Documentation: COVID-19 Rural Impacts Survey

Abstract

An effective policy response to the economic impacts of COVID-19 pandemic requires an enormous range of data to inform the design and response of programs. Public health measures require data on the spread of the disease, beliefs in the population, and capacity of the health system. Relief efforts depend on an understanding of hardships being faced by various segments of the population. Food policy requires measurement of agricultural production and hunger. In such a rapidly evolving pandemic, these data must be collected at a high frequency. Given the unexpected nature of the shock and urgency with which a response was required, Indian policymakers needed to formulate policies affecting India's 1.4 billion people, without the detailed evidence required to construct effective programs. To help overcome this evidence gap, the World Bank, IDinsight, the Development Data Lab and John Hopkins University sought to produce rigorous and responsive data for policymakers across six states in India: Jharkhand, Rajasthan, Uttar Pradesh, Andhra Pradesh, Bihar, and Madhya Pradesh.

Sampling

This dataset includes observations covering six states (AP, Bihar, Jharkhand, MP, Rajasthan, UP) and three survey rounds. We did not have a single, unified frame from which to sample phone numbers. The final sample was assembled from several different sample frames, and the choice of frame sample frames varied across states and survey rounds. These frames comprise four prior IDinsight projects, and from an impact evaluation of the National Rural Livelihoods project conducted by the Ministry of Rural Development Each of these surveys sought to represent distinct populations, and employed idiosyncratic sample designs and weighting schemes. Key features of each sample frame are summarised below:

- **Poshan Abhiyan Monitoring (PAM)**
    - **Description:** Two-stage stratified cluster random sample.
    - **Sample size (with phones/total):**
    - **Representativeness:** full rural populations of Rajasthan & Jharkhand
    - **Time frame:** drawn in winter 2019-2020
    - **Sample frame:** voter rolls
    - **Stratified:** yes, at the AC level
    - **Clustering:** at the polling station (primary sampling unit) level
    - **Weights:** probability weights defined for original sample

- **Poshan Abhiyan SBCC Monitoring (SBCC)**
    - **Description:** Two-stage stratified cluster random sample.

    - **Sample size (with phones/total):**
    - **Representativeness:** representative at the state level of all households with pregnant or lactating mother listed by an ASHA/AWW
    - **Time frame:** drawn in fall 2018
    - **Sample frame:** Frontline workers' (ASHA & AWW) registries
    - **Clustered:** at the village/ward (primary sample unit) level
    - **Weights:** probability weights defined for the original sample

- **State of Aadhaar Report (SOAR)**
    - **Description:** Three-stage stratified cluster random sample.
    - **Sample size (with phones/total):** AP (456/706)
    - **Representativeness:** Representative of AP at the state level
    - **Sample frame:** Voter rolls
    - **Clustered:** at the polling station (primary sampling unit) level
    - **Weights:** probability weights defined for the original sample

- **National Rural Livelihoods Programme (NRLP)**
    - **Description:** Sample was collected by the World Bank and covers 9 states that were part of the National Rural Livelihoods Programme.
    - **Sample size (with phones/total):** Bihar (5,005/5,872), MP (2,621/2,877), UP (2,175/2400)
    - **Representativeness:** Representative of SHG membership in states. Rural districts selected from strata to reflect a range of outcomes.
    - **Sample frame:** Village listings
    - **Clustered:** at the village (primary sampling unit) level
    - **Weights:** not defined in the original survey

- **Soil Health Card (SHC) (Used for round 1 only)**
    - **Description:** Within UP, we selected 4 districts that represent literacy and mobile phone penetration heterogeneity, as well as agro climatic variation. We then selected 24 villages from each district, and approximately 20 farmers per village. The farmers were selected by a random walk using WHO Extended Programme on Immunization (EPI) protocol.
    - **Representativeness:** Representative of farmers in UP
    - **Sample frame:** SHC database, random walk of an area frame
    - **Stratified:** N/A
    - **Clustered:** at the village (primary sampling unit) level
    - **Weights:** undefined (final stage was not a probability sample), but assumed to be approximately self-weighting.

For some state/rounds we randomly selected a subset of households from the frame of phone-owning households. Based on prior experiences we anticipated a 60% response rate. We took a subsample in those instances where we expected applying the 60% response rate to the

sample would yield an adequate sample size. For instance, in round 2 we restricted the sample to 1500 households in Jharkhand and Rajasthan, which should have yielded approximately 900 completed surveys.

In other cases we did not sample from phone owning households because the size of the frame of phone owners was small. For instance, in all rounds for both frames in AP we included every available phone number.

Table 1 gives the number of phone interviews completed/attempted for every state and round, and mentions the sample frame used for that state/round combination.

Response Rates

Table 1: ***Phone interviews completed/attempted by state/round/source data***

	AP		Bihar	Jharkhand	MP	Rajasthan	UP
Round 1	280/565 (50%)	201/461 (44%)	207/447 (47%)	1,174/2,014 (58%)	214/443 (48%)	1,790/2,941 (61%)	710/1450 (63%)
Frame	SBCC	SOAR	SBCC	CIFF	SBCC	CIFF	SHC
Round 2	281/549	230/456	1,073/2,658	890/1500	823/2,339	930/1500	778/1,899
Frame	SBCC	SOAR	NRLP	CIFF	NRLP	CIFF	NRLP
Round 3	217/517	178/430	1,030/2,658	1,078/2,756	944,2,337	995/2,003	757/1,899
Frame	SBCC	SOAR	NRLP	CIFF	NRLP	CIFF	NRLP

Weights

In order to create comparable state-level estimates from the successfully interviewed households- as well as to create correctly pooled estimates across the six states- we have applied weights to the information provided by the sampled households.

