

Centre for Rapid Insights - Methodology

Introduction

India has built an impressive data system, from a periodic Census to a growing corpus of digitized administrative information and public data that was once only available on paper. Citizens' experiences are further captured through regular nationally representative surveys describing consumption, demography and health, and the labor force. Careful analysis of new, granular administrative or public data can provide deep insights with precision, but only survey data can provide critical information on citizens' experiences. Unfortunately, nationally representative surveys typically provide information with a significant lag owing to the time and cost of administering them. Effective data-driven policymaking is contingent upon sophisticated analysis of large, complex administrative datasets alongside up-to-date information about citizens' experiences.

The Centre for Rapid Insights (CRI) provides critical evidence-based policy research by:

1. **Harnessing Administrative and Public Data:** Triangulating real-time administrative and public data in tandem with cutting edge statistical techniques and spatial methods to provide nuanced insights.
2. **Rapid Survey Insights:** Filling lacunae in up-to-date information with a novel tool using Interactive Voice Response (IVR) technology to collect survey data rapidly and at scale.

We believe that the most valuable policy insights come from harnessing the granularity of digitized data from modern statistical techniques and combining it with insights from rapid, high frequency, large-scale surveys.

Harnessing Administrative and Public Data

The volume of administrative data and public data in India has increased rapidly, but these data rarely speak to each other, and making inferences from such data requires triangulation and resolving differences between disparate sources of data. Administrative or public data is rarely generated to answer policy or research questions; rather, it is best characterized data that is generated for the purposes of bookkeeping, monitoring or collecting descriptive information. Harnessing the power of administrative data requires developing methods to “glue together” these different sources of data with an array of statistical methods, from spatialization to small area estimation.

A simple example will clarify the challenges faced by policy researchers in India. Today, India is among the most rapidly urbanizing countries in the world, yet the last Indian Census was furnished in 2011, and even the India's vaunted National Sample Survey (NSS) is now quite out of date – to

speaking nothing of the criticisms of how urban space is defined in these sources of data. Most researchers must rely on analysis of satellite data, using either built area or night light algorithms, to capture up-to-date estimates of urbanization. Figure 1 displays the night lights and built area for the National Capital Region around Delhi. As one can see, while the night lights do a good job of picking up the core urban areas, many smaller dense settlements (seen in the built area data) are harder to observe.

But even if one were able to resolve these differences, it would still be an incomplete characterization; it is not clear how satellite imagery compares to characteristics of the underlying population (information that would be available in the Census). To make sense of this muddle, the most appropriate strategy (and one that has been adopted at CRI) is to resolve Census units with satellite data and estimate how the population has changed and urbanized since the last Census using statistical modeling. For instance, one may be able to cross reference the satellite data with the Periodic Labour Force Survey (PLFS) with rapid insights on labor market and consumption preferences to deduce granular measures of economic activity.

CRI comprises policy researchers who regularly published on aspects of government policy in India, and cooperated on policy at the state and national levels. The analytic tools used are tried and tested to get the most out of administrative and public data for evidence-based policy decisions.

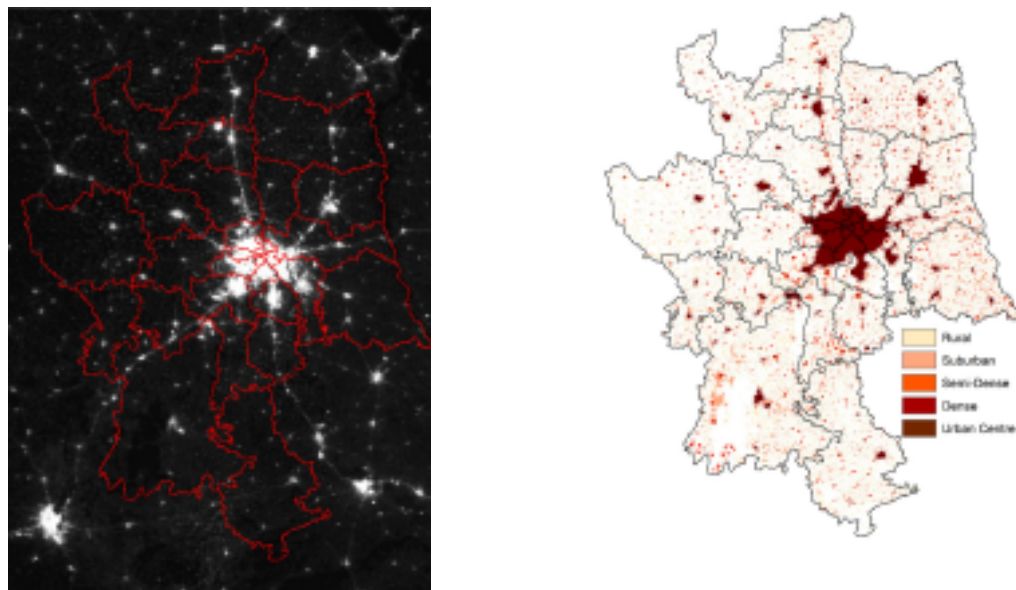


Figure 1: Comparison of Night Lights (left) and Built Area (right) in the National Capital Region

Partnerships

We have partnered with Prashnam.ai to deliver rapid, large scale survey data using Interactive Voice Response (IVR) technology. With more than 30 million individuals in our opt-in panel, the trunk dialing technology infrastructure can survey thousands, even tens of thousands, in a matter of hours.

The system also allows for surveying a focus group through a shareable database with contact details of the group. CRI is committed to maintaining the highest standards when it comes to protecting the privacy of the survey respondents. All insights are generated at an aggregate level and there is no disclosure of sensitive personal information about the respondents.

Naturally, one is concerned that such rapid surveys may generate biases. In order to achieve high quality survey estimates, we process the data through a statistical technique known as multilevel regression and poststratification (MRP), in which we use statistical models to estimate survey responses of subgroups in the population and aggregate these estimates according to the size of the subgroup. This method of analysis allows us fine-grained claims about demographic groups (e.g., women under the age of 30), while making representative claims about the population that is being studied.

This allows us to provide rapid, rigorous, large scale survey insights to fill gaps in administrative and public data for high-quality evidence-based policy research.

Key features	
<p>Speed India's first AI-powered feedback engine delivers insights within 24 hours</p>	<p>Spread 30+ million people on our opt-in panel across India covering</p> <ul style="list-style-type: none"> - 21 states - 92% of population - 519 of 742 districts - 18,289 of 19,856 PIN codes
<p>Scale Outbound calling can dial 200,000 calls an hour</p>	<ul style="list-style-type: none"> - 465 of 543 Parliamentary Constituencies - 2,946 of 4,120 Assembly Constituencies

Figure 2: Key Features of the Sampling Frame for Rapid Survey Insights