Session 1:
From Methods to Analysis
(Recap of Year 1 – Year 3)

C4ED – EUTF

October 2024

Center for Evaluation and Development

European Commission

Welcome to the Training Workshop on 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

Communication during the training

MUTE BUTTON

QUESTIONS

FEEDBACK

Communication during the training

MUTE BUTTON

QUESTIONS

FEEDBACK

Communication during the training

MUTE BUTTON

QUESTIONS

FEEDBACK

Asking Questions

- Please post your questions in the chat room
- Like 👍 the questions of others, so we know they are particularly relevant for you as well
- Carolin will read out all questions and we will answer these at once
- Opportunity to share your experiences
- Use the longer breaks to ask more questions

Communication during the training

MUTE BUTTON

QUESTIONS

FEEDBACK

Giving Feedback

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

Introduction

Day 1 Agenda

10:30 – 11:10	Session 1: Recap of Year 1 (CIE Methods), Year 2 (Data collection) and Year 3 (Data analysis)
11:10 – 11:30	Evaluation quiz for Year 1, 2 and 3
11:40 – 12:10	Session 2: Introduction to basic concepts of evidence synthesis
12:10 – 12:20	Q & A
12:20 – 12:50	Discussion: The need for Evidence synthesis and aligning real-world practices and experiences
12:50 – 14:00	Lunch
14:00 – 14:40	Session 2a: Evidence Synthesis Methodology Part I – Searching & Screening
14:40 – 15:00	Session 2b: Evidence Synthesis Methodology Part II – Data extraction & types of analysis
15:00 – 15:30	Session 2c (Breakout session): Guided walkthrough of an evidence synthesis methodology highlighting 'Searching, Screening, Extraction and Analysis
15:30 – 15:45	Interactive quiz
15:45 – 16:00	Getting hands on: Using an evidence gap map (EGM) and a meta-analysis for future programme and evaluation planning
16:00 – 16:10	Feedback &Closing Day 1

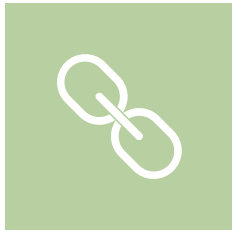
Overall aim of capacity-building trainings

- Improve development effectiveness through impact evaluations
- Share knowledge with partners and stakeholders on necessary tools for impact evaluations
- More specifically knowledge:
  - Impact evaluations with counterfactuals and control groups (*Year 1*)
  - High quality microdata collection (*Year 2*)
  - Data analysis (*Year 3*)
  - Evidence synthesis (*Year 4*)
    - ***Equip partners with knowledge on how to synthesize evidence including EUTF results into future programs***

Objectives of Session 1

Refreshing our knowledge on Counterfactual Impact Evaluation (CIE), data collections for CIE and data analysis of CIE data

Reviewing the basics on counterfactuals, common methods of identifying impact and the importance of high-quality data



Sharing useful external resources and case studies on CIE

RECAP YEAR 1

Counterfactual Impact Evaluation (CIE) Methods

What is a counterfactual impact evaluation?

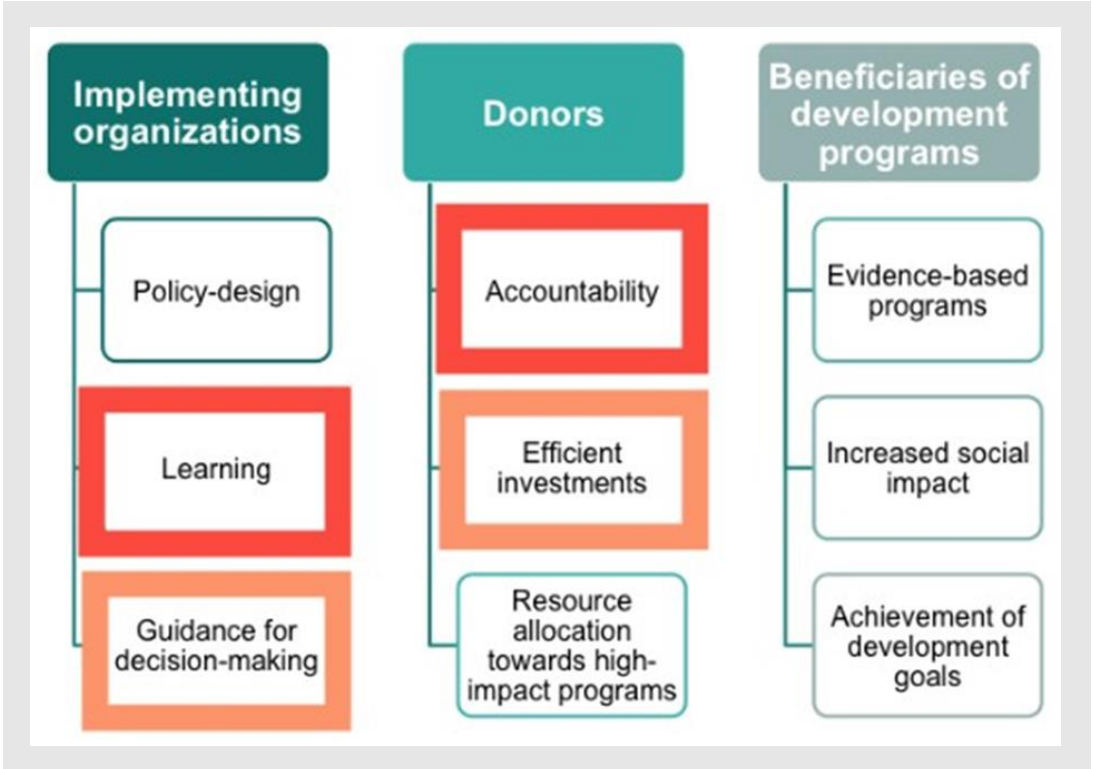
Evaluation	Systematic assessment of program design, implementation or results to support learning or decision-making
Impact	Direct effect of the program on outcomes, clearly attributable to the intervention (causal effect)
Counterfactual	Outcome, measured at that same point in time, had the program not been introduced to the beneficiaries.

Fundamental problem
It is impossible to measure or observe the counterfactual

Solution
Simulate counterfactual situation → creating a control/comparison group

Why do a counterfactual impact evaluation?

- To determine whether an intervention creates an **attributable, causal change** in the outcome, **how (the causal mechanism)** and to what **magnitude**
- To **learn** which intervention strategy works best
- To help make **evidence-based decisions**



How is CIE designed?

Goal: Simulate counterfactual situation with a comparison group

The comparison group:

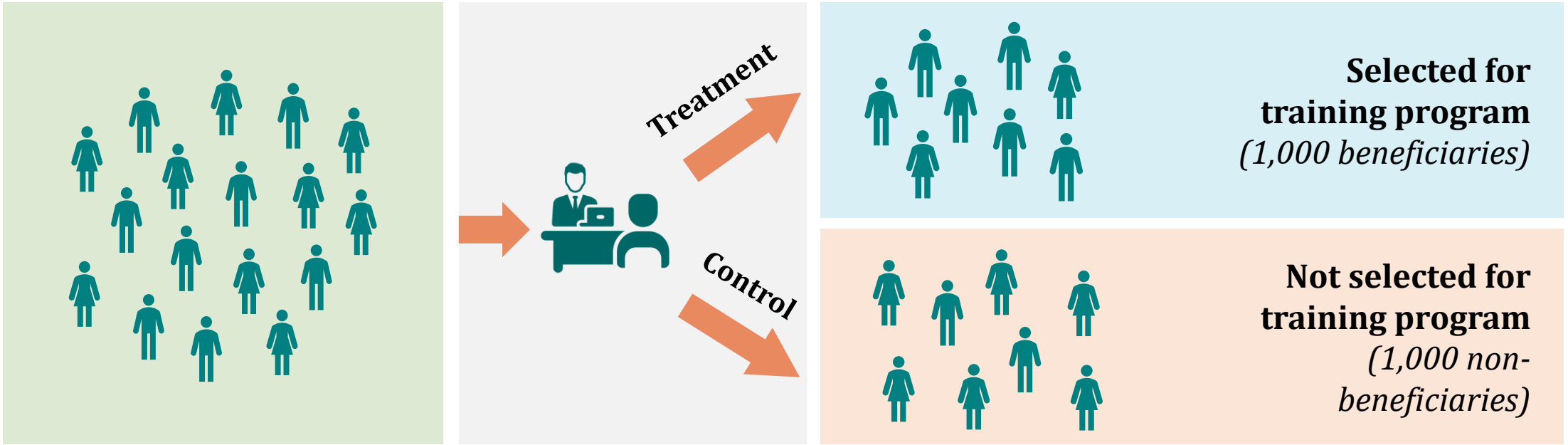
- Has the same characteristics (on average) as the treatment group
- Is not exposed to the program
- Would react similarly to the program as the treatment group (if it were to participate)

Intervention design, context, timeline, data availability and budget determine method:

- Experimental methods
- Quasi-experimental methods

Simulating a counterfactual

Example: Selection for youth vocational training program



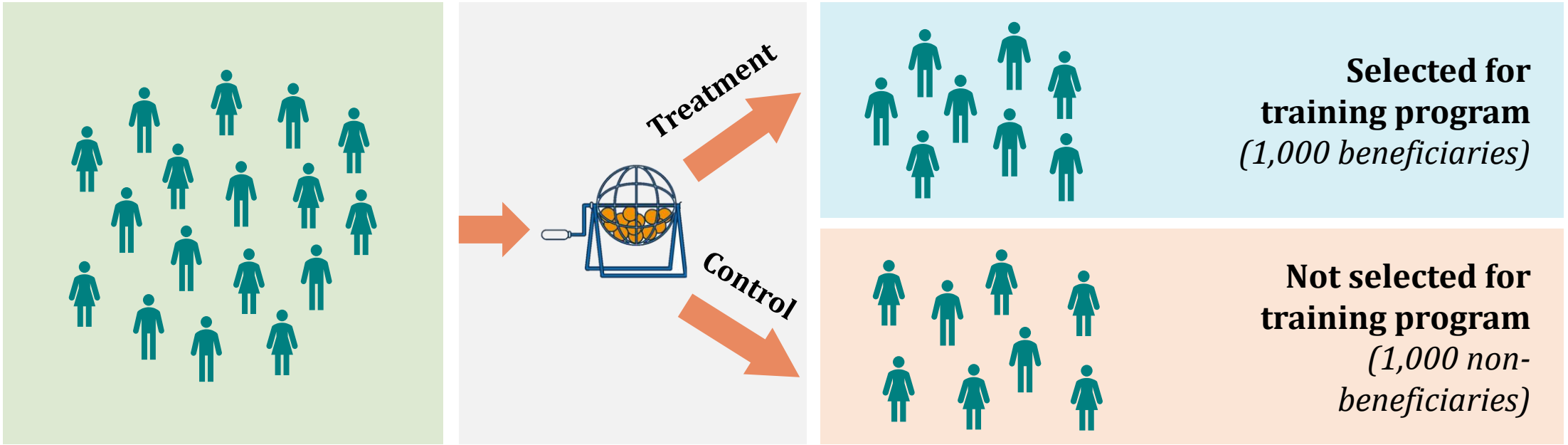
Applicants: 2,000
(1,000 Males / 1,000 Females)

CIE Recap

Experimental
Randomized Control Trial

Experimental Randomized Control Trial

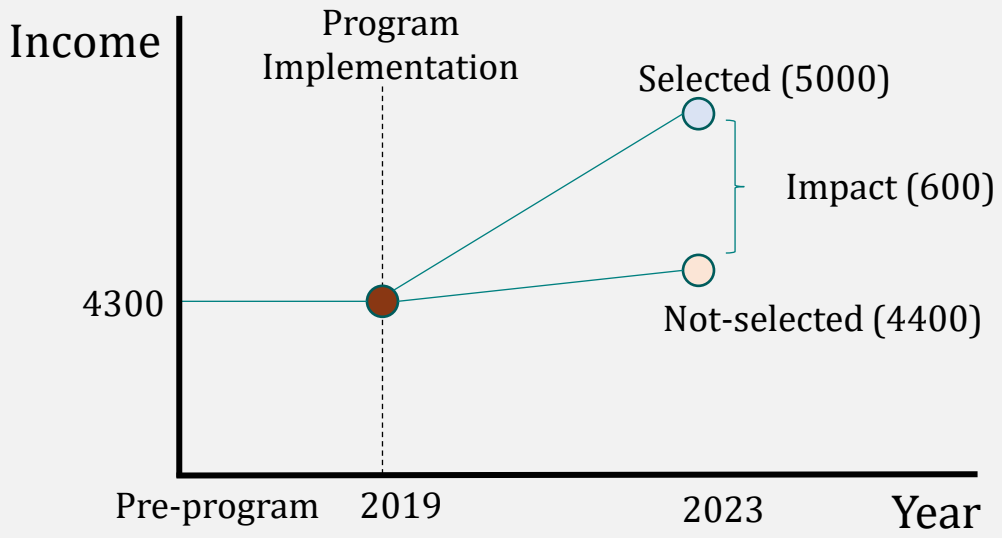
Example: Selection for youth vocational training program



Applicants: 2,000
(1,000 Males / 1,000 Females)

Measuring Impact in an RCT

- As the randomized assignment creates two groups that are (on average) comparable at the beginning of the program, ***impact*** can be measured simply as the difference in the outcome after the program.
- Differences in outcomes between groups can be ***attributed*** to the program



Randomization – What, when, how

What to randomize?

