Improving Survey Quality using Para Data: Lessons from the India Working Survey

Deepti Goel, Rosa Abraham, and Rahul Lahoti

Azim Premji University

Bangalore, India

IMPROVING SURVEY QUALITY USING PARA DATA: LESSONS FROM THE INDIA WORKING SURVEY

Deepti Goel
Associate Professor, Economics
School of Arts and Sciences
Azim Premji University, Bangalore, India
deepti.goel@apu.edu.in

Rosa Abraham
Research Fellow
Centre for Sustainable Employment
Azim Premji University, India
rosa.abraham@apu.edu.in

Rahul Lahoti
Professor
ETH Zurich, Switzerland
rahul.lahoti@gmail.com

Suggested Citation:

Goel, Deepti, Rosa Abraham, and Rahul Lahoti, (2021). "Improving Survey Quality Using Para Data: Lessons from the India Working Survey", *Institutional Research Grant Report Number 03*, Report submitted to NCAER National Data Innovation Centre, New Delhi: National Council of Applied Economic Research.

Table of Contents

1. Introduction to IWS and this Report

The efficacy of survey-based policy recommendations is primarily dictated by the quality of data collected in the first place. Is the survey truly representative of the population it claims to characterise? Are respondents voicing their true opinions or are they playing to the gallery? Did enumerator bias creep into the data? These are questions that most users of surveys have, but are typically brushed aside in the race to get the analyses out. While there are no fool-proof measures to ensure that survey data are authentic, certain steps can be taken to improve their dependability. One such is the use of what is called 'para data' (data about the process of data collection), to streamline enumerator practices, and thereby improve the reliability of the data being collected. This report details our experience of using para data to improve the quality of the India Working Survey (IWS).

The IWS is a household survey in two Indian States, Karnataka and Rajasthan. It aims to understand the working lives of people: specifically, how their social identities in terms of caste, gender, and religion, play out when livelihood choices are being made. The survey was intended to be representative at the State level, with a wave of face-to-face interviews planned from February through April, 2020. However, field operations had to be stopped in mid-March due to COVID-19 and the subsequent national lockdown. Given the uncertainty of the scope of the pandemic, and an interest in understanding the impacts of the lockdown, the principal investigators (PIs)[1] decided to conduct a second wave of phone interviews in September 2020, with the same respondents who had been interviewed earlier. The two IWS waves put together provide a rich picture of the working lives of those interviewed.[2] In this report, we analyse para data and survey data collected during the first wave of IWS field interviews. **Appendix 1** presents the organisation of the IWS questionnaire for the first wave. Specifically, it lists the sections that form the questionnaire. This is useful for following the analyses presented here.

The term 'para data' was coined by Mick Couper (Couper 1998). It refers to data about the *process* of collection of the more substantive survey data.[3] Nicolaas (2011) provides a short review of what constitutes para data. It typically includes data on who conducted the interview, start- and end-time stamps for the full interview and for individual sections, and re-visit information. The richness of para data varies across surveys.

---

[1] IWS has seven PIs including all three authors of this report. 'We', in this report, variously refers to either all the PIs, or to only the three authors of this report.

[2] Despite not being representative of any geography, the IWS contains a wealth of information that could be used to answer important questions. Did gender, caste, and religion matter for employment/earnings recovery after the lockdown? Did social networks play any role in this recovery? Did the means of livelihood change after the lockdown, and did this play out differently across social identities? These are a few examples of the kind of issues that can be studied using IWS.

[3] Para data should not be confused with meta data. Broadly speaking, para data are about the *process* of data collection whereas meta data provide additional information needed to understand the structure of the dataset. Examples of meta data include the questionnaire, sampling frame, sampling methods, variable and value labels, and frequency distribution for each variable.

It could, for example, also include start and end times for each question, keystrokes in case of computer aided interviewing (CAI), global positioning system (GPS) co-ordinates of interview location, interviewer characteristics and interviewer observations, and audio/video recordings of respondent-interviewer interaction. Para data collection has been greatly facilitated by the adoption of CAI, wherein an electronic device prompts the next question based on answers to previous questions. Along with programming of the logic of the questionnaire, the device can also be programmed to record para data. The first IWS wave was conducted using the computer aided personal interviewing (CAPI) technique.

In this report, we share our experience regarding para data collection and its use to: a) monitor survey progress, and b) improve enumerator practices in the field. In doing so, the ultimate objective is to improve the quality of IWS data. Household surveys managed by small teams of individual researchers are gaining traction in India today. The sheer scale and number of tasks involved in the initial stages can overwhelm PIs managing such surveys. Amidst critical tasks such as sampling design, questionnaire design, testing CAI logic, obtaining clearance from the Institutional Review Board (IRB), piloting the survey, making post-pilot changes to the questionnaire and CAI design, and (re)-training of field enumerators, designing effective use of para data is likely to take a back seat. Our experience with para data in the specific context of IWS is, therefore, particularly relevant. While providing a complete prototype of how to operationalise para data use, we hope that other researchers faced with similar budget, time, and manpower constraints will be able to implement and improve on our model.

Finally, we would like to place our work within the existing literature on para data. A rich body of work already exists on the post-survey use of para data to assess and correct non-response error (Kreuter and Olson, 2013), and to a lesser extent measurement error (Yan and Olson, 2013). Contrasting this, the literature on para data use concurrent with survey implementation is still developing. Kreuter, Couper and Lyberg (2010) were early advocates of using para data from a statistical process control perspective, and adopting the methods used in quality control of industrial products to monitor and improve an ongoing survey process. Our paper falls in this realm of para data use. Jans, Sirkis and Morgan (2013) present a detailed exposition of using concepts and tools from statistical process and quality control to reduce process variation in survey data by identifying special cause variation.[4] A remarkable example of statistical process control using para data is presented by Kirgis and Lepkowski (2013). They describe how para data-driven sampling protocol and workload assignments led to a huge increase in the response rate of the National Survey of Family Growth, a nationally representative survey funded by the United States Government. Most examples and illustrations in existing literature of such use of para data are from developed countries. While the underlying statistical theory is portable across contexts, the operationalisation issues faced by smaller teams of practitioners with limited financial, time, and skilled manpower resources, are very different. It is here that we make a contribution to the existing literature. A notable study, set in a similar context as ours, is that by Choumert-Nkolo, Cust, and Taylor (2019). They use examples to illustrate the use of time stamps, GPS coordinates, and other para data to improve data quality before, during, and after data collection in the context of a household survey conducted in Tanzania. Unlike Choumert-Nkolo et al., our focus is limited to the use of para data while the survey is still ongoing, but we venture into greater details of operationalisation, and present a fully fleshed out model of how to use para data to improve enumerator performance in the field. In sharing our experience, we are candid about our oversights and shortcomings, and make clear recommendations drawing from the lessons we learnt. We hope that this report will encourage other researchers planning on primary surveys to use para data to improve their survey quality.

2. IWS Field Operations

---

[4] They focus on: a) how to select key performance indicators; b) illustrate the use of Shewhart chart, a statistical process control tool, to monitor interview duration; c) discuss how to set threshold limits to a process to distinguish special from common cause; d) and list other types of control charts that can be used to monitor and control the survey process.

Before we present the analyses using para data, it is important to get a sense of the scale of IWS field operations and the personnel involved. Data collection for the first wave was outsourced to Institute for Financial Management and Research (IFMR), a private data collection agency with many years of experience in conducting household surveys. IFMR personnel comprised 3 senior managers, about 15 field supervisors, and about 100 enumerators. The actual data collection was done by the supervisors and enumerators who visited the sampled households and administered the survey using CAPI technique. Besides these personnel, the PIs also employed two project managers (one per State), and four independent supervisors (two per State), to oversee field operations and liaison with IFMR personnel on their behalf.

The IWS field visits were conducted between February 3 and March 17, 2020. A total of 6,900 respondents from 3,623 households were contacted over this period. This number includes respondents with incomplete interviews. Every household was visited together by one female and one male enumerator. Typically, this enumerator pair moved together throughout the length of the survey. In keeping with local norms, and given the gender-sensitive nature of some questions, female respondents were interviewed only by female enumerators, and likewise for males. Every male-female enumerator pair was part of a team of 3-4 such pairs, with one field supervisor in charge. The supervisor would periodically accompany an enumerator pair to a household in order to observe their conduct and provide feedback as required.

3. IWS Para Data

Here, we briefly describe what constitutes para data for IWS and how its use was operationalised. During the course of the survey, we would receive para data dumps from IFMR every two/three days. Each dump consisted of two Comma Separated Value (CSV) files, which together contained para data and survey data collected up to that point. In each file, an observation corresponds to a single respondent, and any given respondent has only one observation associated with them. We, therefore, use the terms 'respondent observation' and 'interview' interchangeably, even when the interaction with the respondent may have unfolded over multiple spells/sittings. Table 1 presents an exhaustive list of para data variables pertaining to interview characteristics used in this report. These were captured using CAPI at the time of survey administration. Para data pertaining to enumerator characteristics is shown separately in Table 4.

We hired one programmer whose sole responsibility was to work on para data. He would generate all the graphs and reports based on para data, which would be available to the PIs within a day or two of receiving the data dump. Thus, para data monitoring had a built-in delay of about four days in the case of IWS. While other surveys may be able to reduce this lag time, some delay may still be inevitable. Irrespective of the lag time, we do not recommend relying solely on para data for real-time monitoring. As was done in IWS, other mechanisms to stay in touch with the field in real-time should be employed. In IWS, for the first ten days of the survey, at least one PI was present on the field in each State. Additionally, supervisors would send real-time updates to the PIs using WhatsApp messaging, a popular Internet-based application for instant messaging over mobile phones. A lot of the initial trouble shooting such as last-minute refinements to the questionnaire, corrections to CAPI logic, and glitches in tablet performance were initiated after receiving feedback via WhatsApp. Additionally, in the first week on the field, at the end of each work day, the PIs and supervisors would hold a phone conference to take stock of the day's activities and respond to the issues faced in the field. We cannot over-emphasise the need for active PI participation and presence in the initial days of data collection. Researchers planning on primary surveys should take serious note of this early time commitment.

4. Para Data to Monitor an Ongoing Survey

During most ongoing surveys, a silent tug of war ensues between two objectives: completing a fixed number of interviews per day to avoid cost over-runs, versus requiring that enumerators spend adequate time with each respondent to ensure that meaningful data is collected. Effective use of para data can strike a balance between these competing objectives. For this, we recommend that the following three parameters, constructed using para data, be continuously monitored throughout the survey period:

1. Cumulative count of completed interviews;
2. Average time spent (with the respondent) per completed interview; and
3. Ratio of completed interviews to all initiated interviews.

One key issue that the PIs should be cognisant of is the definition of what is considered as a 'completed' interview. From the agency's perspective, an interview is completed as long as the enumerator went over all the relevant sections with the respondent, whereas, for the PIs, in addition to this, the nature of non-response within each section also matters. As a case in point, in IWS, the enumerators marked 80 per cent of the 6,900 initiated interviews as completed. A stricter definition, which also mandated that a minimum amount of time be spent on select mandatory sections, resulted in only 70 per cent completed interviews, a drop of 10 percentage points. Figure 1 shows the cumulative count of completed interviews over the IWS survey period according to both definitions: 'Visit Result' refers to what enumerators marked as completed interviews, and 'Stricter Definition' refers to the definition set by PIs which required a minimum time threshold for select sections. The discrepancy between the two is clearly seen, reinforcing the need to track the right metric. The stricter definition can be easily coded using para data on section times. We recommend that PIs use this, instead of visit-result, to track completed interviews.

During an ongoing survey, if completed interviews are falling short of the planned number, this may result in cost over-runs. On the other hand, if completed interviews are overshooting the target because the average time spent per interview is less than the minimum value anticipated by the PIs, it is likely that enumerators are rushing through the interview which could adversely affect data quality. Typically, when external agencies are hired for data collection, the latter is a bigger problem. This is because, while the financial burden of cost over-runs is, wholly or partly, borne by the agency, collecting poor quality data does not have direct financial implications on them. We, therefore, recommend that the PIs track the interview duration very closely throughout the survey period. Figure 2 shows the average time per completed interview over the IWS survey period using the stricter definition of completed interviews. Note that this includes only the interaction time between the enumerator and respondent, and excludes, for example, the time between interview spells in instances where the interview with the respondent occurred in multiple sittings. As is typical of most surveys, this metric drops initially and then stabilises to a steady state value as enumerators gain practice and become adept at administering the questionnaire. In IWS, the initial stabilisation period was about two weeks. A completed interview took 61 minutes, on an average, during the first two weeks, and thereafter, for the remaining weeks, this average time reduced by about 20 per cent to 48 minutes. The variance of interview times is large, 44 minutes in the first two weeks, and 37 minutes thereafter. This large variance is a little disconcerting, but it is precisely what we exploit when para data is put to use to improve enumerator performance. Details are explained in a later section.

