big idea

• Data quality matters

• Data problems best addressed in real time

• Lots of potential for tech to improve data quality

• But no one magic solution!
How can technology help?

• Three stages to ensuring data quality:

• PREVENT errors with good instrument design

• MONITOR data as it comes in

• FEEDBACK / CORRECTIONS while team still in the field
PREVENT: how CAPI can help

• Program instruments to catch problems in real time

• The basics
  • Control flow of questionnaire (automate skip patterns)
  • Restrict answer ranges and type (numeric, alpha)
    • Ex: age must be numeric, >0, <115

• More advanced
  • Use internal references for answer choices
    • Ex: choose from dropdown list of HH members
  • “Pre-load” data from previous rounds of surveys
    • + efficiency, -time
MONITOR: how CAPI can help

- Avoid fraud
  - CAPI facilitates backchecks
    - “Pre-load” IDs to simplify dataset merges
    - Compare in close-to-real time
  - Random audio audits of portions of interview
    - Requires informed consent but good low-cost option

- “Soft checks” (warning message)
  - Check actual interview location ≤5m expected location
  - Duration – pop up warning for unusually short interviews
MONITOR: CAPI not enough

• Do high frequency checks of full dataset
  • monitor survey completion
  • check for duplicates
  • Assess performance by enumerator and teams

• Compare data with sample population / previous rounds

• Check geopoints /distances (visualize in maps)
CORRECT: feedback key challenge

• CAPI allows us to fix in real time. Much easier and cheaper than waiting months for data entry!

• But how to communicate errors and issues in a way that is easy to understand/act on, and transparent to document and correct in the final data

• No ideal technology solution I know of. Looking forward to hearing your experiences!
A few notes on tech

- Not all CAPI technologies are equal
  - At DIME we typically work with SurveyCTO
  - Other options out there but be mindful what quality control features are available and the state of the exported dataset

- For high frequency checks, we typically use Stata
  - Reproducible and transparent
  - Can do an increasing amount within SurveyCTO itself
  - R good option as well
Example: ag survey in Rwanda

• System
  • Automated import of data from server every 10 mins
  • Run data checks designed by team after each import
  • Outputs updated in dashboard and shared with teams

• Technology
  • Questionnaire: SurveyCTO
  • Data import: custom api by survey firm (ikapadata)
  • Dashboard: Dropbox Paper
  • Communication to team: Slack
  • Maps: Carto
Example: ag survey in Rwanda

• Research team dashboard
  • Survey completion, disaggregated by team
  • Summary statistics on basic quality checks, past day and in total
  • Graphs comparing key indicators from current and past round

• Field team dashboards
  • Survey completion, by enumerator
  • Summary statistics on basic quality checks, by enumerator

• Instant messaging
  • Supervisor gets daily report on completed surveys and data flags
Dashboard: Quality Checks

- **Duration**
  - Interview <90 minutes

- **Plot unavailable**
  - ≥1 one plot could not be measured

- **GPS Plot**
  - Accuracy of ≥1 geopoint for plot >3m

- **GPS Household:**
  - Accuracy of household geopoint ≥3m

- **Distance**
  - Interview done >5m away from location of interview in previous survey round

- **Outlier**
  - ≥1 variable has a value >3 SDs away from the mean
Total Income

Left: Total Income Boxplot (No Outliers/Outliers)
Right: Total Income Mean (2016 vs FUP3)
slack: feedback to team

ikapabot  APP  12:41 AM
TEAM REPORT D 30nov2016
Completed Interviews: 8
Revisits: 0
Unavailable: 0
GPS Issues: 0%
Short Interviews: 38%
Missing Plot Measurements: 0%
Away from HH: 13%
carto: visualize interview completion
more work needed

- Data security and encryption
  - (SurveyCTO’s new data explorer handles this nicely)

- Transparently resolving data quality issues
Thank you!
the bad old days ...

STEPS TO CHECK-IN AN ADP-SP MIDLINE SURVEY VILLAGE PACKET:

1) Make sure that all of the packet contents listed on the front of the envelope are included. Enter in the contents on the Village Packet Tracking Sheet.

2) Paper clip together each household questionnaire with its spouse questionnaire, and the LF or PF module if the household is a LF or PF household.

3) Enter the HH number, enumerator, reviewer, and list of problems for each household in the Household Tracking Sheet (one per village).

4) Make sure that the household numbers on each questionnaire match the numbers recorded on the Survey Tracking Forms and the List of HHs to Interview.
   a. Note any problems in the Comments section of the Village Tracking Sheet.

5) Put the LF/PF Confirmation Forms, Village Questionnaires and AEDO Questionnaires (if included) in their own envelopes according to district (CF or NM).
Dropbox Paper

• Pros:
  • Displays output from checks developed by team (e.g. in Stata)
    • .do file is reproducible, easy to implement at high frequency
  • Customized dashboards for team members based on role
  • Can compare to other datasets
  • Display maps (but need map software to generate)

• Cons
  • Data security is an issue
  • Requires access to other software (stata/r, carto)
  • Data checks developed from scratch

SurveyCTO data explorer

• Pros
  • User-friendly interface
    • good for users less versed in statistical software
  • Easy way to review all variables
    • good for short surveys
  • Great map integration
  • Solves data security concerns

• Cons:
  • Non-reproducible
  • No way to automate checks
  • No summary / overview
  • Can only share issues with team members one case at a time by sending URL