- Any aspect of the program that the implementation team *fully controls*
- Often requires creativity and a thorough knowledge of the program

When to randomize?

- *Before* program starts, must be included as part of program implementation

How to randomize?

- Simple lottery
- Multiple treatment arms → Can test different treatment modalities
- Phase-in → Delayed treatment for some program beneficiaries
- Encouragement → All have access to program, but some beneficiaries are actively encouraged to participate

Quasi Experimental
Matching & Difference-in-differences (DiD)

Quasi Experimental: Matching

Selected for training program *(1,000 beneficiaries)*

Not selected for training program *(1,000 non-beneficiaries)*

Treatment **Comparison**

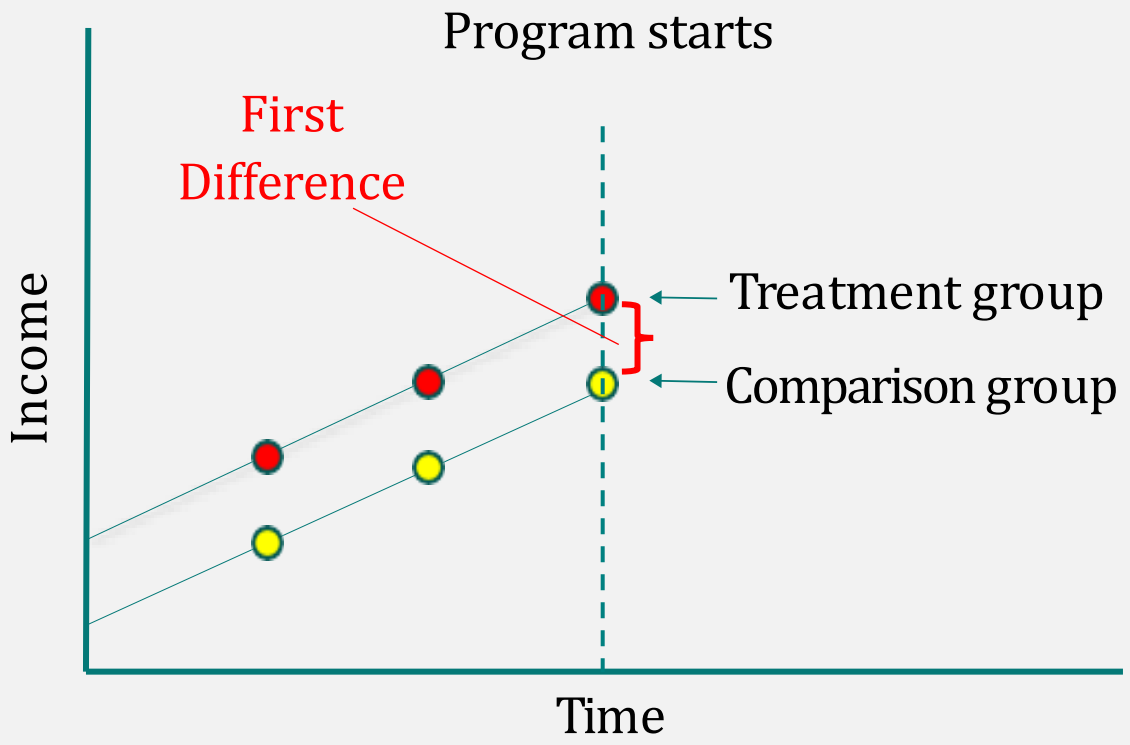
Look for pairs of people that are similar in terms of characteristics

Must only consider pre-program characteristics or characteristics that do not change over time

Quasi Experimental: Difference-in-differences

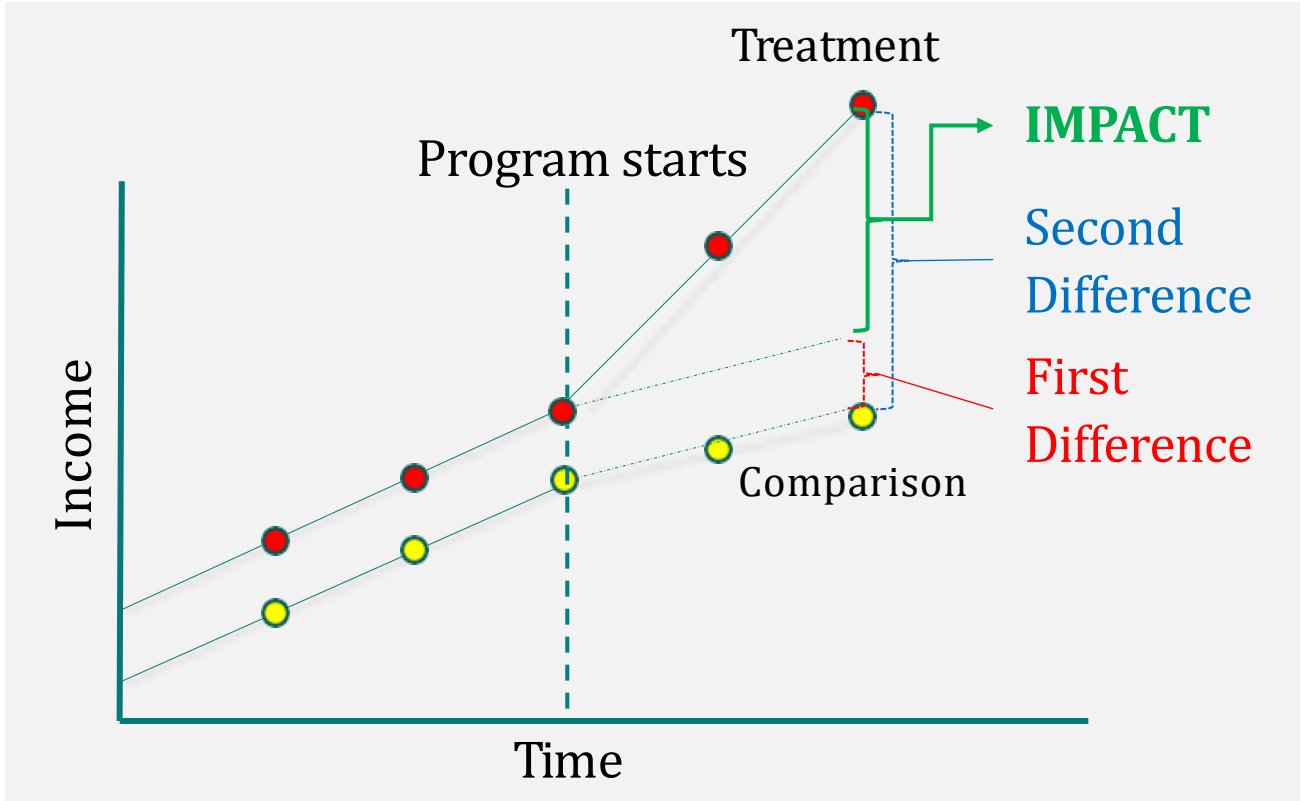
In the difference-in-differences approach, we accept that the Treatment and Comparison groups are *different*.

IMPORTANT: This approach requires to have data on both groups *before* the program starts



Quasi Experimental: Difference-in-differences

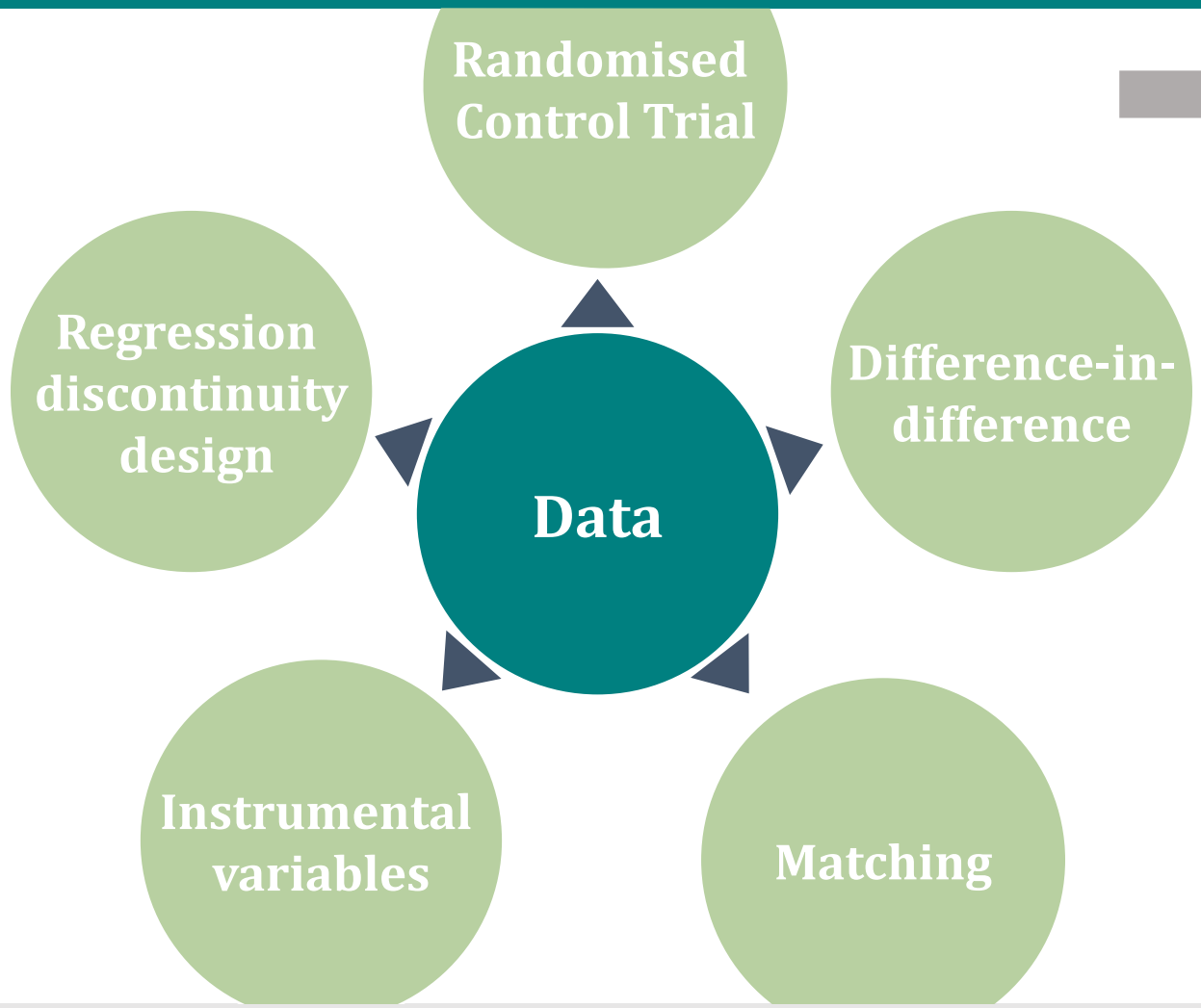
- Data on both groups must be collected at a later point in time, after the program has started
- The difference observed after the program started *(t1)* is adjusted by subtracting the first difference observed *before* the program *(t0)* to yield the impact estimate