Finally, if completed interviews as a share of initiated interviews is low, it suggests that substantial effort by the data collection team is being wasted. Once initiated, an interview could end up being incomplete for multiple reasons: respondents withdrew their initial consent; the respondent was not available during re-visits; the respondent stopped the survey mid-way; and/or the enumerator rushed through the interview and did not meet the minimum time criterion set by the PIs. The PIs must investigate the underlying causes behind the low completion rate and accordingly recommend corrective action. A more effective style of consent administration, fixing prior appointments, better scheduling of re-visits, and sensitising enumerators to spend adequate time with the respondents are possible steps to improve this metric. Figure 3 shows the cumulative fraction of all initiated interviews that were completed over the course of the IWS survey. Like in Figure 2, we used the stricter definition for interview completion. IWS started with a very high completion rate of about 95 per cent, which decreased steadily to about 70 per cent by the time the survey was stopped. This declining trend is largely due to an extraneous factor beyond our control, namely, the nation-wide protests against the Citizenship Amendment Act that were gaining momentum at the time. Given that IWS focuses on social identities, including religion, respondents, especially those from the minority communities, were fearful of participating in our survey, which resulted in lower completion rates towards the end.

The PIs may consider monitoring these three parameters at a more disaggregated level depending on the specifics of their survey. In IWS, each State had its own independent collection team, and it therefore, made sense to monitor each State separately. In fact, we tracked these parameters at the State-sub-region (urban/rural)-gender (of enumerator) level in order to get a better sense of which branches of data collection were working well, and where we needed to intervene.[5]

5. Dashboard to Monitor Survey Progress

A dashboard is basically a one-stop shop were all parameters designed to monitor data collection are displayed at once. We designed our dashboard using 'Shiny', an open-source R-based package for building web applications. Although Shiny was adequate for our needs, we recommend that the PIs spend some time deciding whether they would like to build their own dashboard from scratch (like we did), or use paid applications, such as SurveyCTO, that come with in-built customisable dashboards. Building a dashboard requires specialised coding skills, so the decision would depend upon the pool of talent available and the budget allocation for para data monitoring.

There is no set layout for dashboard design. The PIs need to strike the right balance between wanting to track a large number of parameters to monitor every aspect of the survey, versus, tracking too much, resulting in clutter and obfuscation of information. Figure 4 presents a couple of screen shots of the IWS dashboard. We definitely recommend tracking the three parameters mentioned in the earlier section. We consider these as necessary and adequate for monitoring survey progress. At best, a few more could be added depending on the specifics of the survey, but tracking more than that may result in losing sight of the forest for the trees, especially when supervision time is at a premium (which it generally tends to be).

6. Flagging Errant Enumerators Using Para Data

Our model for tracking errant enumerators essentially falls in the realm of statistical process control, wherein the basic idea is to reduce variation in the process (of data generation), in order to improve the quality of the output (survey data). We used what are called flags, to identify errant (deviant) enumerator practices in the field. Once identified, intervention was in the form of the flagged enumerator's field supervisor talking to them and providing constructive feedback. The basic idea of a flag (explained below), is not novel (Jans, Sirkis, and Morgan 2013; Peng and Karl 2011). However, their use to improve survey quality is not yet widespread. We present a detailed explanation of how we created and implemented para data-based flags, in the hope that others will be inspired to do the same to improve their own surveys.

A flag is a warning or signal that gets triggered whenever enumerator performance deviates substantially from a specified benchmark. It is *suggestive* of a potentially faulty enumerator practice such as very short interview time. A flag should not be construed as conclusive evidence of wrongdoing as it indicates errant behaviour without delving into the reasons for the same. It is possible, though unlikely, that the said errant behaviour was the right response given the situation faced by the enumerator on the field.[6] It is, therefore, important that any intervention based on these flags is not accusatory in nature. We recommend that the PIs spend some time with the field supervisors making sure that they understand this. Under no circumstances should the conversation with a flagged enumerator turn into a finger-pointing exercise.

Essentially, a flag involves comparing enumerators who faced similar field conditions, and identifying (flagging) those enumerators (if any), whose performance deviated substantially from the average or mean performance in their 'comparison' group. Restricting comparison to enumerators *within* a comparison group

---

[5] We kept track of eight branches in all: two States (Karnataka and Rajasthan), times two sub-regions (urban and rural), times two genders (female and male). A branch in our context is a homogenous environment where enumerators are likely to face similar interview conditions. Later on, in this report, we also refer to a branch as a 'comparison group'.
[6] The reason this is unlikely is that when flagging errant behaviour, we take care to only compare across enumerators facing similar field conditions. We expect that, on an average, comparable enumerators facing similar field conditions would display similar practices.

ensures that different processes of data generation are not mixed together. This makes it possible to interpret the group average as the process average in steady state, and deviations from this average as errant behaviour that require intervention.

In the remaining paragraphs of this section, we talk about various design features involved in building para data-based flags. Some considerations require serious deliberation by the PIs, as there are no definitive rules to follow, and the final call would be context-specific. In each instance, we share the decisions we made in IWS. Our actions are to be read as suggestive rather than prescriptive.

6.1. Definition of Comparison Group

In IWS, we defined a comparison group as a specific State (Karnataka/Rajasthan), sub-region (urban/rural), and gender (of enumerator) combination, resulting in eight such groups. Given the significance of properly defining a comparison group in order to generate meaningful flags, we elaborate a little more on this concept. Consider enumerators operating within the same state and sub-region, say in rural Karnataka. Even within this stratum, it would be incorrect to bracket male and female enumerators in the same comparison group as the work profiles of male and female *respondents* are very different and would call for different practices by enumerator gender (recall that a respondent was only interviewed by an enumerator of the same gender). In fact, the IWS questionnaire was itself gender-specific (see Appendix 1), inevitably resulting in different average interview lengths by enumerator gender. If we were to make a single comparison group of all (male and female) enumerators in rural Karnataka, we would not be making an apples-to-apples comparison as gender-specific differences in average field practice are expected by design. Following similar reasoning, one can rationalise the use of state and sub-region for defining a comparison group. We recommend that the PIs think very carefully about what dimensions to use to delineate a comparison group in their specific context. A key principle to keep in mind is that optimally defined comparison groups would maximise between-group variability and minimise within-group variability under stable field conditions.

6.2. Threshold(s) for Deviant Behaviour

Next, we turn to the issue of what constitutes deviant behaviour. How far away from the group mean should a value be in order to be flagged? There is no definitive answer for this. Some studies have referred to the three-sigma rule, that is, three standard deviations away from the mean, as a statistical benchmark (Jans, Sirkis, and Morgan 2013). However, even they have acknowledged that no single rule fits all. In IWS, for most flags, we used a threshold of 1.6 standard deviations from the mean, and for the rest we use a threshold of 1. In general, thresholds could be based on a pilot phase or could be driven by feasibility considerations. For example, if a particular threshold triggers a large number of flags, in turn requiring a large number of interventions, it may be prudent to set a limit that would trigger fewer flags.

6.3. Length of Performance Window

In IWS, we took a conscious decision to look at each week's performance separately, and not compare cumulative performance up until a point in time. By considering each week as a separate window for analysis, performance is flagged as deviant against a *moving* benchmark that accounts for all secular changes over time. For instance, as enumerators gain proficiency, the average interview duration inevitably drops over time. Looking at separate weekly windows accounts for this secular change. In determining the appropriate window length, the following trade-off should be borne in mind: if the window is too long, faulty practices may continue unchecked, adversely affecting survey quality. On the other hand, if it is too short, there may not be enough data points for the underlying statistical theory to operate, invalidating the credibility of the flag-generating process.[7] Additionally, shorter windows dictate more frequent interventions, which come with a

---

[7] The theory is based on the Law of Large Numbers, and requires that there be a reasonably large number of interviews within each comparison group.

time cost for the field supervisor. Looking back at our own IWS experience, a two-week window, instead of a one-week one, would have been more effective in managing the trade-off just described.

6.4. Choice of Flags

Each flag is associated with a specific field practice such as interview duration or the number of times a particular question was skipped. Selecting the list of field practices to monitor is a complex and non-trivial exercise. Given that the marginal cost of creating one more flag is very small, there is a tendency to create a long list of flags without recognising that intervention is costly. Monitoring too many aspects may jeopardise the effectiveness of the whole flagging exercise. To elaborate, imagine a scenario where a field supervisor receives a report that asks them to talk to four different enumerators, sounding out each enumerator on seven different dimensions specific to their performance. Further imagine, having to do such an exercise on a weekly basis, a task over and above the conventional work that falls under the supervisor's purview. First, there is a non-negligible chance that much of the information will be lost in translation between the supervisor and the enumerator. Even if everything were to be communicated correctly, if the list of flagged practices is long, the enumerator may be at a loss to decide which ones to prioritise and focus on first in terms of taking remedial measures. Worse still, an enumerator may feel completely dejected at receiving a long list of flags, and may give up entirely. While four and seven in the above example are arbitrary numbers, the basic point is to warn the other PIs to be judicious in their choice of flags. The choice should be guided by PI priorities and operational constraints. Some competing priorities could be ensuring survey timeliness, survey representativeness, and adherence to interview protocol. In IWS, our choice of flags was driven by a focus on data quality rather than survey timeliness, the rationale being that enumerators were already under pressure from the survey agency to complete the survey on time, and we did not feel the need to reinforce it. We were more concerned about enumerators taking short cuts and compromising on proper interview protocols in order to meet their internal (set by the survey agency) productivity targets.

Table 2 presents the fifteen flags we monitored in IWS, along with detailed information on how each flag was created. In column three, against each flag, we specify the main performance dimension(s) it evaluates. Our classification consists of three dimensions. The first is content knowledge, which refers to a sound understanding of the concepts and definitions used in the questionnaire. The second is the effort exerted by the enumerator, which is proxied by the amount of time spent interacting with the respondent. And the third is adherence to ethics concerning the interview protocol. In columns four and five, we describe each flag in terms of the specific field practice it monitors and the underlying concern it addresses.[8] The last three columns provide details about how each flag was operationalised: a) whether it was constructed using para data variables or the main survey data; b) the criteria used for flagging interviews/enumerators including specifying the thresholds where applicable; c) the method for ranking enumerators for intervention purposes. Our list of flags is neither prescriptive nor exhaustive. It is intended as a guide to help other researchers choose their own set of flags. In fact, if we were to do this exercise again, we would have a different list ourselves. For example, we would add some flags based on the text length of crucial descriptive data and on re-visit information. On the other hand, we would cut down the number of section-wise time flags and only have two flags based on duration; one for the whole interview (Survey Time), and the other for section 4 (Section 4 Time), the core section of the IWS survey.[9]

Another aspect for the PIs to consider, particularly where the survey extends over a very long time period and the initial set of flags seem to have stabilised, is to change the set of flags being monitored.

---

[8] Section time flags for Sections 6, 7, 11, 12, and 13 were not created for the following reasons: Section 6 was administered on paper and therefore, para data for this section is not available; for Section 7, the enumerators did not always follow a common practice in terms of whether or not they recorded section-specific consent time on their tablets; Sections 11 through 13 got infrequently fielded, and, therefore, there would be insufficient data points to create meaningful flags.

[9] A rationale for this is given in Appendix 3. Please read Appendix 3 after the section on 'Analysis of Para Data-based Interventions', as it uses terms first described in this section.

7. Analysis of Para Data-based Interventions

We first describe the interventions aimed to improve enumerator practices in the field. Next, we examine the effectiveness of the first set of interventions in actually impacting enumerator behaviour.

7.1. Description and Timeline of Interventions

Once the flags were generated, they were collated into weekly reports, one for each State. Appendix 2 contains a sample report for Karnataka. Each report was shared with the respective State level supervisor. The supervisor then emailed it to all field supervisors, and followed this up with a phone conversation with each field supervisor where the relevant enumerator-specific information was highlighted. The final step involved a one-on-one conversation between the field supervisor and a flagged enumerator. During these conversations, the field supervisor was advised to take a 'trust but monitor' approach, gently pointing out the deviant behaviour without being accusatory, and nudging the enumerator to take corrective action.

During the survey period, two reports were shared with the field personnel. The first was based on enumerator performance in the week between February 17 and February 23, and the second on performance between February 24 and March 8. The first two weeks of the survey were not targeted for intervention as it inevitably takes some time before the processes stabilise. This is also seen in Figure 2, which shows that the average time per completed interview reached a steady value after two weeks.

The first report was shared on March 4 and March 3 in Karnataka and Rajasthan, respectively; while the second report was shared on March 10 and March 14 in the respective States.[10] The survey was officially stopped on March 17, but no new interviews were closed after March 14, making March 14 the effective end date for the first wave of IWS.

In the analysis that follows, we examine only the interventions based on the first report for the following reasons: a) Just around the time that the second report was shared, other external events, such as the scare generated by Coronavirus, were beginning to impact enumerator performance.[11] It would not be possible to separate the effect of our interventions from that of these events; b) The second report would interact with the first one in affecting enumerator behaviour, making it impossible to separate out the independent effects of each report.[12] c) Finally, there is no post-intervention period for the second report in Rajasthan as the survey had to be stopped on March 17.

7.2. Analysis of the First Set of Interventions

We use Ordinary Least Square regressions, with enumerator fixed effects, to analyse the impact of interventions based on the first report. We estimate the following regression equation in data:

$$Performance_{ij}^k$$
$$= \beta_0 + \beta_1\big(Flag\_Same_j^k * Post_i\big) + \beta_2\big(Flag\_Other_j^k * Post_i\big) + \beta_3 Date_i$$
$$+ \beta_4 Date\_Squared_i + \{Enumerator_j\} + \varepsilon_{ij}^k \qquad (1)$$

Here, i stands for interview, j for enumerator, and k for a specific flag such as Survey Time and Section4 Skip (see Table 2 for details of all the fifteen flags analysed in this report). $Performance$ refers to the particular field practice that a flag measures as described in Column (4) of Table 2. For example, in case of Survey Time, it is the interview duration in minutes; and for Section4 Skip, it is an indicator variable, which takes the value

---

[10] There was a pause in field operations in Rajasthan from March 8 to March 13 on account of Holi, a festival mainly celebrated in north India. The second report was shared after this break in Rajasthan.

