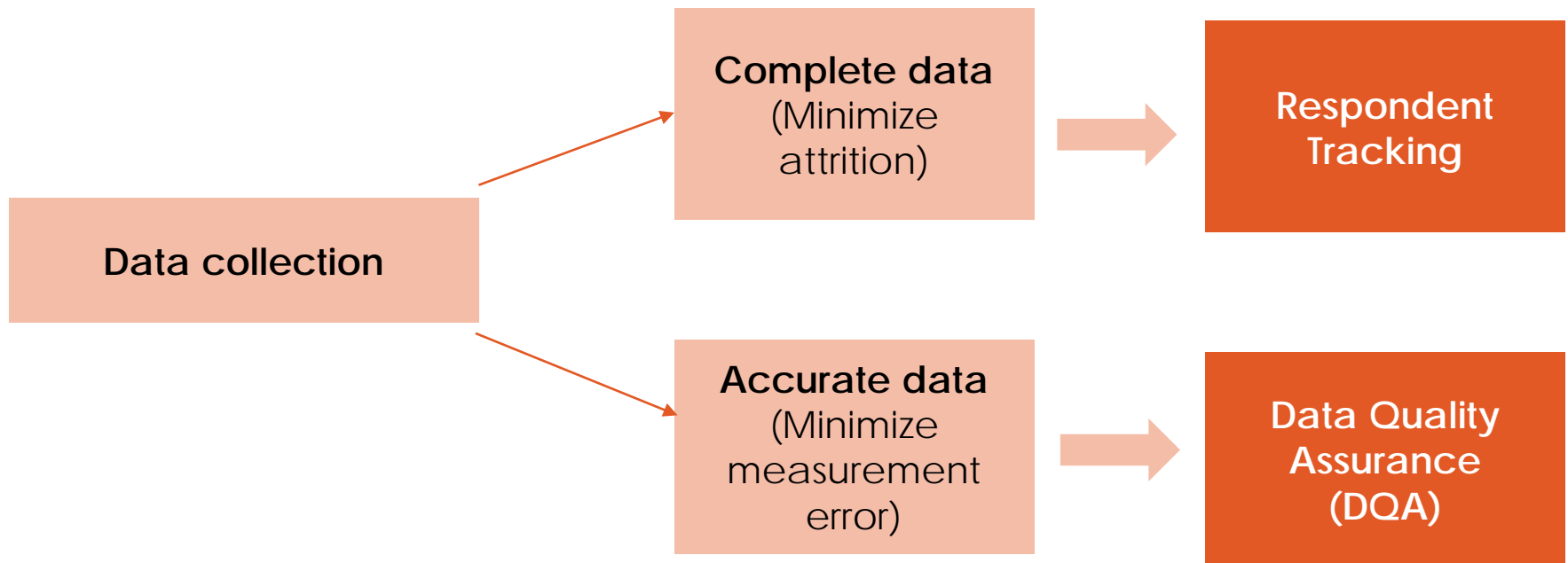




Collecting High Quality Data: Accurate Data

Dr. Christopher Robert

Collecting High-Quality Data



Lecture Overview

- Introduction
- Planning
- Implementation
- Reconciliation

Lecture Overview

- Introduction
- Planning
- Implementation
- Reconciliation

Why is quality control important?

- Only high quality data is truly useful for informing decisions
- Consistent challenges:
 - Enumerator jobs are hard
 - Rain, snow, extreme heat
 - Travel, finding respondents
 - When to ask questions
 - How to ask questions
 - The temptations are many
 - Faking an interview
 - Interviewing the wrong respondent
 - Auto-answering
 - Just know the answer
 - Incentive to answer one way

Data Quality Assurance (DQA)

- Series of checks done both on-field and off-field to ensure accurate data
- On-field, it can happen:
 - When original surveyor collects data
 - After original surveyor collects data
- Off-field, it can happen:
 - In the office after surveyor collects data
- Some DQA processes have both on-field and off-field components

What are we checking for?

- Problems with questionnaire
 - Individual questions
 - Questionnaire structure
- Errors in survey execution
 - Adherence to organization and study protocols
 - Opportunities for retraining
- Fraud

What kinds of checks are common?

