Virtual Training Workshop on Data Collection for Counterfactual Impact Evaluation (CIE)

Data Collection of Microdata in Difficult and Hard to Reach Areas

C4ED – EUTF

September 2022

Welcome to the Training Workshop on Data Collection for Counterfactual Impact Evaluation (CIE)

The material of this workshop was produced with the financial support of the European Union. Its contents are the sole responsibility of C4ED and do not necessarily reflect the views of the European Union

Center for Evaluation and Development

Communication during the training

MUTE BUTTON

QUESTIONS

FEEDBACK

Center for Evaluation and Development

Communication during the training

MUTE BUTTON

QUESTIONS

FEEDBACK

Communication during the training

MUTE BUTTON

QUESTIONS

FEEDBACK

Center for Evaluation and Development

Asking Questions

- Please post your questions in the chat room
- Like questions of others, so we know they are particularly relevant or urgent

- Carolin will read out all questions and we will answer these at once
- Use the longer breaks to ask more questions

Center for Evaluation and Development

Communication during the training

MUTE BUTTON

QUESTIONS

FEEDBACK

Center for Evaluation and Development

Asking Questions

- Please make suggestions
- Feel free to share your comments
- More feedback and questions (especially for the Q&A session):

Center for Evaluation and Development

Day 2 Agenda

Local Time	Proposed activities
10:00 – 10:30 (30 min)	Welcome words by EUTF and C4ED
10:30 – 11:45 (75 min)	Session 5: Data Quality in CIE and Monitoring Systems - Data Collection Methodology (Incl. Breakout Session 3)
11.45 – 12.15 (30 min)	Break
12:15 – 13:15 (60 min)	Session 6a: Leveraging Technology for High-Data quality for CIE
13.15 – 14.00 (45 min)	Lunch Break
14:00 – 14:30 30 min	Session 7: Monitoring Systems as Inputs in CIE
14:30 – 15:00 30 min	Session 8: Research Ethics and Data Protection
15.00 – 15.30 (30 min)	Breakout Session 4
15:30 – 16:00 (30 min)	End of Day Quiz and QA

Center for Evaluation and Development

Overview of Day 2

- Monitoring Systems (MS) as inputs in Counterfactual Impact Evaluation (CIE).
- Data Quality in Counterfactual Impact Evaluation and Monitoring Systems.
- Leveraging technology for High data quality in CIE and Monitoring Systems (MS).
- MS as inputs in Counterfactual Impact Evaluation (CIE).
- Ethical Considerations in data collection
  - Data Security and Data Protection

Center for Evaluation and Development

Session 5a: Data Quality in CIE and Monitoring Systems

Data Collection Methodology

C4ED – EUTF

September 2022

Center for Evaluation and Development

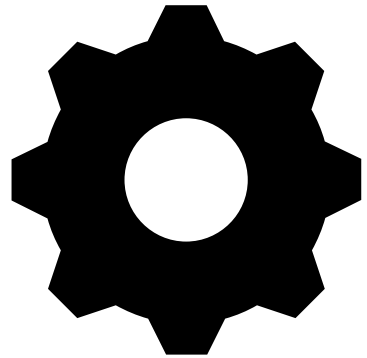
Session 5a Learning Objectives

- Participants will be able to describe the concept of non-response error
- Participants will learn approaches to minimise non-response error
- Participants will be able to explain how effective fieldwork protocols can enhance data quality

Center for Evaluation and Development

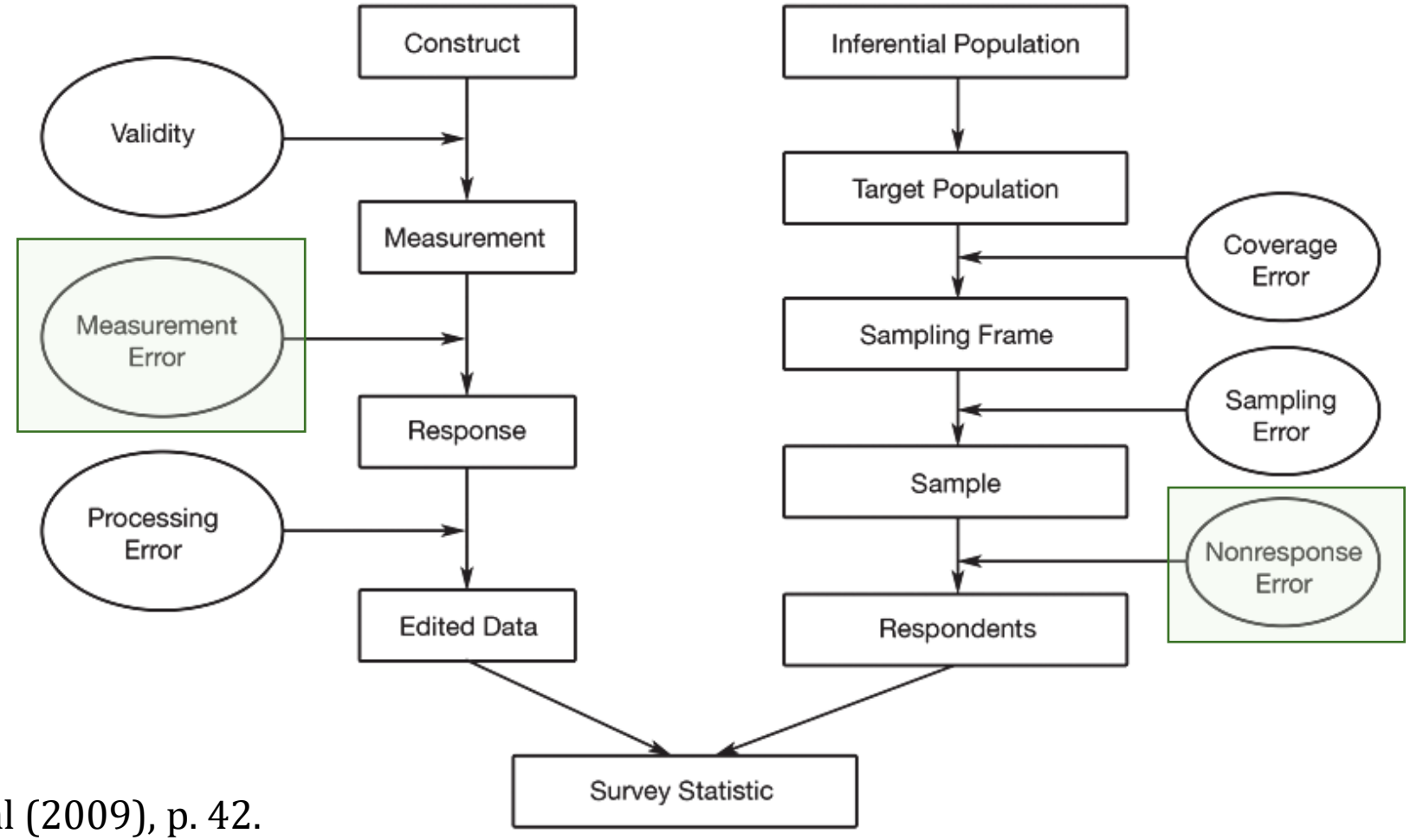
Recap from Day 1

- 1. Using the program theory of change and evaluation questions develop a list of indicators to measure
- 2. Find existing tools or write questions to capture data that will be used for each indicator
- 3. Obtain a sampling frame that contains units of the population of interest
- 4. Calculate the required sample size and take a random sample to select respondents



Center for Evaluation and Development

Survey Life from a Quality Perspective



Groves et al (2009), p. 42.

Center for Evaluation and Development

Factors affecting data quality

- Non-response: Failure to obtain intended info from respondents
  - Attrition
    - E.g. Respondents that were part of the program have moved away and changed their phone number so survey teams cannot find them.
    - Teachers that have left teaching since the baseline and are therefore no longer in the population of interest for a teacher training program.
  - Refusals
  - Poor questionnaire design

- Systematic non-response of sampled respondents
  - If non-random then may lead to biased data
  - Loss of sample size and power of analysis

Factors affecting data quality

- There are two types of non-response
  - **Unit non-response**:
    - Certain sample unit is missing.
- Respondents refusing to take part
- Respondents not available for interview
- Respondents moved away

Center for Evaluation and Development

Non-response

- Non-response due to unavailability at the time of interview
- Can be common with working respondents – particularly agricultural workers who are working at plots during the day
- Key practical remedies
  - If possible with contact details contact respondents ahead of time to make an appointment
  - Not overly burdening enumerators with daily targets to allow for flexibility during the day
  - Clear logging of interview attempts with notes from enumerator
  - Utilising other household members who are knowledgeable

Center for Evaluation and Development

Non-response

- Clear revisit protocols developed and trained to the teams
  - Number of visits to the household
  - Data to be gathered from other household members/neighbours
  - Contact numbers attempted
  - Spread of visits during a day – e.g. one in the morning and one in the evening (take timestamps and GPS)

Center for Evaluation and Development

Non-response

- Non-response due to refusal
- In normal circumstances, this is often avoidable with a well trained enumerator and a clear introductory script/informed consent note
  - Role-plays of introducing the study should be a key part of any data collection training
- With hard-to-reach groups there may be a greater risk of refusal:
  - Wary of providing information when wishing to remain “hidden”
  - Reluctant to sign any consent note
  - Lack of trust of outsiders

Center for Evaluation and Development

Non-response

- Non-response due to refusal
- In normal circumstances, this is often avoidable with a well trained enumerator and a clear introductory script/informed consent note
  - Role-plays of introducing the study should be a key part of any data collection training
- With hard-to-reach groups there may be a greater risk of refusal:
  - Wary of providing information when wishing to remain “hidden”
  - Reluctant to sign any consent note
  - Lack of trust of outsiders

Center for Evaluation and Development

Non-response

- Using oral consent
  - Confirm with IRB
- Buy-in from local leaders and support from a cultural insider in the process
- Hire enumerators from the targeted group

Center for Evaluation and Development

Non-response

- Non-response due to migration
- If the study involves mobile populations then the study is at a higher risk of attrition
- Important to collect good tracking data
  - Multiple contact numbers
  - Use of social media apps such as WhatsApp, Telegram
  - Social media groups (for program participants)
  - Alternative contact details of family and friends
- If there are resources, regular check-ins with the respondent via phone to ensure that information is up to date
- Incentives

Non-response

- Can you think of other ways to reduce non-response due to respondents migrating away from the study and being no longer traceable?

