



GENDER Impact  
Platform

# Integrating Gender into Causal Impact Assessments: understanding the scope of causal impact assessments and how to incorporate gender effectively

## Causal Impact Assessment Methodologies

March 2025

Diana Lopez-Avila  
davila@cgiar.org  
Senior Scientist  
Gender and Inclusion Accelerator-CGIAR



GENDER  
EQUALITY



SOCIAL  
INCLUSION



## IMPACT ASSESSMENT:

### CAUSAL IMPACT ASSESSMENT:

It focuses on identifying the **CAUSAL RELATIONSHIP** between the intervention and the observed changes. It is designed to **attribute the changes** observed to the innovation or intervention studied

### NON-CAUSAL IMPACT ASSESSMENT:

It assesses the **contribution of the program/innovation/intervention to the observed changes**. It does **not establish a direct causal link** between the program and the changes. The **observed changes cannot be definitively attributed** to the intervention itself.

#### QUANTITATIVE:

Effects-of-causes  
approach

#### QUALITATIVE:

Causes-of-effect  
approach



# Causal Impact Assessments

Through a causal impact assessment, the basic question we want to answer is “What is the causal effect of a program/intervention (P) on an outcome of interest (Y)”?

$$\Delta = (Y|P = 1) - (Y|P = 0)$$



Counterfactual



# Causal Impact Assessments-Counterfactual

To implement a causal impact assessment, we need to have a **comparison group (or COUNTERFACTUAL)**. The way we choose the counterfactual is crucial for understanding the true effect of a program, as they help isolate the effects of the intervention from other external factors.

- It represents hypothetical scenarios that illustrate what would have happened to participants had they not received the intervention.
- Because the counterfactual is impossible to measure directly, methods of causal impact assessments try to mimic the counterfactual by selecting a comparison group of non-participants who closely resemble the participants had they not been treated.
  - The way the counterfactual is chosen is what differentiates experimental from quasi-experimental methods



# Causal Impact Assessments-Counterfactual

A comparison group is a **good counterfactual** if:

- ✓ The **average characteristics** of the treatment group and the comparison group must be the **same in the absence of the program**.
- ✓ The **treatment should not affect the comparison group** either directly or indirectly (contamination)
- ✓ The only **difference between the treatment and control groups**, then, is their **participation in the intervention itself**, and the difference in their outcomes therefore represents the impact of the intervention or program.



# Inappropriate estimates of a counterfactual

- *Before-and-after comparisons (also known as pre-post comparisons):*
  - Compare the outcomes of the same group before and after participating in a program/receiving an intervention.
- *Enrolled-and-non enrolled (or self-selected) comparisons:*
  - Compare the outcomes of a group that chooses to participate in a program with those of a group that chooses not to participate.

Why do you think these comparisons don't have a good counterfactual?

*Please type your answers in the chat*



# Inappropriate estimates of a counterfactual

## Before & After:

- The counterfactual is the outcome of the treated group before the intervention started.
- This comparison assumes that, without the program, participants' outcomes (Y) would be the same as before the program.
- Interventions are implemented over several months or years and many things can change over that period

## Enrolled vs Non-Enrolled:

- The counterfactual is a group of individuals who decided NOT to participate in the intervention.
- **Self-selected comparison groups provide biased estimates due to unobserved characteristics influencing participation.**
- Those who opted to not enroll in the program may be different to those who enroll in unobservable traits like motivation, self-efficacy, or other personal factors that influence their decision to participate.

If the comparison group poorly represents the counterfactual, the estimate of the causal effect will be biased and invalid



# Experimental Methods

## Randomized Controlled Trials (RCT)

- All eligible units in a sample are randomly assigned to treatment and control groups (e.g., there can be more than one treatment group)
- Random assignment ensures treatment and control units are, on average, similar in both observed and unobserved characteristics.
- Compare outcomes between randomly assigned participants and non-participants after the program.
- The only difference between the treatment and control groups is their participation in the intervention, so the outcome difference reflects the program's effect.

