



"Gandhi Setu Bridge in Patna, India." by Chandravir Singh

DATA COLLECTION METHODS

KENYATTA UNIVERSITY WORKSHOP

JENNA ALLARD

MAY 14, 2021

WHAT WE WILL COVER

- **Field Planning**

- What type of data do we need?
- How will we recruit respondents?
- How will we collect data?
- How do we structure our teams?

- **Survey Operations**

- Productivity
- Data quality Checks
- Tracking data collection process



FIELD PLANNING

WHAT TYPE OF DATA?

Many observations,
Few indicators

Examples:

- Informal business creation
- Nighttime lights
- Unemployment rates

Key Questions:

- Can you use administrative datasets?
- Can you modify or augment implementation data

As many units as feasible
Indicators and covariates

Examples:

- Household surveys on gender norms
- Business surveys on pollution control equipment

Key Questions:

- Power calculations determine cluster size
- Using sample frames to select respondents

Few observations,
Detailed information

Examples:

- Focus groups on program take-up and usage
- Qualitatively coded interviews on empowerment

Key Questions:

- Do I plan to weight for representativeness?
- Do I need administrative data to sample?

RECRUITING RESPONDENTS

Interview Sites

- Institutions, households, work sites
- Can I collect data remotely?
- Will I be doing it myself?

Other Considerations

- Privacy
- Coercion
- Respondent attention
- Expense and logistics



SETTING KEY PROTOCOLS

- **Response Rates**
 - Respondent follow-ups
 - Respondent replacements
 - Survey status codes for tracking
- **Survey Quality and Targeting**
 - Defining qualified respondents
 - Defining key concepts in the survey
 - Setting protocols for specific types of data collected



CHOOSING A METHOD OF DATA COLLECTION



PAPER-ASSISTED PERSONAL INTERVIEWS

- Minimal equipment
- Difficult to monitor data quality
- Delays in data access
- Surveyor handwriting errors



COMPUTER-ASSISTED PERSONAL INTERVIEWS

- Data quality assurance capacity
- Potentially challenging to set-up
- Need to secure equipment
- Surveyor typing errors



COMPUTER-ASSISTED TELEPHONIC INTERVIEWS

- Remote field teams
- Geographical spread of sample
- Limited take-up and response
- Variation in interview settings



OTHER METHODS OF DATA COLLECTION

- Health or physical tests
- Technical measurements
- Direct observation
- Games

POST-COVID: MOVING FROM CAPI TO CATI

■ CAPI

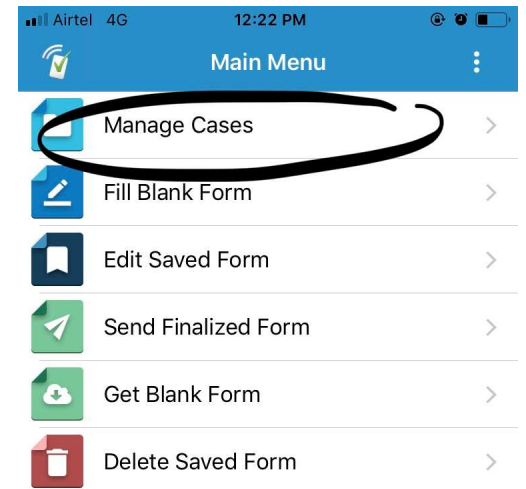
- Assignment by hard- or soft-copy tracking sheets
- Data collection on tablets that are stored and maintained by the project
- In-person staff training and monitoring

■ CATI

- Assignment using “case management”
- Data collection on personal smartphones
- Phone-based staff training and monitoring

■ IVR or digital surveys

- We manage survey data collection in-house on an online portal or app
- An external vendor or partner ensures that surveys reach respondents



A screenshot of a mobile application showing a table of survey data. The table is titled 'Collect > Manage Cases' and has three columns: 'ID', 'callback_time', and 'last_call_status'. The table contains ten rows of data. The first row is highlighted in blue. The status 'Unreachable/no coverage' appears in the 'last_call_status' column for the second row.