Center for Evaluation and Development

Non-response

- All the previously discussed types of non-response have two consequences on the study
  - Attrition is correlated with factor affecting outcome – biases the study
  - Reduced sample size leading to lower credibility of study
- Potential Solution: Use of replacement units
  - Only possible if there are remaining units in the sampling frame that can be selected as replacements
  - Increases the sample size but does not solve bias issue
  - If cases are replaced too easily then it is likely to increase the potential for attrition bias
- Careful replacement protocols should be developed

Center for Evaluation and Development

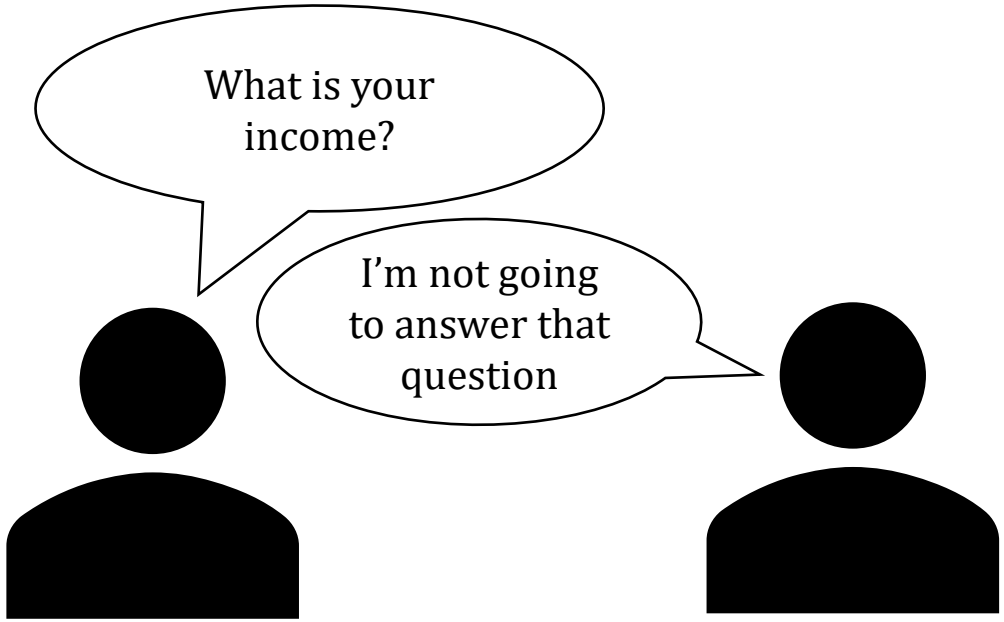
Field Work Protocol

- **Replacement protocols**
- How many visits need to be made before a household is replaced?
- If the field team is informed they have moved
  - Moved outside a radius of X km?
  - Moved to another district?
- Field staff should have clear rules on how to do replacements
  - Avoid bias (do not just go to the nearest replacement household) so replacements should be randomly selected – listed in order on the village lists

Center for Evaluation and Development

Factors affecting data quality

- There are two types of non-response
  - **Item non-response**:
    - Information required from a sample unit is missing or partially collected



Center for Evaluation and Development

Item Non-response

- Non-response can occur when respondents indicate that they either don't know or refuse to answer a question
- Question may be cognitively challenging (Recall session 3)
- A large risk of item non-response comes from asking potentially sensitive questions
  - May range from questions about income to sexual behaviour

Center for Evaluation and Development

Item Non-response

- Tools to address item non-response
  - For items such as income it is important to effectively communicate that the survey is not connected in any way to the government
  - Re-iterate confidentiality of responses
  - Privacy during interview
- Other sensitive questions may require more sophisticated tools

Center for Evaluation and Development

Item Non-response

- Tools to address item non-response
  - For items such as income it is important to effectively communicate that the survey is not connected in any way to the government
  - Re-iterate confidentiality of responses
  - Privacy during interview
- Other sensitive questions may require more sophisticated tools

Center for Evaluation and Development

Randomized response technique

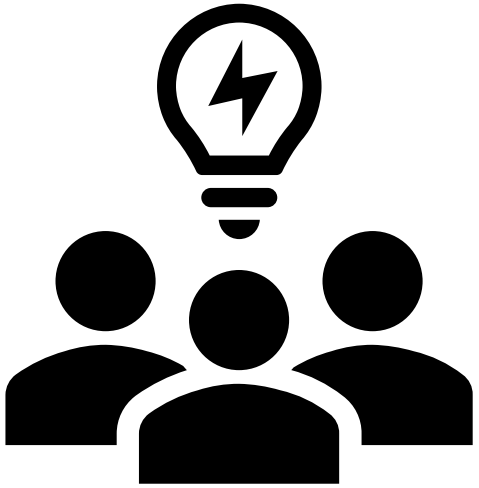
- Before being asked a sensitive question, respondents may be asked to roll a die. If the roll lands on a 5 or 6 the respondent should answer untruthfully.
- Because we know the probability of answering truthfully and untruthfully we can estimate the incidence
- This is particularly useful for sensitive questions because it does not reveal to the interviewer what the respondent answered
  - Reduce social desirability bias
  - Induce higher response rates
- This type of technique was used to study the prevalence of civilian cooperation with militant groups in south-eastern Nigeria (Blair 2014)

Self-Administered Questions

- Respondents are provided with the tool and are asked to complete sensitive modules themselves
- Particularly useful in using audio variant (audio computer-assisted self-interviewing, ACASI), where the respondent listens to questions via headphones and records his/her responses on the keyboard or screen.
- Avoids issues relating to education and literacy
- Review of the method found that it was more likely to capture socially undesirable responses (Phillips et al., 2010)

Breakout Session

Breakout Session 3



Center for Evaluation and Development

Breakout Session 3

- **Scenario**: Your program includes refugee participants who are based in refugee settlements, and you wish to collect data on them after the end of the program. Consider – given the context of the room country – the typical challenges that may occur in regard to the following areas:
  - Access to settlements
  - Tracking participants 6 months after the end of the program
  - Data quality issues when interviewing refugees
  - Contamination/spillovers (non-selected participants receiving the program)
- **Task:** For each of the areas discuss one **potential challenge** and discuss ways you can **overcome or mitigate** it. Once you have completed - copy and paste in the main chat with the name of your group. After the time is up we will return an review some examples as a whole group

Center for Evaluation and Development

Data Quality in Program Monitoring

- All of the previously discussed data quality issues are also relevant to a monitoring system.
- For program monitoring it is particularly important that data is timely and available:
  - Data must be available frequently enough to influence decision making
  - Data should be sufficiently up to date to be useful in decision-making

Center for Evaluation and Development

How do we ensure data quality?

- **How can data quality be ensured?**
- 1. Questionnaire Design (reviewed yesterday)
  - Question wording (Brief, objective, specific, simple)
  - Question sequence (Opening questions, Flow of questions, Location of sensitive questions)
  - Questionnaire length (Paced well but not too long to result in fatigue)

Center for Evaluation and Development

How do we ensure data quality?