RECAP YEAR 2

Data collection of microdata in hard-to-reach areas

Relevance of data collection in CIE



Data collection and quality data enable the construction of a valid and reliable counterfactual

Quality data is essential for answering evaluation questions and measuring program impacts

Evaluation Question to Data Collection

CONSTRUCT	INDICATOR	RESPONSE
What effects do the interventions have on livelihood in terms of **economic wellbeing** of refugee and host communities?	• Average income • **Employment Status** • Security in employment • Business Ownership • Business Performance • Asset Ownership	• A paid employee • A paid worker on household farm • An employer • Unpaid worker • Internship • None of the above

Tool development

Data collection tool development is critical for reliable CIE

- Leverage existing literature and tools
- Pretest tools (desk and field) and refine questions and responses before use in data collections
- Important to avoid measurement errors
  - Poorly designed questions and survey
  - Cognitive challenges in answering the question
  - Social desirability bias

CIE Sample and Sampling Frame

- In most CIE, data cannot be collected from all units → *sample*
- The sampling frame ideally includes all units from the population that the evaluation is focused on → *census*
- For a CIE, the sampling frame is usually a list of all the units that:
  - Received the program (Treatment group)
  - Did not receive the program and are identified as the counterfactual group (Control group)
- The sample is drawn from the sampling frame

Sampling

Types of sampling

Probability sampling methods reduce the possibility of bias as the possibility of someone being selected as part of the sample relies completely on chance

- Simple random sampling; Stratified random sampling; Clustered sampling
- Best suited approaches for quantitative part of a CIE

Non-random sampling means that the selection of the sample is not driven by chance → convenience sampling, purposive sampling (sometimes used for qualitative studies)

- Study and sample can be guided by findings
- Can help if the sampling frame is not clear

A big sample size is critical for increasing the chance of correctly identifying impacts

Data quality

How can data quality be ensured?

- Thorough questionnaire design
- Data collection methodology
  - Fieldwork protocols
  - Training of field staff
  - Method of administering the survey
    - Pen and Paper Personal Interviews (PAPI)
    - Computer Assisted Personal Interview (CAPI): SurveyCTO, Kobotoolbox, Survey Solutions
- Data collection monitoring
  - Daily (automatized) checks to identify potentially problematic data and/or potentially poorly performing enumerators

Monitoring systems

Monitoring is not the same as CIE

Monitoring = Does the program/intervention work as planned?

- Similar to CIE, monitoring depends heavily on data collection
- Monitoring systems are critical for impact evaluation
- Monitoring systems provide information on available resources, outputs, and need for backstopping and correction

Data for CIE - Ethics

- No CIE is worth risking the safety of participants in data collections
- Protection of the rights to privacy and well-being of participants must be the ultimate guiding principle in all data collections
- Obtain informed consent before collection of data
- Satisfy all ethical requirements from the relevant Ethics Board and obtain Institutional Review Board (IRB) approvals and permits before the start of field data collection
- Data security and protection is critical in all data collections

RECAP YEAR 3

Data analysis for monitoring and CIE

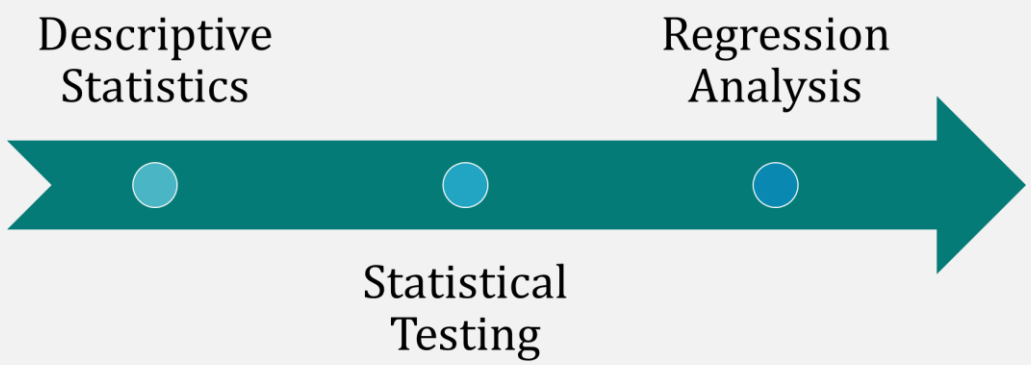
Overview Year 3

Focus Year 3:

Year 3 workshop focused on **what to do with the data**

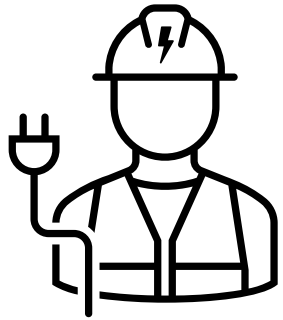
Overall objective Year 3:

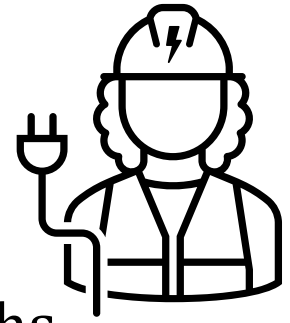
The aim was to get a sense of how different analyses for CIE works at an intuitive level



Data analysis for CIE: *Example - setup*

- Organization (your client) designs and implements a vocational training programme in TVET centres
- **Overarching goal**: Economically empower disadvantaged youth to engage in employment and livelihood strategies
- **Specific goals**:
  - Enrol 1,000 young people to participate in the vocational training
  - Ensure a 50:50 gender split across all participants
  - Achieve a 90% graduation rate
  - Increase monthly income of graduates by 800 units six months after completing training





Data analysis for CIE: *Example - objective*

- You are provided with a monitoring data set with the following information on all participants:
  - Gender
  - TVET centre enrolled at
  - Graduation status
  - Average income of graduates six months after completing training

- In terms of OECD evaluation criteria, was it *effective*?
- Did the program have an *impact* on income? How confident are we that this difference reflects the *true* impact of the program?

Descriptive Statistics

Descriptive statistics

- Simplifies large amounts of data into a sensible summary, helping identify patterns, trends, and insights

Common descriptives

- Mean, median, mode, range, variance, standard deviation, counts & percentages

Applications:

- Assessing program coverage
- Evaluating service delivery
- Identifying areas and gaps for improvements

Statistical Testing in CIE

Statistical testing:

- Helps determine if observed differences or relationships in data are statistically significant through hypotheses validation in impact evaluations

Types of tests:

- Parametric tests
- Non-parametric tests

Applications:

- Ensure robustness and credibility of evaluation results
- Helps in making informed policy decisions based on empirical evidence

Statistical Testing in CIE

Caution:

- A statistical test is a formula, it does not know the **context** of the analysis, where or how you got your data
- Researchers need to do the hard work to provide the formula with strong data and contextualize the findings
- The first step in bringing credibility/confidence to your tests is to carefully select the sample!
- Simply testing the difference in the average income tells you there *exists* a difference between two groups
- A statistical test alone does ***not*** establish causality

Regression Analysis

Regression Analysis

1. *Quantify* the relationship between two variables – by how much does income increase when years of education increase by 1?
2. Gauge whether the relationship is due to chance or whether it is *statistically significant* → statistical testing

Why use regression analysis?

Why use regression analysis? – Confounders

- We cannot say that completing TVET training *completely* explains the difference in income, or that it directly causes it
- There could be other important factors driving the difference

Remark 1 Such factors are called **confounders** or **confounding variables**, because they blur/disturb the relationship of interest between education and income

Remark 2 **Confounders** cause problems because they are correlated with both education *and* income at the same time

Regression Analysis and Dissemination

To help increase the impact of CIEs we can map out a basic theory of change as we do for an intervention itself

CIE measures the impact of a vocational training programme on livelihoods	Decision-makers such as donors, policy-makers, programme directors, governments etc. take decisions informed by findings	New policies, programmes and interventions are designed based on the latest evidence	Policies, programmes and interventions are more effective at reaching their goals	Population of interest improves in the outcomes of interest

Assumption: Intended users find, read and fully understand results of CIE

Resources

- Training Workshop on Counterfactual Impact Evaluation (CIE) – PowerPoint Slides
- Training Workshop on Data Collection of Micro Data in Hard-to-Reach Areas– PowerPoint Slides
- Training Workshop on Data Analysis for Monitoring and CIE – PowerPoint Slides

Books

- World Bank, Impact Evaluation in Practice - Second Edition (Book)

Videos

- InterAction, Introduction to Impact Evaluation
- Esther Duflo, Randomized Controlled Trials and Policy Making in Developing Countries

Podcasts

- IEU Talks Episode 2: The Power of Impact Evaluation in Development Cooperation
- Evidencing impact (parts 1)
- Evidencing impact (parts 2)

Resources

- Year 1: Counterfactual Impact Evaluations
- Year 2: Data Collection of microdata in hard-to-reach areas
  - https://youtu.be/SQpD7Ma2Jm4?si=oTIuA4ISY9uZMSZt
  - https://youtu.be/4pqb5liyENM?si=v8_vfKy7fk6WyRQ5
- Year 3: Data analysis for monitoring and CIE
  - https://youtu.be/mW_rmHbez1Y?si=BFADmz_bz7t8L-Fg
  - https://youtu.be/ZziJrPAzpYQ?si=jbb8wiB6cFXUgJr-
- Year 1 – Year 3 Slides: Wetransfer

END OF SESSION 1

Session 2:

Understanding evidence synthesis – Introduction to basic concepts of evidence synthesis

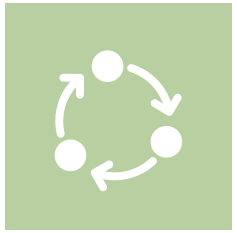
C4ED – EUTF

October 2024

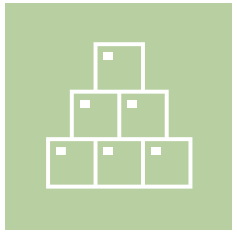
Center for Evaluation and Development

European Commission

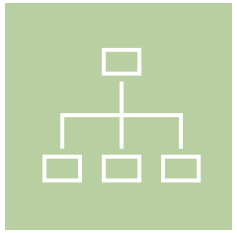
Objectives of Session 2



Define evidence synthesis and its significance in informing decision-making process in programming



Identify the key components of evidence synthesis, including literature review, data synthesis and meta-analysis



Explain the principles of evidence hierarchy and its relevance in evidence-based practice