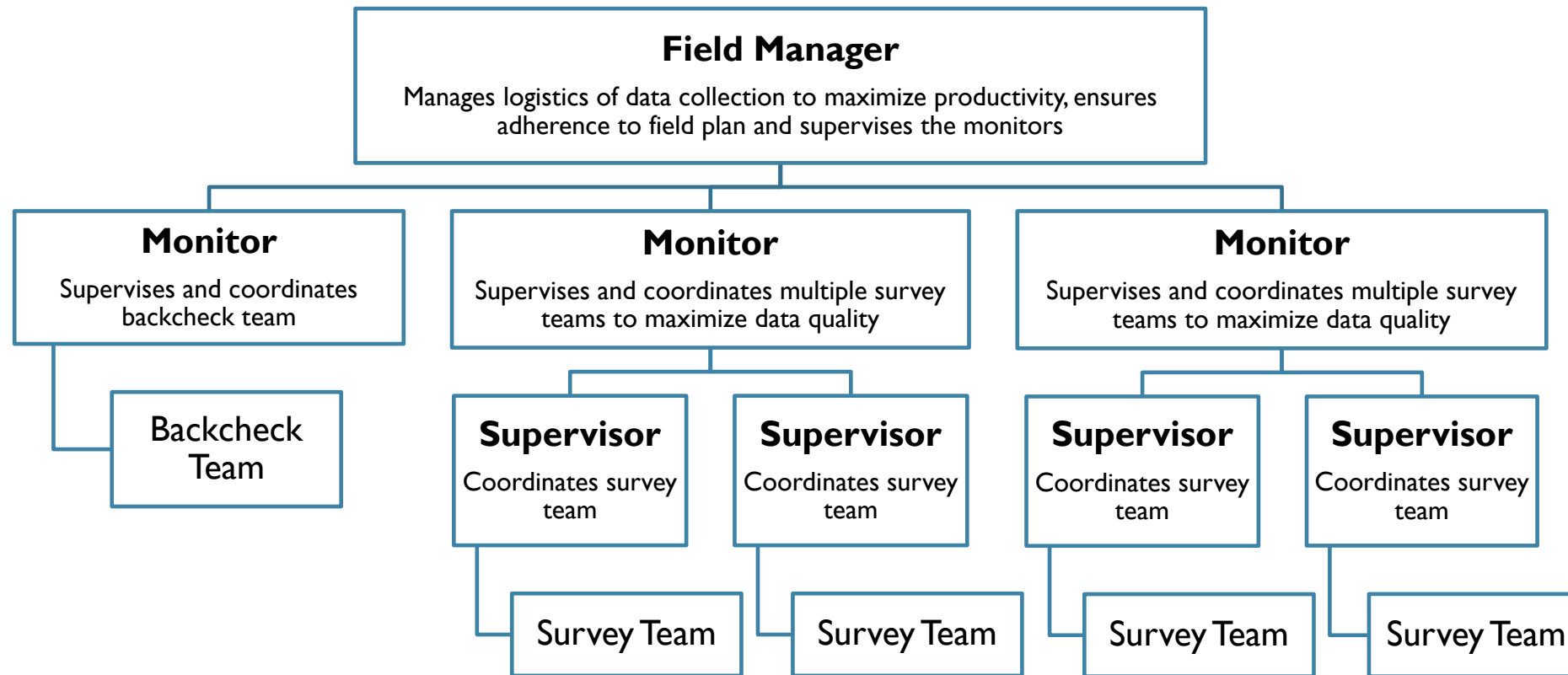
ID	callback_time	last_call_status
300242	2020-04-19 14-15	Switched off
98379	2020-04-19 14-31	Unreachable/no coverage
53555	2020-04-19 15-20	Switched off
152202	2020-04-19 15-23	Does not exist
154853	2020-04-19 15-25	Switched off
57026	2020-04-20 13-00	Rescheduled
159073	2020-04-20 16-00	Rescheduled
226435	2020-04-21 12-12	Does not exist
16355		

