



Virtual Training Workshop on Counterfactual Impact Evaluation (CIE)



C4ED – EUTF

September 2021





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





Welcome EU Enrique de Loma







Welcome C4ED Prof. Dr. Markus Frölich







Please follow the code to participate in the questionnaire





- 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, which will be answered at once
- use the longer breaks to ask more questions
- suggest improvements if you can't follow or disagree (we are open to criticism and constructive suggestions for improvement)
- more feedback and questions (especially for the Q&A session): Send an email to Zahra Sharafi (<u>z.sharafi@c4ed.org</u>) or Dr. Giulia Montresor (<u>g.montresor@c4ed.org</u>)



09:00 - 09:30	Welcome
9:30 - 10:45	Session 1: Introduction To Counterfactual Impact Evaluation (CIE)
10:45 - 11:15	Break
11:15 - 12:15	Session 2: Disentangling program needs, Theory of Change, Evaluation Questions, and Indicators
12:15 - 13:00	Lunch Break
13:00 - 14:30	Session 3: Introduction to Experimental Methods Evaluation Design, CIE Methods
14:30 - 15:00	Q&A



- What is the Counterfactual Impact Evaluation (CIE) and what are its benefits?
- Theory of change and evaluation questions, and Indicators
- Evaluation designs, (C)IE Methods
 - Introduction to Experimental Methods
 - Introduction to Quasi-Experimental Methods
- Setting the expectations right data (data sources, indicators and sample size) and timelines





Session 1: Introduction To Counterfactual Impact Evaluation (CIE)



C4ED – EUTF

September 2021



- Identification of potential and clear objectives of what an IE can achieve
- Identification of suitable evaluation questions for impact evaluations



What is impact evaluation?

• Evaluation of an intervention, a policy and/or a project

The difference between outcomes with and without the intervention

Why evaluate the impact?

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



The difference between outcomes with and without the intervention

Counterfactual: The outcome at that same point in time that the program had not been introduced on the same sample of individuals

Fundamental problem: it is impossible to measure or observe the counterfactual

Solution: "mimicking" the counterfactual → creating a control/comparison group







Why Counterfactual Impact Evaluations?



Sufficient to establish correlations / associations of the program and the outcomes of interest Necessary to establish impacts / a causal link between the program and the outcomes of interest



Sufficient to establish correlations/ associations of the program and the

outcomes of interest

Necessary to establish **impacts/ a causal link** of the program and the outcomes of interest





throughtant

Link











What is the **impact or causal effect** of a policy/ a program/ an implementation approach on an **outcome of interest**?











What is the ULTIMATE goal of a Counterfactual Impact Evaluation?

Greater accountability

Greater learning



What is the ULTIMATE goal of a Counterfactual Impact Evaluation?





What is the ULTIMATE goal of a Counterfactual Impact Evaluation?





What is the ULTIMATE goal of a Counterfactual Impact Evaluation?





In practice, what is roughly the difference between accountability and learning evaluation questions in the CIE implementation stage?





Learning: How does the program work best?





For greater learning: before you start off with a new CIE... check existing evidence



a a a a













Where **plenty** evidence exists:

- Inform important policy **decisions** (where there is a lack of evidence)
 - Uptake of evidence: Make evidence actionable and scale up
 - Test "**robustness**" of results and reduce publication bias

Where **little** evidence exists: Identify **gaps in evidence**

- to guide new research efforts in new settings/ with other groups
- to resolve **conflicting** evidence



Useful tools...



- AI: Literature reviews via Connected Papers
- 3ie's "Evidence Gap Maps"
- International Rescue Committee's "Outcomes and Evidence Framework"
- Education Endowment's Foundation "Teaching and Learning Toolkit"

Center for Evaluation and Development AI: Literature reviews via Connected Papers



<u>(Link)</u>



\rightarrow	C	â	gap maps. 3 ieimpact.org/ evidence-maps/map-evidence-maps-relating-sustainable-development-lmics and the second



Last modified date: 17 April 2017

(D) 🚳 🕇

Q 🕁 💽 💁 🕝 🧳

A map of evidence maps relating to sustainable development in L&MICs

Evidence map	About	
HOVER OVER a legend at the bo	bubble to s	ee details with links to studies. CLICK ON a link in the axes to see an explanation of the Intervention / Outcome. SELECT an area of the chart to zoom in. TOGGLE study categories on and off using the e chart. EXPORT the chart using the menu button at the top right of the chart.