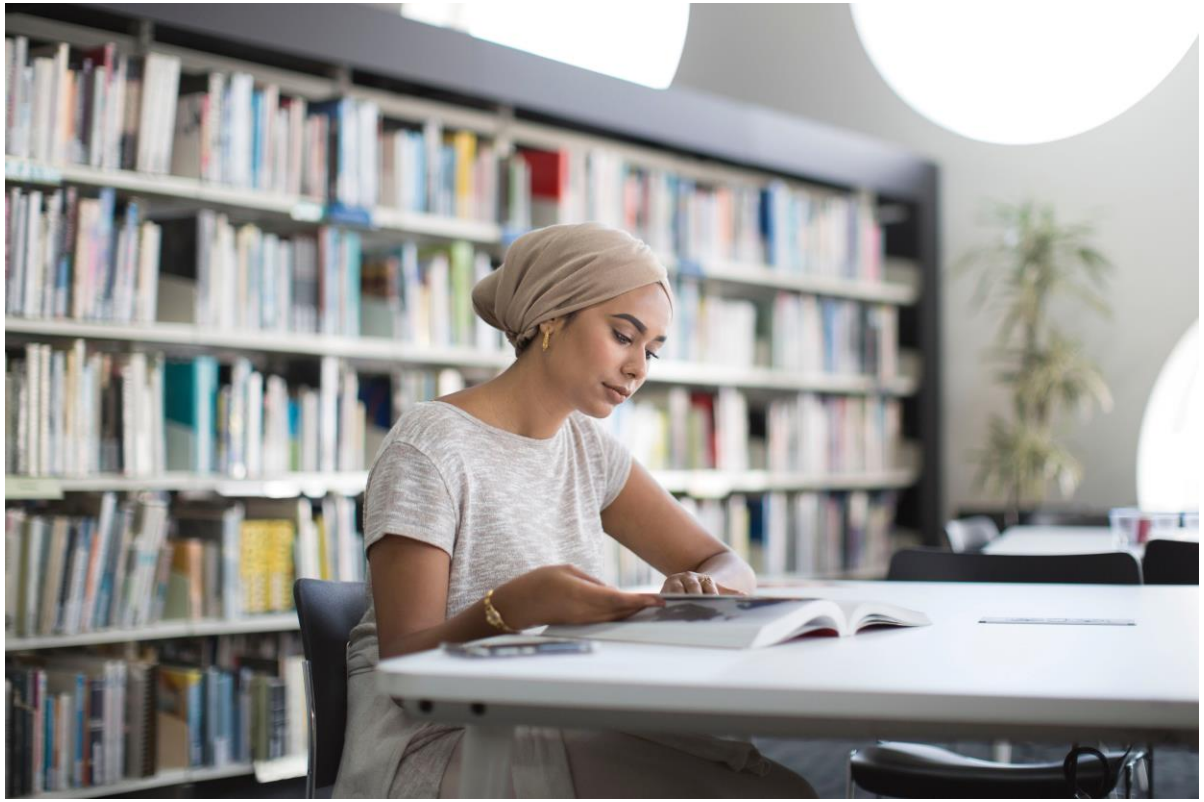
Conceptualization of evidence synthesis

…What evidence drives development policy?

Types of reviews

1. Reviewing reports, research papers on trainings and their impact on income

→ **Literature review**



Types of reviews

2. Speaking with experts in income and/or training

→ **Expert review**

Types of reviews

3. Systematically reviewing and collating evidence from various types of data sources (project reports, research papers) on vocational trainings

→ **Systematic review**

Types of reviews

Evidence synthesis

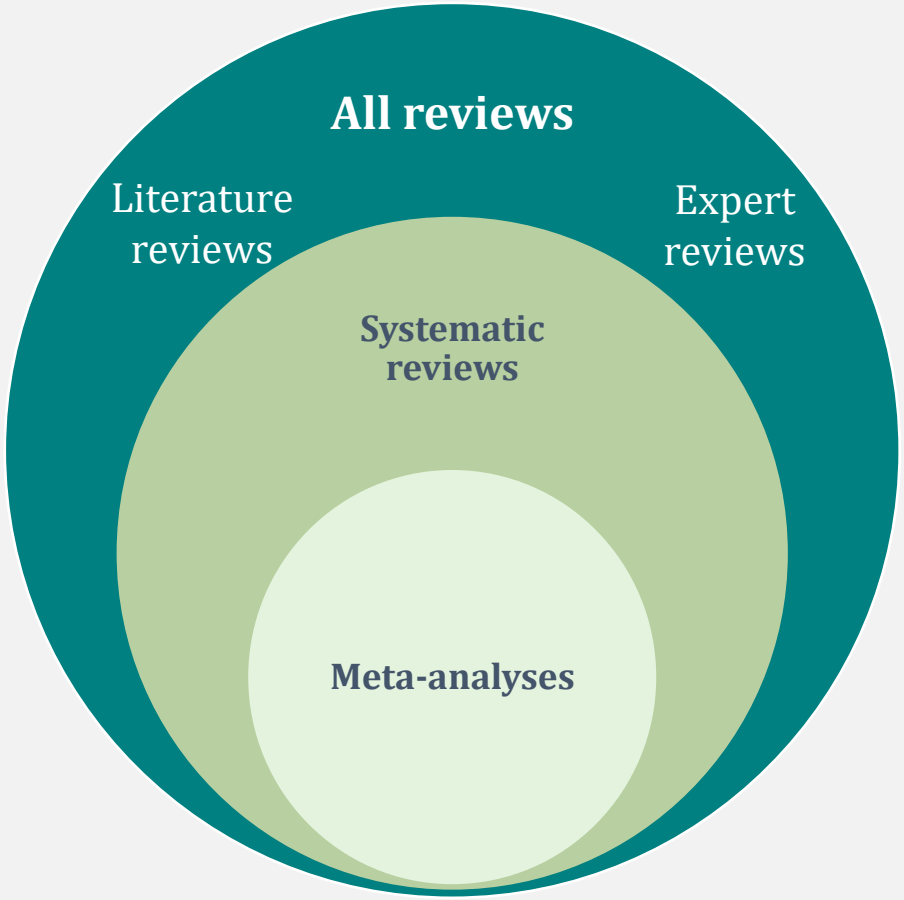
Most rigorous approach is systematic reviews:
Identify, ***evaluate*** and ***synthesize*** relevant research

Types of reviews

Evidence synthesis

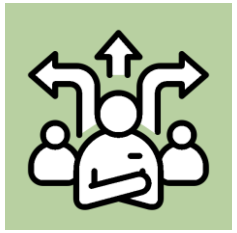
Most rigorous approach is systematic reviews:
Identify, ***evaluate*** and ***synthesize*** relevant research

Meta-analyses:
Statistical techniques for synthesizing results of relevant



What is evidence synthesis?

Process of systematically ***combining and analysing information*** from multiple sources that investigate the same topic to learn about what is known about that topic and what remains unknown



Policy makers: Summarized results on interventions that work

Researchers: Identified knowledge gap informs further research

Why evidence synthesis?

Overwhelming volume of research

- Search for “TVET and Youth Employment” results in 1.4 million results
- Review of all studies infeasible and may include irrelevant studies

→ Need to focus search to derive smaller set of relevant studies

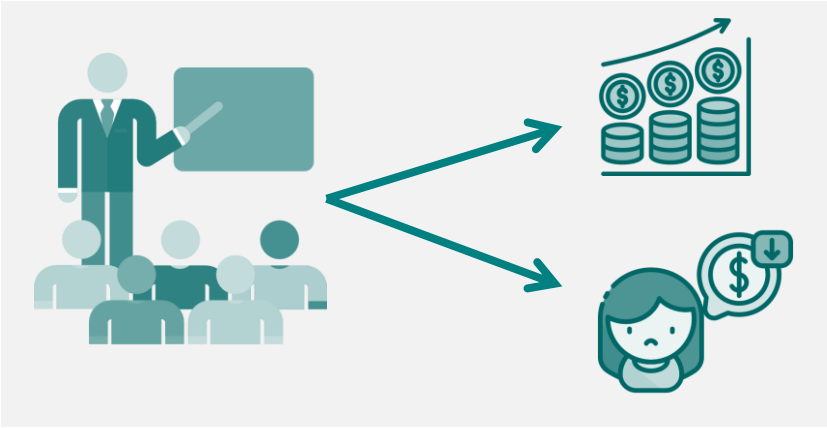
Combine small studies together for more powerful findings

- One study on youth employment in Uganda might lack enough evidence impacts of TVET training but combined with other studies showing impact will be more generalizable across countries and contexts.

Significance and role of evidence synthesis

Examine whether findings vary under different conditions

- *Different interventions work differently for different people under different conditions*
- *Example: Training programmes do not consistently result in increased income across all contexts*

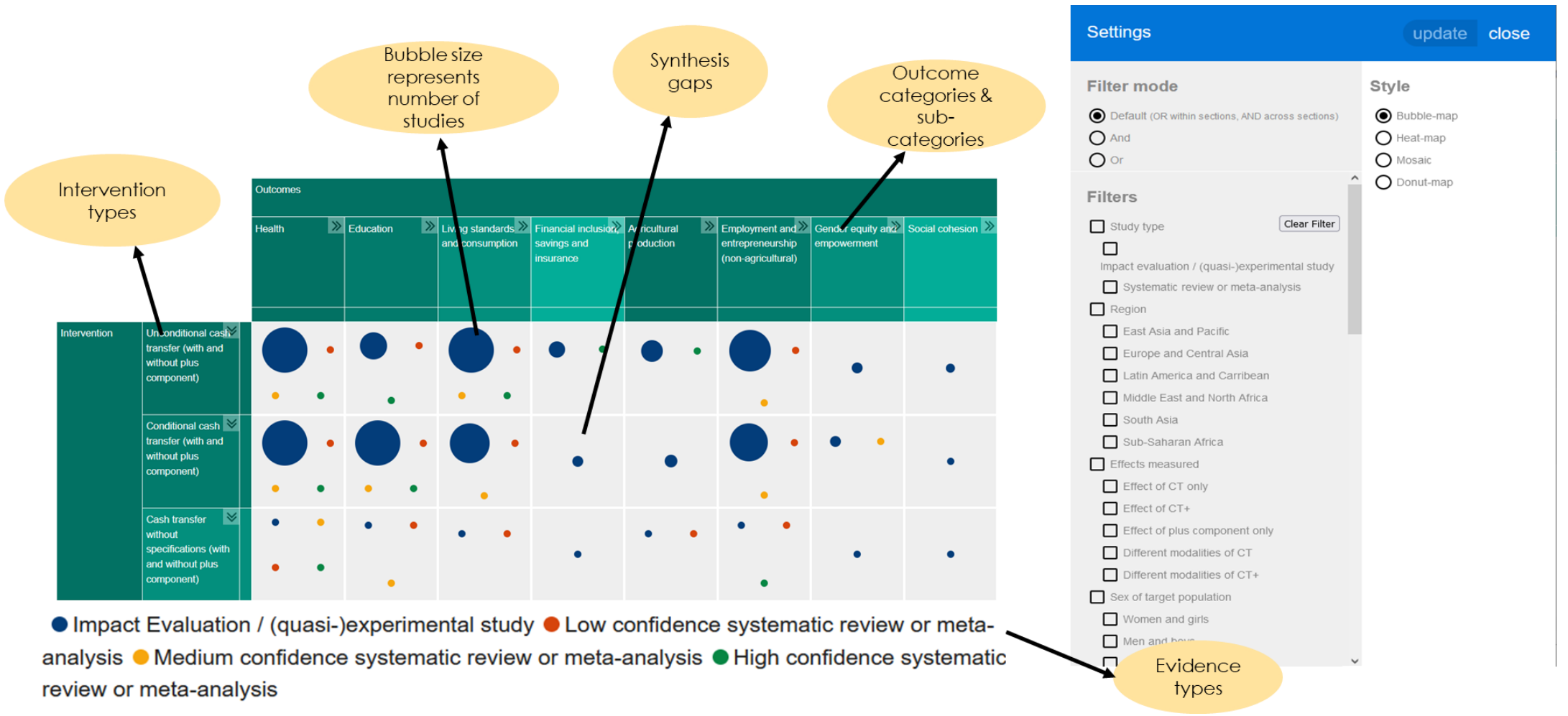


- Being up to date as information arises
- Identify research gaps and needs for future studies

How does evidence synthesis differ from literature reviews?