- Spot checks
- Accompaniments
- Audits
 - Back-checks
 - Audio audits
- High frequency checks

Spot checks and Accompaniments

- Happen in the field, when enumerators are collecting data
- Spot checks
 - *Unannounced* visits to observe a survey
 - Done by supervisors or field managers (higher-level managers may get involved if required)
- Accompaniments
 - Visit to observe a survey that *the surveyor knows about*
 - Done by supervisors or field managers (higher-level managers may get involved if required)
- Ideally neither done by international staff

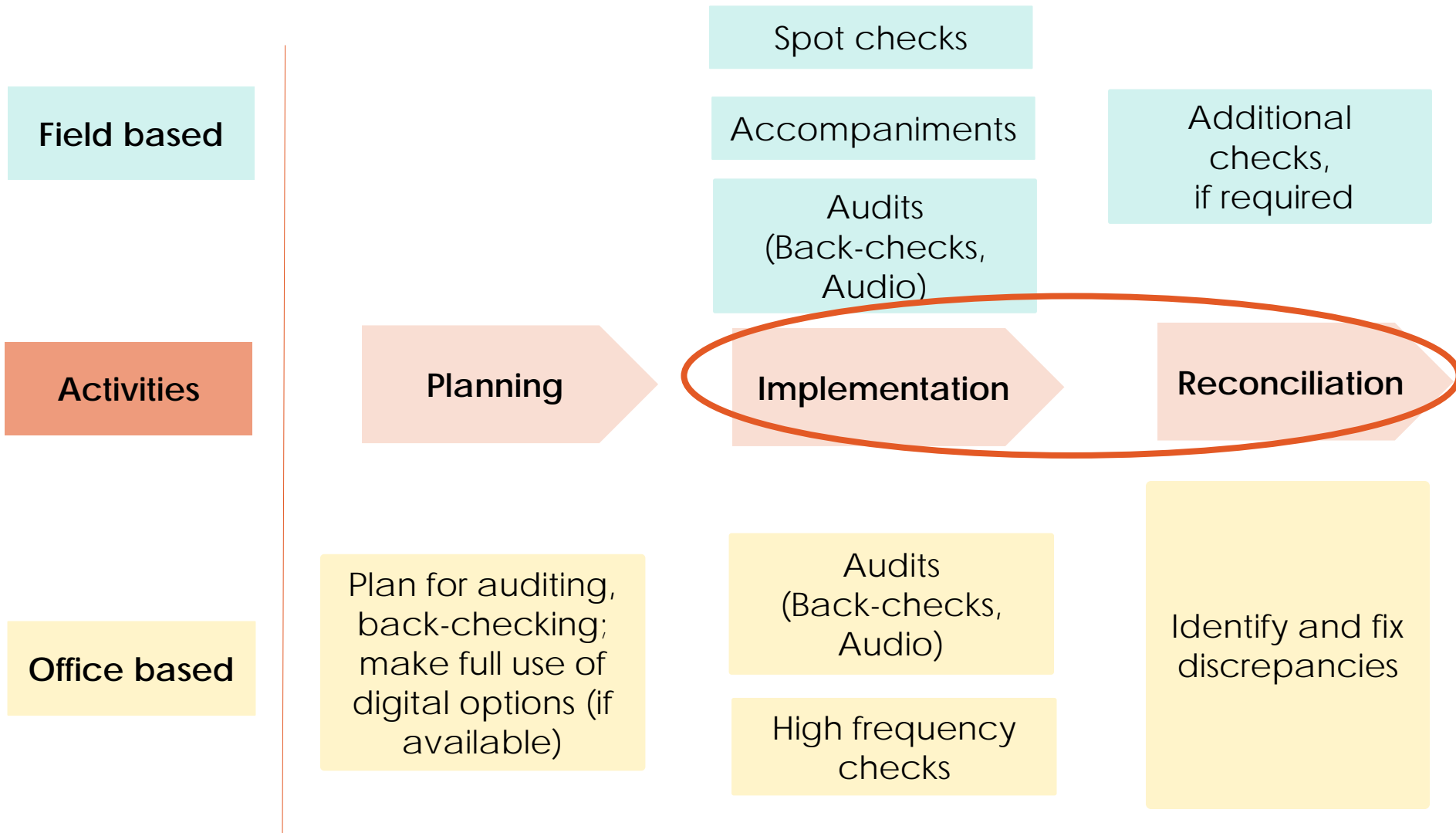
Audits

- Partly in the field, partly in the office
- Back-checks
 - Different enumerator (or supervisor) revisits sub-sample of already-interviewed respondents
 - Administers audit questionnaire
- Audio audits
 - Original surveyor is simultaneously administering survey *and* being audited through recording device
 - Digital data: Software is programmed to randomly or fully audio record as survey is being administered
- In the office, audit data is reconciled with surveyor-collected data

What are we checking for?