Region	Study design		Population		
All	✓ All.	-	All	•	Update chart

								Sustainable I)evelopment	Goals (SDGs)							
Sectors	SDG1 - Poverty	SDG2 - Agriculture nutrition and food security	SDG3 - Health	SDG4 - Education and Learning	SDG5 - Gender equality and empowerment	HSAM - 2066	SDG7 - Energy	SDG8 - Economic growth and employment	SDG9 - Infrastructure	SDG10 - Inequality	SDG11 - Urban and rural development	SDG12 - Consumption	SDG13 - Climate change	SDG14 - Marine	SDG15 - Ecosystem related outcomes	SDG16 - Peaceful and inclusive societies	SDG17 - Global Partnership
riculture and rural velopment	0		•			۲	۰	•	۲		۲	۲	۰			•	
imate change and wironment	0		0	۰		۲			0	۲	۲		•		0	0	
onflict management and ost-conflict reconstruction	۲	۲	۲	۲	٠	۲		۲		۲		۲			۲	۲	
oss-sectoral	•				0			•				0					






(Link)



Toaching an	Big Picture Evidence summaries Practical Tools Projects and Evaluation	Support for schools School In	nprovement Planning News	More 🗸 🔍
Teaching an				
Teaching an				
I Gaulli la all	d Learning Toolkit			
An accessible summary of the internation	al evidence on teaching 5-16 year-olds			
Filter Toolkit	Toolkit Strand ~	Cost~	Evidence Strength ~	Impact (months) -
Filter results by keywords	Arts participation			
	Low impact for low cost, based on moderate evidence.	$(\underline{\mathbf{f}})(\underline{\mathbf{f}})(\underline{\mathbf{f}})(\underline{\mathbf{f}})(\underline{\mathbf{f}})$		2
£	Aspiration interventions			
	Very low or no impact for moderate cost, based on very limited evidence.	£££££		•
	Kenaviour Interventions			

(Link)



...More on methods of counterfactual impact evaluations



• The selection of highly motivated program participants into the program is a challenge for an CIE

yes, "self-selection" into
the program makes it
harder to establish the
sole effects of the program

no



• The selection of highly motivated program participants into the program is a challenge for an CIE

yes, "self-selection" into the program makes it harder to establish the sole effects of the program

no









Do you need a control/ comparison group for a CIE?

yes

no



what would have happened to NON-beneficiaries if they had participated in the program

what would have happened to program beneficiaries if they had NOT participated in the program



what would have happened to NON-beneficiaries if they had participated in the program what would have happened to program beneficiaries if they had NOT participated in the program





yes, we need a pure control/ comparison group (a group of people who receive no program at all) no pure control/ comparison group is needed

Do you need a PURE control/ comparison group for a CIE?

yes, we need a pure control/ comparison group (a group of people who receive no program at all) no pure control/ comparison group is needed









yes

no, but an important add-on to CIE



yes, it is impossible to "control" other changes in the environment no, it is possible with the right method



yes, it is impossible to "control" other changes in the environment no, it is possible with the right method



yes, if the tools (the questionnaires) consider other potential outcomes

no











... more on requirements for Counterfactual Impact Evaluations



ex ante (before the program to be evaluated starts) ex post (after the program to be evaluated ends)













many units of observations (e.g. people, schools, firms, etc.) few units of observations



Think of the sample size as the **accuracy of our measurement**. The more observations you have:

- the more precise is your assessment
- the more confident you are about the conclusions of your evaluation





A small implementing organization cannot afford a CIE

yes, its too expensive

no, its manageable if data availability allows it (expensive are scale-ups of programs that do not work out)



several points of time with measurements

If assignment to program is (credibly) random, theoretically one can only use one wave of observations. Though this is less optimal.