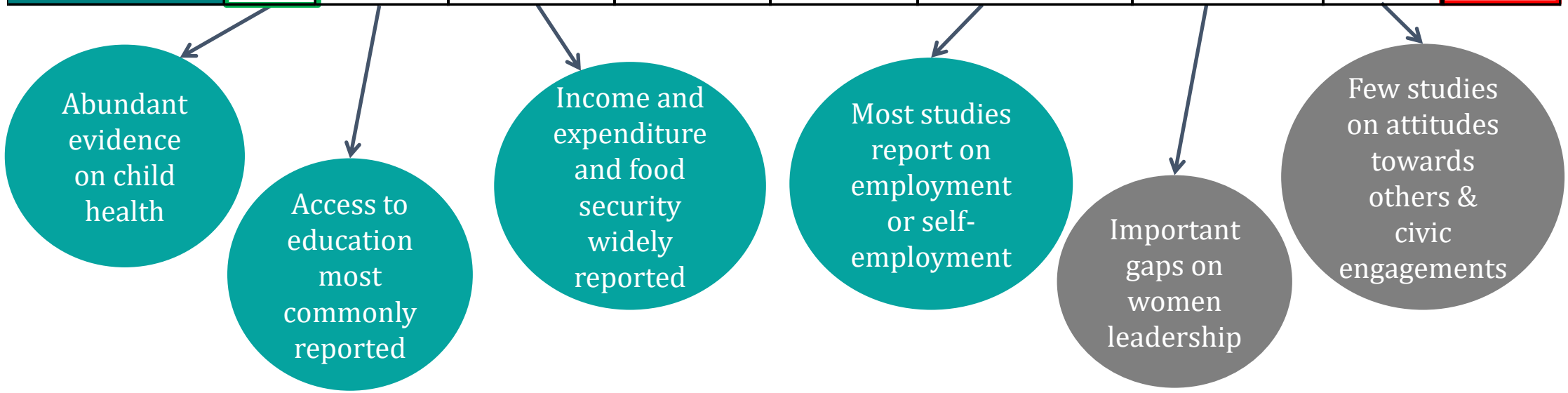
- Aims to identify and synthesize ***all*** related research ***published and unpublished***
    - Uses ***systematic methods for each step***
    - Begins with ***well-defined research question***
    - ***Uses a criteria for inclusion and exclusion of study in the review***
    - Critically ***appraises included studies***
- ***Unbiased*** and ***reproducible*** approach

Examples of systematic evidence synthesis - EGM

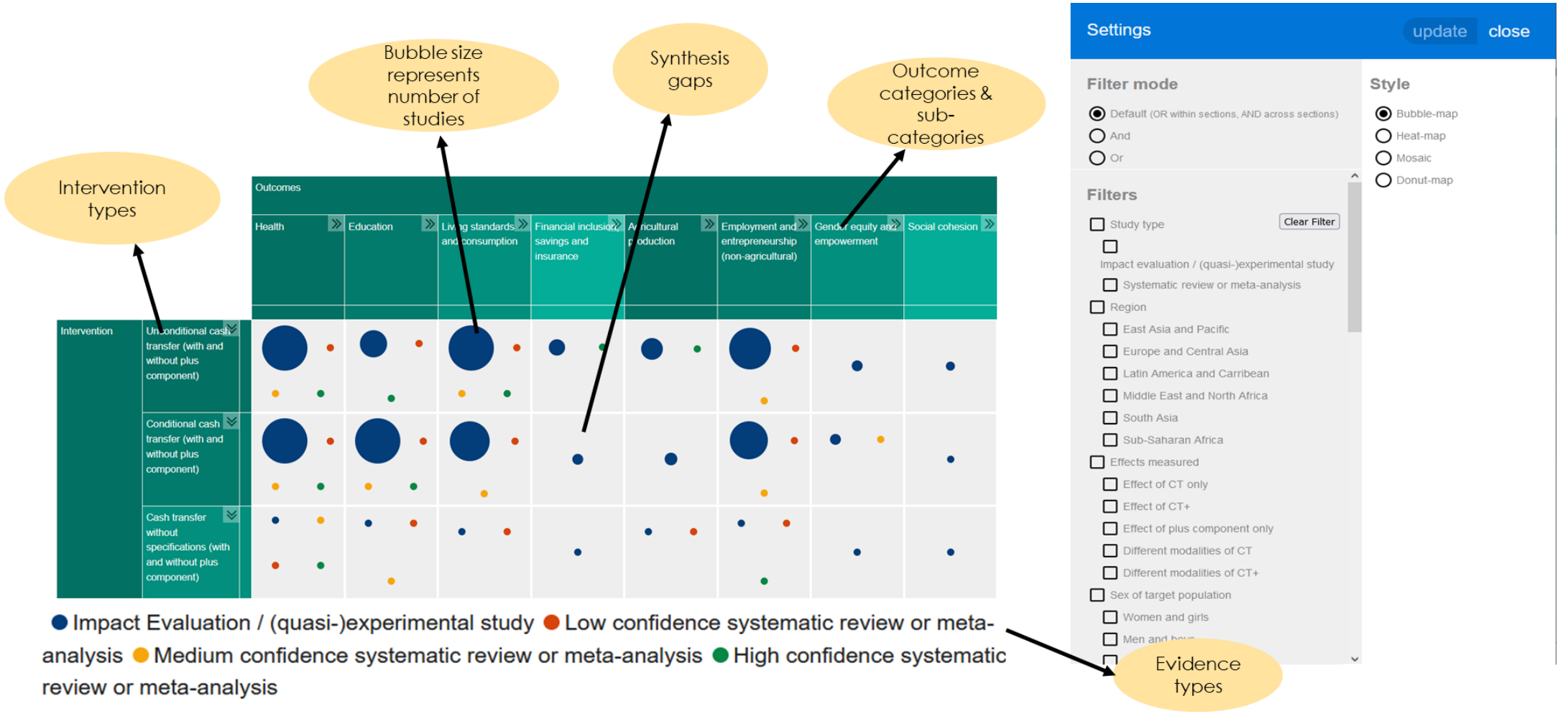


Examples of evidence synthesis - EGM

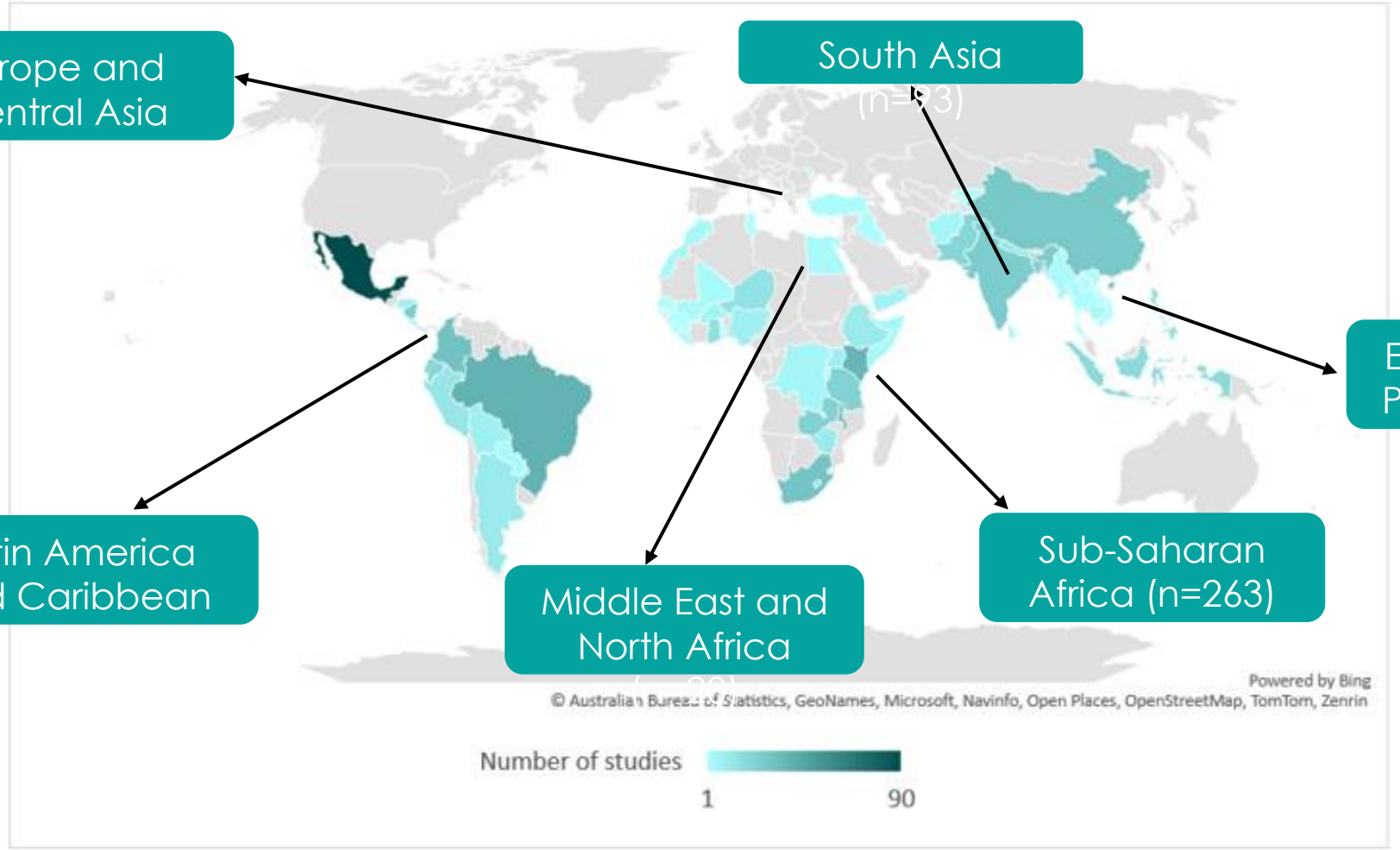
	Health	Education	Living standards and consumption	Financial inclusion, savings and insurance	Agricultural production	Employment and entrepreneurship	Gender equality and empowerment	Social cohesion	**Total**
Unconditional cash transfer	198	76	198	44	59	115	30	25	318
Conditional cash transfer	234	131	109	30	35	105	29	17	384
Total	434	199	315	75	96	217	59	46	**709**



Examples of systematic evidence synthesis - EGM



Examples of evidence synthesis - EGM



Benefits of evidence synthesis for decisions

Comprehensive overview

Consolidates findings from multiple sources

Accessible research

Useful for non-technical audiences

Efficient resource use

Identify and prioritizes key evidence gap

Policy formulation

Informs effective development interventions

Evidence-based practice

Highlights relevant and robust findings

Transparency and reproducibility

Promotes research transparency

Key components of evidence synthesis

Key steps in evidence synthesis

Step 1: *Formulating the research question*

Helps define the scope of the review

Lumping or splitting:

Lumping	Broader reviews that allow comparison of range of interventions across wider range of contexts, population
Splitting	Appropriate to compare studies similar in design, population, intervention characteristics or outcome

Key steps in evidence synthesis

Components parts of review question - PICOS

- **P**opulation
- **I**ntervention
- **C**omparator
- **O**utcome
- **S**tudy design

Example research question: *What are the effects of TVET training on monthly income among young adolescent girls in Horn of Africa*

Key steps in evidence synthesis

Step 2: *Systematic literature search*

- Provides systematically derived list of studies used for the review
  i. Generating **key words** related to PICOS to search – search terms
  ii. Searching **electronic databases** and hand searches using key words – search strategy
  iii. Identifying both **published and unpublished documents** (unpublished government and organizational reports) to avoid biases – grey literature

