

Request for Proposals (RFP) on Methodological Innovations to Improve Data Quality

August 31, 2019

*RFP launch date: July 22, 2019
The last date for submission of proposals has lapsed*

Background:

The National Council of Applied Economic Research (NCAER) is one of India's oldest and largest independent, non-profit, economic and social research institutes. It undertakes grant-funded research and commissioned studies for governments and industry, and is one of the few think tanks globally that also collects primary data. NCAER has set up a National Data Innovation Centre (NDIC) to serve as a laboratory for experiments in data collection, interfacing with partners in think tanks, Indian and international universities, and government. NDIC forms an important core of NCAER's long-standing data collection activities. NCAER, has partnered with the Universities of Maryland and Michigan for the NDIC. Initial funding for NDIC is being provided by the Bill & Melinda Gates Foundation.

Current Request for Proposals:

The focus on data-oriented research and the dynamic policy environment impose a great demand for rapid, high-quality, and policy-relevant data. Changing socio-economic conditions and technological innovations necessitate rethinking of both the kind of data being collected and how they are collected. In that context, quality assurance mechanisms play an important role in influencing the credibility of the data being collected.

Therefore, the focus of the current RFP is to seek proposals on methods to improve data quality across the following themes: ***gender equity, women's time use, health system, health insurance and healthcare expenditure, employment and unemployment, family planning, and financial inclusion.***

The proposals on these themes should ideally focus on generating evidence pertaining to the following specific areas relevant to data quality:

Sampling frame for selection of ultimate sampling units: The list of sampling units or the sampling frame plays an important role in probability sampling and the accuracy and completeness of the frame dictate the magnitude of the coverage bias in survey estimates. Often the sampling frame for the ultimate sampling units is not readily available. Although house listing has traditionally been used for the selection of households, it can be quite resource-intensive.

This RFP seeks to identify alternative ways of constructing sampling frames for different modes of data collection and of validating the quality of the frames for its potential use in surveys. The proposals on these themes should ideally focus on the following specific areas relevant to the following themes:

Use of paradata as a quality assurance mechanism: In the case of surveys based on computer-assisted personal interviewing (CAPI), a lot of process data (paradata) are being generated, almost real-time, throughout the survey. This may include interviewer productivity indicators, call records, number of attempts made per interview to interview the targeted respondent, interview length, question-level time stamp data based on key strokes, use of question-specific remarks, GPS coordinates, and, audio recording of interviews. The RFP is soliciting proposals for innovative ways of using such data for monitoring data collection activities and intervening as per the findings of such a monitoring mechanism. Can remote monitoring based on paradata provide an alternative to field-based monitoring and supervision as traditionally used in surveys in India? Is this targeted monitoring and supervision method more effective as compared to the random back checks method?

Controlling interviewer bias and variability in outcomes: Computer-assisted modes of data collection ensure the availability of survey data on a real-time basis. This RFP also seeks to promote innovative assessments of such data in order to reduce interviewer bias and variability. Examples include measuring interviewer effort based on different indicators, such as household roster size, number of cases having zero value in consumption expenditure items, number of illnesses or hospitalisation episodes recorded, and number of formal and informal loans taken by the household, and comparing them across interviewers, identifying interviewers opting for more negative screening in order to reduce their burden of the interview relative to other interviewers or having a conceptual misunderstanding about specific questions. How can one use such early signs to minimise interviewer bias and variability in the outcome of interest?

Real time survey data and consistency checks: Real-time survey data can be used to identify intentional or unintentional mistakes in data collection and to take necessary actions for preventing such errors going forward. The RFP looks for demonstration of consistency checks based on a single variable or multiple variables in order to identify errors at the nascent stage of data collection, and to determine how regular tracking can reduce errors in the long run. In this context, it also seeks an answer to the following question: What is the most effective mode and hierarchy of communication that makes it a feasible quality assurance mechanism?

Use of machine learning techniques to identify patterns automatically: Depending on the extent and coverage of the survey, and the CAPI software used for data collection, the volume of paradata generated can be so huge that traditional methods of analysing data may not work. Examples of such paradata include key stroke data, and audio files generated in surveys recording the interviews. In such a situation, the following questions need to be addressed: How can one use machine learning techniques to identify interviewers not performing up-to-the mark? Can this process be automated? How can one evaluate machine learning techniques for achieving this objective?

Secondary data analysis to quantify variability between interviewers: This RFP also looks for application of statistical/econometric modelling in existing survey data for quantifying variability among interviewers across different outcomes, such as, sensitive versus non-sensitive data; straightforward versus complicated questions; and questions leading to a substantial skip versus questions without any skip patterns. It is also important to identify ways of interpreting the results, and the potential explanations behind the various findings.

Eligibility

Applicants affiliated to any academic or research institutes, non-profit organisations, and private companies that have experience in primary data collection and work out of offices located within India are eligible to apply. We hope that the successful applicants will be able to collaborate with NCAER researchers in future activities, allowing NCAER and the Centre to expand both its network as well as the skill sets of professionals associated with it.

Funding

The Centre will support a budget of up to Rs. 20,00,000/- (inclusive of all applicable taxes) for a period of 12 months. The budget should clearly indicate the actual needs and modes of utilisation of the funding for the proposed project. There is a provision for two such grants of up to Rs. 20,00,000/- (inclusive of all applicable taxes). However, only one grant for each applicant will be considered for funding.

Application Procedure

All applications must be emailed to Ms. Arpita Kayal, Program Manager, NDIC (akayal@ncaer.org) in a single PDF document (font 'Georgia', size 12) with the following components:

A) The proposal (no longer than 6 pages in single space) on research work falling under the focus areas outlined above. The proposal should include the following sections:

1. Project Summary
2. Specific Aim(s)
3. Research Strategy, which would further specify:
 - a. Significance
 - b. Innovation
 - c. Approach and Implementation Plan
4. Expected Outcomes
5. Potential Challenges and Alternative Strategies
6. Timeline
7. Budget and Budget Justification
8. Institutional Background

(The page limit for the proposal is exclusive of details on the budget and applicant's institutional background.)

B) Curriculum Vitae of the key research staff who will undertake the proposed work.

Expression of interest

Applicants interested in participating in this RFP may inform Ms. Arpita Kayal (akayal@ncaer.org) of their interest. We expect to set up an information sharing phone call with potential applicants during early August 2019.

Selection Criteria

The selection of proposals will be based on the merit of the proposal and the CVs of members of the respective research teams. And, the merit of the proposal will be judged on the basis of the following criteria:

- Alignment of the proposal with the RFP;
- Innovativeness in methods outlined in the proposal;
- Rigour and feasibility of approach;
- Clarity of thought; and
- Clarity in writing.

Expected Output

The grantees will be expected to submit a report on the methodology and results to the Centre, which will be posted on NCAER's website. Successful applicants will be encouraged to submit their results for publications in journals. The study instruments and anonymised data sets will also be placed in the public domain, allowing for free online downloads.

Key dates

Proposal submission due date: **31st August, 2019.**

Announcement of the successful candidates: **1st October, 2019.**

Contract signing: **1st November, 2019.**

Project period: **1st November, 2019 to 31st October, 2020.**