RANDOMLY sampling from a sub-population for interviews

RANDOMLY selecting/ assigning who receives the program and who does not


RANDOMLY sampling from a sub-population for interviews

RANDOMLY selecting/ assigning who receives the program and who does not



very important, relevant results

a low probability that the CIE concludes (from the data) that the project worked when, in reality, it did not



very important, relevant results

a low probability that the CIE concludes (from the data) that the project worked when, in reality, it did not



When CIE researchers speak about a "wellpowered" study, they primarily refer to

a study well supported by policy makers and implementers

a low probability that the (CIE) study does not detect the impacts of the program, when in reality there has been a change due to the program



When CIE researchers speak about a "wellpowered" study, they primarily refer to

a study well supported by policy makers and implementers a low probability that the (CIE) study does not detect the impacts of the program, when in reality there has been a change due to the program



... a risk to all Impact Evaluations







Null hypothesis: No effect of the COVID-19 prevention program.Type II error: The program had an effect (the children wear masks), but we are not able to detect it.





Assume that a program has a positive impact on beneficiaries:

- If the evaluation sample is too small, you might not be able to detect this positive impact
- "Type II error": The risk of failing to conclude that your program has an impact even when it does
- Could lead to policy decisions to eliminate the program, which would be detrimental to beneficiaries and society



... for example by sharing resources, information or mimicking behavior ...

yes

no (can be measured, but then needs a specific design)



Is it a problem when the treatment group indirectly affects those who have not been treated?

- Spillovers contaminate the control group
 - physical (sharing of resources/environment)
 - information ("social learning")
 - market-wide (e.g. displacement effects) read more in "Running Randomized Evaluations" by Glennerster and Takavarasha
- Spillovers can cause impact to be underestimated or overestimated



















Session 2: Disentangling program needs, Theory of Change, Evaluation Questions, and Indicators



C4ED – EUTF

September 2021







- tells the story of a program and its vision for change
- is a conceptual map of a program towards its goal
- articulates the program and its underlying assumptions
- supports monitoring and evaluation



- The Theory of Change is a structured approach used in the design and evaluation of programs. It maps the logical chain of how program inputs achieve changes in outcomes.
- For each step in the ToC, we specify **indicators** that we will measure to help us understand whether the program has worked.
- If the program is not successful, having indicators for intermediate steps helps us understand at which step in the chain the program didn't operate as expected.



- Theories of change should be developed through a collaborative and consultative process
- A key element of the Theory of Change is the **identification of assumptions**. Analyzing and talking through these assumptions can help you to identify and **plan for potential risks**.



Theories of change help to prioritise which projects get evaluated and which do not

ToC also support institutional learning, accountability and increase the credibility and governance of an institution





- The Theory of Change describes all building blocks required to bring about a given goal asking core questions:
 - Why will the program work?
 - What are the channels or mechanisms through which the implementer assumes that the desired outcomes will be reached?
 - What assumptions are made?
 - What uncertainties exist?



Four steps:

- 1. Construction a Theory of Change;
- 2. Developing a results chain;
- 3. Specifying Evaluation Questions (EQs);
- 4. Selecting Outcome and Performance Indicators
- EQs structured as testable hypotheses deriving from the ToC
- EQs need to be guided by the core policy interest at hand

E.g.: Impact of the program on outcomes? Or testing whether one program modality is more cost effective than another? Or introducing a program design innovation expected to change behaviors?

Gertler et al. (2016)



• **Mechanism experiments**: IEs that do not test a program, but test a causal mechanism that underlies the choice of a program

Example: subsidies to greengrocers to fight obesity in poor neighborhoods

- Underlying assumption: Increasing access to healthy food will increase its consumption
- Mechanism experiment: Do households receiving free baskets of fruits and vegetables increase their intake in fruits and vegetables?

If not, we cannot expect the program to work!

