Sampling Design — Beneficiary Survey

Introduction

We propose to conduct a survey on two government schemes, *Ammavodi* and *YSR Cheyutha*. Our focus is to provide detailed feedback and analysis of **process, perception,** and **impact** of these two government schemes.

All told, we aim to sample 825 secretariats, with 20 respondents per secretariat, for a total population of 16,500. We plan for a 10% rate of attrition for a total target sample size of 15,000.

Population of Interest

In order to understand our populations of interest, it is helpful to review some of our previous work with the government of Andhra Pradesh. There is a major design difference between the two schemes (Ammavodi and YSR Cheyutha) in that Ammavodi is directly calculated from government data (and enrollment data shared by schools), while YSR Cheyutha requires the potential beneficiary to file an application.

Our previous work showed two key points:

- The so-called "six step parameter verification" to determine program inclusion admits upwards of 90% of all households in Andhra Pradesh (due to the permissive nature of the criteria and low quality government data for certain parameters).

- The biggest barrier to scheme inclusion in YSR Cheyutha is the failure to file an application. More than 10% of eligible households fail to file an application, although this has gotten much better than last year. In order to make these calculations, we showed that beneficiaries of YSR Cheyutha can be selected with a near certainty from existing government data (without application)

Using these considerations, we determined that it would be best to explicitly sample 4 key populations in each secretariat:

1. **Ammavodi Beneficiaries.** We consider beneficiaries of Ammavodi meeting the eligibility criteria. One of the major issues we flagged with the government of Andhra Pradesh is a mismatch between the unique ID (Aadhar) of the beneficiary and the stated criteria of the scheme, plausibly leading to significant leakage. Because we seek to speak to mothers of school-going children, we have (conservatively) restricted the sample to females who benefitted from the scheme and are between the ages of 21 to 55 according to GSWS data. This population forms 91% of those who were paid under Ammavodi.

2. **Always Filed Application (YSR Cheyutha).** This is the population that always filed YSR Cheyutha application and received money from the scheme. This is group that has been paid each time it is deemed eligible in the past two years (according to direct calculation from government data).

3. **Missed Application Last Year (YSR Cheyutha).** This is the population that filed an application this year (and received money from the scheme) but not last year, even though they were eligible in both years (according to direct calculation from government data).

4. **Has Not Filed Application (YSR Cheyutha).** This is the population that did not file an application this year but would have been eligible this year (according to direct calculation from government data).

Sampling Secretariats

Within each secretariat we have calculated the total number of beneficiary households across these two schemes. A district-level value is shown in the column "beneficiary_wt" (split by rural and urban segments) in the district-level file shared with this note. We have sampled each secretariat proportional to this weight, and stratified the population for sampling across space. We have planned to sample 825 secretariats with 20 respondents per secretariat. Using the beneficiary weights ("beneficiary_wt") in rural and urban Andhra Pradesh, we have calculated a target number of secretariats per district, split by urban and rural secretariats. (See "secretariats_total", "secretariats_urban", "secretariats_rural"). This effectively creates stratification by district.

Sampling Respondents

We note that because each household can receive a benefit once, each household can only be sampled once in the population. Nonetheless, we seek to sample the beneficiary for the survey. For Ammavodi, this is a mother of school-going children listed between the ages of 21 and 55, and female, in the GSWS data. For YSR Cheyutha, this is a female between the ages of 45 and 59 who is not general caste and/or is from a non-Hindu minority. (We note that, unlike Ammavodi, we see much lower rates of data mismatch and leakage for YSR Cheyutha.)

Within each household, we are able to calculate each of the categories in the population of interest. In each secretariat, we select 10 Ammavodi recipients. For the other 3 categories, which refer to YSR Cheyutha, we split the 10 respondents proportional to population in the secretariat.

For instance, if 10% of those who would be included this year did not file an application, 20% did not file an application last year (but were eligible), and the rest have always filed an application, then we would select 1 respondent from the "no application" category, 2 respondents from the "missed application last year" category, and 7 respondents from the "always filed application category." We note that each of these 3 categories is shown aggregated up to the district level, split by urban and rural population, in the shared file ("noapp", "missedlastyr", "thisyr").

Statistical Power

We have estimated (conservatively) the 95% confidence interval with a variance inflation factor of 2 for *each* scheme (Ammavodi and YSR Cheyutha). For each district, our intervals range between 7-10 percentage points. The rural and urban estimates have intervals that range between 2-3 percentage points. The (NSS) region-wise estimates also have intervals that range between 2-3 percentage points.