- **How can data quality be ensured?**
- 2. Methodology of Data Collection:
  - Fieldwork protocols
  - Training of field staff
  - Method of administering the survey

Center for Evaluation and Development

Field Work Protocol

- *Regardless of the tool to be used, Data Collection Protocols are necessary.*

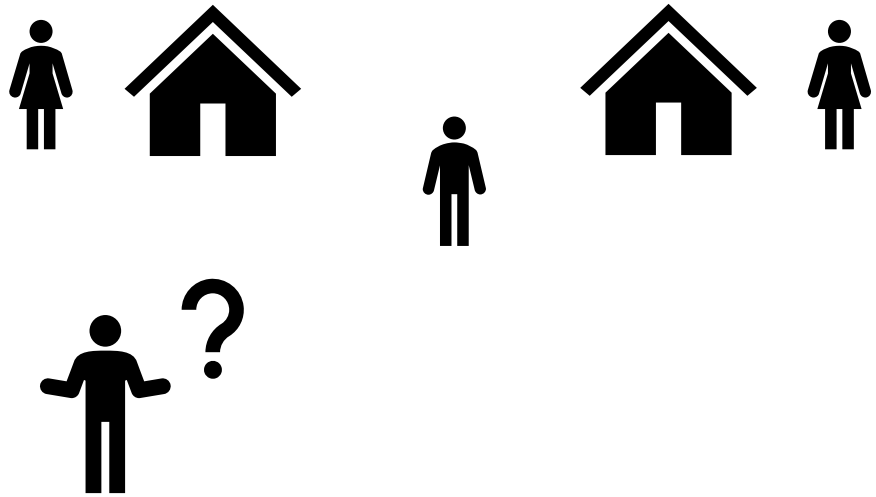
Center for Evaluation and Development

Field Work Protocol

- Fieldwork protocols are a set of rules that need to be followed during data collection.
- They should be guided by both technical considerations, budget constraints and realities in the field.
- They may include:
  - Tracking protocols
  - Interview target protocols
  - Replacement protocols
  - Ethical protocols
- They should be clear, complete and easy to follow

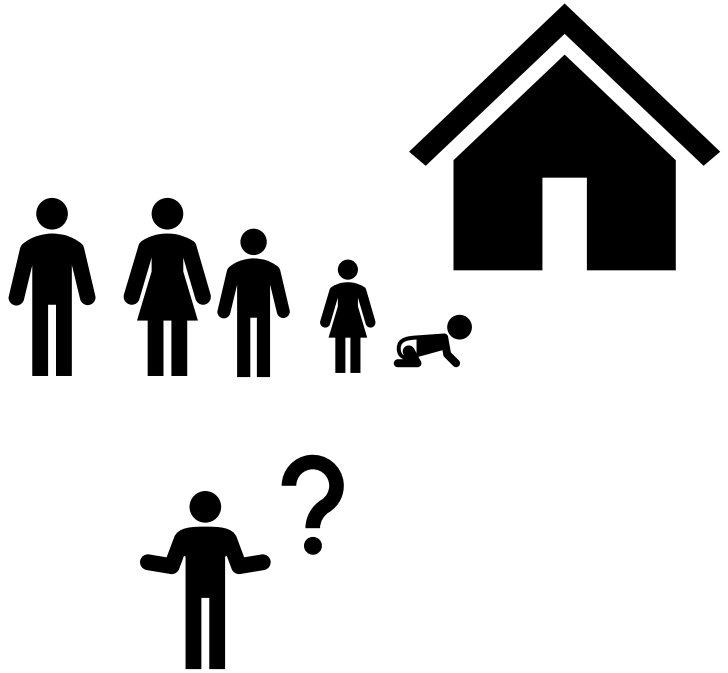
Field Work Protocol

- **Interview target protocols**
- Field staff need guidance on who exactly they interview.
- Often it is quite clear – the interview target is the participant in the programme so the enumerator needs to find their current household.
- Sometimes it's not so clear – if the respondent is polygamously married and has two households – which one do they interview?
- If this is a likely occurrence in the study area then it'll need to be outlined in the field work protocols



Field Work Protocol

- **Interview target protocols**
- Even if the enumerator is sure of the household, they need protocols for who exactly to conduct the interview with.
- If it's a simple household survey about members, education, expenditure you may allow anyone that is knowledgeable about that stuff
- You may have a preference system:
  - 1. Household head
  - 2. Spouse of household head
  - 3. Other knowledgeable household member
- Or you may say that only the head or the spouse can answer and if they are not available then there must be a revisit when they are



Center for Evaluation and Development

Field Work Protocol

- **Interview target protocols**
- Do responses depend on who is answering the survey?
- Many studies have shown that you may get different responses based on the household position of the respondent

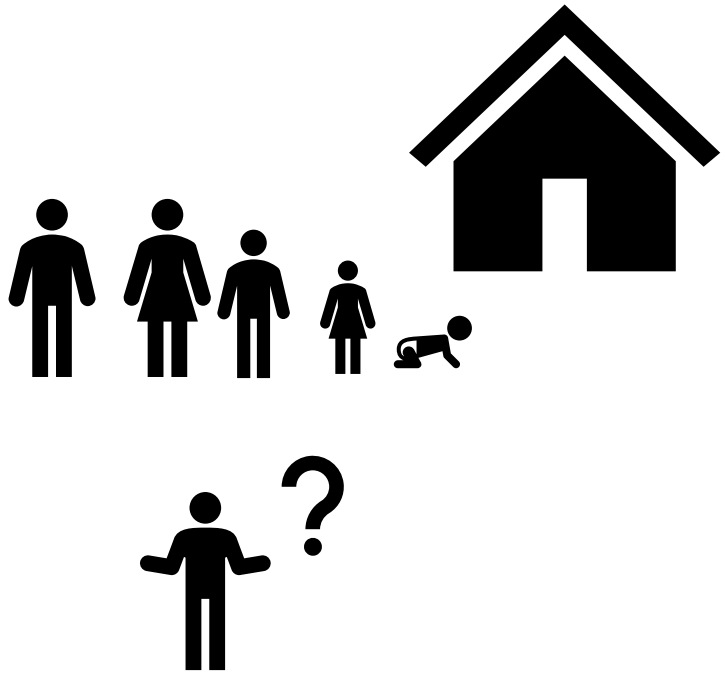
Center for Evaluation and Development

Field Work Protocol

- **Interview target protocols**
- One example was a study in Tanzania by Anderson et al. (2017)
- In the study spouses within a household were asked the same questions
- It was found that they gave different responses on
  - Who takes key decisions on farming
  - Livelihood measures of the household
- In another study in rural Malawi by Fisher et al. (2010) found that in 2/3rds of households Husband's underestimated their wife's income

Field Work Protocol

- **Interview target protocols**
- In other cases the survey may need to be answered by a specific person
- For example if you're interested in how the program changes perceptions or knowledge then you want to interview only the person who took part in the program



Center for Evaluation and Development

Field Work Protocol

- **Ethical Protocols**
- Typically for safeguarding rather than data quality
- If the research involves children or other vulnerable populations then this should be taken into account how field staff interact with them.

Center for Evaluation and Development

Field Work Protocol

- **Other things to consider for inclusion:**
  - Clear definition of roles and responsibilities
  - Communication channels
  - Data workflow:
    - Upstream: Enumerator reviews interviews after completion and then at the end of the day the supervisor reviews all interviews before submitting.
    - Downstream: Researcher informs supervisors of data issues who then in turn take up the issue with the relevant enumerator(s)
  - Data Protection Protocols

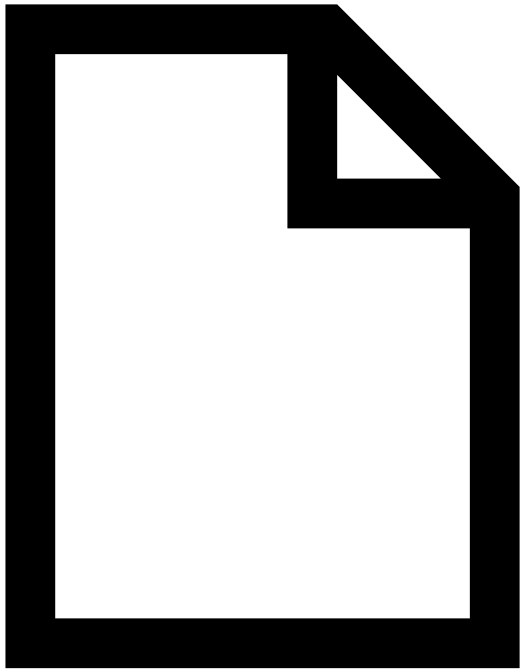
Center for Evaluation and Development

Field Work Protocol

- **Field work protocol documentation**
  - Document outlining how survey will be implemented.
  - Prepare this document well in advance and not the day or week before data collection begins.
  - Anticipate every single thing that could go wrong during data collection and plan around it.
  - List down each scenarios and suggest an appropriate line of action to remedy the situation
  - You might know what to do as a researcher but the field team might not know.
  - Aim is that rules are followed consistently across the whole field team

Field Work Protocol

- **Field work protocol documentation**
- Consider the best format for the field work protocols:
  - Programmed into the survey tool
  - Abridged 'cheat-sheet' style pages for various protocols for enumerators (Supervisors hold more comprehensive versions)
  - Clear integration into training
  - Avoid huge documents that may never even be read



Center for Evaluation and Development

Enumerator Training

- Enumerator Training
  - Provide field staff with in-depth knowledge on content of the survey with particular focus on reasons behind each question (Refer to research questions).
    - Focus on complex concepts and definitions
  - Training on interviewing skills
  - Continuous assessment of trainees (exam, monitoring during sessions, mock interviews etc.) and select only the most able (Invite more trainees than required)
  - Train enumerators on validations and consistency checks programmed into data collection tool to flag unlikely or incorrect responses.

