COVID-19 Shocks in Rural India (Round 1)

Sampling Frame Description

This document outlines the various sampling frames that were used for the survey conducted by the World Bank in collaboration with IDinsight and Development Data Lab (DDL) and the World Bank.

Overall Summary

STATE	ROUNDS[1]	SAMPLE SIZE	FRAME
Jharkhand	3	2,014	CIFF
Rajasthan	3	2,942	CIFF
Uttar Pradesh	3	1,450	SHC
Andhra Pradesh	3	1,026	SOAR, SBCC
Bihar	3	448	SBCC
Madhya Pradesh	3	443	SBCC
TOTAL	**3**	**8,323**	

Description of Frames

1. Aadhar Report Sample (SOAR)

- ***Date:*** January 2018
- ***Level of representation:*** State[2]
- ***Demographic data available:*** Household roster (including education), religion, caste category, poverty line status
- ***States covered:*** Andhra Pradesh
- ***Sampling methodology:***
    - We used the Election Commission's public, online voter ID database as our sampling frame for a multi-stage cluster sampling design- first sampling districts, then assembly constituencies (ACs) and then polling stations (PSs) before finally selecting households. The number of units sampled at each level was determined by minimizing the variances of our estimates subject to time and budget constraints. To calculate the intracluster correlation coefficients needed for estimating the variances

[1] Over 4 months

[2] Rural areas only

in outcomes (which we selected to be access to photo ID and participation in NREGA), we used data from the Indian Human Development Survey).
    - We used probability proportional to size (PPS) sampling without replacement levels to increase the precision of our estimates (where size is defined as the number of voters in each unit according to the voter rolls). We dropped all urban Assembly Constituencies (ACs) from our sample before beginning the selection process to ensure our sample population was representative of rural populations.
    - We selected 6-8 districts within each state using PPS, stratified by region. Within each district, 3 (rural) ACs were selected using PPS. Within each AC, we selected 2-3 polling stations using PPS.
    - After using PPS to select the Districts, ACs, and PSs, we downloaded the voter rolls for selected polling stations from the official website of the Chief Electoral Officer of each state. Within each polling station, simple random sampling was used to select 20 households and one respondent
- ***Caveats:***
    - The survey covers rural assembly constituencies, but not necessarily agrarian households. Thus, we could be underpowered for agriculture indicators.
    - The survey was conducted in late 2017-early 2018. Thus, a significant proportion of respondents' phone numbers may have become inaccurate.

2. Health and Nutrition Messaging Sample (SBCC)

- ***Date:*** September 2019
- ***Level of representation:*** State
- ***Demographic data available:*** Literacy, asset ownership, religion, caste
- ***States covered:*** Bihar, Andhra Pradesh, Madhya Pradesh
- ***Sampling methodology:***
    - The final sample is representative of pregnant women and mothers of children 1-24 months who are registered on ASHA or Anganwadi lists in each state.
    - A stratified random sample of 200 villages/wards per state
    - The characteristics defining this population may differ from the general population of women targeted by SBCC (which include those not registered with ASHA or AWW). Compared to the general population (NFHS), the population in this study includes a lesser proportion of first-trimester pregnant women and a higher proportion of third-trimester pregnant women. This population also has higher TV ownership and higher completion of primary and secondary education. As such, some of our point estimates may differ from the general population - this is particularly true in dense
      urban areas, where registration with frontline workers is less common. Nevertheless, FHW-registered women represent a large and policy-relevant group: 70 per cent of pregnancies across the four states are registered with a frontline health worker, according to NFHS 2015-16. As such, their experiences with nutrition messaging may serve as a useful benchmark for understanding the reach of these messages.
- ***Caveats:***

- Exclusively covers households with pregnant/lactating women- thus, not necessarily agrarian
- Phone numbers are of the mother, who may or may not be the agricultural decision-maker or head of household

3. Agriculture RCT Sample (SHC)

- ***Date:*** March 2019
- ***Level of representation:*** State
- ***Demographic data available:*** Religion, caste, poverty line status, literacy, agronomic characteristics (cropping pattern, area cultivated etc)
- ***States covered:*** Uttar Pradesh
- ***Sampling methodology:***
  - We selected four districts to be representative of the 71 districts in UP. While there are several dimensions of demographic, socioeconomic, and agronomic heterogeneity within UP, we considered relative literacy rates to be the most relevant to the intervention and prognostic of outcomes. In addition, we prioritized agro-economic diversity. Therefore we stratified the list of districts by literacy quartile, then randomly selected one district from each bin (redrawing if a selected district's agro-economic zone had already been represented).
  - Data on the villages in the sampled districts where cards had been distributed in the 2017-19 cycle of the scheme was collected from the Soil Health Card website. We filtered this list to only those villages with enough farmers to meet the sample size requirement or 60 farmers.
  - Within the sampled villages, due to the lack of comprehensive household lists, we used the WHOs Extended Programme on Immunization (EPI) strategy to sample households. Under this strategy, the field team started at the centre of the village. From there, each enumerator was randomly assigned a unique street. He/she walked along that street till the edge of the village while counting the number of households on the route. A random household on this route was then selected as the starting point, and every *n*th household after the first house was sampled. A household was considered eligible for our sample if anyone residing there takes decisions about the inputs that are to be applied to a plot of land where crops are cyclically grown.
- ***Caveats:***
  - Exclusively covers households that take agricultural decisions (i.e- people who own or rent land for cultivation)

4. Health and Nutrition Convergence Sample (CIFF)

- ***Date:*** January 2020
- ***Level of representation:*** State

- ***Demographic data available (for the "eligible" subset only):*** Household roster (including occupation and education) , caste, religion, assets
- ***States covered:*** Rajasthan & Jharkhand
- ***Sampling methodology:***
    - The sample covers Rajasthan (seven districts) and Jharkhand (five districts)
    - District selection was stratified by geographical region and Poshan Abhiyaan implementation status (a proxy for development status), and is intended to be state representative.
    - Voter rolls are the sample frame. For both states, rolls were updated in late 2019. Phone numbers were collected in January/February 2020.
    - Within districts, polling stations are the primary sampling unit and were selected with probability proportional to size (PPS). Within polling stations, 15 households were selected by simple random sampling from the voter roll.
    - IDinsight collected a very abbreviated roster to determine eligibility for the Poshan Abhiyaan study (screened for presence of a pregnant woman or a child under 36 months). The percentage enrolled varied by district from approximately 25-30%
- ***Caveats:***
    - The sample includes full roster data *only for the eligible/enrolled households*. For the unenrolled households, we only have an abbreviated roster that includes gender and age. Further, we did not collect demographic data (caste, religion, assets) on the unenrolled households.
    - The implications of the missing demographic data are twofold:
        - It will make weighting adjustments for nonresponse more difficult since we will not know the characteristics of nonrespondents. However, post stratification weights are still viable.
        - A phone survey would have to include at least a few demographic questions.
    - Nevertheless, the nice properties of this set (e.g. the recency of the frame, the freshness of the phone numbers, the size and representativeness of the sample) outweigh the limitations.