- Problems with questionnaire
 - Individual questions
 - Questionnaire structure
- Errors in survey execution
 - Adherence to organization and study protocols
 - Opportunities for retraining
- Fraud

Data Quality Assurance: Summary



Lecture Overview

- Introduction
- **Planning**
- Implementation
- Reconciliation

Preparing for success in the field

- Pre-select respondents for DQA checks (spot checks, accompaniments, back-checks)
 - Have a list of households and schedule for visits ready
- Identify questions to back-check
 - Data on the interview and surveyor
 - Repeating questions from the original interview
 - Re-ask important questions, consequential questions
 - Consider randomizing among longer list

Questions to include in a back-check questionnaire

**What to detect?
(Problems)**

Fraud

**Errors in survey
execution**

**Problems with
questionnaire**



**How to detect?
(Question type)**

#1—Questions where the answer should never change, consequential Q's

#2—Questions that are difficult to administer (complicated skips or surveyor instructions)

#3— Questions for key outcomes or for which variance is of interest

Preparing for success in the office

- Program templates or set up software checks for analyzing data as it comes in
 - high-frequency checks

Planning for HFCs

- **Logic checks**

- Check that all interviews were completed
- Check that there are no duplicate observations
- Check that all surveys have consent
- Check that certain critical variables have no missing values
- Check that follow-up record IDs match original
- Check skip patterns and constraints
- Check that no variable has all missing values
- Check specify-other vars for items that can be included
- Check that date values fall within survey range
- Check outliers for numeric fields

Source: <https://github.com/PovertyAction/high-frequency-checks/wiki/Background>

Planning for HFCs

- **Enumerator checks**

- Check the percentage of “don’t know” and “refusal” responses by enumerator
- Check the distribution of responses for key questions by enumerator
- Check the number of surveys per day by enumerator
- Check the average interview duration by enumerator
- Check the duration of consent by enumerator
- Check the duration of other modules by enumerator (e.g., anthropometrics, games, etc.)

Source: <https://github.com/PovertyAction/high-frequency-checks/wiki/Background>

Planning for HFCs

- **Project checks**

- Overall survey progress relative to planned sample
- Summaries of key research variables
- Two-way summaries of survey variables by demographic/geographic characteristics
- Attrition rates by type and treatment status
- Comparisons of variables with known distributions
- Maps/GIS: all observations where they're meant to be?



Source: <https://github.com/PovertyAction/high-frequency-checks/wiki/Background>

Digital: Before you get started

Digital data collection tools offer specific opportunities to design your data-collection instrument for quality...

Designing for Quality: Automatic Capture

Automatically capture the survey start and end date and time, version number, and information about the device used. Also capture **GPS locations**.

form_title:	Test form
form_id:	test_form_for_presentation
form_version:	 Automatic (based on formula)
encryption:	 Off
languages:	english default_language: english
meta-data:	starttime, endtime, deviceid, subscriberid, simid, devicephonenumber, username, duration, caseid
attachments:	Files: apple1.jpg, mango1.jpg, banana1.jpg Server datasets: (none)

× Close

⚙ Edit form settings

Designing for Quality: Skip Patterns

Skip irrelevant questions and eliminate skip-pattern errors.

The screenshot shows a configuration window titled "Add select one field" with a close button in the top right. On the left, there are "Save now" and "Cancel" buttons, and a "Shortcuts:" section with links for "Required options", "Labels", "Relevance", "Constraint", and "Other options". The main area is titled "Relevance" and contains a sub-section "Add relevance - Step 1" with a help icon and a "Save" button. Below this, a text box explains: "A field will only display when it is deemed 'relevant'. Use the + and - buttons below to add and remove conditions." The configuration is set to "Field is relevant if All of the following conditions are true:". Two conditions are listed: "Field: gender is equal to the text value F" and "Field: age is greater than the numeric value 10". Each condition has a "+" and "-" button. At the bottom right, there is a "+ Add condition" button, and at the bottom left, a "Cancel" button, and at the bottom right, a "Save" button.