Gertler et al. (2016)



- Main project stakeholders should agree on primary outcomes and anticipated changes
- Indicators need to be SMART (Specific, Measurable, Attributable, Realistic, Targeted)
- Identify indicators all along the results chain: key to know about Why an intervention may or may not work



Theory of Change of a School Voucher Program





Theory of Change of a School Voucher Program



Students and parents actually prefer public schools over private schools e.g. due to distance

Theory of Change of a School Voucher Program





For the Choice of IE questions....

- MEASURE WHAT? follow the theory of change and try to capture changes at each stage
- MEASURE WHEN? be realistic of not only what changes should be measured realistically, but also whether you can capture them at the time of measurement



Choice of IE questions





Choice of IE questions





Choice of IE questions








Choice of IE questions





Choice of IE questions



Are teachers informed?



1. Assessing impact requires identifying the right evaluation questions to investigate

2. Measuring impact of an intervention requires identifying relevant indicators

• Learning goals:

 Develop evaluation questions
 Develop and refine SMART (Specific, Measurable, Attributable, Realistic, Targeted) indicators





- Step 1: Identify project interventions and their aims
- *Step 2: Identify existing evidence on these interventions*
- Step 3: Check the underlying assumptions of the ToC
- Step 4: Define the main objectives of the impact evaluation
- Step 5: Validate and refine evaluation questions with stakeholders







EUTF economic project impact evaluation research in the Gambia: EQs and indicators

Main EQ No.	Country Level Evaluation & Sub- Evaluation Questions	Dimensions	Indicators	Evaluation method	Source of information	DAC Criteria
EQ0. Programme-specific monitoring EQ						
0.1. Implementation	0.1.GMB.a. Did Tekki Fii train the intended number of individuals?	Individuals trained	 Number of individuals trained Number of women trained Number of returnees trained 	Quantitative	Monitoring data	Effectiveness
EQ1. To what extent did EUTF interventions contribute to employment, job creation, and skills?						
To what extent did the Tekki Fii programme contribute to employment, job creation, and skills?						
1.1. What impact do trainings have on employability of beneficiaries and access to (decent) employment?	1.1.GMB.a. What effects does the Tekki Fii programme have on (decent) employment?	Employment	 Employment Employment status Formality Hourly productivity 	Quantitative	Youth questionnaire ¹	Impact
	1.1.GMB.b. What effects does the Tekki Fii programme have on employability?	Employability	- Perceived Employability	Quantitative	Youth questionnaire ¹	Impact
1.5. To what extent are training facilities 'fit-for- purpose' in delivering skills training to final beneficiaries?	1.5.GMB. To what extent are training facilities 'fit-for-purpose' in delivering skills training to Tekki Fii trainees?	Trainer background	 Years of education/field of education Years of experience in the trade Years of teaching experience 	Quantitative	Monitoring data	Relevance
		Trainee evaluation/ feedback	- Perceived training relevance	Quantitative	Youth questionnaire ¹	Relevance
EQ2. To what extent did EUTF interventions change resilience and livelihoods for beneficiaries?						
To what extent did the Tekki Fii programme change resilience and livelihoods for beneficiaries?						
2.1. What effects do trainings have on livelihoods and resilience?	2.1.GMB.a. What effects does the Tekki Fii programme have on livelihood, in terms of income?	Income	- Monthly income	Quantitative	Youth questionnaire ¹	Impact
	2.1.GMB.b. What effects does the Tekki Fii programme have on resilience?	Resilience	 Economic resilience: Has multiple activities Lowest level of income Income variation Income diversification 	Quantitative	Youth questionnaire ¹	Impact





Session 3: Introduction to Experimental Methods Evaluation Design, CIE Methods



C4ED – EUTF

September 2021





Abhijit Banerjee, Esther Duflo, Michael Kremer



2019 Sveriges Riksbank Prize in Economic Sciences in Memory of Alfred Nobel



116



"One of humanity's most urgent issues is the reduction of global poverty, in all its forms. Abhijit Banerjee, Esther Duflo, and Michael Kremer have introduced a new approach to obtaining reliable answers about the best ways to fight global poverty. It involves dividing this issue into smaller, more manageable, questions. Since the mid-1990s, they have been able to test a range of interventions in different areas using field experiments, for example for educational outcomes or child health."

