Survey Process – IHDS Wave 3

IHDS is a multi-topic panel survey, conducted across the country. The panel nature of the survey made it distinct from other national level surveys like NFHS or NSS, etc. At the same time, it increased the complexity of the survey operations manyfold. Survey operations of IHDS wave 3 was much different from the earlier two waves. Being a panel study, it does not offer much opportunity to change the whole questionnaire as that distorts the comparability of questions between different rounds. However, to accommodate questions related to new emerging areas, we have modified a few questions, dropped a few themes like perspective of the youths and included a few new themes, like availability of various health insurance schemes to the households, availability of healthcare, education and livelihood opportunities during Covid pandemic in India, etc.

The early IHDS waves were conducted in Pen and Paper Interview (PAPI) mode of data collection, which provided certain benefits during survey preparation, fieldwork and physical monitoring of the filled-in questionnaires. However, PAPI had certain limitations like long post-processing time, difficult to track interviewer's behaviour, relies more on physical manpower, etc., which were possible to overcome using Computer Assisted Personal Interviewing (CAPI) techniques. But, transitioning from PAPI to CAPI was not easy, and required very careful planning and implementation. Please find below the experience of the IHDS team related to this transition.

1. Transitioning from PAPI to CAPI: Learning from IHDS-III

-Bijay Chouhan, Dinesh Kumar Tiwari, Om Prakash Sharma

The shift from Paper and Pencil Interviewing (PAPI) to Computer-Assisted Personal Interviewing (CAPI) offers numerous advantages, including improved data quality, reduced errors due to in-built consistency and range checks, instruction manual with each variable, soft checks and enhanced efficiency along with real time data availability and centralized monitoring. However, moving from PAPI to CAPI adds significant complexity in three areas: (1) Survey design; (2) Interviewer training and support; and, (3) Data security.

Survey Design: CAPI allows for more complex question types and branching logic compared to PAPI which can minimize errors and inconsistencies. While this flexibility can improve data richness, it's crucial to ensure the questionnaire remains clear and respondent-friendly to avoid confusion and careful review and testing of various branching logic is required to ensure that there is no inadvertent data loss. This is particularly important in multi-lingual and heterogeneous societies where social structures vary regionally. For example, the system must be flexible to handle rare cases

of polygamy.

Training and Support: CAPI necessitates investment in devices and software. Organizations must ensure compatibility with existing systems and provide adequate training to core team and interviewers on the new technology. It is also necessary to ensure that interviewers are fully trained and have access to a helpdesk when facing problems in the field. Initial training must include handling the devices and secure data transfer systems. Field team must understand the protocol for reporting software or hardware issues and receiving timely assistance.

Data Security and Confidentiality: CAPI systems often handle sensitive personal data, so organizations must implement robust data security measures to protect respondent confidentiality. This may include encryption, secure storage, and access controls. In addition, interviewers should be trained to adhere strictly to device security protocols aimed at preventing unauthorized access or data breaches.

The IHDS-III team began with excellent technological partnership with University of Michigan's Survey Research Center but also learned that complex structure of Indian society and field conditions required considerable adaptation.

Source: IHDS Research Digest, April, 2024

Technological Transition:

While we started much early for the technological transition from PAPI to CAPI and partnered with one of most recognized global leaders, the University of Michigan, who has designed and fielded the renowned Panel Study of Income Dynamics (PSID) for a longer time period, the journey was not easy. The data model was developed using Blaise software. Blaise software was developed by the Statistics Netherland and the Survey Research Centre (SRC) of the University of Michigan had designed numerous survey instruments using Blaise in the past. However, that wasn't enough for them to design various instruments for IHDS. IHDS had 13 different questionnaires and 12 instruments were for noting interviewers' observation. The questionnaires were translated into 11 different languages and shared among 14 survey agencies. Most of the instruments were inter-linked and some were pre-filled with information from previous waves. These altogether made the task challenging to them.