[11] Another such event occurred on March 12, when a Hindu-Muslim enumerator pair was accosted by villagers in Karnataka. The pair were allowed to go only after the local administration intervened on our behalf. It is very likely that this unfortunate incident adversely affected the psyche of many enumerators.

[12] On hindsight, we should have delayed sharing the second report so as to give us a longer post intervention period for the first report. In our case, this did not matter as the survey had to be stopped for extraneous reasons anyway.

1 when the respondent is reported as 'Not Working' and 0 otherwise. $Flag\_Same$ and $Flag\_Other$ are indicator variables for whether the enumerator was flagged for flag k and for some other flag ($\sim k$), respectively. $Post$ is also an indicator variable for whether the interview was closed in the post-intervention period, that is, after the first report was shared with the enumerators. $Date$ and $Date\_Squared$ form a quadratic in time, and $Enumerator$ is an enumerator-specific fixed effect. ε is a catch all for all idiosyncratic factors that affected performance.

The primary coefficient of interest is β_1. It indicates the magnitude of change in performance as a result of talking to an enumerator to correct that specific practice. β_2 is also of interest as it tells us whether intervening to correct some other practice had an effect. The time controls, $Date$ and $Date\_Squared$, account for secular changes over time that affect all enumerators. Finally, by including enumerator fixed effects, we identify the effect of interventions by looking at whether they changed behaviour relative to an enumerator's own behaviour prior to being flagged.

Table 3 presents some descriptive statistics on flags, along with the regression results for studying intervention effectiveness. As mentioned earlier, the first report flagged enumerators based on their performance in the week from February 17 to February 23. The table examines a longer period, between February 17 and March 10 for Karnataka, and between February 17 and March 14 for Rajasthan. We refer to this as the analysis period. Of this, the pre-intervention period is before March 6 for Karnataka, and before March 5 for Rajasthan, and the remaining constitutes post-intervention. A total of 88 enumerators completed at least one interview (going by the strict definition of a completed interview), during the analysis period, of which 46 were women. Only those flags are listed for which at least one enumerator was flagged in the report. The third column in Table 3 shows the number of enumerators flagged against each flag. Columns 4 and 5 present the mean value of the field practice being monitored for all the enumerators and for the flagged enumerators, respectively, during the pre-intervention period. The regression results are shown in Columns 6 through 10. In order to improve precision of the estimated coefficients, the regressions are restricted to enumerators with at least ten completed interviews. Columns 7 and 8 present estimates for β_1 and β_2, respectively.

A look at our main coefficient of interest, β_1, indicates that there is some evidence that our interventions had the intended effect for two flags, namely Section4 Time and Alone Section7. For Section4 Time, they resulted in increasing the interview time for Section 4 by 0.7 minutes, an increase of 18 per cent of the average section time in the pre-intervention period. For Alone Section7, intervening reduced cases where the enumerator reported being all alone with the respondent when administering the Discrimination section by 5.2 percentage points, an effect size of 6 per cent of the average reporting level in the pre-intervention period. However, both these effects are statistically significant only at the 10 per cent level, implying that we cannot make definitive conclusions as our estimates are imprecise. Our own conjectures for the reasons behind imprecision are: a) Our post-intervention period is too short, four days for Karnataka and seven days for Rajasthan, resulting in fewer observations during this period, and b) We had too many flags, which could have adversely affected the effectiveness of communication between the field supervisors and the flagged enumerators. This second conjecture is partly strengthened by some significant estimates for β_2. The table shows that flagging for some other practice resulted in an improvement in performance in terms of increasing the time spent on Section 5 (Section5 Time), and lowering the number of respondents reported as 'Not Working' (Section4 Skip). It suggests that intervening, in general, helped, but perhaps not always along targeted lines.

8. Lessons Learnt

Here we share some important lessons that we learnt from our experience of using para data. Some of these are specific to the IWS context where data collection was outsourced to an external agency and was implemented using the CAPI technique.

8.1. Lesson 1: Understanding the Structure and Composition of Para Data
The way in which raw para data is organised can be complex, and varies across projects and data agencies.

In our case, the structure was fairly simple: para data was collected on the same unit of analysis as survey data, both were organised around the individual respondent. There could, however, be more complex structures. For example, para data could be structured around each respondent session (interview spell), in which case a respondent could have multiple entries in the para data file. While we are agnostic about which structure is better, it is important that the PIs fully understand the structure in which para data would be *handed over* to them. This is essential to be able to design para data use. It is equally important to know the exact para data variables that will be generated *and shared*. For example, knowing the granularity of time stamp data, whether it is at the interview, section, or question level, is important when deciding what flags to create and monitor. This aspect becomes especially crucial when data collection is outsourced. The external agency may not be forthcoming in sharing detailed para data information in order to avoid a close scrutiny of its enumerators. It is important to engage in a dialogue with the agency and have them fully on board with all aspects of para data monitoring right from the design stage.

Recommendations:

A) Make a list of para data variables, similar to Table 1, and share it with the data collection agency, perhaps at the time of sharing the survey instrument. It would be best if para data requirements could be included as deliverables in the contract drawn with the data collection agency.

B) Pilot the para data along with the main survey data. Besides the provision of a dry run for para data use, the pilot para data can itself be used to prune the questionnaire. For example, data on question durations can be used to decide which questions to modify or drop altogether. For instance, if a particular question is taking too long and is not justified by its research value, it can be removed from the main survey (for details see Choumert-Nkolo, Cust, and Taylor 2019).

8.2. Lesson 2: Choosing between Dashboard and Printed Reports for Flagging

Our advice on this is contrary to the push towards dashboards found in contemporary writing. Highly sophisticated dashboards with capabilities of generating automated reports in almost real time, and which contain customised information for each end-user, are undoubtedly preferred to a system generating manual reports with a lag. However, effective dashboard design requires specialised talent and time, and when these are scarce, it may be prudent to use manual reports rather than spend time building a sub-optimal dashboard. In IWS, we found the dashboard very useful for tracking the overall progress of the survey, but we were less successful in building a customised dashboard for each field supervisor where the content shown to them would be limited to only the enumerators under their supervision. In fact, we relied on paper reports directed to a State supervisor who would, in turn, provide customised feedback to each field supervisor. R scripts were used to generate the statistics that went into these reports, but manual intervention was needed to run the scripts, generate the reports, and provide customised summaries to each field supervisor. These worked well for our purposes.

Recommendation: The PIs must take a call on whether they are in a position to design an effective dashboard keeping the money, talent, and time constraints in mind. Dashboard, at the end of the day, is a tool towards an end, and if, other, more effective tools are available, dashboard design should not be considered a necessary ingredient for para data-based interventions.

8.3. Lesson 3: Principle for Dashboard Design/Report Generation, or 'More Is Not Always Better'

Mohadjer and Edwards (2018) present a detailed account of dashboard design in the context of collecting data for the Programme for the International Assessment of Adult Competencies (PIAAC) in the United States. They share many useful lessons and we would like to highlight two of these. First, a dashboard should be designed keeping only *one type* of end-user in mind. This is because the information requirements of each user type are different, and targeting multiple users using a single dashboard would make it harder for each type to access the particular information they need. In case of surveys like the IWS, two dashboards should be created—one for the PIs to monitor the overall progress of the survey and another one for the field supervisors, with customised views to enable each supervisor to track only their own enumerators. Combining the requirements of these two distinct users into a single dashboard is not a good idea. Second,

less may be better than more when it comes to dashboard design. There is a tendency to want to track all aspects of the survey at once. Using a layered dashboard design that highlights a few salient aspects in the first view, with inner views providing details (only if necessary), is better than a flat design which displays all aspects at once. These lessons are equally applicable when reports act as a substitute for dashboards.

Recommendation: Dashboard/report design should be driven keeping a single end-user type in mind, and one must avoid the temptation to clutter the displays/pages with too much information.

8.4. Lesson 4: Ensuring that the Para Data Itself Is of High Quality

It may be necessary to go into the details of how important variables are captured to ensure that the para data itself is of high quality. We draw from our experience to talk about a few specific variables.

Time stamps: As far as possible, the time stamps should be captured automatically by the device and the device settings should be checked before data collection starts. In order to avoid tampering, it should not be possible to re-enter this information once it has been captured. It is worthwhile for the PIs to spend some time understanding how time stamps are generated in the data.

Link between question, Section, and interview durations: Ideally, if all time stamps are recorded up to the lowest level, question-times should add up to section-time, and section-times should add up to interview-time. However, when certain activities are not recorded, such as administration of consent, or times for off-CAPI sections, this may not be the case. The PIs must clearly understand the composition of all variables related to time durations and what parts of respondent-interviewer interaction are covered between the start and end timestamps.

Visit result: At the time of closing a respondent case, an enumerator has to mark the status of the interview as complete, incomplete, door refusal, or not available. In our experience, enumerators either do not realise the importance of this variable or are not clear about the criteria to code it correctly. The PIs must ensure that the options under 'visit result' are mutually exclusive and exhaustive, each option is well-defined using objective criteria, and enumerators understand how to code this variable correctly. This should be discussed at some length during enumerator training.

Enumerator identifier: Undoubtedly, this is an important variable when it comes to para data-based monitoring of enumerators. This variable should preferably be a single variable (and not constructed using a combination of other variables), and it should be selected from a drop-down list of enumerator names (rather than codes) at the time of data entry. This will help avoid cases where enumerators mis-spell their names or use two different numeric codes.

8.5. Lesson 5: Content and Frequency of Intervention, Again, 'More Is Not Always Better'

Intervening to change enumerator behaviour is costly, not so much in terms of explicit financial costs, but more in terms of field supervisors' engagement. In IWS, the basis for intervention was a weekly report that listed out the names of flagged enumerators. As seen in Appendix 2, our report is organised flag-wise. Given that the unit of intervention is an enumerator, a more effective report style would be enumerator-wise. This would directly inform the field supervisor about whether any action is required for a particular enumerator, and, if so, which field practices need a review. Additionally, a report should not contain a lot of information (too many flags), or be too complicated (comprising nuanced flags that are hard to understand and talk about). This would jeopardise effective communication between the supervisors and enumerators. Furthermore, frequent interventions would put increased pressure on the supervisor, and also leave insufficient time for the enumerator to absorb feedback, introspect, and take corrective action. It is very important that the PIs deliberate on the entire process of intervention to make it more effective.

Recommendation:

A separate session for field supervisors should be included as part of their training where they are shown how to read para data-based dashboards/reports and also how to communicate feedback effectively to enumerators.

9. Markers of a Good Enumerator

In addition to para data on interview characteristics, we also recorded enumerator characteristics using a voluntary, self-administered CAPI survey of enumerators. This was administered on the last day of the field training, just before the main IWS survey began. The enumerator survey collected information on demographic characteristics such as age, caste, education, and work experience; and also included a test to measure mastery over key concepts and definitions used in the IWS questionnaire. The intention was to examine whether these could predict performance in the field. Such an analysis could potentially inform hiring (and training) of enumerators, with the important caveat that 'no discrimination on the basis of social identity' is practised.

About 90 enumerators initiated at least one interview during the pre-intervention period, that is, the window during which we examine enumerator performance. We exclude the post-intervention period, as we do not want to confound the effect of characteristics with that of our interventions. Of the 90, 65 returned the filled-up survey questionnaire. We are only able to study the influence of enumerator characteristics using this select sample. Table 4 presents findings from our enumerator survey. As seen, while the two States had a roughly equal share of active enumerators during the pre-intervention period (48.9 per cent from Karnataka), more from Karnataka submitted the enumerator survey (64.6 per cent from Karnataka). Female enumerators are also slightly over-represented among enumerators who submitted the survey (46.7 per cent female among active enumerators, and 50.8 per cent among those who submitted the survey). About a quarter of our enumerators belong to the so-called lower castes (Scheduled Castes and Scheduled Tribes, and 12 per cent belong to the highest caste group, namely, Brahmins. We prefer to categorise caste by separating the lowest and highest caste groups and clubbing the remaining into a single category, namely, Other Backward Classes and Non-Brahmins. Only 1 of the 65 enumerators who submitted the survey is Muslim, while the remaining are Hindus. We, therefore, exclude religion from our subsequent regression analysis as there isn't much religious variation. 60 per cent have studied beyond the Bachelor's level, 25 per cent are proficient in English, and 69 per cent had worked as enumerators in at least three other projects before they joined IWS. The average score in the test designed to measure mastery over the IWS questionnaire was 59.4 per cent, with a standard deviation of 10 percentage points.