Center for Evaluation and Development

Data Collection Methods

- Several ways to administer questionnaires: 2 popular ways
  - Pen and Paper Personal Interviews (PAPI):
    - Traditional “Pen and Paper based”.
    - Questions are answered onto paper questionnaire and entered later into a computer
  - Computer Assisted Personal Interview (CAPI):
    - Collecting data using electronic devices
    - Questionnaire is programmed on a computer and administered using either a computer, tablets, or phones.

END OF SESSION 5

Session 6: Leveraging Technology for High-Data quality for CIE and MS

C4ED – EUTF

September 2022

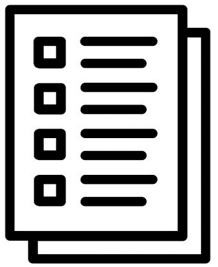
Center for Evaluation and Development

Session 6 Learning Objectives

- Participants will be able to describe the different methods of administering a survey
- Participants will be able to explain the benefits and challenges of each approach
- Participants will be able to describe the benefits and challenges of using telephone interviewing for data collection in hard-to-reach areas.
- Participants will be able to describe how GIS can be used to help prepare for a data collection
- Participants will be able to describe how GIS can be used to assess data quality

Session 6 Learning Objectives

- PAPI: Paper Based Personal Interviewing

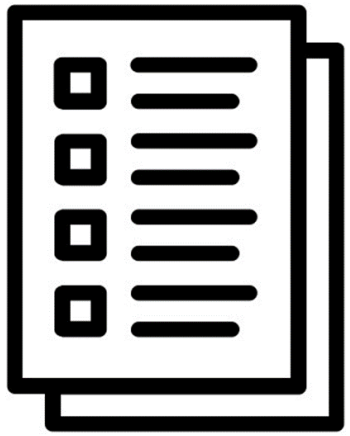


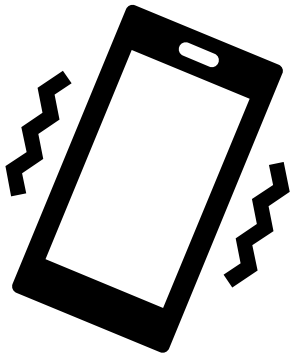
- CAPI: Computer Based Personal Interviewing

- CATI: Computer Based Telephone Interviewing

Center for Evaluation and Development

Decision: Pen-and-paper or Electronic





Center for Evaluation and Development

Paper Tools

Pros

- **No programming skills to develop**
- **No technical equipment required**
- **Requires no technical skill to administer**
- **Permanent physical copies**
  - **No threat of data loss due to corrupt files**

Cons

- Human error in following questionnaire routing
- Researchers typically only receive the data well after the end of data collection.
- Requires budget for data entry
- Permanent physical copies
  - More vulnerable to data protection issues
  - Paper copies can be damaged or lost in field

Center for Evaluation and Development

Electronic Tools

Pros

- **Data monitoring and quality checks in real time**
- **Automated questionnaire routing**
- **Real time consistency checks on data**
- **Data can be password protected and encrypted**

Cons

- Requires some expertise in programming
- Technical equipment can be costly
- Need to budget for power banks and mobile internet access
- Enumerators need to be trained on using the software
- Study participants may be suspicious of being recorded/photographed

Center for Evaluation and Development

Electronic v Paper

- RCT conducted where data collectors were randomly assigned to conduct household interviews with either PAPI or CAPI (Caeyers et al. 2012).
- The study found that interviews conducted with CAPI had:
  - Fewer routing errors
  - Fewer missing entries
  - Fewer unlikely and impossible answers
  - Interview time reduced by 10%
- Good for reducing item non-response and efficiency (and budget considerations!)

Center for Evaluation and Development

CAPI Software

Survey Solutions	SurveyCTO	KoboToolbox
Developed by the World Bank	Developed by Dobility using ODK Open source tool	Developed by KoBo Inc. using ODK Open source tool
Requires user to setup/have their own cloud or local server	Subscription fee	Free (with usage limits)
Extremely in-depth paradata	Range of plug-ins developed to enhance surveys	Less developed plug-ins
Simple design of complex questionnaires	Requires strong programming skills for complex questionnaires	Requires strong programming skills for complex questionnaires

Data Collection: Leveraging Tech.

- If you go ahead with using electronic tools, the next step is to consider how to administer the tool
- Due to the Covid-19 pandemic, the use of Computer Assisted Telephone Interviews (CATI) has increased significantly
- Other examples include:
  - Interactive Voice Response (IVR) surveys - automated voice surveys
  - Short Message Service (SMS) surveys - use text messaging.

Data Collection: Leveraging Tech.

- Sometimes data collection is not always as easy as it is imagined.
  - Conflict
  - Remote areas
  - Mobile populations
  - Pandemics (Covid-19)

https://www.dw.com/en/oecd-german-schools-falling-behind-in-digitalization/a-57209127

Data Collection: Leveraging Tech.

- Computer Assisted Telephone Interview:
  - Set up a centralised or ‘virtual’ call centre
  - Interviewers call respondents to collect data.
  - Responses are entered into interactive survey questionnaire digitally:
    - Computer
    - Smartphone
    - Tablet
- CAPI Software such as SurveyCTO and Survey Solutions have built in CATI functionalities to integrate phone calls into the app

Data Collection: Leveraging Tech.

- CATI data collections can be done at a relatively low cost compared to in-person data collections – can be done between 5 USD – 15 USD per respondent depending on context
- Expanded possibilities for data quality monitoring:
  - Phone Logs
  - Recordings of phone calls (consent and ethics)
  - Centralized location allows supervisor to monitor team members
  - Centralized location in city allows clearer communication between research team and field staff

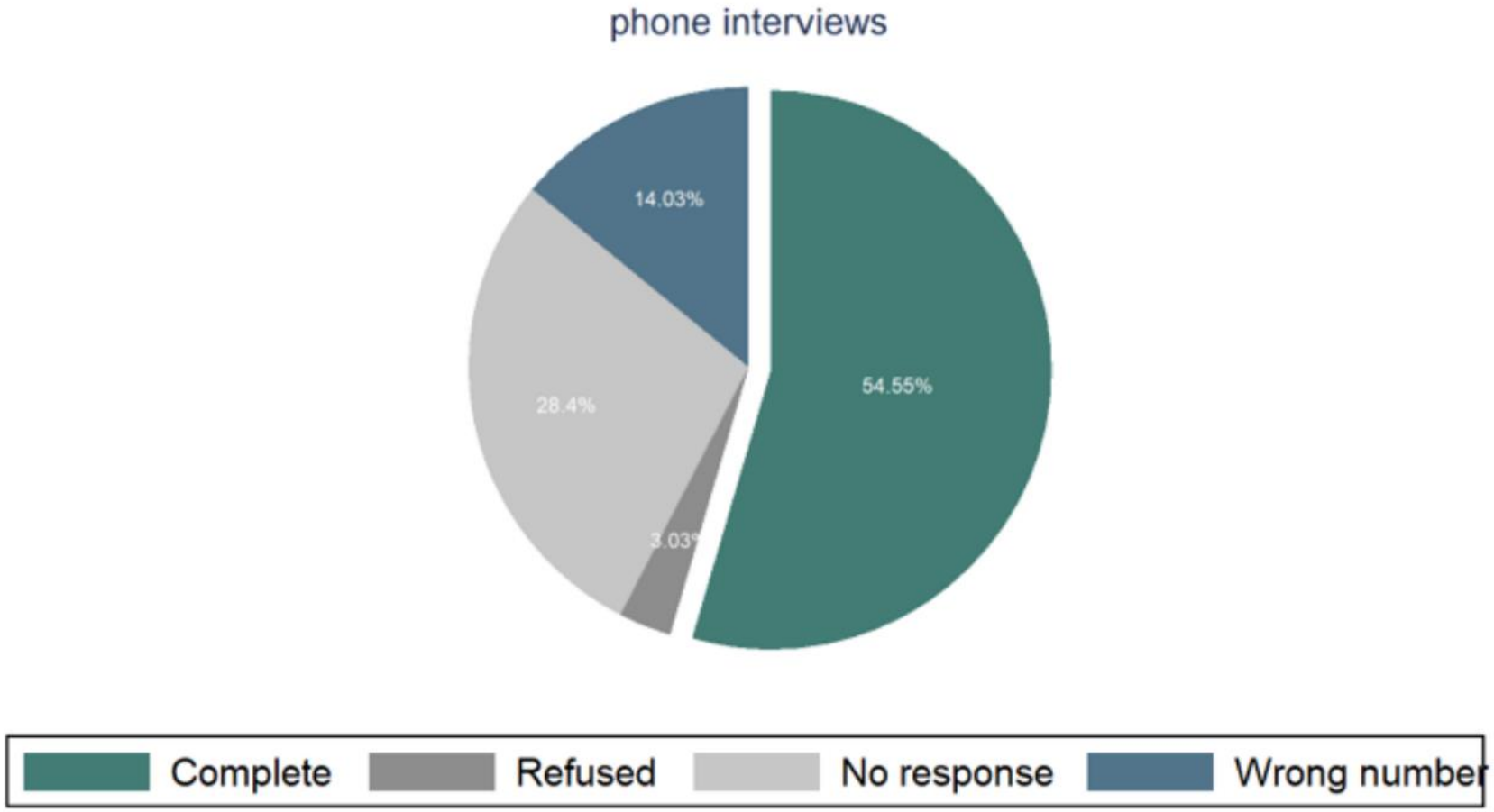
CATI

- Sounds great! So this is the end of in-person interviews?
- Not quite – there are some clear drawbacks of using CATI
  - Even a best-case scenario, phone surveys will be representative of the population that has a working phone – coverage error!
  - Attrition:
    - Phone numbers change/don't work – a NGO in Liberia attempted to contact respondents during the Ebola outbreak reported that six months after selection 43% of phone numbers were permanently switched off and 23% did not ring[i].
    - Other non-response such as not answering the phone or refusing to take part