Designing for Quality: Validation

Restrict allowed responses for individual questions to prevent entry of impossible or contradictory answers.

The screenshot shows a software interface for adding an integer field constraint. The main window is titled "Add integer field" and has a blue header bar. On the left side, there are two buttons: "Save now" (blue) and "Cancel" (red). Below these is a "Shortcuts:" section with links for "Required options", "Labels", "Relevance", "Constraint", and "Other options". The main content area is titled "Add constraint - Step 2: Constraint parameters" and has a light blue header with a "Help" icon and a question mark. It contains several input fields and checkboxes:

- Allow decimals?
- Minimum value:** (leave blank for none) with an input field containing "18".
- Maximum value:** (leave blank for none) with an input field containing "120".
- Allowable exceptions:** (comma-separated list) with an input field containing "-888, -999".
- Please enter the error message to display when the constraint is violated:** with a text area containing "Please enter a valid age for this respondent (18-120)."

At the bottom of the dialog, there are "Cancel" (red), "Back" (blue), and "Save" (blue) buttons.

Designing for Quality: Preloading Data

Preload external data and use it to set default values so that enumerators don't make mistakes re-entering information already collected earlier.

The screenshot shows a web interface for creating a new dataset. At the top, the title is "Server datasets for pre-loading data into forms". Below the title is a navigation bar with three icons: a plus sign for "New dataset", a refresh icon for "Refresh", and a question mark for "Help". The main content area is titled "Create a new dataset" and contains two input fields: "Dataset title:" with a placeholder "The dataset title here..." and "Dataset ID:" with a placeholder "The dataset id...". Below these fields is a section for uploading a file, with the text "Please upload a file with initial dataset contents (.csv file). You can upload it directly from your computer or use Google Drive:". There are two buttons: "Upload from computer" and "Upload from Google Drive". Below this is another section titled "Please choose a dataset file to upload (.csv):" with a blue button labeled "+ Select file...". At the bottom right, there are two buttons: a red "Cancel" button and a blue "Create dataset" button.

Monitoring Quality with Digital tools

Digital data collection tools also offer opportunities to monitor survey administration in specific ways...

- Have to plan – *and design your digital instrument* – accordingly

Monitoring Quality: Survey Timing








Text audits invisibly record each enumerator's path through each survey, capturing how much time they spent on every question.

+ Add visible field

+ Add a group

+ Add hidden field

What type of hidden field would you like to add?

calculate		comments		text audit		audio audit	
speed violations audit		speed violations count		speed violations list		calculate here	

randomly record
survey timing

Cancel

Configure...

Monitoring Quality: Audio Audits

Audio audits allow you to listen in, invisibly recording audio clips at random or specific points in your surveys.

+ Add visible field + Add a group + Add hidden field

What type of hidden field would you like to add?

calculate	?	comments	?	text audit	?	audio audit	?
speed violations audit	?	speed violations count	?	speed violations list	?	calculate here	?

randomly record
survey audio

Cancel

Configure...

Planning for Audio Audits

- Inform IRB that you will record respondents
- Include audio recording in informed consent
- Think about which questions or sections to audio audit (or if you prefer to audio audit the entire questionnaire)
 - Consider the bandwidth for sending data

Monitoring Quality: Speed Limits

Speed limits trigger invisible audio recordings if your platform detects an enumerator recording answers suspiciously fast.

Edit select_one field

[Save now](#) [Cancel](#)

Shortcuts:

- [Required options](#)
- [Labels](#)
- [Relevance](#)
- [Constraint](#)
- [Other options](#)

appearance: (none) - just an ordinary select_one field

Display options for your field.

required message:

Custom "a response is required" message (if any).

publishable: Yes No

For encrypted forms, leave this field publishable on the server?

minimum_seconds:

If using speed limits, how many seconds should user spend on this field?

media:image: [x](#) [Edit](#)

Display an image as part of this field.

Monitoring Quality: Reporting

Configure automated daily reports via email that flag critical errors, suspicious surveys, or enumerators whose responses differ significantly from the rest of the team.

   
Options Checks Run now Report

Quality checks: overall options

The below options apply to the set of quality checks configured for this form.

Run all checks nightly (uncheck to pause)

(Nightly checks will run only when data have changed since last run.)

Send email summary of quality-check reports to emails specified below?