Apart from the survey instrument design, SRC was responsible for developing a monitoring framework using their own-developed software called the SurveyTrak, which was used to manage sample lines, run the survey, generate process data, and also transfer data to the main server. Please find the experience of the SRC team, who worked with us in execution of the whole project.

2. SurveyTrack: IHDS-III Sample Management Tool

-Stephanie Chardoul, Sarah Broumand

NCAER partnered with the University of Michigan's Survey Research Center (Ann Arbor, MIUSA) to help transition the IHDS-Wave III from paper questionnaire administration to computer-assisted interviewing (CAI). CAI means that interviewers are using an electronic questionnaire application to read the survey to participants, and to enter responses directly into the computer. CAI requires not only electronic surveys, but also an electronic sample management system. The sample management system SurveyTrak, developed by the Survey Research Center (SRC), is robust enough to manage the extremely complex protocol of the IHDS– with almost 48,000 study households in approximately 3,000 communities, 13 different questionnaires, and 12 interviewer observation instruments, all translated into 11 different languages resulting in 321,000 completed surveys.

We modified SurveyTrak to meet the needs of IHDS-III, which included the ability to import all IHDS household data, allow for locating activity and recording of updated contact information, completing a roster of each household to identify eligible members, and the generation of appropriate new questionnaires. For IHDS, there were several sub-projects (Household, Community, Migrant), each with multiple respondents and separate questionnaires. SurveyTrak was required to track the generation of new respondents, the status of each questionnaire, and link them all together under the original IHDS household. SurveyTrak allowed for the allocation of original households to 14 separate agencies who were able to monitor and report on the work of their interviewing teams, while the NCAER management team was able to oversee the entire project. SurveyTrak managed the distribution of cases to individual interviewers, the transfer of cases between interviewers, and the secure transmission of data from the field to IHDS project servers at the University of Michigan. SurveyTrak reporting was used to know the current status of every single component of every single case.

Source: IHDS Research Digest, May 2024

Major Challenges:

One of the major challenges faced during the survey process was to translate the survey questionnaires into 11 different major languages, spoken across the country. Assigning the work to some consulting farm was easier and are frequently preferred. However, we don't have a pleasant experience with these farms for language translations. Since, IHDS is a PAN India survey, we required to administered the same questions across all states without distortion of the meaning of the question due to language. So, we decided to translate the questionnaires in-house with close supervision of the IHDS team. This leads to numerous back and forth revision of the languages to stick to the original meaning and intent of the questions. Find below the excerpt from the January 2024

IHDS Research Digest, where Ranjima K.V. and Bipasa Banerjee shared their translation experience during IHDS wave 3 survey preparations.

3. Challenges of Translating the Survey Questionnaire

-Ranjima K.V., Bipasa Banerjee

One of the distinctive features of IHDS 3 is that it is conducted in 11 Indian languages thus bringing in many challenges. We joined the IHDS team as summer interns working to translate the questionnaire in Malayalam and Bengali. We stayed as research associates helping in training and quality control. Only then did we realize how translating questionnaires is different from general translation. Interesting examples:

- Using Google translate we would get literal translation such as head would get translated as 'matha' in Bengali or 'sar' in Hindi making the intervention of someone with the subject and language knowledge mandatory.
- Cultural and political variation of India made it difficult to find analogous terms in various languages. Gauna (cohabitation ceremony) seemed to have no analog in South India where marriage and cohabitation occur at the same time. Similarly, MGNREGA is commonly known as is commonly called 'Job card' in West Bengal and 'Thozhilurappu' in Kerala.
- During interviewer training we were able to refine wording based on local knowledge from interviewers.

Source: IHDS Research Digest, January 2024

Apart from language translation, we had to undergo several rounds of data model testing to ensure that the programming done by the SRC team at University of Michigan runs on the desired directions. We used to perform those testing in-house in a dummy interviewer-respondent set-up. Once we became sure about the functioning of the programming, we took those data models to the field for pre-testing in a real set-up. We spent several days in the field during the data-model pre-testing phase. After the pre-testing, we consolidated the feedback received from the interviewers from the field on functioning of the program, length of the questionnaire, questionnaire wordings, etc. and modified them for the next round of the testing. After numerous testings, the final data model was adopted for administration in the field.