i. https://elisamaffioli.files.wordpress.com/2020/03/maffioli_method.pdf

Center for Evaluation and Development

CATI

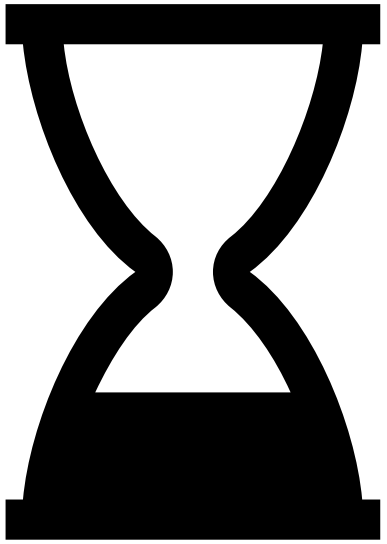


Recent C4ED CATI Survey by outcome

Center for Evaluation and Development

CATI

- Phone interviews should not last more than 30 minutes.
- A typical questionnaire should include 20-35 questions.
- This limits the amount of information that can be collected for a survey



Center for Evaluation and Development

CATI

- Consider the issue of sensitive questions and measurement error
- Telephone interviewing may exacerbate any concern over revealing the information.
- Harder to build a rapport and trust
- Easier to avoid and potentially jeopardize whole interview (hanging up on survey)

Center for Evaluation and Development

CATI

- When to use and when to avoid collecting data by phone

When to avoid…	When to use…
Large comprehensive household surveys	Collecting specific information on an individual
Surveys with modules that are complex, time consuming or highly sensitive	Relatively short surveys with easy to conduct modules
Data that requires physical measurement such as plot measurement, GPS, anthropometrics	More frequent surveys (monthly or quarterly)
Surveys that require visual aids	Reaching people that would otherwise be inaccessible

Adapted from: Mywish Maredia (2019). Using Mobile Phones to Do Research in the Time of COVID-19.CGIAR

Center for Evaluation and Development

Self-administration

- As discussed in Session 5, when asking sensitive questions modules can be self-administered to remove social desirability bias
- It is also possible to collect all data using self administration such as through setting up a web survey
  - Reduces the cost of data collection
  - Allows for more regular reporting
- Challenges:
  - Participation rates may be low
  - Participation may skew to a certain type of respondent -> Bias
  - Potential for misunderstanding or inconsistent understanding of questions without trained enumerator

Center for Evaluation and Development

Use of Apps

- Bespoke apps can be developed for assisting real-time reporting for participants
- Typical example is the use of GBV incident reporting mobile applications – Flone Initative's "Report It Stop It"
- Respondents enter their levels of perceived safety as well as experiences harassment and sexual violence when they occur
- App-embedded nudges can be utilized to incentivize higher reporting in the app

Center for Evaluation and Development

GIS

- Using GIS to plan surveys
  - Geographic Information Systems: Computer system for capturing, storing, checking and displaying data related to positions on Earth's surface.
  - Shows different kinds of data:
    - Streets, buildings, vegetation, population density, demographics
  - This information can be useful in analysing and understanding patterns and relationships.
  - Geographical information can play an integral role in data collection

Center for Evaluation and Development

GIS

- GIS system allows for comparison of different kinds of data that can be integrated into it.

….Geospatial information on distribution of a sample can be superimposed on information on the location of vocational and technical training institutes to understand better how proximity to an institute can improve uptake of interventions…

Center for Evaluation and Development

GIS

- How can integrating GIS into data collection facilitate high data quality from hard to reach areas?

- Data collections always need careful planning:
  - Better understand the population being surveyed
  - Sampling
  - Tracking and planning surveys
  - Data checks

Center for Evaluation and Development

GIS

- Better understand the population being surveyed
  - Envisionining the data collection demands that you have a lot of information about the target population.
  - Most useful at the planning stage.
  - Avoiding the pitfall of capturing the wrong respondents as part of the sample.
  - Understanding the population being surveyed provides enough information for the research team to anticipate how the sample might respond to the survey.

Center for Evaluation and Development

GIS

Can we just not use a simple map?

….Yes in theory we can use just a map and our local knowledge about study communities but GIS offers a wealth of information.

Center for Evaluation and Development

GIS

- Tracking survey respondents:
  - Mostly during follow-up surveys tracking respondents becomes absolutely necessary.
  - Not being able to track increases attrition levels
  - GIS can be used as a starting point to tracking respndents
  - Geographic location of beneficiaries can be superimposed on maps in GIS softwares.
  - Based on the assumption that either a baseline study, listing or monitoring data captures the location of beneficiaries.
  - The survey team can then plan its data collection accordingly

Center for Evaluation and Development

GIS

- This includes
  - Selecting the right enumerators and field supervisors.
  - Deciding how to assign respondents to enumerators.
  - Estimate travel time and other resources that would be needed in total for the data collection.

- The accuracy of the information from GIS for tracking will be largely dependent on what kind of information is fed into it

Center for Evaluation and Development

GIS

- Cons:
  - This is highly dependent on what data you feed into the system.
  - Accuracy can sometimes be reduced due to unavailability of GIS information on some extreme remote areas.
  - Highly dependent on the information provided by beneficiaries which cannot be always verified.
  - Learning to use GIS software can be a steep learning curve.

Center for Evaluation and Development

GIS

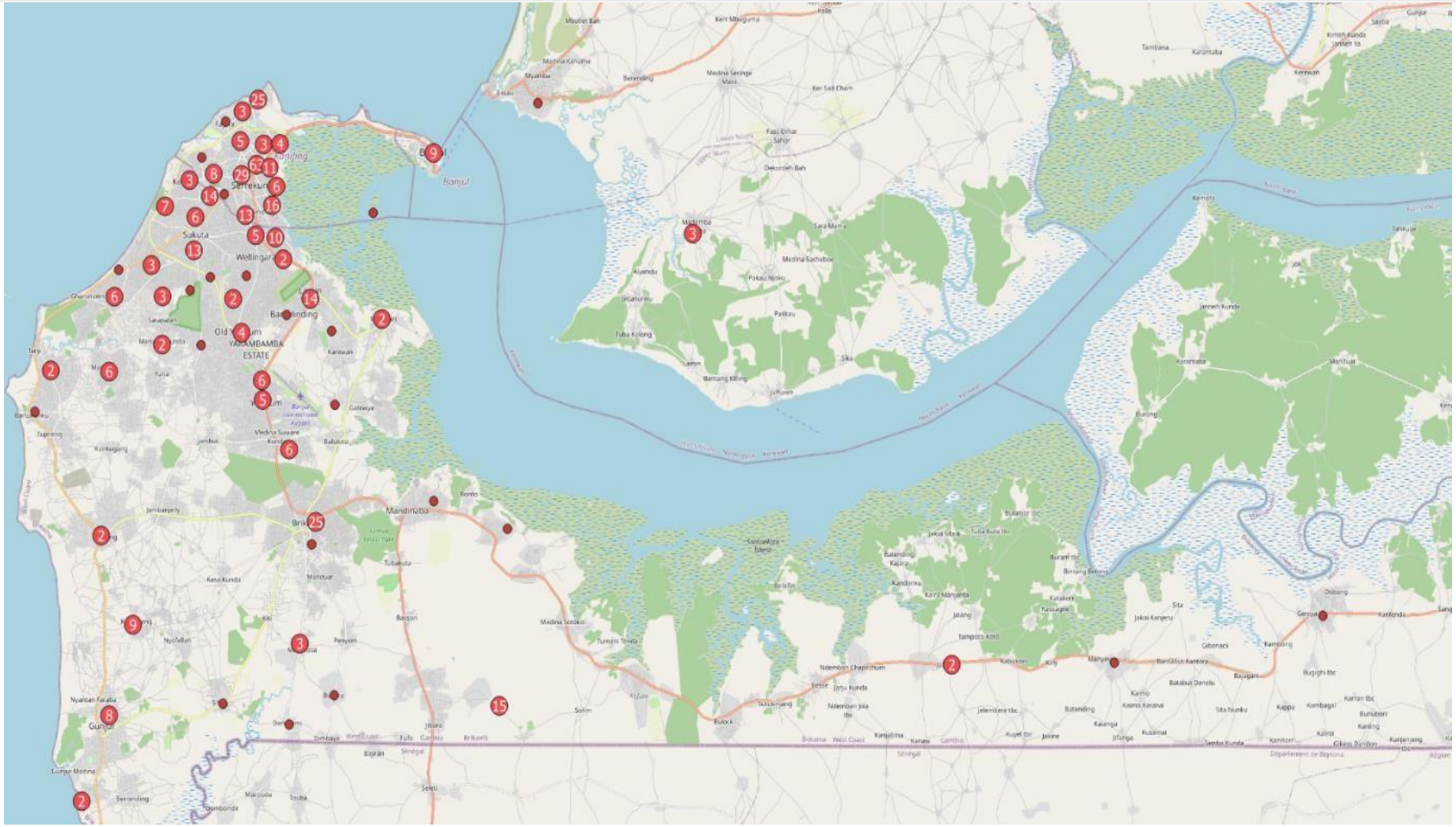


Fig: Map showing distribution of beneficiaries of an intervention in the Gambia

Center for Evaluation and Development

GIS

- Data quality checks
  - Most CAPI softwares have GPS functionality
  - GPS functionality allows for electronic devices being used for surveys to capture the geographical location of the point of data collection interviews.
  - Provides information on location where interview took place.
  - Provides monitoring information to ascertain if enumerators are following the sampling protocol.