Table 5 presents regression results to identify markers of a good enumerator. The unit of observation is an enumerator. Panel A shows the enumerator-level mean values for six measures based on performance in the pre-intervention period. These are: Interview Completion Rate (fraction of initiated interviews completed); Survey Time, Section4 Time, Network Size, Skip4 Section, and At least one Flag (whether or not the enumerator was flagged at least once in the first report). These specific performance measures have been selected, either because they are comprehensive measures of overall performance (Interview Completion rate, Survey Time, and At least One Flag), or, they test content knowledge that was emphasised during training (Section4 Time, Network Size, and Skip4 Section). Before we discuss the regression results, we would like to emphasize that these regressions only establish partial correlations between enumerator characteristics and performance. Given the absence of any identification strategy, and the select sample in panel C, regression coefficients should not be given a causal interpretation. Furthermore, given our interest in identifying the predictors of good performance, we do not discuss regressions with an overall low predictive power (as measured by a low R-squared value) even when some coefficients are shown to be statistically significant.

We find that submitting the enumerator survey positively predicts a higher interview completion rate: an increase of 21 percentage points from a base of 69 per cent. Belonging to the 'Brahmin' caste is also positively correlated with a higher completion rate. Belonging to the 'Other Backward Classes or Non-Brahmin caste' and being more educated, positively predict Network size, the number of persons in the respondent's social network. Performing well in the enumerator test positively predicts Network Size, and negatively predicts Section4 Skip (indicator for whether respondent is 'not working'), though the coefficients are not significant at the 5 per cent level of significance. However, the p-value is 0.068 in both cases, and the coefficient magnitudes are large. These two concepts, that is, composition of the social

network and the criteria for categorising a person as 'not working', were difficult concepts to comprehend and during the training, there was a tendency to incorrectly exclude individuals from being part of the social network, and to mis-categorise people as not working, especially women. These results suggest that those who grasped these concepts well during the training (as shown in better test scores), did a better job of capturing this information on the field.

10. Concluding Remarks

In this report, we share our experience of using para data to improve the India Working Survey (IWS), a large household survey conducted in two Indian States. The report presents a complete and detailed prototype of how to design and operationalise para data use to monitor survey progress, as well as to improve enumerator practices while the survey is still ongoing. It highlights the trade-offs involved at various junctures that would require the PIs to deliberate and make informed decisions to suit their own specific circumstances. The report is also full of recommendations drawn from the lessons we learnt along the way. We hope that this report will encourage other researchers to use para data to improve their own surveys.

We found para data to be extremely effective in monitoring survey progress. The use of dashboards in this context proved to be an efficient way for the PIs to track crucial parameters. In terms of using para data to improve the enumerator field practices, ex-post regression analyses suggests that we had some success in influencing enumerator behaviour, though our estimates are imprecise. The most important aspect for us from this entire exercise was in the form of valuable lessons for the future. We plan to make use of our experience in future primary data collection endeavours.

While we have clarified many important issues regarding para data use, there is more ground to be covered. More work is needed to arrive at an economical list of flags which would ensure that enumerator practices along multiple dimensions are being adequately monitored. We have made some suggestions about how to choose flags but more work is needed to arrive at a standardised list of flags that is applicable in varied contexts. As para data use becomes widespread, this would help in establishing empirical thresholds to identify deviant behaviour. Finally, we have been silent about many important forms of para data such as GPS coordinates, and it would be good if other practitioners share their experience with these.

We would like to end with some suggestions for donor agencies that fund primary data collection efforts. One way to encourage the use of para data to improve surveys is for funders to: a) mandate their use, b) provide a budget specifically earmarked for it, and c) require that some aspects of para data be made public. Interview length for each completed case is a good example of a para data item which should be included in the main survey data that is made available for use. Given suggestive evidence that enumerators who had a better understanding of the survey instrument also did better in the field (though our estimates are imprecise), funders could also mandate that an enumerator test (of the kind discussed in this report), be conducted at the end of the training, and only those candidates who clear a minimum bar be part of the final survey team. This last recommendation needs more careful thinking in terms of who sets and evaluates the test, and who bears the cost of compensating personnel who did not clear the test and had to be let go for that reason. A more nuanced consideration in instituting such a test is that it may change the applicant pool of potential enumerators, but we anticipate that this would be in the direction of positive selection on the basis of skill. It may also entail a second round of hiring and training if there is substantial attrition, but these steps may be well worth the cost if they result in substantial improvement in survey quality. If, because of a push from funders, using para data to improve survey quality becomes a standard practice, it would also result in data collection agencies viewing para data not as a threat to their commercial interests, but as an integral tool to improve their business.

References

Choumert-Nkolo, J., H. Cust, and C. Taylor. (2019). "Using paradata to collect better survey data: Evidence from a household survey in Tanzania", *Review of Development Economics*, 23: 598– 618. https://doi.org/10.1111/rode.12583.

Couper, M. (1998). "Measuring survey quality in a CASIC environment" In Proceedings of the Section on Survey Research Methods at the American Statistical Association.

Jans, M., S. Sirkis, and D. Morgan. (2013). "Managing data quality indicators with paradata based statistical quality control tools: The keys to survey performance", in F. Kreuter (ed.), *Improving Surveys with Paradata: Analytic Uses of Process Information*, Hoboken, NJ: John Wiley and Sons.

Kirgis, N.G. and J.M. Lepkowski. (2013). "Design and management strategies for para data-driven responsive design: Illustrations from the 2006-2010 National Survey of Family Growth", in F. Kreuter (ed.), *Improving Surveys with Paradata: Analytic Uses of Process Information*, Hoboken, NJ: John Wiley and Sons.

Kreuter, F., M. Couper, and L. Lyberg. (2010). "The use of paradata to monitor and manage survey data collection", Section on Survey Research Methods–JSM 2010.

Kreuter, F. and K. Olson. (2013). "Para data for non-response error investigation", in F. Kreuter (ed.), *Improving Surveys with Paradata: Analytic Uses of Process Information*, Hoboken, NJ: John Wiley and Sons.

Mohadjer, L. and B. Edwards (2018). "Para data and dashboards in PIAAC", *Quality Assurance in Education*, 26(2): 263-277. https://doi.org/10.1108/QAE-06-2017-0031.

Nicolaas, G. (2011). "Survey para data: A review", Economic and Social Research Council (ESRC), *National Centre for Research Methods Review Papers No. 17*, Swindon, U.K.

Peng, D. and F. Karl (2011). "Quality Control in Survey Research Today", *Survey Practice*, April. www.surveypractice.org

Yan, T. and K. Olson (2013). "Analyzing para data to investigate measurement error", in F. Kreuter (ed.), *Improving Surveys with Paradata: Analytic Uses of Process Information*, Hoboken, NJ: John Wiley and Sons.

Tables

[1] Variable	[2] Description
enumerator.id	Unique identifier (code/name) associated with each enumerator
enumerator.gender	Gender (male/female) of the enumerator
interview.id (respondent.id)	Unique variable (or set of variables) associated with each interview/respondent
state	State (Karnataka/Rajasthan) of the respondent
region	Region of residence (rural/urban) of the respondent
consent	Whether or not the respondent consented to the interview
interview.start.stamp	Date (dd-mm-yyyy) and time (hrs: mins) when the interview started
interview.end.stamp	Date (dd-mm-yyyy) and time (hrs: mins) when the interview ended. Note that incomplete interviews also have an end stamp.
interview.duration	Time between start and end of the interview. This only includes the time that the enumerator spent with the respondent administering the survey questions. If the interview was conducted in multiple spells, it does not include the time between spells.
section.duration	Time between the start and end of each section of the questionnaire. There is one such variable for each section.
revisits	Number of additional visits made to interview the respondent
visit.result	The final completion status of the interview at the time of ending it, as marked by the enumerator
This list does not include variables from survey data that were also used to generate flags.	

Table 1: Para Data on Interview Characteristics, India Working Survey

Source: IWS Para Data.

[1] S. No.	[2] Flag Name	[3] Performance Dimension Being Evaluated	[4] Description of Field Practice Being Monitored	[5] Description of Underlying Concern	[6] Constructed Using [Para Data/Survey Data]	[7] Flag Criteria	[8] Criteria for Ranking Enumerators to Initiate Intervention
				Table 2: Para Data Flags to Monitor Enumerator Performance, India Working Survey (1/5)			
1	Survey Time	Effort	Time taken to field select sections: 1 Demographic Characteristics, 5 Household Production, 7A Discrimination, 8 Decision Making, and 10 Networks.[1]	Data quality is unlikely to be good in very short interviews as the enumerator would not have spent enough time delivering consent, answering respondent questions, and reading out all the instructions, response options, and transition statements.	Para data Based	An interview is flagged when standardised survey time[2] is below -1.6 OR survey time is below 10 minutes.	Consider all completed interviews in that week. For each enumerator, calculate their flagged interviews as a share of interviews completed by them. Within each state-gender category, report and intervene on the top three enumerators with positive shares (if any).
2	Section0 Time	Effort	Time taken to field Section 0, Household Register.	Data quality of Household Register section may be poor if it was hurriedly administered.	Para data Based	An interview is flagged when its standardized section time is below -1.6.	Consider all completed interviews in that week. For each enumerator, calculate their flagged interviews as a share of interviews completed by them. Within each State-gender category, report and intervene on the top three enumerators with positive shares (if any).
3	Section1 Time	Effort	Time taken to field Section 1 Demographic Characteristics.	Data quality of Demographic Characteristics section may be poor if it was hurriedly administered.	Para data Based	An interview is flagged when its standardised section time is below -1.6.	Consider all completed interviews in that week. For each enumerator, calculate their flagged interviews as a share of interviews completed by them. Within each State-gender category, report and intervene on the top three enumerators with positive shares (if any).
4	Section2 Time	Effort	Time taken to field Section 2, Household Living Standards.	Data quality of Household Living Standards section may be poor if it was hurriedly administered.	Para data Based	An interview is flagged when its standardised section time is below -1.6.	Consider all completed interviews in that week. For each enumerator, calculate their flagged interviews as a share of interviews completed by them. Within each State-gender category, report and intervene on the top three enumerators with positive shares (if any).

[1] S. No.	[2] Flag name	[3] Performance Dimension Being Evaluated	[4] Description of Field Practice Being Monitored	[5] Description of Underlying Concern	[6] Constructed Using [Para Data/Survey Data]	[7] Flag Criteria	[8] Criteria for Ranking Enumerators to Initiate Intervention
				Table 2: Para Data Flags to Monitor Enumerator Performance, India Working Survey (2/5)			
5	Section3 Time	Effort	Time taken to field Section 3, Activity Profile for the Last Year.	Data quality of Activity Profile for the Last Year section may be poor if it was hurriedly administered.	Para data Based	An interview is flagged when its standardised section time is below -1.6.	Consider all completed interviews in that week. For each enumerator, calculate their flagged interviews as a share of interviews completed by them. Within each State-gender category, report and intervene on the top three enumerators with positive shares (if any).
6	Section4 Time	Effort	Time taken to field Section 4, Weekly Labour Force Status.	Data quality of Weekly Labour Force Status section may be poor if it was hurriedly administered.	Para data Based	An interview is flagged when its standardised section time is below -1.6.	Consider all completed interviews in that week. For each enumerator, calculate their flagged interviews as a share of interviews completed by them. Within each State-gender category, report and intervene on the top three enumerators with positive shares (if any).
7	Section5 Time	Effort	Time taken to field Section 5, Household Production Activities.	Data quality of Household Production Activities section may be poor if it was hurriedly administered.	Para data Based	An interview is flagged when its standardised section time is below -1.6.	Consider all completed interviews in that week. For each enumerator, calculate their flagged interviews as a share of interviews completed by them. Within each State-gender category, report and intervene on the top three enumerators with positive shares (if any).
8	Section8 Time	Effort	Time taken to field Section 8, Decision Making.	Data quality of Decision Making section may be poor if it was hurriedly administered.	Para data Based	An interview is flagged when its standardised section time is below -1.6.	Consider all completed interviews in that week. For each enumerator, calculate their flagged interviews as a share of interviews completed by them. Within each State-gender category, report and intervene on the top three enumerators with positive shares (if any).

[1] S. No.	[2] Flag name	[3] Performance Dimension Being Evaluated	[4] Description of Field Practice Being Monitored	[5] Description of Underlying Concern	[6] Constructed Using [Para Data/Survey Data]	[7] Flag Criteria	[8] Criteria for Ranking Enumerators to Initiate Intervention
				Table 2: Para Data Flags to Monitor Enumerator Performance, India Working Survey (3/5)			
9	Section9 Time	Effort	Time taken to field Section 9, Intergenerational Mobility.	Data quality of Intergenerational Mobility section may be poor if it was hurriedly administered.	Para data based	An interview is flagged when its standardised section time is below -1.6.	Consider all completed interviews in that week. For each enumerator, calculate their flagged interviews as a share of interviews completed by them. Within each State-gender category, report and intervene on the top three enumerators with positive shares (if any).
10	Section10 Time	Effort	Time taken to field Section 10, Social Networks.	Data quality of Social Networks section may be poor if it was hurriedly administered.	Para data based	An interview is flagged when its standardised section time is below -1.6.	Consider all completed interviews in that week. For each enumerator, calculate their flagged interviews as a share of interviews completed by them. Within each State-gender category, report and intervene on the top three enumerators with positive shares (if any).
11	Roster Size	Ethics	Number of household members recorded in the household roster.	The enumerator may have deliberately left out some household members if other considerations, such as availability for interview, were taken into account. This would adversely affect survey representativeness.[3]	Survey data based	An enumerator is flagged when their standardised average roster size is below -1.	Consider all completed interviews in that week. Report and intervene on all enumerators (if any), whose average roster size is one standard deviation below the average for the comparison group.
12	Network Size	Content	Count of persons in the respondent's social network.	Some persons in the respondent's social network may have been left out if the enumerator misunderstood the definition of social network or is not able to explain the concept properly to the respondent. This would bias any analysis done using this section.	Survey data Based	An enumerator is flagged when their standardised average network size is below -1.	Consider all completed interviews in that week. Report and intervene on all enumerators (if any), whose average network size is one standard deviation below the average for the comparison group.