Caution: Publication and reporting bias in systematic differences in findings 

Key steps in evidence synthesis

Step 3: *Screening studies - Inclusion and exclusion criteria*

- Establish screening (usually) based on PICOS:
  - **P**opulation: Does it stem from continent of interest?
  - **I**ntervention: Does it include a training component?
  - **O**utcomes: Is income measured as an outcome?
  - **T**ime: Is study too old to be relevant?
- Exclude studies not meeting criteria
- Systematic process for resolve discrepancy (if more than 2 reviewers)

Key steps in evidence synthesis

Step 4: *Data extraction*

- Consider and define types of evidence to be extracted:
  - PICOS + duration of training + type of training
- Standardized tools for ***systematic extraction of evidence***
  - Excel, Revman, EPPI Reviewer software, ODK, SurveyCTO
- **Quality assessment of studies and data**
  - Detailed critical appraisal of included study designs based on '***risk of bias***' evaluation criteria
  - Ideally two reviewers working independently do the quality assessment

Key steps in evidence synthesis

Step 5: *Data analysis*

- **Qualitative** using a narrative synthesis and descriptive analysis
  - Contextualize types and modalities of interventions out there
  - Understand barriers and facilitators that may improve impact of training on income
- **Quantitative** using statistical techniques (e.g. meta-analysis)
  - More suitable when analysing quantitative effects
  - Calculate overall ***effect size*** and overall variance
  - Improved statistical power to overcome sampling errors
  - Policy makers can see 'signal' and 'noise' associated with policy interventions

Key characteristics of evidence synthesis

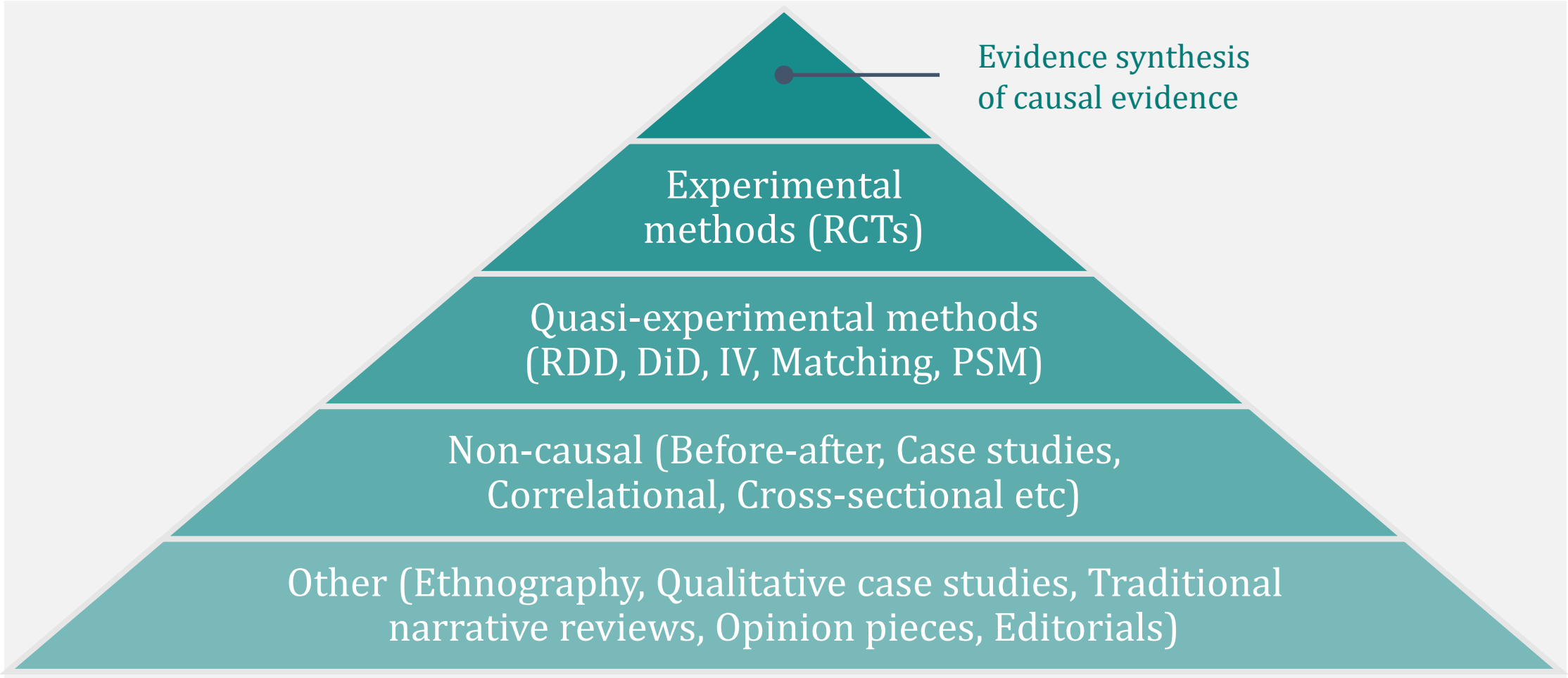
- Clearly stated objectives with *predefined eligibility criteria*
- Reproducible methodology (Protocols)
- Systematic search for *all eligible studies*
- A systematic synthesis and presentation of findings
- Assessment of *validity* of findings (risk of bias)

Evidence hierarchy

Relevance of evidence hierarchy

- Evidence hierarchy refers to a system for ranking evidence based on its reliability and validity
- Hierarchy helps policymakers and practitioners prioritize types of existing evidence
- Evidence synthesis of experimental methods at the top of the hierarchy followed by quasi-experimental methods
  - Randomized Control Trials (RCTs) are the gold standard!
- Higher levels of evidence likely to have lower bias and provide more robust evidence

Evidence hierarchy



END OF SESSION 2

Q&A

Discussion

In your usual line of work, how do you approach generating evidence for your projects?

What challenges do you typically encounter when generating evidence for your projects?

Session 2a:

Evidence Synthesis Methodology Part I – Searching & Screening

C4ED – EUTF

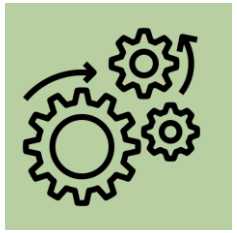
October 2024

Center for Evaluation and Development

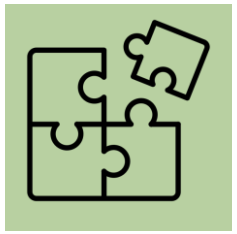
European Commission

Objectives of Session 2a

Learn how to define a research question for evidence synthesis



Understand the process of systematic literature searches
Be aware of steps to identify relevant studies



Grasp the concept of how to rate studies.
Similarities and differences between quality assessment and critical appraisal

Defining the research question

Research Question

- The research question is the foundation of any evidence synthesis.
- Use PICO(S) methodology to ensure comprehensive and well-defined research question:
  - **P**opulation
  - **I**ntervention
  - **C**omparator
  - **O**utcome
  - **S**tudy design
- Crucial step, as both search strategy and screening process will be based on these criteria

Research Question

What is good and bad about these research questions:

1. What do TVET programmes do?
   - It asks about a clear **i**ntervention
   - But doesn't specify anything else about outcomes or target population
2. What RCTs have been done in Senegal?
   - It tells us a **p**opulation (geographic location) and a **s**tudy design
   - No intervention or outcome defined, or whether only a certain type of people in Senegal are of interest (e.g. pregnant women)
3. Do young women increase their income compared to old men?
   - **C**omparison is clear, **o**utcome is stated
   - Through which intervention? Where in the world? What types of evidence?

PICO(S) methodology – step by step

PICO(S) - Population

- Question: 'Who?'
- Describes characteristics of people: gender, age, occupation, health …
- Geographic location if relevant, e.g. living in Horn of Africa, low- and middle-income countries (LMICs)

Example:
Youth (ages 18-35) living in Africa

PICO(S) methodology – step by step

PICO(S) - Intervention 

- Treatment or action considered for the population
  - Cash transfers, trainings, mentoring, increased access to health centers, organizing internships, provision of food baskets, school clothes, …
- Amounts and frequencies may play a role, e.g. for cash/ in-kind transfers or trainings

Example:
TVET training

PICO(S) methodology – step by step

PICO(S) - Comparator

- What is the intervention being compared against? Status-quo, no intervention, other intervention?

Example:
No participation in TVET training

PICO(S) methodology – step by step

PICO(S) - Outcome

- How is the effectiveness of the intervention measured?
- Multiple outcomes possible for broader evidence synthesis

Example:
Income

PICO(S) methodology – step by step

PICO(S) - Study design

- Study design is optional, but can help define where in evidence hierarchy evidence should be gathered from
- For causal systematic reviews, this would only be experimental, quasi-experimental or mixed-methods studies

Example:
Experiments, quasi-experiments, mixed-methods (including Q-ED)

PICO(S) – Full research question

Population:	*Youth (ages 18-35) living in Africa*
Intervention:	*TVET training*
Comparator:	*No participation in TVET training*
Outcome:	*Income*
Study design:	*Experimental studies, quasi-experimental studies, mixed-methods studies*

Research Question:

What is the

PICO(S) – Full research question

Population:	*Youth (ages 18-35) living in Africa*
Intervention:	*TVET training*
Comparator:	*No participation in TVET training*
Outcome:	*Income*
Study design:	

Research Question:

What is the causal effect of

PICO(S) – Full research question

Population:	*Youth (ages 18-35) living in Africa*
Intervention:	
Comparator:	
Outcome:	*Income*
Study design:	

Research Question:

What is the causal effect of TVET training

PICO(S) – Full research question

Population:	*Youth (ages 18-35) living in Africa*
Intervention:	
Comparator:	
Outcome:	
Study design:	

Research Question:

What is the causal effect of TVET training on income

PICO(S) – Full research question

Population:

Intervention:

Comparator:

Outcome:

Study design:

Research Question:

What is the causal effect of TVET training on income on youth living in Africa?

Systematic literature searches

Process for determining search strategy

1. Define *benchmark studies* through **scoping search**
2. Use PICOS to determine relevant *search terms* as starting point
3. Build *search query* in blocks by using operators
4. *Search* databases and other sources
5. *Refine search terms* based on results (check for benchmark studies)

   ↺ Iterative loop
6. Save *final search results* for screening

Scoping search

- Conduct a short manual search to identify a set of relevant studies for the review

→ **Define benchmark studies**

- Should ideally represent the full breadth of studies to be included
- List should be saved, as it will be used to test the search strategy
- Can be reviewed and expanded by topic experts

Search strategy components

A good search strategy consists of the following:

a. *Search terms* based on PICO(S)

b. Set of *databases & websites* to search
   - Good coverage is important for comprehensiveness and representativeness
   - Including both published journal articles and grey literature reduces bias

c. *Search query* combining search terms with *logical operators (AND/OR)*
   - Different databases & websites have slightly different search rules
   - Search has to be adapted to each individually
   - Additional limitations can be included, e.g. date of publication or language

d. *Manual searches* may be necessary where no advanced search supported in databases

Set search terms

Example: Find search terms based on PICO(S)

PICO(S)	Definition	Search terms
Population	Youth aged 18-35 living in Africa	Youth, Africa, SSA, young, [list of relevant country names]
Intervention	TVET training	TVET, training, vocational training, technical education, job training, apprenticeship
Comparator	No TVET training	-
Outcome	Income	Income, salary, wealth
Study design	Experimental studies, quasi-experimental studies, mixed-methods studies	RCT, CRCT, randomized, DID, difference-in-difference, difference-in-differences, RDD, regression discontinuity, instrumental variable, IV, PSM, matching

Blocks of search terms- create search query

Population:	(Youth OR young)
	AND
	(Africa* OR SSA OR sub-Saharan)
AND	
Intervention:	TVET OR training OR vocation* OR "technical education" OR "job training" OR apprenticeship
AND	
Outcome	income OR salary OR wealth
AND	
Study	RCT OR CRCT OR randomized OR randomised

Literature search

- Save results from each search (one per database), noting the query used, the date of the search, the number of results
  - References of results can be retrieved in different formats such as CSV, RIS, BibTeX, JSON

- Once all results from all queries are aggregated: remove duplicate studies (same study found in various data sources)
  - Often inbuilt tool identifies and removes duplicates
  - Important to track number of duplicates removed

Challenges and pitfalls in the searching process

Challenge	Mitigation strategy
Badly-defined research question (RQ)	→ Use PICO(S) to structure and refine RQ
Selecting databases & sources	→ Include multiple databases & grey literature → Consult information specialist or expert
Develop balanced search strategy	→ Pilot search queries → Use iterative process to refine
Number of results (once search strategy is finalized)	→ Use reference software to keep track → Identify (and remove) duplicates.

