randomization inference literature has shown that uncertainty in small samples decreases dramatically, when the sample size is more than 100 (Imbens & Rubin, 2015).

fall within one standard deviation of the mean of population and distance-to-town of MG villages. We obtained a list of 6314 villages as our sampling frame, which have i) population in the range of 2,000-12,000, and ii) distance to nearest town ranging between 5 and 36 km. From this list, we randomly sampled 50 non-MG villages, also blocked by district and distributed in same proportion as MG villages across sample districts

However, we also need to ensure that our sampled villages have enough enterprises (shops). But as this data is not easily available, this could only be ascertained through field verification. So, we created a waiting list of non-MG villages, to allow replacement of villages which do not have at least one market area, with a minimum of 20 enterprises/shops.

C. Sampling DF villages:

The sampling for these villages was done in manner similar to that of non-MG villages. Eight districts in Odisha are currently being served by Distribution Franchises (see Appendix B). Using the Census 2011 data for villages falling under these districts, we first created a sampling frame of villages, which have i) population in the range of 2,000-12,000, and ii) distance to nearest town ranging between 5 and 36 km. From this sampling frame, villages, we randomly sampled 50 DF villages, also blocked by district and distributed in proportion to the number of villages in sample frame per district. In order to ensure that our sampled villages have least one market area, with a minimum of 20 enterprises/shops, we created a waiting list of DF villages, to allow replacement of villages which do not meet this criterion.

D. Sampling non-DF villages

20 districts across Bihar, Rajasthan, Odisha and Uttar Pradesh, were identified as prospective districts for operation of DFs in future (see Appendix B). We sampled 50 non-DFs from these districts in manner similar to that of DFs. Using the Census 2011 data for villages falling under these districts, we first created a sampling frame of villages, which have i) population in the range of 2,000-10,000, and ii) distance to town ranging between 5 and 36 km. From this sampling frame, we randomly sampled 50 DF villages, also blocked by district and distributed in proportion to the number of villages in sample frame per district. In order to ensure that our sampled villages have least one market area, with a minimum of 20 enterprises/shops, we created a waiting list of non-DFs, to allow replacement of villages which do not meet this criterion.

Overall, we sampled 100 villages (MG, non-MG) in 19 districts of Uttar Pradesh and Bihar, and 100 villages (DF and non-DF) in 28 districts of Bihar, Rajasthan, Odisha and Uttar Pradesh. Our sampled villages are distributed across states as shown in table below.

State	Category of village				
	MG	non-MG	DF	non-DF	Grand Total
Bihar	11	11	0	6	28

Uttar Pradesh	39	39	0	17	95
Odisha	0	0	50	12	62
Rajasthan	0	0	0	15	15
Total	50	50	50	50	200

2. Sampling within revenue villages

Within each revenue village, we sampled 50 households and 10 enterprises (shops).

A. Strategy adopted for sampling households:

- The survey team enters the revenue village and consults panchayat officials or the village elders to obtain information about up to six biggest hamlets in the village and number of households in each of them.
- The sample of 50 survey households is allocated to these hamlets in proportion to the total household numbers in each.
- Within each hamlet, the survey team selects a random spot (say a temple, water tank, etc.). From there, the team selects a random direction and samples every 10th household in that direction.
- If a household refuses to participate, the next neighbouring household is surveyed.

B. Strategy adopted for sampling enterprises (shops):

There exist a variety of commercial entities in villages, such as retail shops (grocery, fruits/ juice, etc.), service shops (photocopying, cyber café, mobile repairing, etc.) and commercial processing entities (flour mills, etc.). In order to ensure a representative sampling, while capturing different types of shops, we devised an innovative strategy for sampling enterprises within villages.

- The survey team first conducts a mapping exercise to count different types of shops available in the village, categorised into three broad groups. In villages having more than 300 shops, instead of full listing of shops, every 10th shop was counted to do the mapping exercise.
- Using this information, the sample of 10 survey shops is allocated to each of the three groups in proportion of total shops encountered within that each group.
- The team then selects a random spot in the market and choose a random direction, in which every 10th shop is identified and assessed for sampling. The shop is surveyed, if it belongs to the group, in which the number of shops to be surveyed is more than the shops yet surveyed from the group. (see Appendix C for details and an example.)
- If a shop ownerrefuses to participate in the survey, then the neighbouring shop is assessed for sampling.

3. Survey tools

Separate questionnaires were developed for household and enterprise surveys, and were designed to collect detailed data related to:

i) Social, economic and demographic profile of the households and enterprises, along with psychographic profiling of the respondents.

ii) Ownership and use of different sources of electricity and lighting; expenditures on and satisfaction with sources available.

iii) Appliance ownership, their power ratings, number of hours for which each appliance is used, and time of the day for appliance use.

iv) Awareness, attitude and preferences towards alternative electricity sources, particularly mini-grid electricity.