Separate multiple email addresses with commas (e.g.: "manager@surveycto.com, teamlead@surveycto.com")

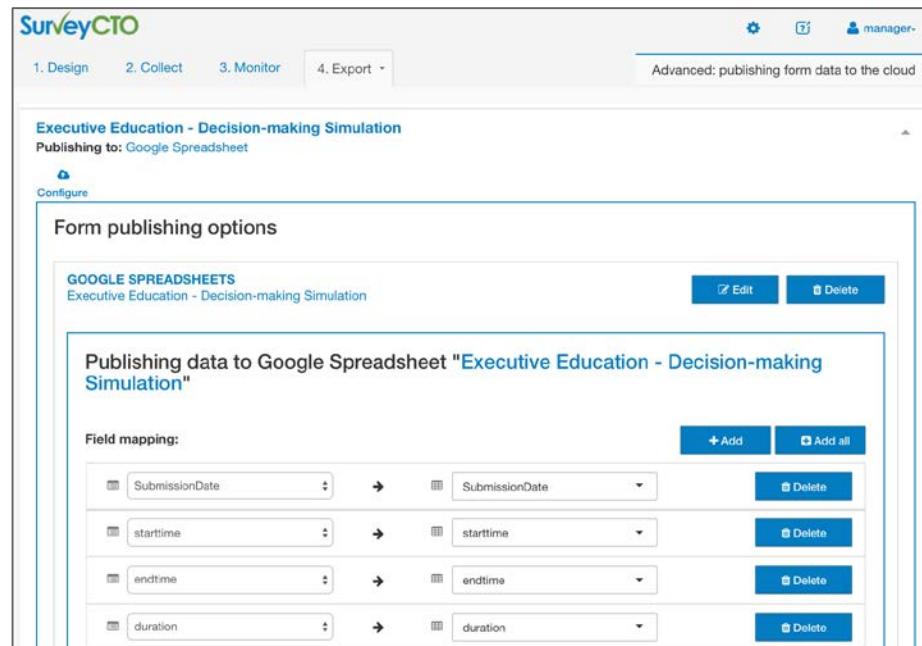
manager@surveycto.com

Cancel

Save

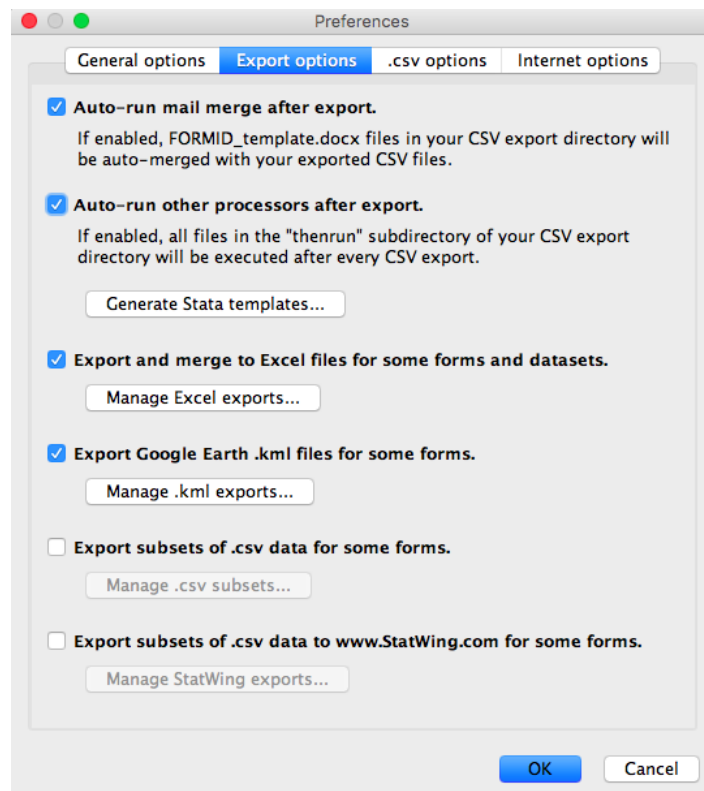
Monitoring Quality: Reporting

Configure publishing of incoming data to Google Sheets or Excel spreadsheets to easily review surveys and visualize key information – perhaps using pre-configured dashboards.