STRUCTURING YOUR FIELD TEAM

- **How long does it take to administer a survey?**
- **How many people are required to:**
 - Identify respondents and administer surveys
 - Supervise field operations
 - Quality check the surveys
 - Make necessary revisits
- For paper surveys, who will scrutinize and data-enter surveys?
- For digital surveys, who will secure equipment and maintain soft-copy tracking?



AN EXAMPLE TEAM STRUCTURE





DATA COLLECTION

ENSURING SURVEY PRODUCTIVITY

- Survey productivity and survey quality can be at odds, and it is important to strike a balance between the two
- Survey timelines may be determined by partner organizations, budget constraints, or growing seasons, and so it can be critical to stick to productivity targets
- Prizes, bonuses, and recognition can be important to motivate members of the field team
- Good field planning and productivity assumptions can help to uncover fraud



DATA QUALITY PROTOCOLS

Data Tracking and Validation

- Protocols to assign surveys
- Protocols to validate surveys

Field Quality Checks

- Spot or shadow-checking a set percentage of surveys
- Accompanying or back-checking a set percentage of surveys

Office Quality Checks

- Scrutinizing completed surveys
- Conducting additional quality checks on data

Error Resolution

- Revisiting households
- Resolving discrepancies in the field

ARE SURVEYORS SURVEYING THE INTENDED SAMPLE?

SURVEY TRACKING

- **Ensuring that assigned surveys are completed and can be uniquely identified. This involves matching:**
 - Assignment Sheets (*who you intend to survey*)
 - Tracking Sheets (*daily field activities*)
 - Completed Surveys (*data in hand*)
- *For paper surveys, this involves tracking sheet scrutiny*
- *For digital surveys, this can be coded*



ARE SURVEYORS ASKING QUESTIONS CONSISTENTLY?

IN-FIELD QUALITY CHECKS

Ensuring that surveyors are administering the survey correctly

Spot Checks	A monitor listens as the survey is administered	Set a spot check target, ensure that monitors take notes on key performance indicators and offer feedback after the survey has completed
Shadow Checks	A monitor listens as the survey is administered, unobserved by the respondent and the surveyor	Can yield a good picture of surveyor behavior when they believe they are not being observed, but logistically challenging to implement

Ensuring that surveyors are filling in questionnaires correctly

Accompaniments	A monitor listens as the survey is administered, and fills in a duplicate questionnaire.	Requires that monitors observe a substantial portion of the survey, and that you set up processes to scrutinize duplicate responses
Back-checks	A monitor returns to ask the respondent a subset of questions and fills in a partial questionnaire.	Requires that monitors return to a household a second time, and that you set up processes to scrutinize duplicate responses. It could introduce recall bias.

DO THE RESPONSES MAKE SENSE?

IN-OFFICE QUALITY CHECKS



- **Scrutiny**

- Checking for completeness, skip patterns, reasonable responses
- Allows us to make revisions before data is finalized

- **High-Frequency Checks**

- Outliers in assets, income, or key indicators
- Consistency checks for rosters, skip patterns, in-survey calculations, metadata

- **Cross-validation**

- Are there other measures of outcomes in admin data
- Is there baseline, village-level, or treatment data that informs responses?