[1] S. No.	[2] Flag Name	[3] Performance Dimension Being Evaluated	[4] Description of Field Practice Being Monitored	[5] Description of Underlying Concern	[6] Constructed Using [Para Data/Survey Data]	[7] Flag Criteria	[8] Criteria for Ranking Enumerators to Initiate Intervention
				Table 2: Para Data Flags to Monitor Enumerator Performance, India Working Survey (4/5)			
13	Odd Start	Ethics	Whether interview started at a time outside the usual survey hours.	May indicate falsification of data as it is not possible that the respondent was present at the odd start time of the interview.	Para data based	An interview is flagged if it started between 9 pm and 6 am.	Consider all (complete and incomplete) interviews in that week. Report and intervene on all enumerators with even a single interview starting at an odd time. The concept of a comparison group is irrelevant for this flag.
14	Alone Section7	Ethics	Whether the enumerator interviewed the respondent in private when fielding the Discrimination section.[4]	Enumerators were asked to record if they were successful in ensuring complete privacy when administering the Discrimination section. If an enumerator is recording being always alone or never alone with the respondent, then it is most likely that they are not documenting the privacy status correctly, rendering this information useless for analysis.	Survey data based	An enumerator is flagged if the share of interviews they completed where the respondent is reported as being interviewed in private, is either 1 (always alone) or 0 (never alone).	Consider all complete interviews in that week. Report and intervene on all enumerators who state being always alone or never alone with their respondents when administering the Discrimination section. The concept of a comparison group is irrelevant for this flag.
15	Section4 Skip	Content, Effort, Ethics	Whether the respondent is reported as 'Not Working'.[5]	A battery of probing questions was included to ensure that no work activity goes unreported. The concern is that the enumerator is not clear about what constitutes work, or did not probe enough, or deliberately recorded no work to avoid subsequent sections.	Survey based	An interview is flagged if the respondent is reported as not working: Question 411a is NOT missing.	Consider all complete interviews in that week. For each enumerator, calculate flagged interviews as a share of interviews they completed. Within each State-gender, report and intervene on the top three enumerators with positive shares (if any).[6]

Table 2: Para Data Flags to Monitor Enumerator Performance, India Working Survey; Table Notes (5/5)
[1]All sections that fulfil the following two conditions are included in Survey Time: it must be applicable to every respondent and it must NOT be linked to the respondent's work profile.
[2]Standardised survey time refers to the z-score of Survey Time created using the mean and standard deviation in the enumerator's comparison group.
All other standardised variables in the table are similarly defined as z-scores constructed using the respective comparison group mean and standard deviation.
[3]From each household, the two main respondents were randomly selected from amongst adult male and female members recorded in the household roster.
[4]The most preferred scenario is for the entire interview to be administered privately to the respondent. But, given field realities, this was not always feasible. Given the particularly
sensitive nature of the Discrimination section, enumerators were asked to do their best to administer this section without others being within earshot. They were also asked to
record whether they were successful in doing so. The idea was to use this information to get a sense of how likely it is that the data represent the true views of the respondent.
[5]One important objective of the IWS is to correctly capture respondent's work status, especially that of women whose work tends to be under-reported.
[6]Effectively, for this flag, the comparison group is State-gender and not State-region-gender.

Source: IWS Para Data.

		Descriptive Statistics				Regression Results			
title									

Table 3: Effect of Para Data-based Interventions on Enumerator Performance (1/2)

[1] S. No.	[2] Flag Name	[3] Number Flagged of All Enumerators with at least 1 Completed Interview	[4] Mean over Interviews of All Enumerators in Pre-Intervention Period	[5] Mean over Interviews of Flagged Enumerators in Pre-Intervention Period	[6] Number Flagged of Enumerators with at least 10 Completed Interviews	[7] Enumerator Flagged for Same Field Practice	[8] Enumerator Flagged for at least One Other Field Practice	[9] R squared	[10] No. of Observations (Completed Interviews)
1	Survey Time (minutes)	12 of 88	14.5	12.2	12 of 75	0.029	0.614*	0.27	3013
			(5.8)	(4.3)		(0.548)	(0.342)		
2	Section2 Time (minutes)	5 of 46[1]	2.2	2.1	5 of 39[1]	0.463	0.104	0.18	1779
			(0.9)	(1.0)		(0.367)	(0.087)		
3	Section4 Time (minutes)	2 of 88	4.1	2.3	1 of 75	0.732*	0.370*	0.05	3009
			(6.4)	(1.8)		(0.375)	(0.192)		
4	Section5 Time (minutes)	1 of 88	1.9	1.6	1 of 75	-0.025	0.228***	0.21	3009
			(1.1)	(1.1)		(0.270)	(0.085)		
5	Section8 Time (minutes)	1 of 88	1.6	2.1[2]	1 of 75	-0.237	0.047	0.18	3009
			(1.2)	(0.8)		(0.191)	(0.077)		
6	Section9 Time (minutes)	6 of 88	2.0	1.8	5 of 75	0.134	-0.092	0.15	3008
			(1.0)	(0.9)		(0.170)	(0.073)		
7	Section10 Time (minutes)	7 of 88	4.2	3.8	5 of 75	0.059	0.048	0.24	3009
			(2.5)	(2.3)		(0.326)	(0.159)		

[1] S. No.	[2] Flag Name	Descriptive Statistics				Regression Results			
		[3] Number Flagged of All Enumerators with at least 1 Completed Interview	[4] Mean over Interviews of All Enumerators in Pre-Intervention Period	[5] Mean over Interviews of Flagged Enumerators in Pre-Intervention Period	[6] Number Flagged of Enumerators with at least 10 Completed Interviews	[7] Enumerator Flagged for Same Field Practice	[8] Enumerator Flagged for at least one Other Field Practice	[9] R squared	[10] No. of Observations (Completed Interviews)
8	Network Size (members)	7 of 88	3.2	2.4	6 of 75	0.201	-0.006	0.47	2994
			(1.5)	(1.2)		(0.220)	(0.087)		
9	Alone Section7 (1 if alone, 0 otherwise)	37 of 88	0.87	0.93	36 of 75	-0.052*	0.013	0.35	3009
						(0.031)	(0.031)		
10	Section4 Skip (1 if follow up section not needed, 0 otherwise)	12 of 88	0.32	0.42	11 of 75	-0.071	-0.070**	0.17	3009
						(0.055)	(0.034)		

The table examines the set of interventions based on the first report, which flagged enumerators based on their performance in the week from

February 17 to February 23. Section0 Time, Section1 Time, Section3 Time, Roster Size and Odd Start are omitted from the table as none of the

enumerators was flagged for these in the first report.

The descriptive statistics are limited to enumerators who completed at least 1 interview (by the strict definition of a completed interview), while

the regressions are limited to enumerators who completed at least 10 interviews.

The regression analysis is based on enumerator performance between February 17 and March 10 for Karnataka, and between February 17 and

March 14 for Rajasthan. Regressions are at the interview level, and the dependent variable is indicated under the column Flag Name.

Standard deviations (for descriptive statistics)/robust standard errors (for regression coefficients) are shown in parentheses.

* stands for statistical significance at the 10 per cent level of significance, ** at 5 per cent, and *** at 1 per cent.

[1]Section2 was only administered by female enumerators.

[2]The mean for flagged enumerators could be higher than the mean for all enumerators because flags were generated based on performance between

February 17-23, whereas the means are based on performance over a longer time period, namely, the pre-intervention period. The pre-intervention

period starts February 17 and goes all the way till the date of intervention (March 4 in Rajasthan and March 5 in Karnataka).

Source: IWS Para Data.

| Table 4: Para Data on Enumerator Characteristics, Pre-intervention Period	
Number of enumerators who *initiated* at least 1 interview in pre-intervention period	90
of which	
Karnataka (in %)	48.9
Female (in %)	46.7
Number of enumerators who submitted the voluntary enumerator survey	65
Enumerator Characteristics Conditional on Submission of Enumerator Survey, (65 enumerators)	
Karnataka (in %)	64.6
Female (in %)	50.8
Mean Age in years (Standard Deviation)	27.4 (4.5)
Caste (in %)	
Scheduled Caste/ Scheduled Tribe (SC/ST)	26.2
Other Backward Classes & Non-Brahmin	61.5
Brahmin	12.3
Religion (in %)	
Hindu	98.5
Muslim	1.5
Education (in %)	
Below Bachelor's	7.7
Bachelor's	32.3
Above Bachelor's	60.0
Proficient in (speaking and understanding) English (in %)	24.6
Conducted 3 or more surveys prior to IWS (in %)	69.2
Mean Test Score, per cent answered correctly[1] (Standard Deviation)	59.4 (0.10)
[1]The test was designed to measure mastery over the IWS questionnaire and consisted of 22 questions.	

Source: IWS Para Data.

Table 5: Predictors of Enumerator Performance, Analysis based on Pre-Intervention Period						
	[1] Interview Completion Rate	[2] Survey Time	[3] Section 4 Time	[4] Network Size	[5] Section4 Skip	[6] At least one Flag
Panel A: Enumerator Performance, Pre-Intervention period						
Mean (Std. dev.)	0.69 (0.25)	16.0 (12.4)	4.4 (1.4)	3.0 (1.0)	0.28 (0.21)	0.62
Units for mean value	fraction of initiated interviews	minutes	minutes	persons	fraction of completed interviews	fraction of enumerators
No of observations/enumerators[1]	90	87	87	87	87	87
Panel B: Regression Results on whether Submission of Enumerator Survey Matters for Performance						
Submitted Enumerator Survey	0.209*** (0.063)	-2.079** (1.023)	-0.871* (0.445)	0.481* (0.254)	-0.051 (0.049)	0.026 (0.140)
No of observations/enumerators[1]	90	87	87	87	87	87
R-squared	0.41	0.01	0.18	0.36	0.37	0.02
Panel C: Regression Results on Predictors of Enumerator Performance Conditional on Participation in Enumerator Survey						
Age	0.008* (0.005)	-0.128 (0.220)	0.063* (0.036)	-0.001 (0.026)	-0.003 (0.005)	-0.008 (0.015)
Caste (SC/ST omitted)						
Other Backward Classes and Non-Brahmin	-0.001 (0.046)	2.306 (2.394)	0.535 (0.337)	0.627***(0.223)	-0.016 (0.050)	-0.120 (0.158)
Brahmin	0.168*** (0.060)	-0.422 (2.808)	0.067 (0.645)	0.116 (0.313)	-0.003 (0.072)	-0.433* (0.239)
Education (Below Bachelor's omitted)						
Bachelor's	0.084 (0.061)	4.733 (5.452)	-0.864 (0.529)	0.891*** (0.296)	0.056 (0.075)	0.335 (0.238)
Above Bachelor's	0.109* (0.058)	6.061 (6.656)	-0.466 (0.532)	0.700** (0.261)	0.012 (0.070)	0.309 (0.221)
Proficient in English	-0.053 (0.048)	6.309 (6.528)	-0.021 (0.351)	-0.032 (0.196)	-0.036 (0.042)	0.195 (0.135)
Conducted 3 or more surveys prior to IWS	-0.046 (0.041)	-2.496 (4.014)	-0.234 (0.435)	0.357* (0.190)	-0.051 (0.053)	0.139 (0.138)
Test Score	0.083 (0.165)	6.367 (12.493)	1.787 (1.317)	1.611* (0.864)	-0.537* (0.288)	-0.306 (0.643)
No of observations/enumerators[1]	65	65	65	65	65	65
R-squared	0.53	0.09	0.28	0.58	0.47	0.13
See Table 2 for definitions of Survey time, Section4 time, Network size and Section4 skip, and for a list of all fifteen flags studied in this report. The means shown						
in panel A are over enumerators (and not over interviews), and are therefore different from those shown in Table 3. Both sets of regressions in panels B and C are						
at the enumerator level and include enumerator's gender and State as additional controls. Robust standard errors are shown in parentheses. * stands for						

statistical significance at the 10 per cent level of significance, ** at 5 per cent and *** at 1 per cent.
[1]Column 1 of panels A and B includes all enumerators who *initiated* at least one interview in the pre-intervention period (90). The remaining columns include those
who *completed* at least one interview in the pre-intervention period (87). Panel C is limited to enumerators who submitted the enumerator survey (65).

Source: IWS Para Data.