The next major challenge related to survey preparation was Interviewers' training. It usually takes around 3:00-3:30 hours to complete all modules of the main household instrument. This is a pretty long time-span, sometimes makes it difficult for the respondents to answer at one go. So, we distributed the instruments among the male and female interviewers in a way that reduced the response burden drastically. We

instructed the interviewers to split the interview time and revisit households, allowing the respondents to continue with their usual activities.
Each interviewers' training during IHDS wave 3 lasted for about 21 days. Given the nature of complexity of the survey and to reduce potential information loss due to training of trainers (ToT), we decided that the IHDS core research team will provide training to the interviewers and supervisors directly. We conducted a total of 17 such trainings altogether. Please find the training experience that Mr. Om Prakash Sharma and Dr. Dinesh Tiwari shared in the August 2024 IHDS Research Digest.

4. Comprehensive Training Approach for the India Human Development Survey 3: Ensuring Robust Data Quality

-Om Prakash Sharma, Dinesh Kumar Tiwari

Training for the IHDS-III data collection was a challenging task for the team because of the panel nature and digital collection (Blaise software) and sample management (SurveyTrak). For IHDS-III data collection a total of 17 trainings comprising of 21 days each were organised across various states in India. Given the complexity of the survey, all trainings were organised by the core research faculty (theme experts) of the IHDS team. This is in contrast to the Training of Trainers (ToT) model. Each training session comprised both classroom sessions as well as training in the field. We focused on both the content and the processes (technical and other aspects) of the survey. For the specific instruments, separate parallel sessions were also organized. For instance, setting up training for Eligible Women's instruments for females only and that for community instruments for Supervisors only.

Classroom training consisted of a detailed presentation with an overview of the major concepts across various survey instruments, followed by hands-on training on the survey data model. During the classroom training mock interview sessions were also organized to help interviewers get acquainted with the questions and the associated instructions, along with the skip patterns in the data model. For major sections, we provided interviewers with both online and offline quizzes to solve. Thereafter, the results of the quizzes were discussed focusing on concepts that required further clarification. For several states, we also used regional language translators during training sessions for a better understanding of the colloquial dialects.

For field practice, teams were sent to households in selected villages to collect data from the field itself. A total of 2 such field trainings were done for all training sessions. After each of the field trips, feedback sessions were organised for the interviewers, where, the NCAER faculty shared their observations from the field visit.

Training for the procedural part of the survey, such as recording of all interview attempts for each household that was made to complete the survey, sub-sample generation, offline transfer of the sample to the partner interviewer, technical troubleshooting and data synchronization with the server, was done by the IT personnel of the core team. Practice sessions for the same were also organized towards the end of the training to build interviewers' confidence.

Source: IHDS Research Digest, August 2024

The transition from PAPI to CAPI was aimed at adopting the cutting-age digital data collection technology to ease the survey processes, more accurate and efficient data collection, and reducing the post processing time. It also provided an opportunity to monitor the survey data collection process centrally and on a timely, and efficient manner using various processed data (para data).

Some of the significant benefits of using CAPI mode of data collection were – 1) assigning range to avoid data entry errors, 2) pre-defined response codes, which do not allow other codes, 3) pre-populated roster member to ease remembering names and better interviewing, etc. PAPI mode of data collection never allowed such flexibilities. This often leads to some biased estimates as reported by Sonalde Desai and colleagues in the February 2023 version of the IHDS Research Digest, reposted below.