<u>Source</u>



- Understand the richness of the method and explore different method
- Understand the limitations of the method
- You will need a coin...



Goal: Simulate "counterfactual" situation

- What would have happened if the program did not operate?
- Best way to find a counterfactual: Randomized Control Trial (RCT)
- Intuition: Randomly assign program to a treatment group and control/ comparison group control/ comparison group "mimics" counterfactual outcome of the treatment group



Why is the RCT considered "the **gold standard**" of quantitative impact evaluations?

- No assumptions such as outcome levels or trends between groups necessary if randomized within groups
- Delivers strong evidence
- Easy to interpret























White, H., & Raitzer, D. A. (2017)

- 1. Identify the evaluation question
- 2. Isolate treatment(s) of interest
- 3. Discuss spillover effects
- 4. Determine level of randomization, treatment and analysis
- 5. Decide on the type of randomization
- 6. Identify your eligibility group
- 7. Draw the sample for analysis
- 8. Randomize
- 9. Collect baseline & check balance
- 10. Ensure the integrity of the design & monitor



- RCTs give the clearest possible causal evidence, if welldesigned
- RCTs are **easy to analyse**
- Randomization is a fair and transparent way to allocate benefits

Still not convinced?

- Designs without pure control group exist
- Your roll-out might already fit the RCT design (Phase-in)
- Randomization can be applied to only a subsample of the target group



- The level of observation at which treatment and comparison groups are randomly assigned.
- e.g., individual, household, school, village



- 1. When demand outstrips supply
- 2. When an innovation cannot be delivered to all units at once
- 3. When experimental units can be temporally isolated
- 4. When experimental units are spatially separated or interunit communication is low
- 5. When change is mandated and solutions are acknowledged to be unknown
- 6. When a tie can be broken or ambiguity about need can be resolved
- 7. When some persons express no preference among alternatives
- 8. When you can create your own organization
- 9. When you have control over experimental units
- 10. When lotteries are expected

SHADISH, W.R., COOK, T.D. AND CAMPBELL, D.T., (2002)



RCTs in PRACTICE



- A. RANDOMLY sample sub-population for study NO
- **B. RANDOMLY select/ assign who receives the program and who not** YES!





J-PAL, Poverty Action Lab, L4, 2019

RANDOMLY select/assign who receives the program and who not



Random assignment into a program (treatment group)





J-PAL, Poverty Action Lab, L4, 2019











Collect baseline data for eligible population

UNICEF-IKEAF program Collect endline data Calculate impact

















Source: White et al. (2014) Randomized Controlled Trials (RCTs), Methodological Briefs: Impact Evaluation No. 7, UNICEF Office of Research, Florence. Modified by authors.



RCTs: Different approaches



How to RANDOMIZE into treatment and control/ comparison group?

- 1. Lottery
- 2. Multiple treatment arms
- 3. Randomized phase-in
- 4. Randomized intensities/ multiple-stage RCTs
- 5. Encouragement design



- Randomly select treatment units (people, streets, communities) by a lottery
- Two types of lottery:
 - Public: Fair and transparent but may lead to behavioral changes in control/ comparison group (pull out of a hat/bucket)
 - Private: Less transparent, but mitigates behavioral responses in control/ comparison group



• one-off cash transfer worth 1,000 USD in a rural refugee settlement







Consumption (calories)








Consumption (calories)

146



- Advantage: Transparent and fair
- **Disadvantage**: Control/ comparison group likely to also change behavior if public; need big sample size; not practical to stratify



- Building clusters for easier implementation and to provide a buffer
- Assign program at a higher level, even when we collect data on a lower level









- Assign units (people, communities, etc,) randomly to different treatment arms (versions of the program; types of interventions) and control/ comparison group
 - Analyze different versions and components of a program against each other and against the control/ comparison group
- Also possible without a control/ comparison group



Educational intervention of any kind delivered via two possibilities:

- Personal visit
- Phone call



• Possible combinations of interventions

	Phone call	No Phone Call
Personal visit	\$ T	ŕ
No personal visit	ŝ.	