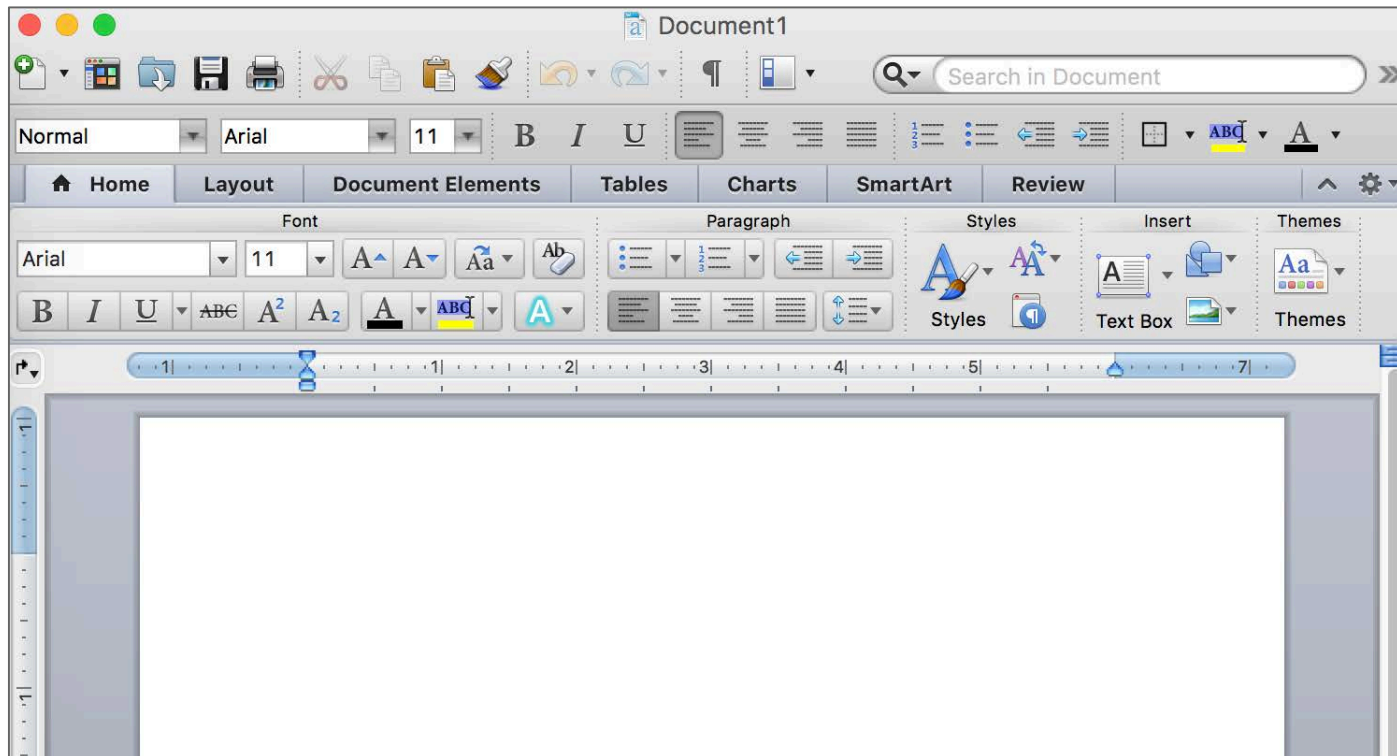
Monitoring Quality: Reporting

Practice viewing **GPS-tagged data** in Google Earth so you can later confirm where surveys were done.