BEST PRACTICES

PAPER SURVEY SCRUTINY



- Have a senior member of the field team review each questionnaire, question-by-question
- Assign different pen colors to different members of the field team, to easily identify the work of each member
- Set up physical space to separate surveys that have not been scrutinized, and that have not been data-entered
- Decide in advance on a protocol for dealing with surveys that need to be re-checked

AN EXAMPLE:

DATA QUALITY CHECKS

- **Outliers**

- **Inventory of Assets:** Do respondents have unusual numbers of assets?
- **Income and Expenditure:** Do respondents have incredibly high or low income or expenditure
- **Key Indicators:** Are survey responses matching administrative data we have about the area or respondent

- **Consistency Checks**

- **Roster Questions:** Are respondents answering questions unusually according to their stated age or gender?
- **Skip Patterns:** Are respondents not answering survey sections that are applicable based on previous answers?
- **Time and Date:** Are start times and dates always earlier than end times and dates?
- **Monthly and Annual Calculations:** Is monthly income or expenditure greater than yearly income or expenditure?

Integrated data tracking
systems allow us to ensure:

High Quality Data

Staff Accountability

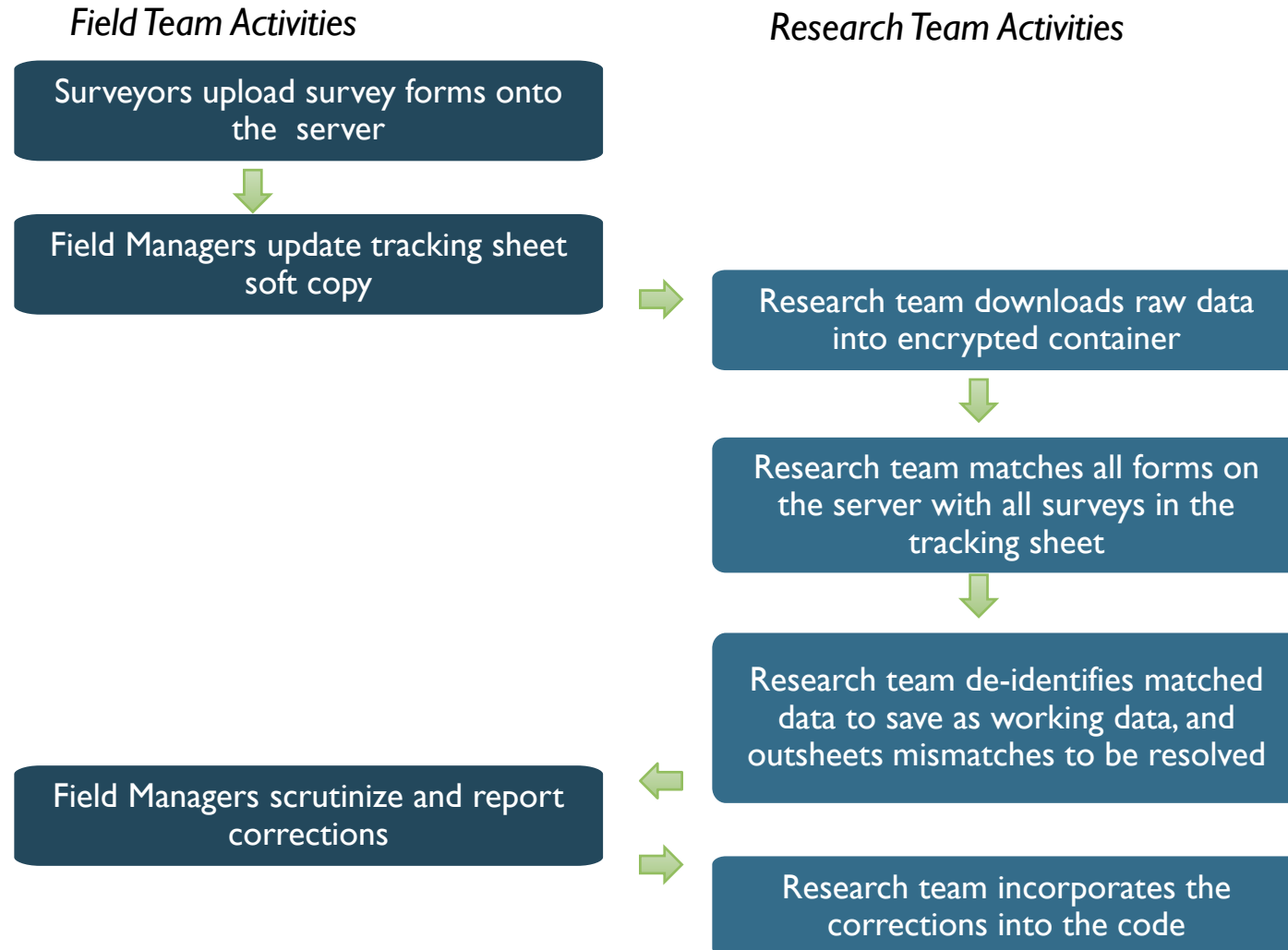
Efficient Field Plans

On-Budget Data Collection

Continual Research Improvement

A DIGITAL DATA PROCESS FLOW

DATA VALIDATION: PROCESS FLOW



DATA VALIDATION:AN EXAMPLE

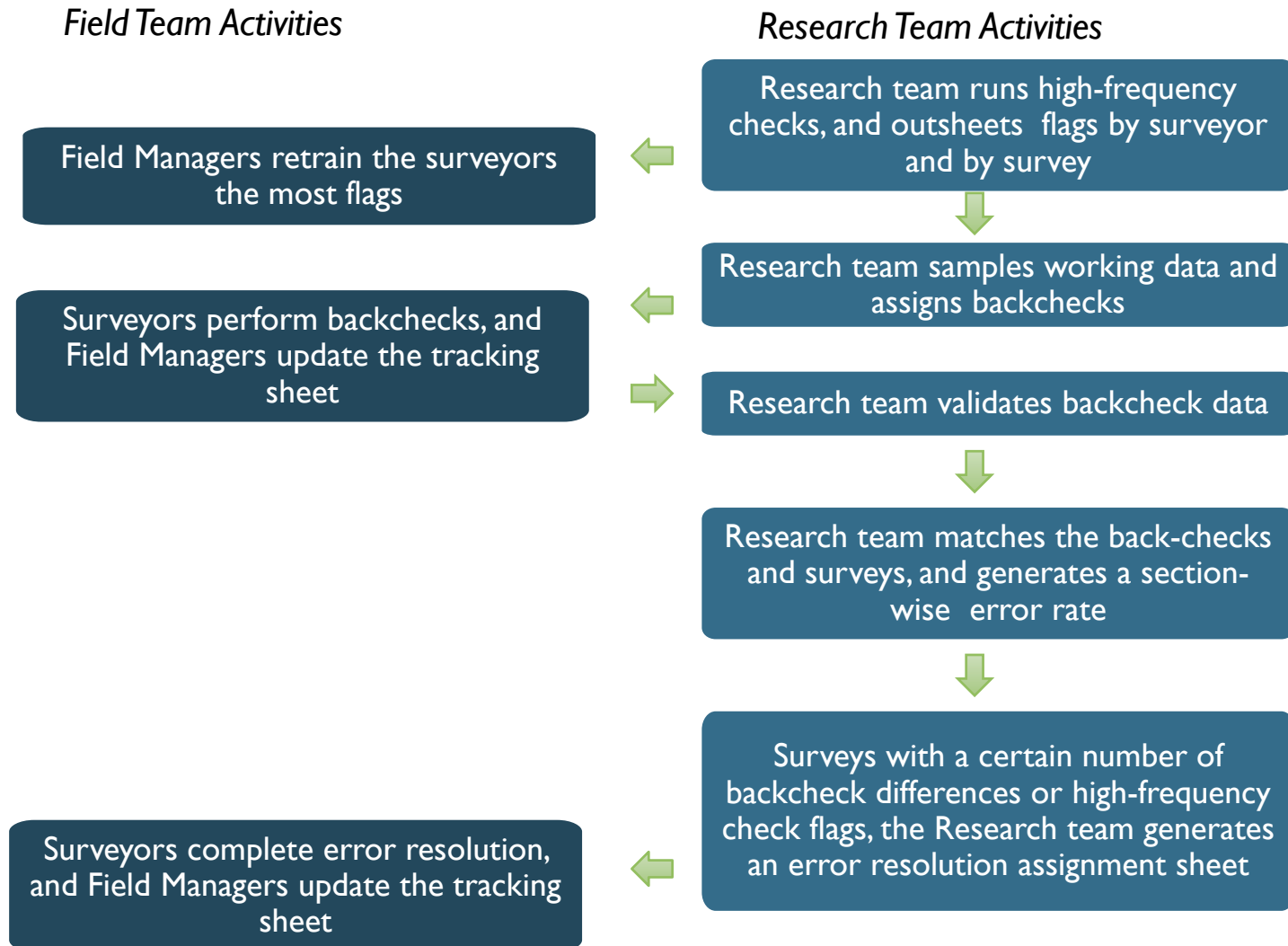
- Identifying unique identifier mismatches