Most useful when under uncertainty, especially if you want to identify which component is the best way to achieve your goals

- Advantage: Gives better insight to channels of impact rather than just answering if it works or not
- **Disadvantage**: Needs larger sample size



Social norms, motivation and habit play crucial roles in adherence to safe hygiene behavior

• SMSs with prominent public figure: Improved compliance with national recommendations of the recipient and surrounding (Banerjee et al. 2020; West Bengal, India)







Famous academic sharing additional COVID-19 messages

- Encouraging to report symptoms to the local public health workers
- One health preserving behavior emphasized (2-meter, mask, hygiene)

One message stressing externalities (illness could be damaging to others)

Explicitly stating that ostracism of infection is unacceptable







Important role for messaging by credible individuals

- behavior and knowledge are not one-to-one: large shifts in distancing, hygiene, and mask-wearing despite little-to-no shifts in knowledge
- the exact content of the message may not matter; important to nudge to pay attention to it



Social norms, motivation and habit play crucial roles in adherence to safe hygiene behavior

• Awareness campaigns via phone and loudspeakers (Avdeenko et al. 2020; Pakistan)











14	55.1% know about asymptomatic spread of COVID-19	
	86.8% know symptoms	
1	21.6% believe its better to keep COVID-19 status a secret	
	41.1% believe traditional healers can treat the COVID-19 succe	

lly











Everybody receives treatment, but at different time periods

- Those who receive treatment last is control/ comparison group
- Should be done with enough time in between for impacts to develop



• Phase 1: Treat only $\frac{1}{4}$, $\frac{3}{4}$ remains control





• Phase 1: Treat only 2/4, 2/4 remains control





• Phase 1: Treat only $\frac{3}{4}$, $\frac{1}{4}$ remains control





• Phase 1: Treat only $\frac{4}{4}$, 0 remains control





- Most useful when: Budget/time/capacity constraints prevent roll-out of program for everybody at the same time
- Advantages: Everyone receives treatment. Transparent decision, fairness. Allows to analyze effects with different intensity (duration) of treatment
- **Disadvantages**: Long-term impacts not possible to capture















Randomly treat in some communities 100% of eligible people, in other communities only part (here e.g. 50%), and in other community nobody (control/ comparison group)



Example: Intervention on hygiene awareness to fight infectious diseases

- Village 1: Control (no treatment)
- Village 2: 100% if individuals get treatment (information on hygiene)





Now only 50% of individuals in village 2 get treatment:

- Hypothesis: People share the information with each other
- Compare outcomes of treated AND untreated in village 2 with control/ comparison group in village 1





- **Most useful when**: You want to analyze a program, where spillovers are likely. For example, spreading of information in an informational program
- Advantage: Analyze possible spillovers
- **Disadvantage:** Bigger sample size necessary



Situation: It is impossible to deny a treatment (free food stamp program)

- Solution: Randomize encouragement → Give advertisement and information to some, and none to others. People can still take-up the intervention, even without encouragement
- Encouragement works like an instrument (IV)
- Compare those who received encouragement with those who did not



• Solution: "Treatment" is encouragement to use food stamps





- Compare whole treatment group with whole control/ comparison group
- Analyze the effect of encouraging food-stamps (instrumental variable) and indirectly the impact of food-stamps on nutrition itself




- Advantage: Allows to analyze programs where everybody must have access
- **Disadvantage:** Only part of the impact is measured; could be problematic for estimation if all control/ comparison group is highly encouraged (e.g. if encouragement/ mobilization via TV, radio/ not targeted)



Assignment



- A household survey in Fantasia Land revealed that 60% of the rural population still cooks with fire wood. Cooking with wood adds to the current problem of deforestation in the country and increases the indoor pollution.
- The government has therefore decided to distribute clean cooking stoves
- Eligible are rural households, with less than 1ha of land, in two of the ten departments of the country
- The cooking stove can be collected at the local extension worker offices, which are widespread and hold records of all farming households in their catchment area.
- The implementation period is 2022-2025.