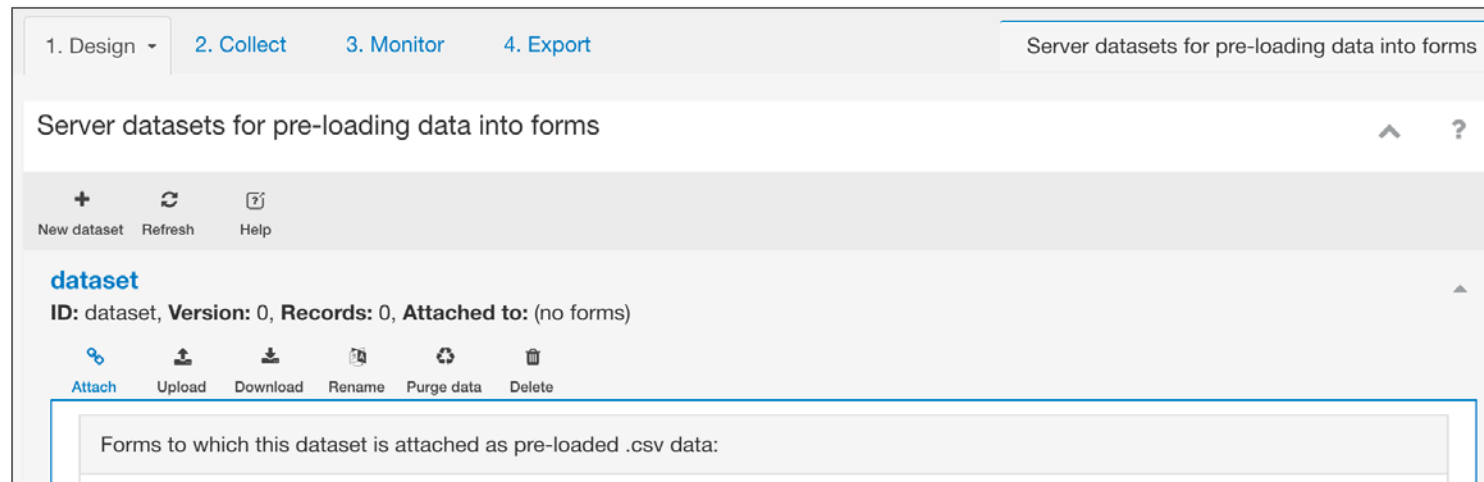
Monitoring Quality: Reporting

Configure mail merge with Microsoft Word templates so that staff can more easily review incoming data.



Monitoring Quality: Back-checks

Configure two-way data streaming to automatically select surveys for back-checks and send data from the selected surveys to back-checkers' devices. No need to print out back-check assignments or download, export, clean, and re-upload data.



Lecture Overview

- Introduction
- Planning
- **Implementation**
- Reconciliation

Implementation

- Types of checks:
 - Spot checks
 - Accompaniments
 - Audits
 - Back-checks
 - Audio audits
 - High frequency checks

Spot Checks and Accompaniments

- Happen in the field, when enumerator collects data
- Observation of interviews by supervisors or field managers (and sometimes by more senior managers)
- Spot-checks unannounced, accompaniments pre-planned

Implementing Spot Checks and Accompaniments

- Visit the household during survey
 - Recommended: Spot-check and/or accompany at least 10-15% of surveys
 - Front-loaded in the survey period
- Ensure that the enumerator is comfortable
 - Every enumerator should be spot-checked and accompanied
- Think about if and how observers need to be introduced to the respondent
 - Concern: increasing social desirability bias
 - International staff, in particular, are potentially disruptive
- Fill in a spot-check or accompaniment form during the interview


QA Template: Paper

Monitor/FM/ Name		District_Name			Block_Name					
Date:										
Spot checks/Accompaniments sheet										
HH ID	Name	Surveyor code	Type Of Quality Check 1-Accompaniment 2-Spot Check	Format of check		If Partially , timing hh, mm		Shadow 1- Yes 2- No	Overall rating 1-Very Bad 2- Bad 3- Average 4- Good 5- Very Good	Comment
				Fully 1- Yes 2- No	Partially 1- Yes 2- No					

QA Template: Excel

IPA's Spot check/ accompaniment form

Spot-Check and Accompaniment Form (Front)



1 Start, end time of the observation (24hr format); Date

Start time: hh mm End time: hh mm Date: dd mm yy

2 The spotchecker's (your) FULL name. Use CAPS

3 The spotchecker (your) ID, respondent ID and household ID (if applicable)

Spotchecker ID Respondent ID Household ID

4 The spotchecker's (your) position

Supervisor / Teamleader Monitor / Field Manager Backchecker Project Associate

Evaluation Coordinator Survey Coordinator Research Manager Project Coordinator

Back-checks

- Partly in the field, partly in the office
- Different enumerator, field monitor, or supervisor revisits sample of already-interviewed respondents
- Administers back-check questionnaire
- Back-check data is reconciled with the enumerator-collected data in the office

Implementing Back-checks

- Randomly assign by enumerator
- Front-load during survey period (to catch errors early on)
- Do not disclose to original enumerator which questions or respondents will be back-checked
- Visit the household and administer the back-check survey
 - Explain the purpose of the visit to respondent
 - Respondent may be wary as to why s/he is being asked the same information twice
 - Conduct within 1-3 days of original visit

Audio Audits

- Partly in the field, partly in the office
- Enumerator is simultaneously administering survey *and* being audited through recording device
- Digital surveys: Software is configured to audio record entire survey or specific questions in the survey instrument
 - Recordings can be random and invisible
 - Enumerator never knows when he or she is being observed!
- Audit data is reconciled with the surveyor-collected data in the office

Implementing Audio Audits

- Do not disclose to enumerators (or even field supervisors) what is audited when
- Ensure that the respondent is aware that his/her interview may be recorded
 - Ideally also allow for an opt-out – but then watch opt-out rates!