Documentation is key!

Reasons for systematic literature searching

Comprehensive coverage of evidence on topic

- Reduction of bias compared evidence derived from few studies
- Best for evidence-based decision making

Reproducible

- Transparent criteria
- Methodologically rigorous

→ Enhanced quality of evidence synthesis

Identifying relevant studies

Set up screening

- Collect search results in one ‘place’: specialized software or reference management software
- Set up screening ‘codes’ to document reason for inclusion and exclusion for each round of screening
  - Follow PICO(S) and other criteria set up in protocol, e.g. publication date or language for exclusion codes
  - Screening codes may differ by screening round
  - Allow for ‘Maybe’ option, for second opinion or discussion within team

Example of screening codes and exclusion

- A screening code indicates that a particular criteria in PICO(S) is not met by study
- For the following study, which criteria are not met?

  Impact of agricultural training on farmers' technological knowledge and crop production in Bandarawela agricultural zone

- The training is agricultural and not TVET→ *Intervention is not relevant*
- What other codes can be given when excluding this study?
- Remember PICO(S)!!
  - Hint: Bandarawela agricultural zone in Sri Lanka

Screening

Typically, two rounds of screening:

1. Title-abstract screening
2. Full-text screening

- For each round, a separate pilot phase to test screening codes against studies
- Use detailed inclusion and exclusion criteria based on PICO(S)
- Continue refining throughout→ Make sure to discuss with all screeners
- Between the two rounds, full-texts need to be retrieved
  - Potentially issues with paywalls for certain journal articles

Which study is relevant (based on title)?

- Remember our research question:
  - What is the causal effect of TVET training on income for youth living in Africa?
- Which of these studies can be excluded based on the title of the study?
  a. Investigating the impact of technical and vocational educational education (TVET) on youth unemployment in Ghana
  b. The evidence is in: How should youth employment programs in low-income countries be designed?
  c. The Impact Of Technical Vocational Education and Training (TVET) On Employment Among Youth in Pakistan: A Counterfactual Analysis
- Also think about what codes you give study when excluding studies!

Challenges and pitfalls in the screening process

Challenge	Mitigation strategy
Inconsistent application of inclusion and exclusion criteria	→ Initial training, → Pilot phase for both title-abstract and full-text phases, → Meetings for clarification as necessary, → Double-screening
Challenging cases for inclusion or exclusion	→ Allow for 'Maybe' option at all screening stages for review → Adaptation of inclusion and exclusion criteria by common understanding of entire team
Missing studies due to limited information in title and abstract	→ Lack of information leads to include to avoid bias

Challenges and pitfalls in the screening process

Challenge	Mitigation strategy
Time-consuming and resource-intensive screening process	Use algorithm to prioritize studies based on relevance (e.g. EppiReviewer), use automation tools where feasible
Inter-reviewer reliability for quality or risk of bias assessment	Have initial training, a pilot and use double-rating. Discuss difficult cases with team for better common understanding

Rating studies

Commonalities of approaches to rating studies

- Try to categorize studies according to all applicable criteria
  - Systematic approach for replicability
- Usually both double-assessment and piloting are common, as in screening process
  - Be as objective as possible
  - Consistent ratings across team members
- Quality assessment and critical appraisal of studies are similar but separate concepts

Quality assessment

- Focus on objective measurement of the quality, certainty and reliability of each study
  - Consistent and transparent
  - Frequently used tool: GRADE
- Risk of bias (RoB) is integral part and is often highlighted before incorporating it into overall quality rating (for missing evidence)
  - Frequently used RoB tool: Cochrane Risk of Bias Tool (RoB2), ROBINS-I
  - Can be included in meta-analysis

Quality assessment example - GRADE

- Domains covered:
  - Risk of bias
  - Inconsistency
  - Indirectness
  - Imprecision
  - Publication bias
- Evidence is sorted into four categories: high, moderate, low, very low
- Typically used in healthcare, but can be adapted to socio-economic research studies

Critical appraisal

- Includes relevance and applicability on top of quality assessment
- Used to determine how findings can be applied in practice of policy-making
- More subjective judgements necessary
- Frequently used tool: CASP
- Can influence conclusions

Critical appraisal example – CASP

- CASP has checklists for different types of studies including systematic reviews
- Checklists cover the following components with 10-12 questions:
  - Validity
  - Results
  - Applicability
- The precise questions are tailored to each study type
- Checklists include those for RCTs, cohort studies, systematic reviews, qualitative studies, amongst others.

END OF SESSION 2a

Systematic review similar to our example:

Technical and Vocational Education and Training (TVET) Interventions to Improve the Employability and Employment of Young People in Low- and Middle-Income Countries: A Systematic Review

Tripney, J., Hombrados, J., Newman, M., Hovish, K., Brown, C., Steinka-Fry, K., & Wilkey, E. (2013).

Campbell systematic reviews, 9(1), 1-171.

https://doi.org/10.4073/csr.2013.9

Session 2b:
Evidence Synthesis Methodology Part II – Data extraction & types of analysis

C4ED – EUTF

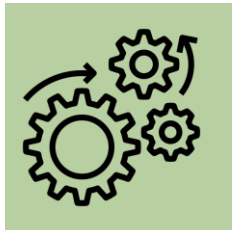
October 2024

Center for Evaluation and Development

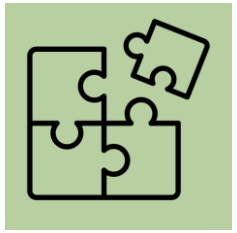
European Commission

Objectives of Session 2b

Understand the difference between qualitative and quantitative data extraction



Get an overview of different analytical approaches: Thematic approach, Evidence gap maps, Descriptive approach, Meta-analytic approach



Learn about the strengths and limitations of each approach

Data extraction

Data extraction

Purpose

- Ensures systematic and consistent collection of information from studies
- Facilitates analysis and comparison across studies
- Supports transparency and reproducibility

Best practices

- Use standardized forms and protocols
- Train data extractors and pilot tools
- Double-check and verify extracted data
- Thorough documentation

Differentiate qualitative and quantitative data extraction

Quantitative data extraction

- Used in meta-analysis and systematic reviews
- Focus on numerical (and categorical) data
- Most interested in effect sizes, variation metrics, sample sizes and study characteristics

Strengths	Limitations
• Objective and efficient • Enables quantitative analysis	• Potential data loss due to lack of flexibility • Limited contextual data

Components in quantitative data extraction

Study identification	Title, authors, publication year, journal or source
Study characteristics	Study design, sample size, setting
Participant characteristics	Demographics, inclusion & exclusion criteria
Intervention	Specifics, intensity, length, amount, etc.
Outcomes	Primary and secondary, definitions and measurement tools
Comparator	What does the control group receive?
Results	Numerical results, measures of variability, adjustments for confounders

Qualitative data extraction

- Used in thematic analysis and narrative synthesis
- Focus on non-numeric data
- Extraction of participant experiences, intervention descriptions and contextual factors

Strengths	Limitations
• Rich contextual information • Flexible	• Subjective • Time consuming

Analysis types

Analysis methods overview

- Thematic approach
- Evidence Gap Maps
- Descriptive approach
- Meta-analytic approach
  - Forest plot
  - Funnel plot
  - Meta-regression

Thematic approach

- Qualitative approach focused on finding underlying themes
- Appropriate for contextualization, finding trends in barriers and exploring breadth of influencing factors across the literature on a topic
- Key steps:
    1. Familiarization with studies
    2. Generate initial codes
    3. Search for themes
    4. Review themes
    5. Define and name themes
    6. Write narrative across themes

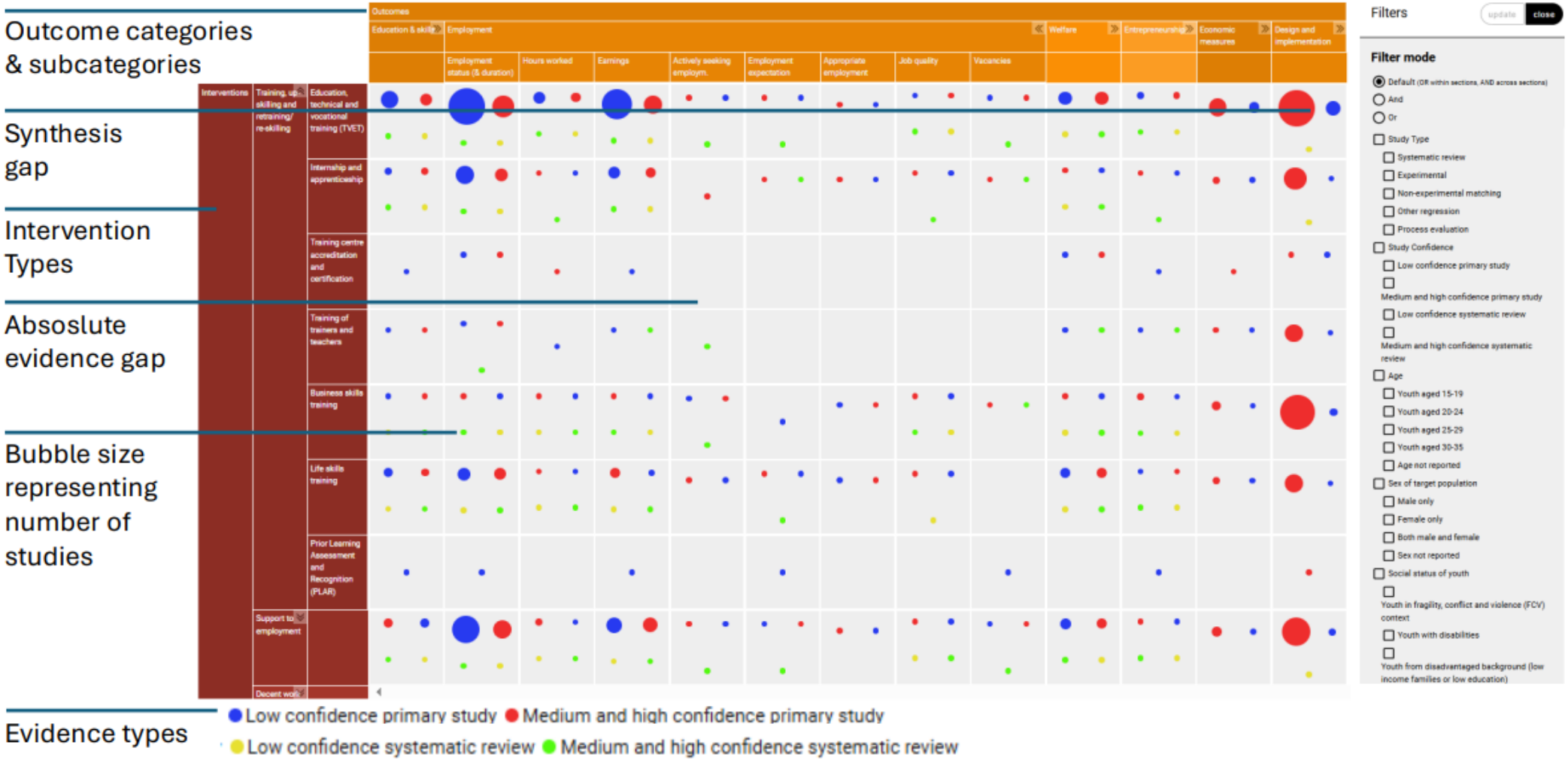
Thematic approach

Strengths	Limitations
• Appropriate for qualitative studies and research questions • In-depth analysis of patterns and meanings • Preserves contextual richness	• Quantitative research questions can be answered • No opportunity to measure objective effectiveness of an intervention • Potentially complex and time-consuming • Subjective, thus ensuring consistency and reliability is challenging • Not easily generalizable as it is context specific