5. Ensuring Data Quality: The Biggest Challenge

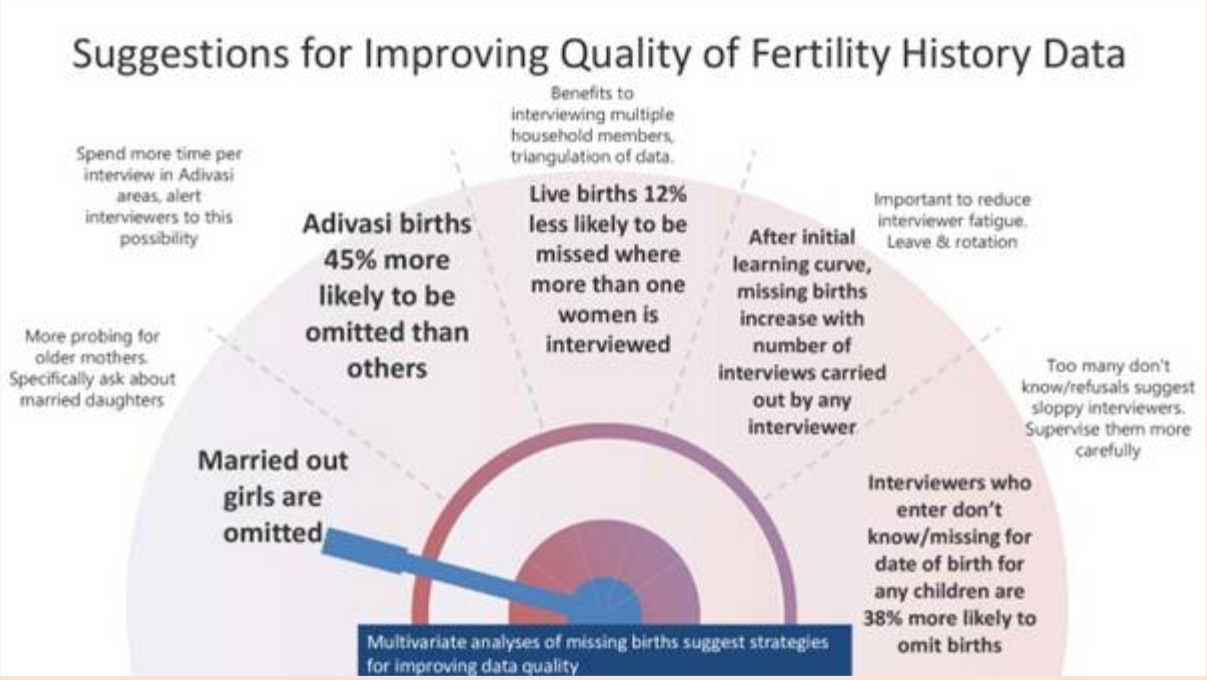
-Sonalde Desai, Sharan Sharma, Om Prakash Sharma, Debasis Barik

Most data users don't realise the primary struggle faced by survey programmes—ensuring data quality. All large survey programmes struggle with it. However, few academic programmes prepare us for strategies to improve data quality. One of the primary goals of IHDS-III is to develop a line of research designed to provide practical guidance to future survey researchers. Research on collecting fertility history, led by Sharan Sharma, O.P. Sharma, and Debasis Barik, under the guidance of Sonalde Desai, provides an interesting example.

IHDS-I and IHDS-II collected fertility history from about 25,000 women who were interviewed in both waves. The IHDS team wanted to check whether births that were reported in Wave I were also reported in Wave II and predictors of mismatch. Matching fertility histories across two waves is challenging; for instance, names may change where a child who was identified by the family name—Munni—in one wave is formally named as Munira in the second wave. Names could also be mis-spelt across interviews, where, for example, Sameer might be called Samir. Most importantly, birth dates could be

approximate or missing. After several rounds of matching algorithms were applied, Om Prakash Sharma, Fellow, NCAER, and his team of coders manually matched thousands of mismatched birth records and identified about 2.5% of the births, reported in Wave I, as being omitted from Wave II.

Most of the omissions were either for children who had died prior to Wave I or for daughters who had probably married out by Wave II. Selective omissions have substantial implications for estimates of sex ratio at birth, infant mortality rate, and other parameters of demographic interest. Analysis of the characteristics of the omissions by Sharan Sharma, Assistant Research Professor, University of Maryland, provides clues to improving data quality in future rounds. These recommendations are summarised below:



Source: IHDS Research Digest, February 2023