High-Frequency Checks (HFC)

- Happen in the office *with the full or partial dataset*
- Checking for validity of the data, possible enumerator effects, possible problems with the data-collection instrument itself
- Implemented using data analysis software such as Stata, R, SAS, etc. – or by digital data collection tool

```
display "Displaying percent DK/RF..."
foreach var of varlist _all {
    capture confirm numeric variable `var'
    if _rc == 0 {
        scalar miss = .
        scalar dk = .a
        scalar rf = .b
    }
    else {
        scalar miss = ""
        scalar dk = "don't know"
        scalar rf = "refusal"
    }
}

quietly count if `var' != miss
local nmiss = r(N)
quietly count if `var' == dk
local ndk = r(N)
local dkrate = `ndk' / `nmiss'
quietly count if `var' == rf
local nrf = r(N)
local rfrate = `nrf' / `nmiss'

bysort enumerator: egen totdkrf = total(inlist(`var', dk, rf))
quietly count if totdkrf >= 3

if `dkrate' >= 0.025 | `rfrate' >= 0.025 | r(N) > 0 {
    describe `var'
    display %20 = string(100 * `dkrate', "%1.1f") "%1"
    display %20 = string(100 * `rfrate', "%1.1f") "%1"
    tabulate enumerator `var' if inlist(`var', dk, rf) == 1, missing
}

drop totdkrf
}
```

Edit a quality check

Select the type of quality check:

Value is too frequent

Use this to check for a particular value's frequency and warn whenever it is above the threshold you specify.

Select the fields to be checked:

- Q1_1
- Q1_2
- Q1_3
- Q1_4

Warn when this field value:

2

Leave blank to check for missing values.

Occurs > this % of the time:

10

Critical quality check: flag warnings as high-priority

But ignore these values (separate with a comma):

Cancel Save

Implementing HFCs

- All of the hard work was done during the planning stage, so now you just have to run your checks as frequently as is feasible – and follow up on whatever issues emerge

Lecture Overview

- Introduction
- Planning
- Implementation
- Reconciliation

General Reconciliation Strategy

- Review DQA data throughout the survey process
- Always identify and reconcile issues as soon as possible (ideally within one week of survey)
- Conduct additional, targeted checks if necessary
- Use results for retraining and case studies – and to correct or refine your instruments when possible
- Have a strategy for communicating DQA findings to the survey team
- Correct final working (clean) data, *not raw data*
 - Don't change the raw data, keep audit trail of corrections

Audio Audits (office)

- **Goal 1:** compare recorded enumerator questioning with established scripts/protocols
- **Goal 2:** compare enumerator-entered responses with audio recording
- Options
 - Transcribe audio recording
 - Fill out audit questionnaire scoring and flagging surveys
 - Entire team listening to audio recordings for learning and social pressure

Back-checks (office)

- Compare the back-check data and survey data to identify discrepancies
 - Paper survey: Enter back-check questionnaires as soon as possible
 - Digital survey: Data available for comparisons right away
- Usually done using data analysis software such as Stata, R, SAS, etc.
 - Digital survey: Entire workflow can be at least partly automated

Back-checks (Stata)

- -bcstats- in Stata (IPA-written command) compares back-check data and survey data, producing structured output
- Types of questions identified in the code
 - #1—The answer should never change
 - #2—Questions related to skips or difficult to administer
 - #3—Key outcomes/Questions whose variance is of interest
- -bcstats- completes enumerator checks for type 1 and type 2 variables and stability checks for type 2 and type 3 variables

Dealing with Discrepancies

Discrepancies are likely

- Original and audit entries won't match – but how often, and to what extent?
- Different responses for different types of discrepancy (training, disciplinary action, re-survey)
 - Decision tree for response, disciplinary action (e.g., warning, dismissal)



J-PAL

ABDUL LATIF JAMEEL POVERTY ACTION LAB

END