Figures

Figure 1: Cumulative Count of Completed Interviews

Cumulative Count of Completed Interviews
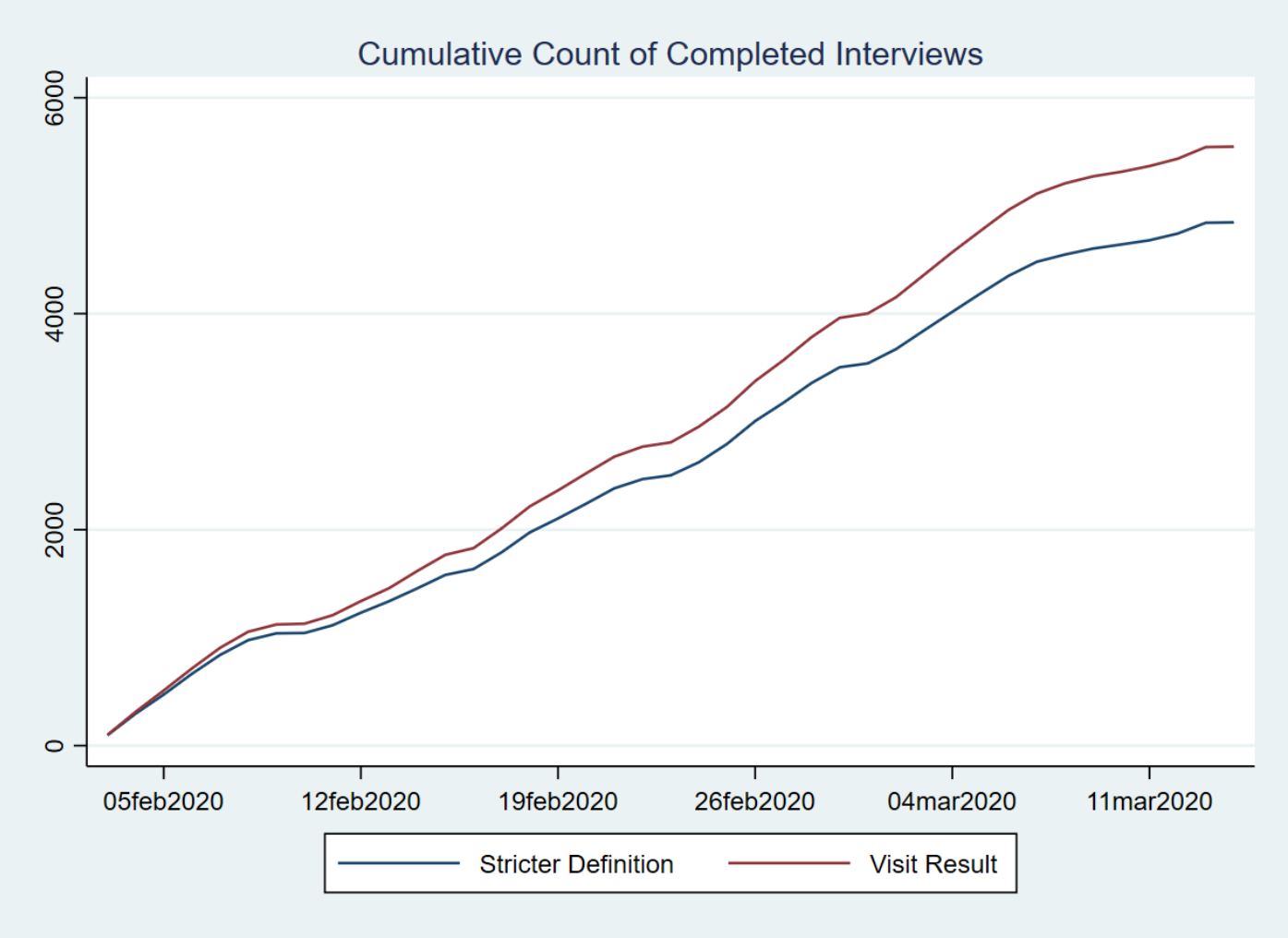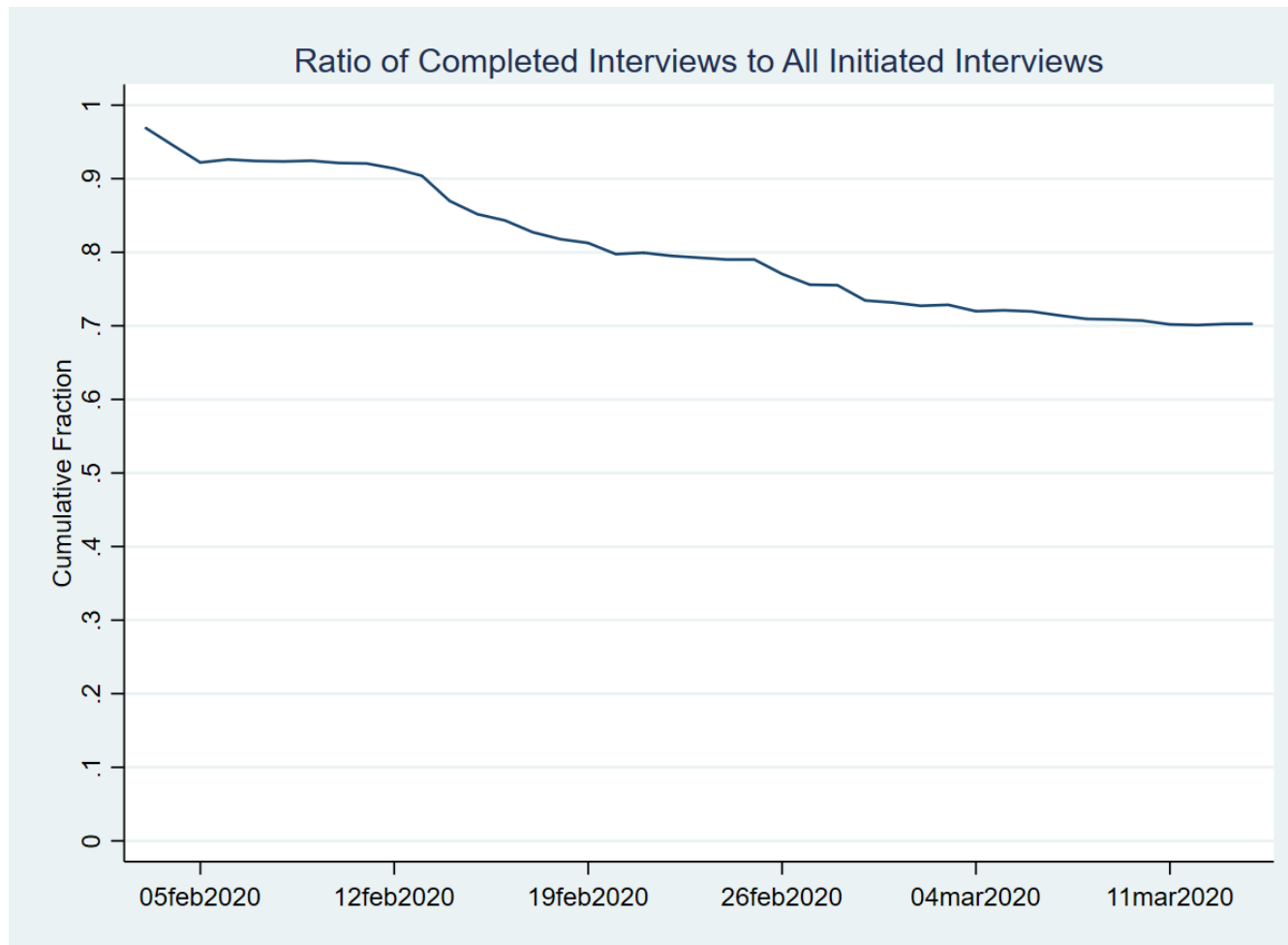
Source: IWS Para Data.

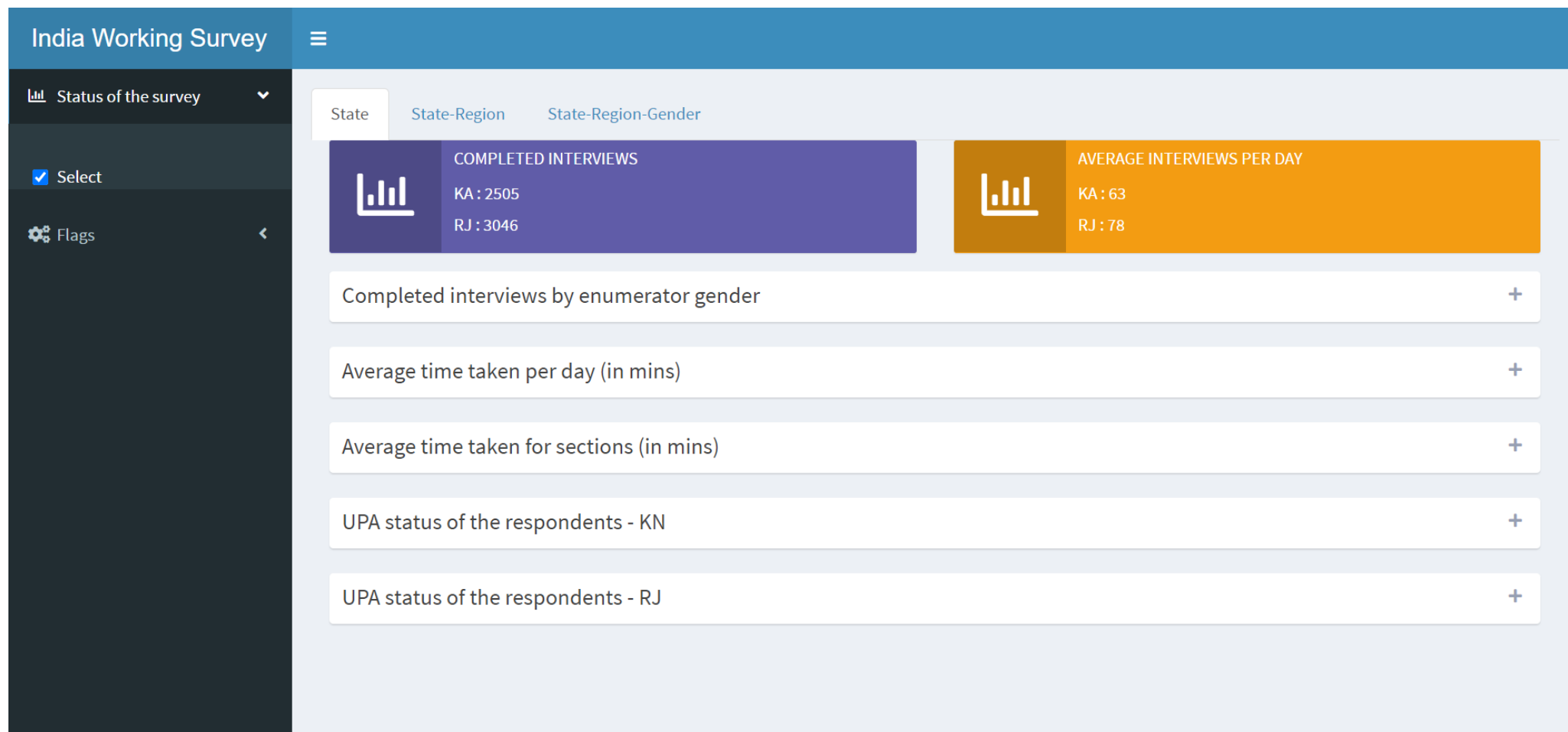**Figure 2: Average Time Spent per Completed Interview**
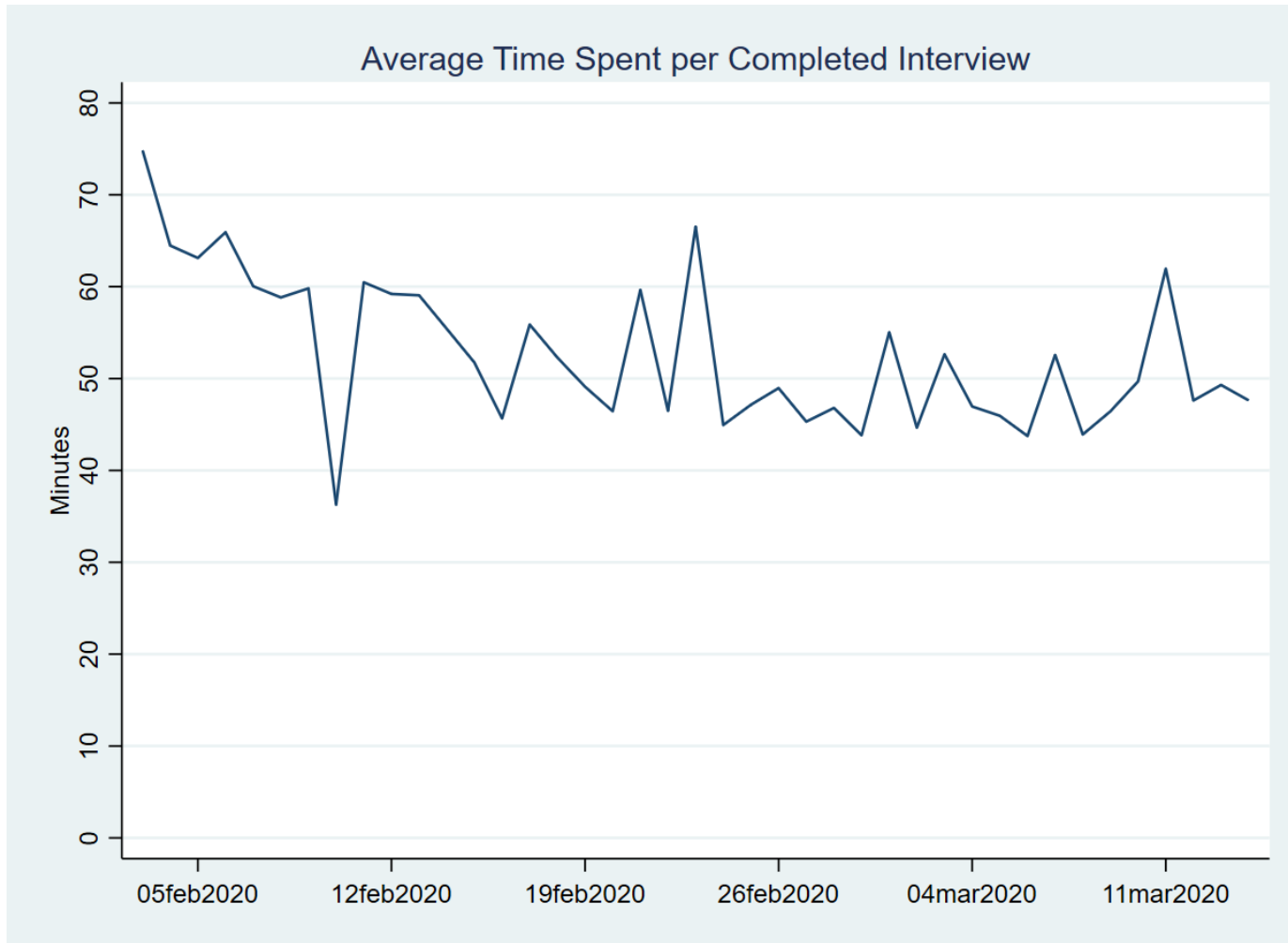
Source: IWS Para Data.