Evidence Gap Map

- Useful for visualizing what areas within a topic have more and less research
- Maps interventions against outcomes

Evidence Gap Map



Evidence Gap Map

Strengths	Limitations

• Visualization of existing evidence and lack thereof
• Data extraction remains simple:
  • only study characteristics
  • no need for effect size extraction and standardization
• Filters allow for context-relevant searches
	• Existence of evidence ≠ evidence of effectiveness
	  • No indication on effect sizes or significance
	  • Similarly, lack of evidence ≠ lack of effectiveness
	• Filters only way to add context, e.g. quality of studies
	• Snapshot in time and quickly outdated unless updated regularly

Descriptive approach

- Often first step in quantitative analysis of evidence synthesis
- Used to understand basic characteristics of included studies
- Typically displayed in graphs or tables
- Consists of descriptive statistics on included studies:
    - Mean and standard deviation
    - Frequencies (geographic distribution, outcomes, etc.)
    - Range, potentially confidence intervals, sample size, etc.
- May need to standardize statistical data for comparison and aggregation
- Similar to descriptives in primary studies

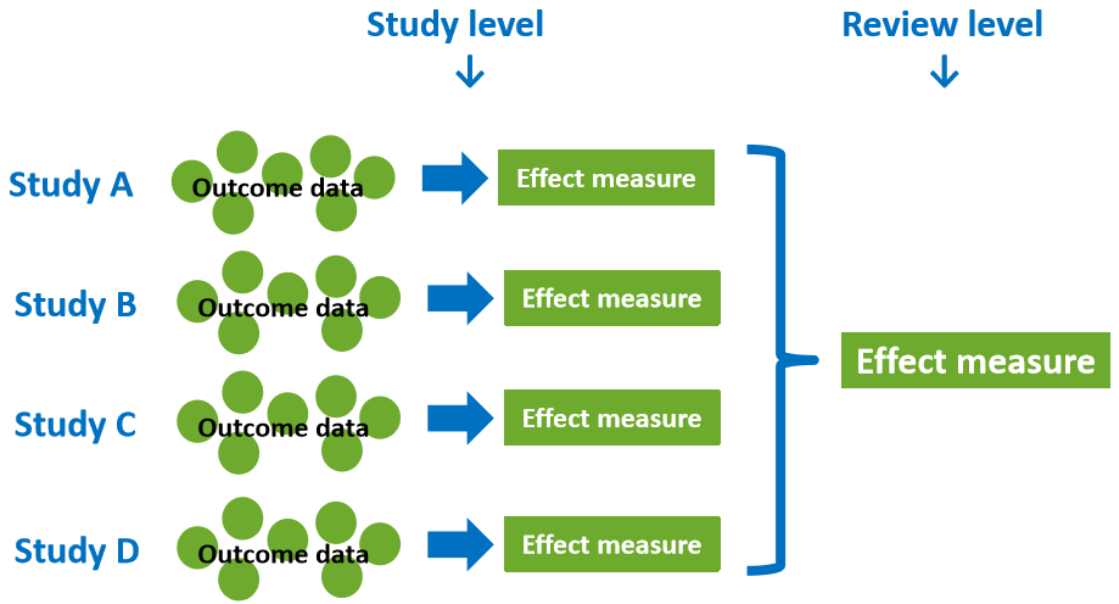
Descriptive approach

Strengths	Limitations
• Easy to conduct and understand • Foundation for further analysis • Suitable for visualization (graphs, tables)	• No causal inference • Simplistic and may miss important nuances or patterns • Potentially misleading – studies are seen as equally informative despite differences in quality, precision and context that may skew results

Meta-analytic approach

- Primary goal is to summarize statistical data from multiple studies into one overall effect size.
- Increases power and precision compared to primary studies.
- Able to explore differences between studies.
- Versatile and makes up for limitations of descriptive approach.

Basic approach:



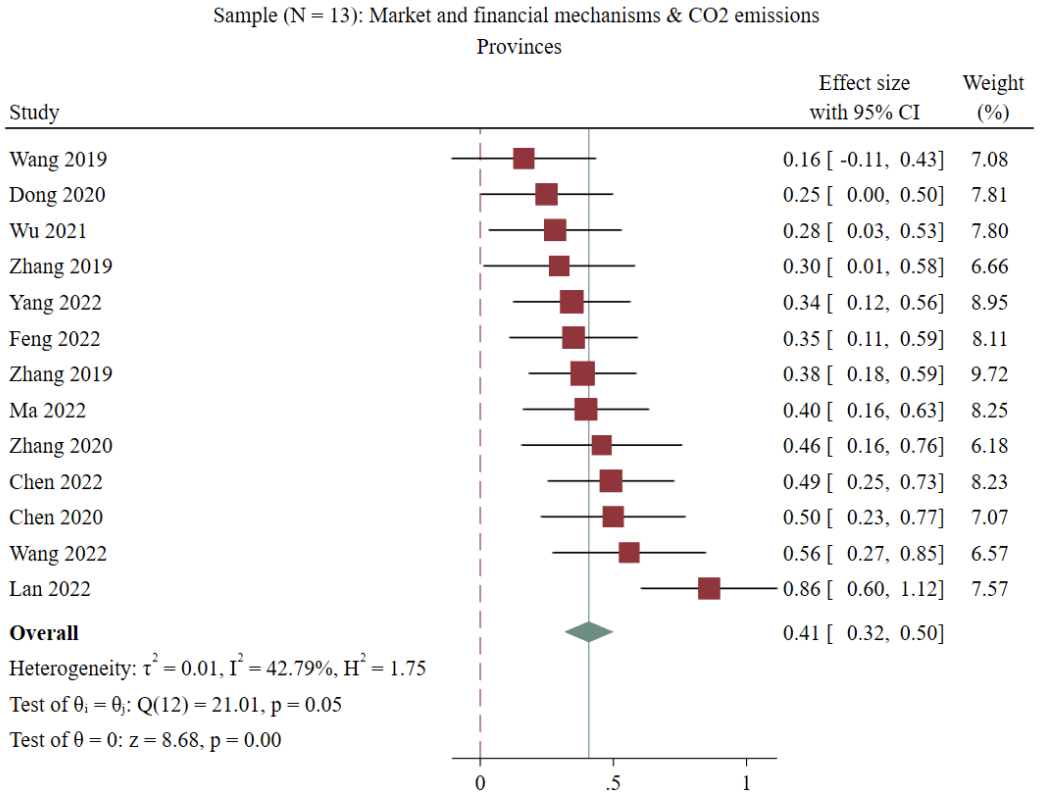
Meta-analytic approach

- Need detailed data extraction, such as statistical information on effect size, sample sizes, and any other variables of interest.
- Common methods include:
  - Funnel plots
  - Forest plots
  - Meta-regression
  - Publication bias analysis (not discussed)

Meta-analysis

Weighted regression of studies, visualized through forest plot

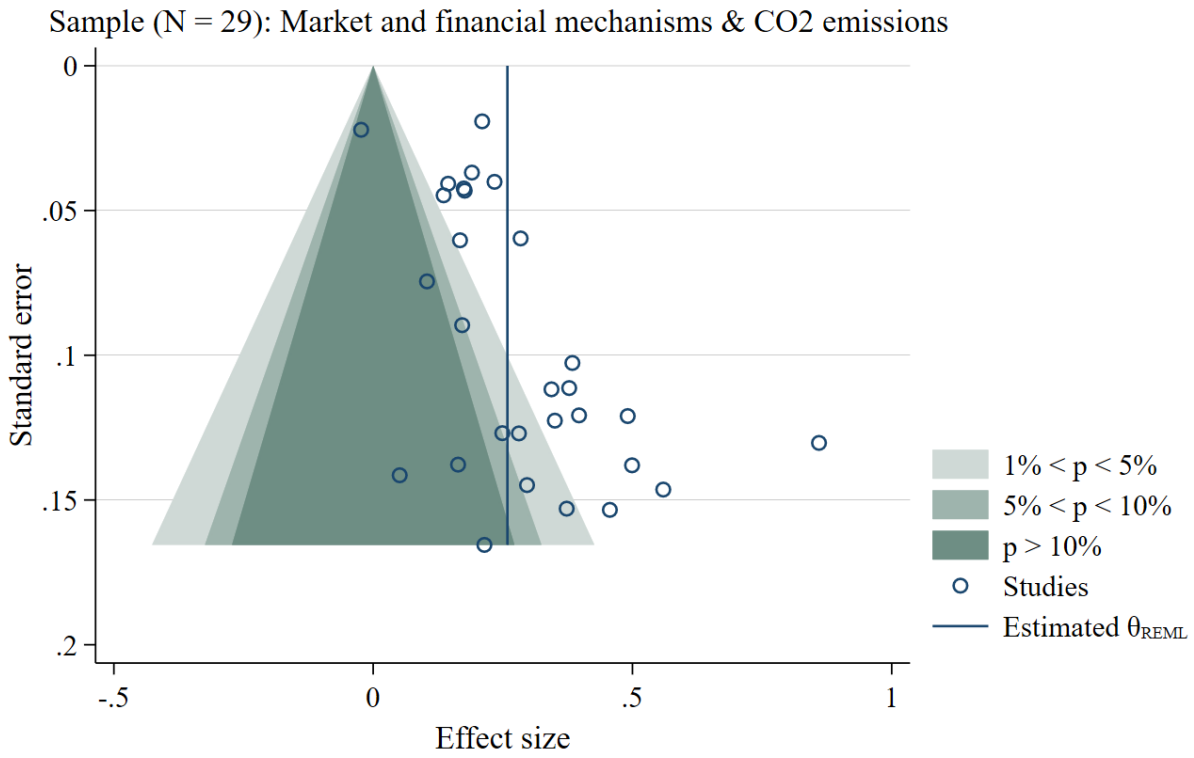
- Shows confidence interval and effect size of each study
  - Sub-group analysis for binary or categorical variables possible
- Result (green diamond) is estimate of overall effect size from weighted regression
  - The diamond edges represent the confidence interval of the overall effect size estimate



Meta-analysis – Funnel plot

Alternative visualization: funnel plot

- X-axis shows the effect size, y-axis shows precision of the study estimate
- The higher in the graph (= the smaller the standard error), the more precise the study estimate
- The estimated overall effect size is the vertical line
  - The example has contours for different significance levels of primary studies



Meta-analysis – Meta-regression

- Meta-regression is useful for determining influencing factors for differences between studies
- Each ‘observation’ is an effect size from a study
- Other variables are study level information or characteristics
- Results are presented like regular regression results

Meta-analysis – strengths & limitations

Strengths	Limitations
• Increased power and precision, • Consistency assessment • Objective • Informs policy • Possibility of assessing and correcting publication bias	• Limited to quantitative data • Results are dependent on quality of input data • Complicated by publication bias, heterogeneity and complexity

END OF SESSION 2b