A week long field visit comprising 20 qualitative interviews was conducted in Hardoi district of Uttar Pradesh, to inform the initial drafts of the questionnaire, which then underwent several sounds of revisions. The questionnaires were then programmed to facilitate survey through tablets. The questionnaires were also piloted in the Saran district of Bihar, to ascertain the ease of administration, average time taken, nature of responses and gaps in data coding. Based on the pilot learnings, the questionnaires as well as their digital versions were revised.

4. Survey

We fielded a 40 minute long surveys with detailed questions on household and enterprise characteristics, awareness and attitude towards different electricity and lighting sources, appliance ownership and usage patterns, and willingness to pay for different electricity packages.

The survey was conducted in the local languages (Hindi in the states of Bihar, Rajasthan and Uttar Pradesh, and Odiya in Odisha), by an experienced survey company Morsel India. Two survey teams comprising 35 enumerators for Hindi speaking states and 20 enumerators for Odisha, were trained by one of researchers along with project managers for four days each. The training involved classroom training on basics of electricity sources and appliances and survey questions, role plays and dry runs in nearby pilot villages.

Appendix A: List of districts for sampling MG and non-MG villages

State	Districts for MG villages	Districts for non-MG villages
Bihar	Araria	Araria
	East Champaran	East Champaran
	Gaya	Gaya
	Gopalganj	Gopalganj
	Saran	Saran
	Siwan	Siwan
	Supaul	Supaul
	West Champaran	West Champaran
Uttar Pradesh	Balrampur	Balrampur
	Farrukhabad	Farrukhabad
	Hardoi	Hardoi
	Kushinagar	Kushinagar
	Lakhimpur Kheri	Lakhimpur Kheri
	Pratapgarh	Pratapgarh
	Rae Bareli	Rae Bareli
	Shajahanpur	Shajahanpur
	Shrawasti	Shrawasti
	Sitapur	Sitapur
	Unnao	Unnao

Appendix B: List of districts for sampling DF and non-DF villages

State	Districts for DF villages
Odisha	Puri
	Nayagarh
	Khorda
	Cuttak
	Jagatsinghpur
	Kendrapara
	Dhenkanal
	Anugul

State	Districts for non-DF villages
Rajasthan	Kota
	Ajmer
	Baran
	Bikaner
	Dhaulpur
	Jhalawar
	Karauli
	Sawai Madhopur
Bihar	Munger
	Rohtas
Odisha	Ganjam
	Balangir
	Deogarh
	Kalahandi
	Kandhamal
	Sambalpur
Uttar Pradesh	Agra
	Azamgarh
	Jhansi

	Mirzapur

Appendix C: Sampling procedure for enterprises

	Type of shops	Count	Group-wise sample size	Sample allocated
1	Grocery shop (kirana store, cosmetics, etc.)	70		6
2	Fruit/juice shop	10		
3	Sweet shops	20		
4	Readymade clothes/bags shops	4		
5	Hardware shop	3		
6	Agricultural inputs (public/private)	2		
7	Medical store	8		
	SUB TOTAL (A)	**117**	117*10/188 = 6.22	
8	Mobile repair shops	7		3
9	Cyber café (photocopier/ printer/ scanner)	10		
10	Photo-studio	5		
11	Coaching centre	3		
12	Tailoring shop	9		
13	Beauty Parlour	1		
14	Carpentry shop	7		
15	Cycle, handcart repair centre	5		
	SUB TOTAL (B)	**47**	47*10/188 = 2.5	
16	Car/ bike service centre	4		
17	Private Clinic/ Nursing home/ Veterinary clinic	6		

18	Flour or oil extraction mill	9		
19	Godowns / warehouse	2		
20	Cold storage	1		
21	Dairy / Milk Chilling Centre	2		1
	SUB TOTAL (C)	24	24*10/188 = 1.3	
	Grand Total (D= A+B+C)	188		

1. Based on enterprise mapping exercise, obtain the total number of each type of shop.
2. Count total shops in each group. Also calculate total shops (D).
3. Calculate sample size for each group, as shown:
    a. Sa = A*10/D
    b. Sb = B*10/D
    c. Sc = C*10/D
4. Round off the numbers.

Example:

Sampling protocol:

1. Identify a public place to start the sampling.
2. Start from your right and choose the very first shop.
3. Now identify the 10th shop from here.
4. Check which group the shop belongs to. If the number of shops surveyed from this group are less than the number of shops to be surveyed, then survey this shop. Otherwse, skip this shop, and assess whether the next shop can be surveyed.

E.g. if first shop belongs to group C and you are supposed to survey only 1 shop from group C, then if the next 10th shop is from Group C, then you skip it. Go to next shop and check if it belongs to the group, from which you still need to survey more shops.

Now, if the next shop is say a grocery store (group A), then survey it. If the next 10th shop from here is also a grocery shop (Group A), then survey it. Do this, till you have surveyed 6 shops from group A. Similarly, survey 3 shops from group B.