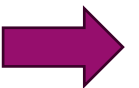




# Randomized Controlled Trials (RCTs)

- Let's think about the following situation:
  - A group of researchers is planning to assess the impact of an extension service intervention, with a focus on **testing different approaches to better target women and increase their participation in agricultural decision-making.**
    - Incorporating a **gender sensitization** component on women's roles in agriculture, targeted at couples.
    - Ensuring that at least 50% of the extension agents are women

To test which approach is more effective, they propose the following randomized controlled trial





# Randomized Controlled Trials (RCTs)

Random allocation of villages across the 3 groups) →

46 villages where agricultural extension services will be implemented

Researchers did a census and found that, on average, there are 60 couples per village

16 villages where extension services will be implemented as usual (ES) (~960 couples)

In 15 villages, in addition to the standard extension services, at least 50% of the extension agents will be women (ES+WA) (~900 couples)

In 15 villages, in addition to the standard extension services, a gender sensitization will be offered to couples (ES+GS) (~900 couples)

ES vs ES+WA

ES+WA vs ES+GS

ES vs ES+GS

Integrating Gender into Cluster Impact Assessments-Diana Lopez-Avila





# Randomized Controlled Trials (RCTs)

- Let's think about the following research questions and discuss which comparison allow researchers to answer each question:

Does a having a good representation of female agents among extension agents increase adoption rates and women's participation in agricultural decisions?

ES vs ES+WA

Does a implementing a gender sensitization component targeted to couples (addressing women's role in agriculture) increase adoption rates and women's participation in agricultural decisions?

ES+WA vs ES+GS

Is a gender sensitization more effective than having a good representation of female agents among extension agents?

ES vs ES+GS



# Randomized Controlled Trials (RCTs)-Timeline

- ✓ The intervention to test is identified and the research questions of interest, as well as the planned experimental design
- ✓ If the target population (e.g., the population that is going to be part of the study) hasn't been identified, a census or identification mapping is conducted
- ✓ Questionnaires are designed to measure key outcomes of interest- key to determine from whom data will be collected (e.g., primary male, primary female, other household members)
- ✓ Power size calculations are conducted to identify the sample size needed to measure an effect (the smaller the effect expected, the larger the sample needed)
- ✓ Baseline data collection takes place
- ✓ Randomization is conducted (through a statistical package) and balance tests are implemented
- ✓ The intervention starts and is implemented
- ✓ Follow up data collections take place depending on the ToC and times in which effects are expected to be observed



# Randomized Controlled Trials (RCTs)-Timeline

## ✓ Power size calculations:

- ✓ Power calculations are essential to determine the minimum sample size required to detect an effect of size  $X$  in the outcome of interest.
- ✓ If multiple key outcomes are being assessed, power calculations should be conducted for each, and the largest required sample size should be used.
- ✓ In many cases, when data from the target sample is unavailable, we must rely on data from a similar study or context.
- ✓ Keep in mind that when randomizing at the group (or cluster) level (e.g., villages), the number of clusters is crucial. While cluster size matters, the total number of clusters is what mostly determines the statistical power.

## ✓ Balance Tests

- Once BL data has been collected and the randomization has been done, we need to compute balance tests, that is to compare the different treatment and control groups across key socio demographic characteristics and outcomes of interest to ensure they are balanced (e.g., we don't observe large and significant differences between groups)



# Randomized Controlled Trials (RCTs)

- Considerations to have in mind when designing and implemented a randomized controlled trial:
  - **Are baseline characteristics balanced?** Based on pre-intervention data, we need to test whether the groups are comparable, ensuring there are no statistically significant differences in baseline characteristics, particularly in those that correlate with the outcomes of interest (e.g., adoption rates and women's participation in agricultural decisions)
  - **Has noncompliance occurred?** Verify that all eligible units received the treatment they were supposed to receive
    - There is no contamination (e.g., farmers in villages assigned to ES were also part of the gender sensitization)
    - Many male spouses in villages in group ES+GS didn't participate in the gender sensitization campaign
  - **Has there been substantial attrition, and does it differ between treatment and control groups?** Attrition may occur between data collection rounds due to factors like migration, death, or refusal to participate. If a significant portion of the sample is lost, it can lead to power issues and compromise the ability to detect an effect. Additionally, if attrition differs between groups, the results may be biased.