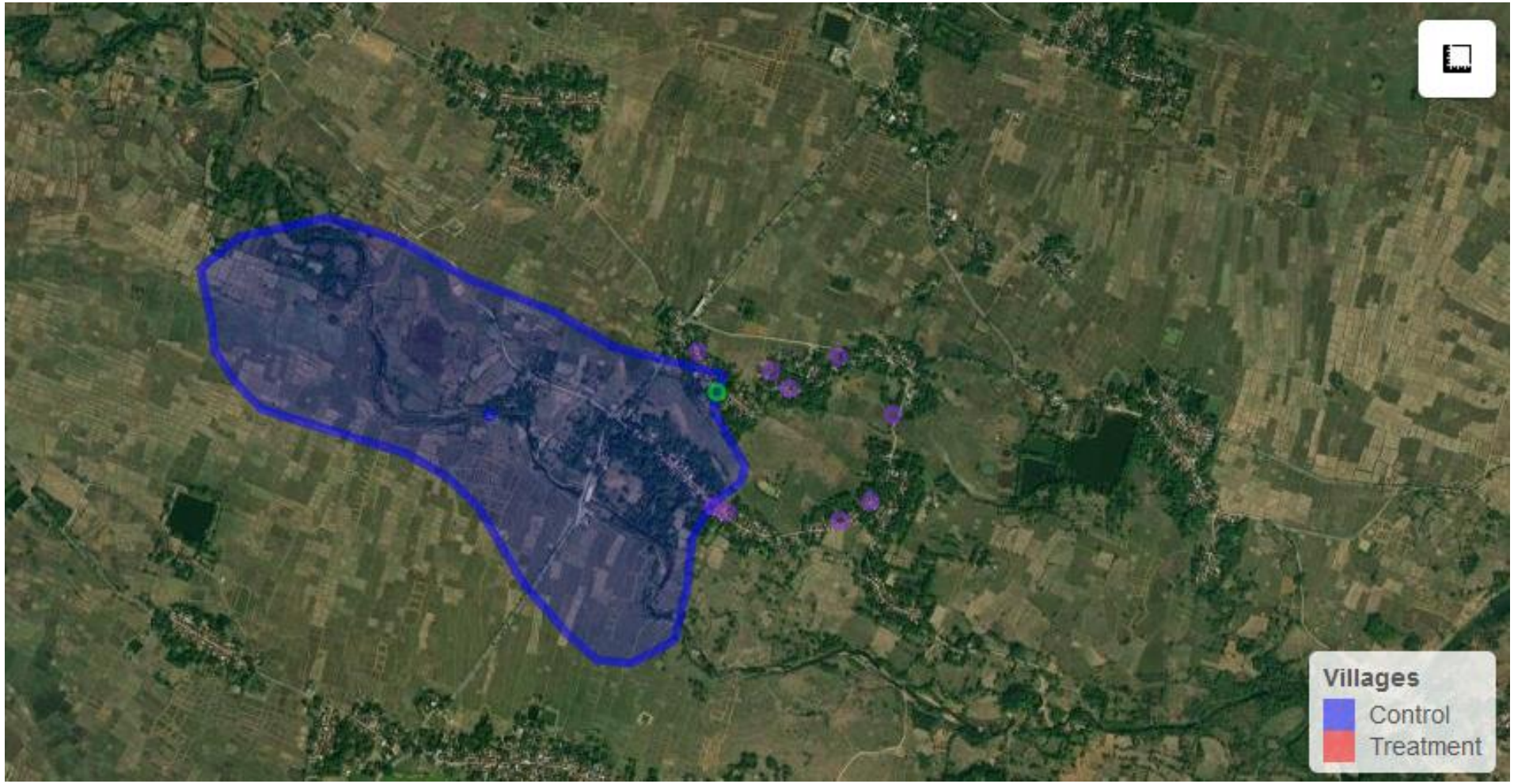
Center for Evaluation and Development

Data Quality Checks: GIS

Example of data collected within village boundaries

Center for Evaluation and Development

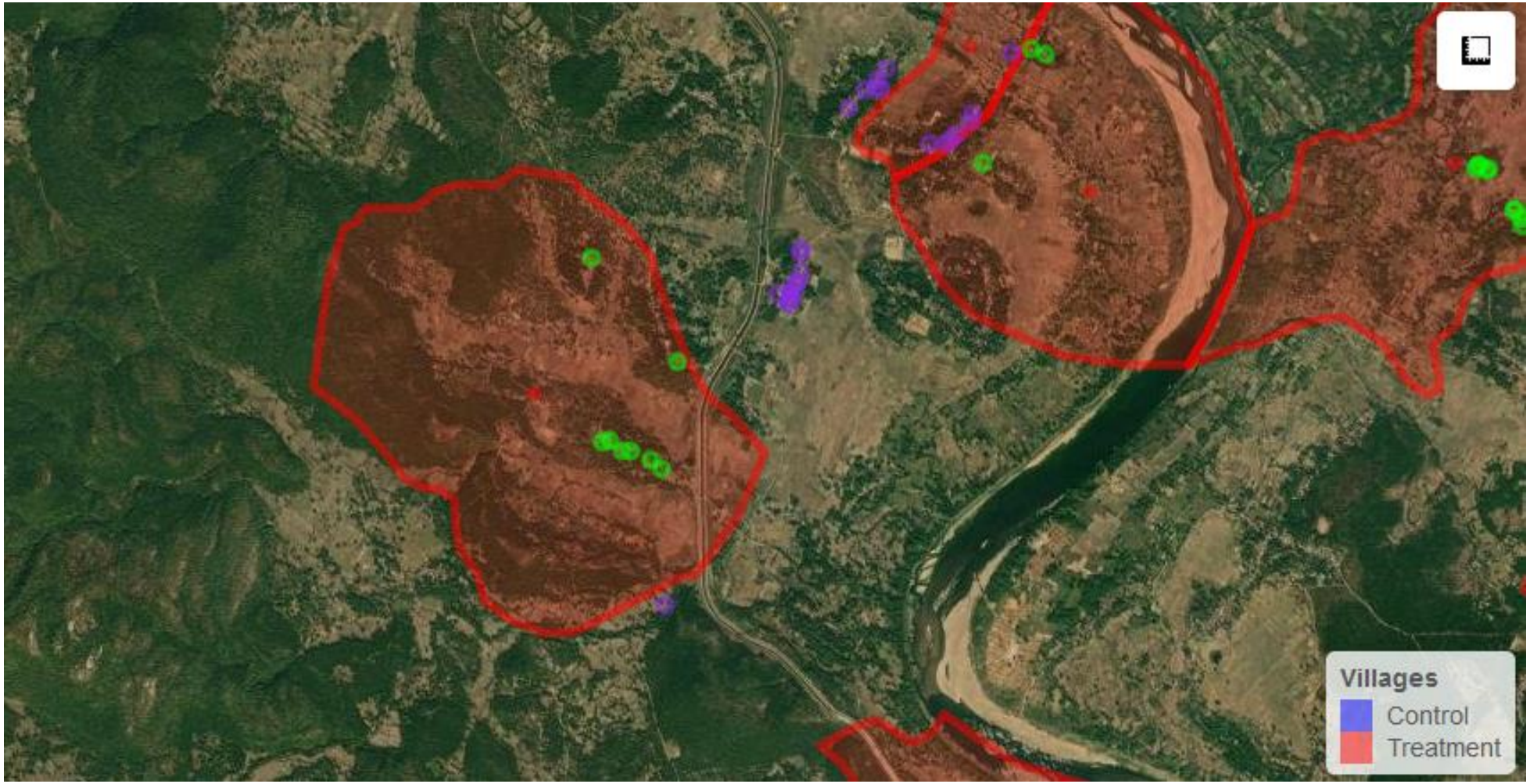
Data Quality Checks: GIS



Example of data not collected within village boundaries

Center for Evaluation and Development

Data Quality Checks: GIS



Example of data not collected within village boundaries

Center for Evaluation and Development

Session 7: Monitoring Systems as Inputs in CIE

C4ED – EUTF

September 2022

Center for Evaluation and Development

Session 7 Learning objectives

- Participants will be able to explain common misconceptions about monitoring systems and CIE
- Participants will be able describe the basic steps to building a monitoring system to integrate into CIE
- Participants will be able to describe innovations in collecting monitoring data

Center for Evaluation and Development

Monitoring Systems

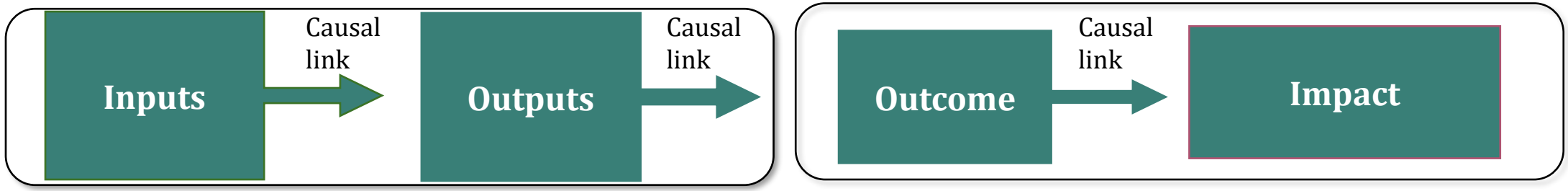
- Is CIE needed if monitoring data is available?
- What are some misconceptions about monitoring and CIE?
  - Is monitoring the same as impact evaluation?
  - Can monitoring be used in place of impact evaluation?
- Overview on building a monitoring system.
- Innovations in collecting monitoring data.