parent_key	hh_id	surveyor_code	supervisor_code	start_date	area_id	Remarks
uuid:c6f8acc2-cc2a-4969-8880-7f8141d0292a	229	224	401	23-Sep-14	3-17-1-3	HH id 338
uuid:95bb874e-6fce-4d00-9e1b-7612bdac5c61	229	224	401	22-Sep-14	3-17-1-3	Right

- Correcting the data after receiving feedback from the field

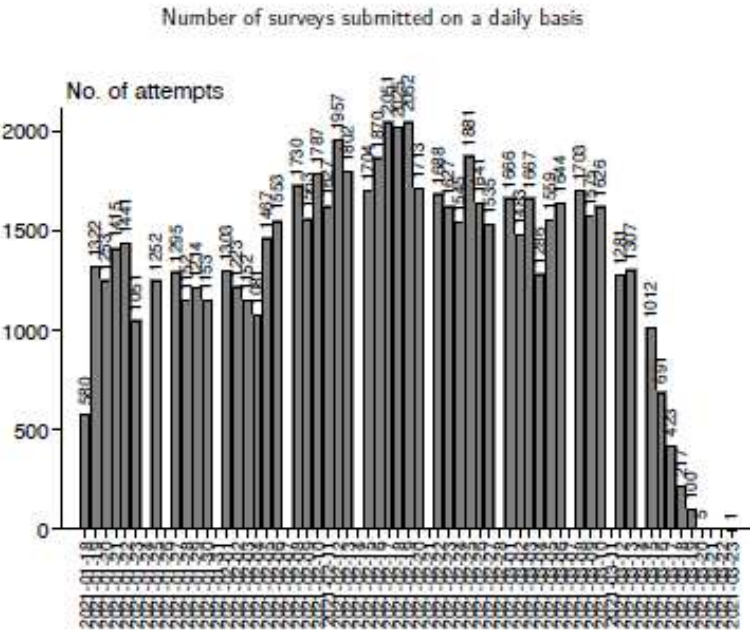
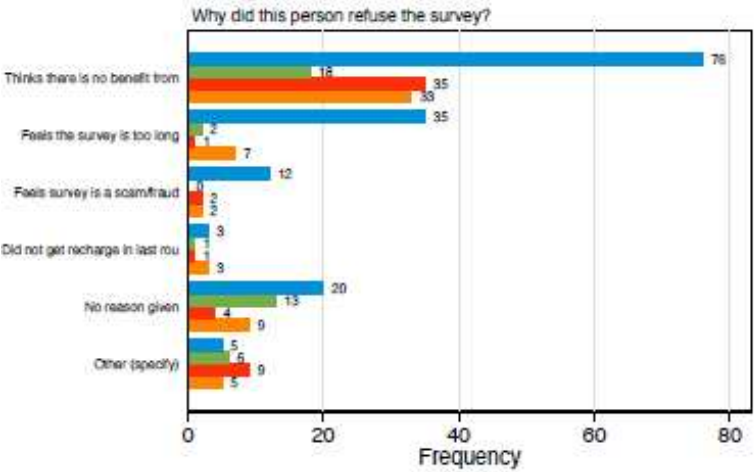
```
//Resolving survey duplicates  
replace hh_id=149 if parent_key=="uuid:fd199523-d887-4497-9ca5-819556148f50"  
drop if parent_key=="uuid:5c4b859d-e417-49a8-8d90-0a8ef8d3f529"
```