Figure 3: Ratio of Completed Interviews to All Initiated Interviews
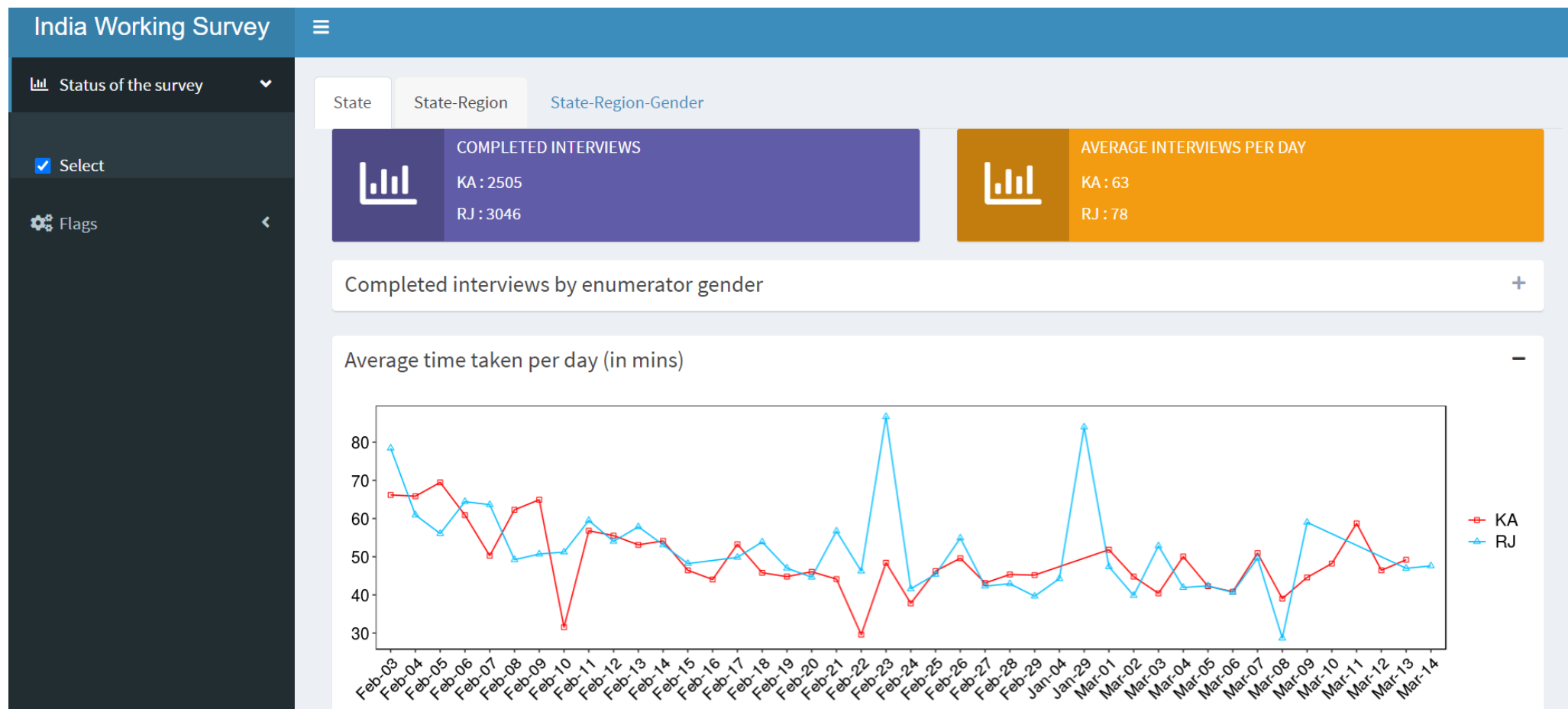

Source: IWS Para Data.

Figure 4A: IWS Dashboard to Monitor Survey Progress

Source: IWS Dash Board.

Figure 4B: IWS Dashboard to Monitor Survey Progress

Source: IWS Dash Board.

Appendices

Appendix 1: Organisation of the First Wave IWS Questionnaire

Here we present the list of sections (or modules) that form the IWS questionnaire for the first wave field survey.

Section Number	Section Name	Description	Question Count Female Enumerator	Question Count Male Enumerator
Section 0	Household Register	Household roster. Only fielded by the female enumerator.	6	
Section 1	Demographic Characteristics	Demographic information such as religion, caste, education, and major work status. The female enumerator recorded this for all household members, while the male enumerator recorded this only for the selected male respondent.	23	13
Section 2	Household Living Standards	Information about the dwelling, household amenities, and assets. Only fielded by the female enumerator.	12	
Section 3	Activity Profile for the Last Year	Major work activity status and skill of the respondent.	26	26
Section 4	Weekly Labour Force Status	Detailed information on respondent's activities in the week prior to the date of interview.	58	58
Section 5	Household Production Activities	Time spent by the respondent on household production activities in the day before the date of interview.	12	12
Section 6	Life History Calendar	This section was administered on paper and not using CAPI. Para data for it is not available.		
Section 7	Discrimination	Perceptions and attitudes on gender, caste and religion related to work; experiences of discrimination by wage workers and non-farm businesses.	30	30
Section 8	Decision Making	How are decisions made within the household.	12	12
Section 9	Intergenerational Mobility	Respondent's parents' education and what they did.	9	9
Section 10	Social Networks	Respondent's social contacts and help extended by them	5	5
Section 11	Women Out of the Work Force	Information about those women respondents who reported 'not working' as their major work status.	8	8
Section 12	Students	Information about respondent who reported 'studying or attending an education institution' as their major work status.	2	2
Section 13	Unemployed	Information about respondent who reported being 'unemployed' as their major work status.	9	9

The questionnaire was administered using CAPI by a male and female enumerator pair. Section 0 was fielded by the female enumerator to any adult member of the household, available and capable of providing information on all members. The remaining sections were variously administered to one adult male, and one adult female, picked at random from among the members of the household. The female enumerator administered the survey to the female respondent, and likewise for the males. The question count is approximate, and is shown to give a rough sense of the length of each section. Whether a particular section is fielded or not, and its length, may vary by the enumerator's gender.

Enumerator Report for Week: 17th February through 23rd February

Find below the list of flagged enumerators. They have been flagged because as compared to the other enumerators from their State, they are doing something very different in some of their interviews. It is, therefore, important for the supervisors to talk to them and figure out why this is the case. **The flagged enumerators may not necessarily be doing something wrong. It is important that the supervisors do not assign blame when talking to enumerators.**

Below the list of flagged enumerators under each flag, you will also find information about the particular interviews for which the enumerator is being flagged. For example, if an enumerator is being flagged for survey time, interviews that he/she conducted which took very little time are shared with you. For the first flag, that is, Survey Time, the data on interviews is given in a separate Excel file, but for all other flags, the interviews are shared in this report itself. The supervisors may want to use the information about specific interviews if it helps them when talking to enumerators.

1) **Survey Time:** These enumerators are taking less time to complete some important sections in the survey. The sections being tracked are— Section 1: Demographic Characteristics, Section 5: Household Production Activities, Section 7A: Discrimination, Section 8: Decision Making, and Section 10: Networks. The concern here is that the flagged enumerators are going through the survey very fast, taking very short times for their interviews, and this may result in poor data quality.

(Average duration per interview for female enumerators is 14.42 minutes.)
(Average duration per interview for male enumerators is 13.95 minutes.)

In the table below, the value in parentheses against each enumerator shows the share of completed interviews by the enumerator that had short survey time.[13]

Gender	KA
Female(s)	Ms. X1 (0.37)
	Ms. X2 (0.33)
	Ms. X3 (0.3)
Male(s)	Mr. Y1 (0.5)
	Mr. Y2 (0.42)
	Mr. Y3 (0.33)

The flagged interviews against each enumerator are given in a separate Excel file called KA_SurveyTime_FlagInterviews**.**

---

[13] Actual enumerator names have been replaced with placeholders to protect the privacy of the flagged enumerators.

2) **Section0 Time:** These enumerators are taking less time to complete Section 0, 'The Household Register'. The concern here is that the flagged enumerators are going through the section very fast, taking very short times for their interviews, and this may result in poor data quality.

(Average duration per interview for this section for female enumerators is 3.5 minutes.)
(Average duration per interview for this section for male enumerators is 3.3 minutes.)

In the table below, the value in parentheses against each enumerator shows the share of completed interviews by the enumerator that had short section time.

Gender	KA
Female(s)	
Male(s)	

3) **Section1 Time:** These enumerators are taking less time to complete Section 1, 'Demographic Characteristics'. The concern here is that the flagged enumerators are going through the section very fast, taking very short times for their interviews, and this may result in poor data quality.

(Average duration per interview for this section for female enumerators is 3.7 minutes.)
(Average duration per interview for this section for male enumerators is 2.1 minutes.)

In the table below, the value in parentheses against each enumerator shows the share of completed interviews by the enumerator that had short section time.

Gender	KA
Female(s)	
Male(s)	

4) **Section2 Time:** These enumerators are taking less time to complete Section 2, 'Household Living Standards'. The concern here is that the flagged enumerators are going through the section very fast, taking very short times for their interviews, and this may result in poor data quality. This flag is only applicable for female enumerators, as male enumerators do not field this section.

(Average duration per interview for this section for female enumerators is 2.1 minutes.)

In the table below, the value in parentheses against each enumerator shows the share of completed interviews by the enumerator that had a short section time.

Gender	KA
Female(s)	Ms. X1 (0.25)
	Ms. X2 (0.16)
Male(s)	

The flagged interviews against each female enumerator are given below.

surveyorname	starttime	endtime	districtname	villagename	hhid	Respondent_Name	headname	sec2_duration (secs)
XXX	18 February 2020 1:40:29 PM	2020-02-19 21:54:15	XXX	XXX	XXX	XXX	XXX	54
XXX	18 February 2020 10:38:11 AM	2020-02-18 11:14:50	XXX	XXX	XXX	XXX	XXX	52

5) **Section3 Time:** These enumerators are taking less time to complete Section 3, 'Activity Profile for the Last Year'. The concern here is that the flagged enumerators are going through the section very fast, taking very short times for their interviews, and this may result in poor data quality.

(Average duration per interview for this section for female enumerators is 2.5 minutes.)
(Average duration per interview for this section for male enumerators is 3.2 minutes.)

In the table below, the value in parentheses against each enumerator shows the share of completed interviews by the enumerator that had a short section time.

Gender	KA
Female(s)	
Male(s)	

6) **Section4 Time:** These enumerators are taking less time to complete Section 4, 'Weekly Labour Force Status'. The concern here is that the flagged enumerators are going through the section very fast, taking very short times for their interviews, and this may result in poor data quality.

(Average duration per interview for this section for female enumerators is 3.4 minutes.)
(Average duration per interview for this section for male enumerators is 4.9 minutes.)

In the table below, the value in parentheses against each enumerator shows the share of completed interviews by the enumerator that had a short section time.

Gender	KA
Female(s)	Ms. X1 (0.14)
Male(s)	Mr. Y1 (0.5)

The flagged interviews against each enumerator are given below.

surveyorname	starttime	endtime	districtname	villagename	hhid	Respondent_Name	headname	sec4_duration(secs)
XXX	15 February 2020 11:08:36	2020-02-17 20:07:07	XXX	XXX	XXX	XXX	XXX	41

XXX	18 February 2020 1:05:12	2020-02-18 18:09:32	XXX	XXX	XXX	XXX	XXX	22

7) **Section5 Time:** These enumerators are taking less time to complete Section 5, 'Time Spent on Household Production Activities'. The concern here is that the flagged enumerators are going through the section very fast, taking very short times for their interviews, and this may result in poor data quality.