The government asks you for help:

They want to learn about the **impact of the intervention**! They are open for experimental methods.



- 1. Please suggest an experimental impact evaluation method that could be used.
 - 1. Explicitly describe how you would identify the treatment and control group.
 - 2. What are the limitations and strengths of each method?
- 2. State the evaluation question that this design can answer and identify suitable indicator(s).
- 3. What kind of data do you need and when should you collect it?



"Household air pollution arising from the combustion of dirty-burning fuels in and around the home for cooking and heating (e.g., wood, crop waste, dung, coal) is estimated by the World Health Organization (WHO) to cause around 4 million premature deaths per year".

World Health Organization. 2014. Indoor air quality guidelines: household fuel combustion.).



Target group: rural households, with less than 1ha of land, in two of the ten departments of the country

Outcomes: respiratory infection, physician-diagnosed pneumonia, cough, forced expiratory volume, premature death, children health, family health, secondary outcomes: level of deforestation in the target area

Evaluation question(s): what are the effect of distributing (what about really using?) clean cooking stoves on the above outcomes?

Method of randomization: Lottery, Randomized phase-in, Multiple treatment arms, Encouragement design

Treatment group: randomly selected HHs/villages in two departments (unit of observation will be HHs in villages) who receive clean cooking stoves

Control group : randomly selected HHs/villages who will not receive clean cooking stove (at least for some period of time)

Data: Baseline (not absolutely required in case of randomization but highly recommended) and Endline





Gertler, P. J., Martinez, S., Premand, P., Rawlings L., Vermeersch, C. M. J. (2011) Impact Evaluation In Practice, First Edition, The World Bank.

Gertler, P. J., Martinez, S., Premand, P., Rawlings L., Vermeersch, C. M. J. (2016) Impact Evaluation In Practice, Second Edition, The World Bank.

White, H., & Raitzer, D. A. (2017). Impact evaluation of development interventions: A practical guide. Asian Development Bank.

Correlations: https://towardsdatascience.com/nicholas-cage-pool-saviour-9c13feafff6f

White, H. (2019). The twenty-first century experimenting society: the four waves of the evidence revolution. Palgrave Communications, 5(1), 1-7. <u>https://www.nature.com/articles/s41599-019-0253-6</u>

Connected Papers: https://www.connectedpapers.com/

3ie's "Evidence Gap Maps": <u>https://gapmaps.3ieimpact.org/evidence-maps/map-evidence-maps-relating-sustainable-development-lmics</u>

International Rescue Committee's "Outcomes and Evidence Framework": <u>http://oef.rescue.org/#/home? k=zhw0nz</u>

Education Endowment's Foundation "Teaching and Learning Toolkit": <u>https://educationendowmentfoundation.org.uk/evidence-summaries/teaching-learning-toolkit/</u>

Gertler, P. J., Martinez, S., Premand, P., Rawlings, L. B., & Vermeersch, C. M. (2016). Impact evaluation in practice. World Bank Publications.

https://www.nobelprize.org/prizes/economic-sciences/2019/press-release/

Shadish, W. R., & Cook, T. D. (86). Campbell. DT (2002). Experimental and quasi-experimental designs for generalized causal inference.

J-PAL, Poverty Action Lab, L4, 2019: https://www.povertyactionlab.org/sites/default/files/L4_How%20to%20Randomize.pdf

IDinsight (2020): <u>https://www.idinsight.org/article/implementing-a-10000-refugee-household-public-lottery-in-eight-days-lessons-from-kiryandongo-uganda/</u>

Banerjee, A., Alsan, M., Breza, E., Chandrasekhar, A. G., Chowdhury, A., Duflo, E., ... & Olken, B. A. (2020). Messages on COVID-19 prevention in India increased symptoms reporting and adherence to preventive behaviors among 25 million recipients with similar effects on non-recipient members of their communities (No. w27496). National Bureau of Economic Research.

Avdeenko et al. 2020: Evidence on Covid-19 Pandemic Control Interventions and their Impacts on Health-Related Outcomes

World Health Organization. 2014. Indoor air quality guidelines: household fuel combustion.