Monitoring Systems

Needs Assessment	Theoretical Program Evaluation	Process Evaluation, **MONITORING**	Impact Evaluation	Cost-Efficiency Analysis
• What is the challenge? • What is the objective of the program?	How, in theory, can we solve the challenge?	**Does the program work as planned?**	Are the impact objectives being met? To what extent?	Given its effect and cost, how does this program compare with other alternatives?

Monitoring Systems

Theory of Change and Monitoring



Project monitoring

CIE→Project's impact

Money spend per output
Material used per output
Human resource allocation per output

Immediate tangible and countable products/services produced

Center for Evaluation and Development

Monitoring Systems

Which questions should a functioning monitoring system answer?

How many resources were spent on which output?
How many resources are still available for which output?
To what extend is the project close to achieving the output target?
How is the project proceeding?
When to backstop and correct!

Center for Evaluation and Development

Why Do We Need Monitoring?

A team of evaluators assess the impact of a training on employment

First observations:
- Highly qualified trainers are conducting the trainings
- Curriculum meets the highest scientific standards
- Curriculum is based on the needs

Final result of the impact evaluation:
- **No effect of the training on the participant's employment is detectable!**

Why?
- After consulting the trainers, it turns out that most of the participants could only attend 4 of the 20 days of training due to extremely bad weather
No monitoring data on attendance was collected during the trainings

Center for Evaluation and Development

Setting up a Monitoring System

- Monitoring and Evaluation Learning (MEL) Plan.
- Performance Management Plan (PMP).
- Data management plan.
- Indicator plan and reporting.
- Monitoring data collection tools.
- Sampling approach.
- Reporting tool (Dashboard).

What do we want to monitor?
How do we want to monitor it?
When do we want to monitor it?
Who does the monitoring?

Center for Evaluation and Development

Setting up Monitoring Plan

Example of a monitoring plan for inputs

Input	Unit of Measurement	Data Source	Frequency of measurement	Budget	Related Output	Person Responsible

Example of a monitoring plan for outputs

Output	Indicator	Type of Indicator	Unit of Measurement	Data Source	Frequency of Measurement	Target	Person Responsible

Center for Evaluation and Development

Innovations: Data Collection for Monitoring

- Collecting data using CAPI.
- Data automatically fed into the monitoring system (barcodes).
- Developing MIS to automize data presentation and reporting.
- Building high frequency checks and logic skips within the MIS to detect inconsistencies.
- Linking MIS to existing systems.
- Other ideas?

Center for Evaluation and Development

Session 8: Research Ethics and Data Protection

C4ED – EUTF

September 2022

Center for Evaluation and Development

Session 8 Learning objectives

- Participants will be able to explain practical requirements for ethical clearance of a CIE
- Participants will be able to describe the principals underpinning ethical research
- Participants will be able to identify types of data that need to be protected
- Participants will be able to describe various methods of ensuring that private data is protected

Center for Evaluation and Development

Introduction to Ethics in Data Collections

- *A study interviewing trainers may discover that they are behaving against the ethics of their profession; reporting this misconduct to the authorities could harm the trainers, while not reporting could harm trainees…….*
- *What would you do?*

Center for Evaluation and Development

Introduction to Ethics in Data Collections

- *….A study on the employment returns of vocational and technical skills for young people in Ghana is currently ongoing. One of the beneficiaries who is 16 years old is approached to be interviewed. The respondent reads the consent section and asks relevant questions about the study and shows a good understanding of the research……..*
- *Is it ok to continue asking the questions?*

Center for Evaluation and Development

Introduction to Ethics in Data Collections

- Protection of human subjects (respondents) and their rights especially to rights to privacy is central to ethical data collections.
- Core Question: Are the research protocols comfortable for respondents and communities as a whole?
- Design an ethical framework for the study both at design stage and implementation.
  - Be proactive (Procedural ethics)
  - Be reactive (Reflexivity)
- Ethical considerations in data collections
  - Informed consent
  - Institutional Review Boards (IRB)
  - State and local research permits

Center for Evaluation and Development

Informed consent

- Respondents have the right to decide not to be part of the research.
- **Informed Consent**
  - Ensure that respondents have a clear understanding of the contents of the consent note.
  - Language is key in the consent section.
  - A clear understanding that they can withdraw their consent midway and that should be respected.
  - Disclose any gifts or rewards that respondents might receive.
  - Informed consent for children must be obtained from parents and guardians.

Center for Evaluation and Development

Informed consent

- Key sections of consent notes
  - Purpose
  - Procedures, risks and benefits
  - Rights
  - Confidentiality
  - Contact information (Further questions)
  - Response

Center for Evaluation and Development

IRBs and Permits

- **Ethical Clearance** from Institutional Review Boards (IRB)
- IRB seek to ensure ethical protection of participants of a research
  - Respect for persons
  - Beneficience
  - Justice
- Most countries have IRB boards now
- Universities and academic institutions have taken the lead in issuing IRB approvals
- Some other countries have state IRB institutions
- IRBs serves as a system for checks and balances in data collections for CIEs

Center for Evaluation and Development

IRBs and Permits

- For most IRB approval processes, the following documents must be submitted
  - Study design including research questions and objectives
  - Data collection methodology
  - Data collection tools (questionnaire)
  - Informed consent note
  - Potential risks and mitigation actions
  - Details of Principal investigators and research assistants
  - In some cases Human Subjects Research Certification from accredited institutes.

IRBs and Permits

- Will an IRB be relevant in an impact evaluation such as an RCT that randomly chooses refugees and people from host communities to be trained and others not chosen to be trained? If yes why? If no why?

- Collecting data from children Grade 4 & 5 on their preparation to professional life. The only variables involved here are self perception, self efficacy, life skills. There is no need for collection of sensitive data. Will an IRB be needed for the this IE? If yes why? If no why?

Center for Evaluation and Development

IRBs and Permits

- IRB is relevant in both cases even more in a harmless and not so sensitive data collection involving children
  - Autonomy of children
- IRB approval processes will look into the whether randomisation is to the detriment of some participants
  - Randomisation would be to the detriment if the intervention is a basic need that participants are entitled to by law or convention

Center for Evaluation and Development

IRBs and Permits

- Challenges with IRB approval processes
- IRB approvals processes can often be a huge hurdle which could delay survey data collections (Average of 4 weeks for expedited approval)
- When in-country IRB approval structure does not exist, out country IRB approvals can be sought
  - Reviewers lack contextual knowledge to ensure a proper review
- IRB processes can be long and cumbersome
  - Prepare ahead of time for submission
  - Submit a draft of questionnaire even if it is yet to be finalized
  - Amendments can always be made if significant changes come up

Center for Evaluation and Development

IRBs and Permits

- IRB standards are written for academic and medical research
  - Demands of academia vary from implementation of development intervention.
- Many IRB approvals do not cover ethical dilemmas that arise in the field.
- IRBs must be complemented with researcher responsibility.

Center for Evaluation and Development

IRBs and Permits

- IRB approvals are not a substitute for local and institutional permissions
- Research permits and local approvals are equally important
- Local and institutional permissions are mostly interested in research objectives and questions, study design, impacts of the study if any and publication plans.
- Demonstrates adherence and respect for local regulations
- Buy in of local authorities into CIE methodology and legitimization of CIE process.

Center for Evaluation and Development

IRBs and Permits

- Obtaining local permits sets up a collaboration between research team and local authorities who most likely are community gatekeepers
- Obtain a letter of support from relevant federal ministry or state institution (Some IRBs require letters of support from relevant state institutions)
- Local level government institutions should also always be informed about data collections even after obtaining permission from federal or national institutions.

Center for Evaluation and Development

IRBs and Permits

- At the local level consider both local level government officials and if possible ***local traditional institutions*** who are custodians of local customs and laws.
- Aside legitimizing CIE methodology and data collection, it opens doors by granting access to respondents.
- Once traditional authorities are aware of data collections participants or respondents are more comfortable to participate.