(Average duration per interview for this section for female enumerators is 2.0 minutes.)
(Average duration per interview for this section for male enumerators is 2.0 minutes.)

In the table below, the value in parentheses against each enumerator shows the share of completed interviews by the enumerator that had a short section time.

Gender	KA
Female(s)	Ms. X1 (0.1)
Male(s)	

The flagged interviews against each enumerator are given below.

surveyorname	starttime	endtime	districtname	villagename	hhid	Respondent_Name	headname	sec5_duration
XXX	4 February 2020 1:56:37 PM	2020-02-22 11:46:46	XXX	XXX	XXX	XXX	XXX	29

8) **Section8 Time:** These enumerators are taking less time to complete section 8, 'Decision Making'. The concern here is that the flagged enumerators are going through the section very fast, taking very short times for their interviews, and this may result in poor data quality.

(Average duration per interview for this section for female enumerators is 1.7 minutes.)
(Average duration per interview for this section for male enumerators is 1.5 minutes.)

In the table below, the value in parentheses against each enumerator shows the share of completed interviews by the enumerator that had a short section time.

Gender	KA
Female(s)	Ms. X1 (0.125)
Male(s)	

The flagged interviews against each enumerator are given below.

surveyorname	starttime	endtime	districtname	villagename	hhid	Respondent_Name	headname	sec8_duration
XXX	18 February 2020 11:17:55 AM	2020-02-18 21:31:07	XXX	XXX	XXX	XXX	XXX	0

9) **Section9 Time:** These enumerators are taking less time to complete Section 9, 'Intergenerational Mobility'. The concern here is that the flagged enumerators are going through the section very fast, taking very short times for their interviews, and this may result in poor data quality.

(Average duration per interview for this section for female enumerators is 2.0 minutes.)
(Average duration per interview for this section for male enumerators is 2.2 minutes.)

In the table below, the value in parentheses against each enumerator shows the share of completed interviews by the enumerator that had a short section time.

Gender	KA
Female(s)	Ms. X1 (0.4)
	Ms. X2 (0.33)
	Ms. X3 (0.25)
Male(s)	Mr. Y1 (0.25)

	Mr. Y2 (0.08)	

The flagged interviews against each enumerator are given below.

surveyorname	starttime	endtime	districtname	villagename	hhid	Respondent_Name	headname	sec9_duration (secs)
XXX	18 February 2020 12:07:13 PM	2020-02-18 17:28:41	XXX	XXX	XXX	XXX	XXX	50
XXX	4 February 2020 1:56:37 PM	2020-02-22 11:46:46	XXX	XXX	XXX	XXX	XXX	14
XXX	6 February 2020 12:22:22 PM	2020-02-22 11:40:02	XXX	XXX	XXX	XXX	XXX	13
XXX	18 February 2020 11:17:55 AM	2020-02-18 21:31:07	XXX	XXX	XXX	XXX	XXX	0
XXX	15 February 2020 1:51:44 PM	2020-02-17 09:00:33	XXX	XXX	XXX	XXX	XXX	37
XXX	3 February 2020 11:41:02 AM	2020-02-22 20:58:19	XXX	XXX	XXX	XXX	XXX	18
XXX	3 February 2020 3:40:03 PM	2020-02-23 18:23:11	XXX	XXX	XXX	XXX	XXX	21
XXX	4 February 2020 10:35:49 AM	2020-02-23 18:12:51	XXX	XXX	XXX	XXX	XXX	27
XXX	5 February 2020 12:03:23 PM	2020-02-23 18:31:43	XXX	XXX	XXX	XXX	XXX	15
XXX	19 February 2020 9:21:30 AM	2020-02-19 22:30:31	XXX	XXX	XXX	XXX	XXX	54
XXX	7 February 2020 4:05:01 PM	2020-02-23 12:35:45	XXX	XXX	XXX	XXX	XXX	32
XXX	7 February 2020 12:11:35 PM	2020-02-23 19:48:12	XXX	XXX	XXX	XXX	XXX	0

10) **Section10 Time:** These enumerators are taking less time to complete Section 10, 'Social Networks'. The concern here is that the flagged enumerators are going through the section very fast, taking very short times for their interviews, and this may result in poor data quality.

(Average duration per interview for this section for female enumerators is 3.0 minutes.)
(Average duration per interview for this section for male enumerators is 4.4 minutes.)

In the table below, the value in parentheses against each enumerator shows the share of completed interviews by the enumerator that had a short section time.

Gender	KA
Female(s)	Ms. X1 (0.2)
	Ms. X2 (0.125)
Male(s)	Mr. Y1 (0.16)
	Mr. Y2 (0.125)

The flagged interviews against each enumerator are given below.

surveyorname	starttime	endtime	districtname	villagename	hhid	Respondent_Name	headname	sec10_duration (secs)
XXX	4 February 2020 1:56:37	2020-02-22 11:46:46	XXX	XXX	XXX	XXX	XXX	0
XXX	6 February 2020 12:22:22	2020-02-22 11:40:02	XXX	XXX	XXX	XXX	XXX	0
XXX	18 February 2020 11:17:55	2020-02-18 21:31:07	XXX	XXX	XXX	XXX	XXX	0
XXX	15 February 2020 9:55:27	2020-02-17 15:19:31	XXX	XXX	XXX	XXX	XXX	35
XXX	15 February 2020 11:23:03	2020-02-17 14:24:21	XXX	XXX	XXX	XXX	XXX	44

11) **Roster Size:** These enumerators are recording a smaller number of individuals within a household as eligible for interview. The concern is that they may be deliberately leaving out some eligible adults and only noting those who are available at the time of the first visit.

(Average roster size per interview for female enumerators is 4.43 members.)
(Average roster size per interview for male enumerators is 4.38 members.)

Gender	KA
Female(s)	
Male(s)	

12) **Network Size:** These enumerators are recording a smaller network size of the main respondent, that is, they are recording that the respondent knows very few people. The concern is that they are not probing enough to get the full network of the respondent.

(Average network size per interview for female enumerators is 2.04 persons.)
(Average network size per interview for male enumerators is 2.60 persons.)

Gender	KA
Female(s)	Ms. X1
Male(s)	

13) **Odd Start Time:** These enumerators have been flagged because they are reporting odd start times (between 9 pm and 6 am) for some of their interviews.

Gender	KA
Female(s)	

Male(s)	

14) Alone Section7: These enumerators are either reporting that they are 'Always Alone' or are 'Never Alone' with the main respondent for all their interviews. This does not sound truthful, as one would expect some variation in being able to find the respondent all alone when asking questions in Section 7 Discrimination. It is important to stress to the enumerators that they should note the true environment in which they interviewed the respondent when asking questions in Section 7.

If the enumerator is having 1, it means that the enumerators are reporting that they are 'Always Alone' with the main respondents for all their interviews. If enumerator is having 0, it means that the enumerators are reporting that they are 'Never Alone' with the main respondents for all their interviews.

Gender	KA
Female(s)	Ms. X1 (1)
	Ms. X2 (1)
	Ms. X3 (1)
	Ms. X4 (1)
	Ms. X5 (1)
	Ms. X6 (1)
	Ms. X7 (1)
	Ms. X8 (1)
	Ms. X9 (0)
Male(s)	Mr. Y1 (1)
	Mr. Y2 (1)
	Mr. Y3 (1)
	Mr. Y4 (1)
	Mr. Y5 (1)
	Mr. Y6 (1)

	Mr. Y7 (1)
	Mr. Y8 (1)
	Mr. Y9 (1)
	Mr. Y10 (1)
	Mr. Y11 (0)

15) **Section4 Skip:** In Section 4 on 'Weekly Labour Force Status', these enumerators are recording that in the last week, the main respondent was not engaged in any work activity. The concern is that they are either not probing enough about work or are recording this so as to skip other questions related to work.

In the table below, the value in parentheses against each enumerator shows the share of completed interviews by the enumerator wherein the latter recorded the respondent as 'Not Working'.

Gender	KA
Female(s)	Ms. X1 (0.8)
	Ms. X2 (0.7)
	Ms. X3 (0.7)
Male(s)	Mr. Y1 (0.7)
	Mr. Y2 (0.5)
	Mr. Y3 (0.5)

The flagged interviews against each enumerator are given below (the enumerator marked the respondent as Not Working).

surveyorname	starttime	endtime	districtname	villagename	hhid	Respondent_Name	headname
XXX	15 February 2020 11:08:36 AM	2020-02-17 20:07:07	XXX	XXX	XXX	XXX	XXX
XXX	15 February 2020 1:51:44 PM	2020-02-17 09:00:33	XXX	XXX	XXX	XXX	XXX
XXX	17 February 2020 2:32:27 PM	2020-02-20 20:13:09	XXX	XXX	XXX	XXX	XXX

XXX	17 February 2020 8:36:45 AM	2020-02-17 20:55:08	XXX	XXX	XXX	XXX	XXX
XXX	18 February 2020 10:36:36 AM	2020-02-18 17:09:02	XXX	XXX	XXX	XXX	XXX
XXX	18 February 2020 11:13:10 AM	2020-02-18 17:16:01	XXX	XXX	XXX	XXX	XXX
XXX	18 February 2020 2:26:01 PM	2020-02-18 17:31:06	XXX	XXX	XXX	XXX	XXX
XXX	18 February 2020 10:21:43 AM	2020-02-18 21:25:17	XXX	XXX	XXX	XXX	XXX
XXX	18 February 2020 11:16:04 AM	2020-02-18 21:22:07	XXX	XXX	XXX	XXX	XXX
XXX	18 February 2020 11:34:21 AM	2020-02-18 17:22:04	XXX	XXX	XXX	XXX	XXX
XXX	19 February 2020 11:10:18 AM	2020-02-19 12:41:01	XXX	XXX	XXX	XXX	XXX
XXX	19 February 2020 8:55:54 AM	2020-02-20 12:16:21	XXX	XXX	XXX	XXX	XXX
XXX	20 February 2020 10:12:45 AM	2020-02-20 12:49:46	XXX	XXX	XXX	XXX	XXX
XXX	18 February 2020 1:05:12 PM	2020-02-18 18:09:32	XXX	XXX	XXX	XXX	XXX
XXX	18 February 2020 11:36:50 AM	2020-02-18 18:03:38	XXX	XXX	XXX	XXX	XXX
XXX	18 February 2020 12:06:18 PM	2020-02-18 17:56:51	XXX	XXX	XXX	XXX	XXX
XXX	18 February 2020 9:53:16 AM	2020-02-18 18:53:03	XXX	XXX	XXX	XXX	XXX
XXX	17 February 2020 12:13:53 PM	2020-02-18 18:58:43	XXX	XXX	XXX	XXX	XXX

Appendix 3: Relationship between Survey Time Flag and Other Section Time Flags

In this section, we explore the relationship between the Survey Time flag, which captures the overall interview duration, and other flags based on individual section times. We only include those sections for which at least one enumerator was flagged in the first report. These are sections 2, 4, 5, 8, 9, and 10. The motivation for this exercise is to examine whether including only the Survey Time flag allows us to do away with flags based on individual section times.

In the first report, 12 enumerators were flagged for the Survey Time Flag.[14] At the interview level, about 16 per cent of the 868 completed interviews were flagged for Survey Time.[15] Table A3.1 presents these numbers for the other section time flags. It also presents the overlap, if any, between Survey Time and the other section time flags. Figure A3.1 shows the overlap between the six flag pairs in terms of the flagged interviews using gar graphs. If Survey Time is to be a good stand-in for the other section time flags, then conditional on being flagged for a particular section, the interview should also be flagged for Survey Time, that is, the red shaded area in the second bar in each sub-plot should be large. Except for Section 2, the Survey Time flag does a pretty good job in this respect. However, the data points are too few to arrive at a decisive conclusion, and we hope that future work will shed more light on this.

For reasons stated in the main text, we recommend that the PIs be parsimonious in the number of flags they monitor. While the overall Survey Time flag should definitely be monitored, we advise against including individual section flags. At best, one or two core sections may be included.

Table A3.1: Overlap between Survey Time Flag and Other Section Time Flags				
Flag Name	Number of Enumerators Flagged	Number of Flagged Enumerators in Common with Survey Time	Number of Interviews Flagged	Number of Flagged Interviews in Common with Survey Time
Survey Time	12	12	141	141
Section2 Time	5	0	5	0
Section4 Time	2	0	2	1
Section5 Time	1	1	1	1
Section8 Time	1	0	1	1
Section9 Time	6	3	13	9
Section10 Time	7	3	9	7

---

[14] Recall that flags in the first report were created based on performance between February 17 and February 23.

[15] This 16 per cent includes flagged interviews of enumerators who may not be among the 12 flagged enumerators as for an enumerator to be flagged, their flagged interviews as a share of their completed interviews must be among the top three such shares. Thus, in spite of having flagged interviews, an enumerator may not be flagged if his share value is not among the top three.

Figure A3.1: Overlap between Flagged Interviews for Survey Time and Particular Section Times

Survey Time Flag conditional on Section 2 Time Flag

Survey Time Flag conditional on Section 4 Time Flag

Survey Time Flag conditional on Section 5 Time Flag

Survey Time Flag conditional on Section 8 Time Flag

Survey Time Flag conditional on Section 9 Time Flag

Survey Time Flag conditional on Section 10 Time Flag

51