# Group exercise

Members of the women's group preparing food. Photo by C. de Bode/CGIAR





# Randomized Controlled Trials (RCTs)-Group Exercise

In the first sessions, we explore research questions and ToC for the following research topics:

1. The role of gender norms in shaping women's participation in livestock rearing and strategies to enhance vaccination uptake among farmers
2. Designing gender-inclusive extension services to promote the adoption of improved rice varieties among farmers
3. Enhancing women's financial literacy and inclusion to improve access to agricultural inputs and boost productivity

We explored descriptive, relational and causal questions





# Randomized Controlled Trials (RCTs)-Group Exercise

Let's break up into groups

- Based on the research questions and ToC discussion, think about how you could design a randomized controlled trial to answer the causal research question of interest
- To guide the discussion, think about:
  - What comparison you need to answer the research question of interest? Do you need only one treatment group or do you need more than one treatment group?
  - Who is your target population (e.g., sample of interest), only female farmers, female and male farmers, female farmers and their spouses, etc.?
  - What will be the level of the randomization? Counties, villages, farmers associations, farmers, etc.? What do you think this is the right level of randomization?
- Please nominate a facilitator and a rapporteur
- At the end of the group exercise, the rapporteur will present the key points discussed (~ 5 min per group)
- Time dedicate for discussion 15 minutes



# Quasi-experimental methods

- Some of the quasi-experimental methods available are:
  - Matching
  - Differences-in-Differences
  - Regression Discontinuity Design

There is a counterfactual group BUT is NOT RANDOMLY chosen



# Quasi-experimental methods : Matching

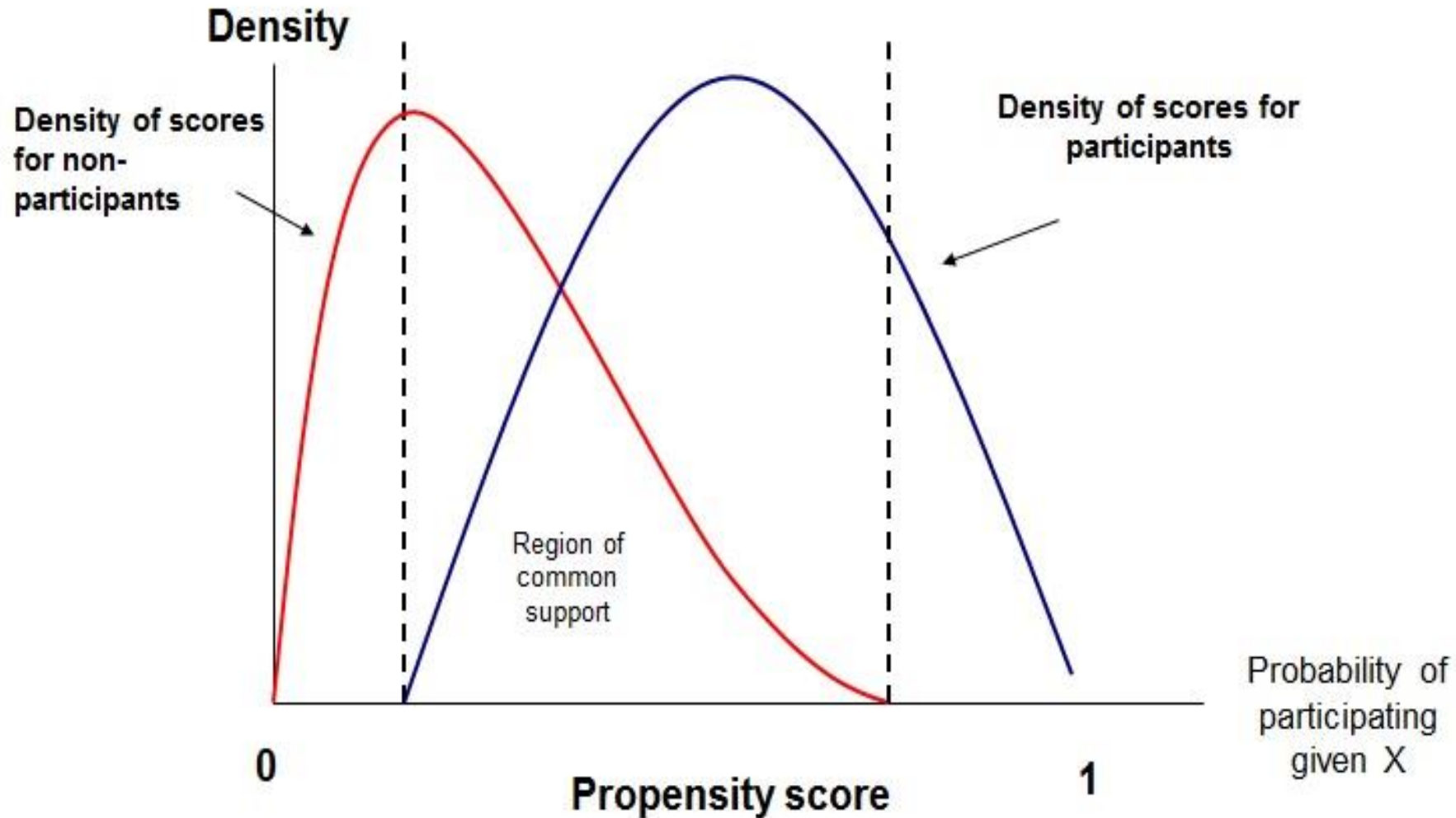
- What does this methodology do?
  - **Matches** individuals in the **treatment group** to those in the **control group** based on **observable characteristics**.
  - Pairs each treated individual with one or more comparable individuals from the potential control group.
  - Estimates the probability of treatment/program participation using observed characteristics.
  - **Retains only treated and control individuals with similar likelihoods of being treated**, ensuring comparability.