Ethical Considerations: Main takeaways

What do all if these mean in practical terms?

Center for Evaluation and Development

Ethical Considerations: Main takeaways

- Ethical approvals
  - Consider getting certification to conduct research
  - Obtain necessary IRB approvals from relevant local or international institutions
  - Also obtain permissions from local authorities

- Requesting informed consent
  - Informed consent must be obtained with full knowledge on use of information obtained from survey.
  - Consider minors and respondents who do not have autonomy

Center for Evaluation and Development

Ethical Considerations: Main takeaways

- Respect right to privacy
  - Confidentiality
  - Privacy
  - Anonymity of respondents
- Be proactive and reactive
  - Outline data collection protocols clearly
  - Train enumerators appropriately to observe ethical processes
  - Put in data checks to ensure ethics protocol are being followed

Center for Evaluation and Development

Ethical Considerations

- Adverse events like this should be clearly defined in the Ethical Protocols.
- It should not be on the individual field staff member to decide what to do in the situation
- Clearly defined steps to take that consider:
  - Local customs or laws
  - Organisational (Donor, Researcher, Local Partner) guidelines
- It should not be on the individual field staff member to decide what to do in the situation

Center for Evaluation and Development

Data Security and Protection

Universal Declaration of Human Rights (Article 12)

*"No one shall be subjected to arbitrary interference with his **privacy**, family, home or correspondence, nor to attacks upon his honour and reputation. Everyone has the right to the protection of the law against such interference or attacks".*

Center for Evaluation and Development

Data Security and Protection

- The right of respondents to their privacy supersedes all research and personal interest.
- Consider data protection and data security from day 1 of planning the CIE or MS.
- If data security cannot be guaranteed , then the data collection might not be worth conducting.
- Data protection protocols must be submitted during IRB applications.
- Some IRBs request for human subject certificate from data collectors.

Center for Evaluation and Development

Data Security and Protection

- What data has to be protected?
  - Determine early on before data collection begins if data is confidential.
  - Almost every survey has confidential data especially Personally Identifiable Information (PII).

- PII is any kind of information that allows identification of a data subject
- PII can come in form of direct and indirect identifiers

Center for Evaluation and Development

Data Security and Protection

- **Direct identifiers** uniquely identify a person
- Examples include:
  - National ID Number
  - Full name
  - Phone number
  - Audio file of person speaking
  - Photograph of person's face
  - Household Location (Descriptive Address, GPS)

Center for Evaluation and Development

Data Security and Protection

- **Indirect identifiers** can be combined to uniquely identify a person
- Examples include:
  - Date/Year Of Birth
  - Household Composition
  - Occupation
  - Education
  - Ethnicity
  - Nationality
  - Workplace/Employer
  - Geographic Area With Less Than 20,000 Inhabitants

Center for Evaluation and Development

Data Security and Protection

- Consider this – if you made a complaint about your colleague and it included a combination of:
  - Job Title
  - Gender
  - Age

- Would you feel comfortable with that information being shared?

Center for Evaluation and Development

Data Security and Protection

- Data being shared over the internet or on electronic devices is vulnerable to being read by people who have the skill to.
- This can be avoided if data is encrypted when being shared
- Ensure that data collection tool (CAPI) exports data in an encrypted form.
- PAPI data present a 2 pronged problem
  - Data on the filled up questionnaires
  - Data entered electronically using the paper based questionnaire

Center for Evaluation and Development

Data Security and Protection

- Data encryption
- Deidentifying data

Center for Evaluation and Development

Data Security and Protection

- Encryption:
  - Data encryption changes the structure of the data and hides its real meaning.
  - Only persons with the decryption key can actually transform the data to reveal its true meaning.
  - The decryption of the data must only be known by individuals who have the authority to this data.
  - Keep encrypted data in a secure location
    - No matter how robust encryption is, not having it in a secured location is a weak link.

Center for Evaluation and Development

Data Security and Protection

- De-identify data
- Firstly try not to collect too much PII
  - A lot of PII means it can always be linked with indirect PII
  - Data minimisation: Only collect what you need and is relevant to research objectives
  - Collecting personal information which is not relevant for the research objectives is unethical
- De-identifying data involves removing all direct and indirect PII.

Center for Evaluation and Development

Data Security and Protection

- De-identifying data involves removing all direct and indirect PII.
  - Pseudonomysation:
  - Anonymisation:
- De-identify data as early as possible:
  - Immediately after data export to computer.

Center for Evaluation and Development

Data Security and Protection

- For data checks and revisits PIIs will still be needed but this should be only accessible to the research data manager whose role must have been spelt out in the protocol.

- Once data is de-identified, keep identified data in a secure location.
- De-identified data does not need to be enrypted.
- When sharing data with partners or clients, ensure every PII has been removed before sharing.
- Protect excel sheets with passwords before sharing.

Center for Evaluation and Development

Data Security and Protection

So many passwords to keep?

…Using your birthday as a password is pretty obvious….

……Using you child's name as a password is also pretty obvious…

……Using your pet's name is also very obvious……

Center for Evaluation and Development

Data Security and Protection

- Password managers:
  - Password managers help you generate unique and hard to decode
  - Storage for passwords
  - But.....

  - **DO NOT FORGET THE PASSWORD TO THE PASSWORD MANAGER.**

  - Only individuals who really need to have access to data with PII should have access to these passwords.

Center for Evaluation and Development

Data Security and Protection

- Make use of two-factor authentication
  - You have to enter a second confirmation using your mobile device when it is logged into any of your data repositories from a new location.

Center for Evaluation and Development

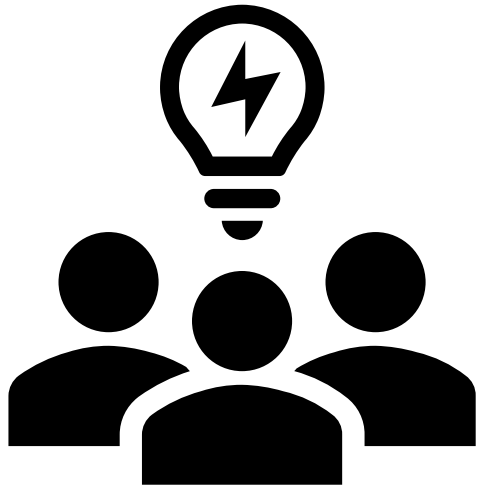
Data Security and Protection

- Data Protection Policy and Plan:
- Document that provides clear guidelines on responsibilities, roles and actions necessary to protect the data obtained from CIE and MS.
- Be proactive in coming up with a plan if none exists already.

Center for Evaluation and Development

Breakout Session

Breakout Session 4



Center for Evaluation and Development

Feedback

- We will like you to give us some feedback
- Please enter this code to answer the feedback questions.
  - Code: **32144851**

Center for Evaluation and Development

Quiz Day 2

- We'll end with a short quiz on CIE Methods

- Kindly look in the chat section for the link to the quiz
- Enter your name and use the code below if required
  - Code: **24641478**

Center for Evaluation and Development

Q&A

References

Wang, R. Y. & Strong, D. M. (1996) Beyond Accuracy: What Data Quality Means to Data Consumers, Journal of Management Information Systems, 12(4), 5-33

Groves, R. M., Fowler, F. J., Couper, M., Lepkowski, J. M., Singer, E., Tourangeau, R. (2009). Survey methodology. Wiley.

Anderson, C. L., Reynolds, T. W., Gugerty M. K. (2017). Husband and Wife Perspectives on Farm Household Decision-Making Authority and Evidence on Intra-Household Accord in Rural Tanzania, World Development, 90, 169-183.

Fisher, M., Reimer, J. J., Carr, E. R. (2010). Who Should Be Interviewed in Surveys of Household Income?, World Development, 38, 966-973.

Caeyers, B., Chalmers, N., De Weerdt, J. (2012). Improving Consumption Measurement and other Survey Data through CAPI: Evidence from a Randomized Experiment, Journal of Development Economics, 98 (1), 19-33.

Himelein ,K. Eckman, S. Lau, C. Mckenzie, D. (2020). Mobile Phone Surveys for Understanding COVID-19 Impacts: Part I Sampling and Mode (Himelein,K. Eckman, S. Lau, C. Mckenzie, D. (2020). Mobile Phone Surveys for Understanding COVID-19 Impacts: Part I Sampling and Mode (https://blogs.worldbank.org/impactevaluations/mobile-phone-surveys-understanding-covid-19-impacts-part-i-sampling-and-mode)

Maffioli, E. M. (https://elisamaffioli.files.wordpress.com/2020/03/maffioli_method.pdf)

Mywish Maredia (2019). Using Mobile Phones to Do Research in the Time of COVID-19. CGIAR (https://pim.cgiar.org/2020/04/11/using-mobile-phones-to-do-research-covid-19)

Guillemin, M., & Gillam, L. (2004). Ethics, Reflexivity, and “Ethically Important Moments” in Research. Qualitative Inquiry, 10(2), 261–280.

United States. (1978). The Belmont report: Ethical principles and guidelines for the protection of human subjects of research. Bethesda, Md.: The Commission.