The weights were calculated in several steps. Due to the variation in sampling frames and sampling procedures across states and across rounds, the precise weight procedures tend to be idiosyncratic to a given state/frame/round combination. The procedure detailed below gives a generalized set of steps, and notes significant state/frame deviations from the process.

1. *Rescale base weights*: Base weights reflect a probability of selection into the original sample, and can be interpreted as an expansion factor to some population. In cases where the sample frame does not include a base weight or a means of calculating a

probability of selection equal weights are assigned to each unit in the frame. State-wise details follow:

a. *Bihar*: The sample frame used in round 1 provides a base weight. The sample frame used for rounds 2 and 3 do not provide a base weight, nor does it provide a means of estimating probability of selection.
b. *Madhya Pradesh*: The sample frame used in round 1 provides a base weight. The sample frame used for rounds 2 and 3 do not provide a base weight, nor does it provide a means of estimating probability of selection.
c. *Uttar Pradesh*: The sample frame used in rounds 1, 2, and 3 do not provide a base weight, nor does it provide a means of estimating probability of selection.
d. *Andhra Pradesh*: The source datasets provide base weights for all three rounds, which expand to the state population (SOAR) or the population of households listed on ASHA/AWW rosters (SBCC).
e. *Jharkhand, Rajasthan*: The source dataset provides base weights which expand to full district populations. These districts were chosen with PPS from geographic strata to represent the rural population of the two states.

2. *Apply an adjustment for noncoverage (i.e. selection into the frame):* Next, create weighting classes within districts based on socioeconomic covariates known from respective sample frames. Within these weighting classes sum the base weights, and divide this into the sum of the weights for "covered" households (i.e. those with a mobile phone). Next, multiply the base weights from by the inverse of these proportions. State-wise details follow:
    a. *Bihar, Madhya Pradesh, Uttar Pradesh*: We define the weighting classes by two categorical variables from the NRLP data: caste (SCST/OBC/General) and household income (five quintiles).
    b. *Andhra Pradesh*: We define weight classes by caste (SCST/OBC/General), and PPI quintile (if the household is from the SBCC sample) or ration card status (if the household is from the SOAR sample).
    c. *Jharkhand, Rajasthan*: Unfortunately weighting class adjustments are not possible in these states due to insufficient covariate data.

3. *Apply an adjustment for selection into the sample:* In certain states (Bihar, Jharkhand, Rajasthan) we randomly selected a subset of the covered households. In these cases we apply an adjustment factor to the weight from (1.b.) to reflect each unit's probability of probability of selection. In other states (AP, MP, UP) we never

4. Rescale the weights from step 3 to account for nonresponse

a. **Nonresponse:** The correction for nonresponse also employs weighting classes, and follows the steps from 2 exactly. Nonresponse adjustments are applied to the output of step 3.

5. Calibrate the weights from step 4
    a. **Process:** *Poststratification* attempts to correct for known differences between a sample and a target population. The process entails adjusting the sample weights such that their sum within each poststratum equals the known population total for that poststratum. Specifically, we use the *raking to margins* method, which recursively recalibrates weights to marginal totals of the chosen poststratification covariates until the weights converge.
    b. **Data:** We define poststrata for SCST status and religion using population totals (household level) for caste and religion from the 2011 population census.

Data collection

- **Dates of data collection**

Round	Sample size	Start date	End date
1	4,576	05/05/20	10/05/20
2	5,006	19/07/20	23/07/20
3	5,200	20/09/20	24/09/20

- **Instrument:** The survey questionnaires covered the following subjects:

1. ***Agriculture:*** COVID-19-related changes in price realisation, acreage decisions, input expenditure, access to credit, access to fertilisers, etc.

2. ***Income and consumption:*** Changes in wage rates, employment duration, consumption expenditure, prices of essential commodities, status of food security etc.

3. ***Migration:*** Rates of in-migration, migrant income and employment status, return migration plans etc.

4. ***Access to relief:*** Access to in-kind, cash and workfare relief, quantities of relief received, and constraints on the access to relief.

5. ***Health:*** Access to health facilities and rates of foregone healthcare, knowledge of COVID-19 related symptoms and protective behaviours.

While a number of indicators were consistent across all three rounds, questions were added and removed as and when necessary to account for seasonal changes (i.e: in the agricultural cycle).

Data processing

The India COVID-19 surveys were conducted using Computer Assisted Telephone Interview (CATI) techniques. The household questionnaire was implemented using the CATI software, SurveyCTO. The software was deployed through surveyors' smartphones, who called respondents via mobile, and recorded their responses over the phone. If unreached, surveyors would attempt to call back respondents up to 7 times, often seeking explicit appointments for suitable times to avoid non-responses.

Validation and consistency checks were incorporated into the SurveyCTO software to avoid human error. Extreme values and outliers were scrutinised through a real time dashboard set up by IDinsight. Surveys were also audio audited by monitors to check for consistency and accuracy of question phrasing and answer recording. Finally, supervisors also randomly back-checked a subset of interviews to further ensure data accuracy.

IDinsight cleaned and labelled the data for further processing and analysis. Johns Hopkins University examined the data for discrepancies and errors and merged the dataset with their proprietary spatial data.

All personally identifiable information has been removed from the datasets.