DATA QUALITY: PROCESS FLOW



DATA QUALITY:AN EXAMPLE

- Identifying refusal reasons
- Tracking completion
- Call outcomes by time



Call outcome by day of week: Both states and genders

Day of Week	Monday	Tuesday	Wednesday	Thursday	Friday	Saturday	Total
Complete	713 (6)	684 (6)	765 (6)	734 (7)	759 (6)	729 (6)	4384 (6)
Incomplete	71 (1)	73 (1)	92 (1)	85 (1)	70 (1)	66 (1)	457 (1)
Refused	93 (1)	82 (1)	64 (1)	67 (1)	58 (0)	84 (1)	448 (1)
WrongNo	179 (1)	136 (1)	152 (1)	125 (1)	176 (1)	165 (1)	933 (1)
Unavail/Unable	138 (1)	140 (1)	167 (1)	106 (1)	152 (1)	112 (1)	815 (1)
Resched/AltNum	1667 (13)	1413 (12)	1447 (11)	1241 (12)	1551 (12)	1268 (11)	8587 (12)
PreconDrop	85 (1)	65 (1)	73 (1)	57 (1)	90 (1)	73 (1)	443 (1)
CallFail	9585 (76)	8658 (76)	9943 (78)	8194 (77)	9715 (76)	9171 (78)	55266 (77)
Unqualified	52 (0)	45 (0)	46 (0)	38 (0)	53 (0)	48 (0)	282 (0)
NoNum	55 (0)	59 (1)	50 (0)	37 (0)	88 (1)	47 (0)	336 (0)
Total	12638 (100)	11355 (100)	12799 (100)	10684 (100)	12712 (100)	11763 (100)	71951 (100)

DATA QUALITY: AN EXAMPLE

■ Generating Error Rates

	Mean	Mean	Mean	Mean	Mean	Mean	Mean
	E1-Elec Any	E5-Genset	E13-Solar	E19-Grid	E20-Metered	E21-Func Meter	E33-Billed
District Code							
3 (n=18)	5.56%	33.33%	5.56%	11.11%	11.11%	11.11%	72.22%
4 (n=9)	0.00%	22.22%	11.11%	22.22%	22.22%	22.22%	11.11%
5 (n=4)	0.00%	0.00%	0.00%	0.00%	0.00%	25.00%	75.00%
6 (n=1)	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	100.00%
Total (n=32)	3.13%	25.00%	6.25%	12.50%	12.50%	15.63%	56.25%

■ Calculating Flags by Surveyor

surveyor_code	total_flags	energy_flags	non_energy_flags
128	92	36	56
123	86	26	60
130	75	21	54
125	56	24	32
122	55	24	31
113	53	7	46
150	51	11	40
112	50	18	32
151	44	14	30
127	41	16	25
121	41	13	28
118	39	29	10
129	39	24	15

FINAL THOUGHTS

- Based on field conditions and survey design, customize your field plan and team
- Pilot to confirm assumptions, budget and plan for safety
- Designate and train supervisory staff for quality checks and productivity management
- Develop data process flows to maintain data quality and to be responsive to changing field conditions
- Allow staff to offer feedback through the data collection process
- Document decisions made and challenges encountered during data collection