**Key assumption:**  
participation into  
treatment is only  
defined by  
observable  
characteristics



# Quasi-experimental methods : Matching





# Quasi-experimental methods : Matching

- Key Considerations for Identifying a Control Group:
  - ✓ **Larger Control Group:** The control group should be bigger than the treatment group, as some units may not serve as good matches and will need to be excluded from the analysis.
  - ✓ **Eligibility Criteria:** The same criteria applied to the treatment group (e.g., land size, household composition, plot characteristics, household income) must also apply to the control group.
  - ✓ **Match Based on Pre-Intervention Characteristics:** Matching should be done using baseline (BL) data collected before the intervention. This requires panel data, where the same units are observed multiple times.
    - ✓ **Addressing Lack of BL Data:** If baseline data is unavailable, you can attempt to match using variables which are (most likely) unaffected by the intervention
      - ✓ **Matching on post-intervention characteristics really compromises the accuracy and validity of the results.**



## The power of the collective empowers women: Evidence from self-help groups in India.

- **Motivation/Problem:** Women's self-help groups (SHGs) in rural South Asia, particularly India, have evolved from savings and credit models to platforms promoting health, governance, and social equity. Despite their success, evidence on their impact on women's empowerment remains mixed.
- **Methods:**
  - This paper uses panel data from 1,470 rural Indian women across five states to assess the impact of SHG membership on women's empowerment in agriculture. It employs the Women's Empowerment in Agriculture Index (pro-WEAI) and the abbreviated A-WEAI to measure empowerment at the individual and household levels.
  - Given that SHG membership is not randomized, the study uses nearest neighbor matching (NNM) methods to address potential self-selection bias and isolate the causal effects of SHG membership on women's empowerment and intrahousehold inequality.
  - They construct a comparison group by matching SHG members to nonmembers based on observable respondent, household and community characteristics.

Kumar, N., Raghunathan, K., Arrieta, A., Jilani, A., & Pandey, S. (2021). The power of the collective empowers women: Evidence from self-help groups in India. *World Development*, 146, 105579.



## The power of the collective empowers women: Evidence from self-help groups in India.

- **Methods:**
  - Since SHGs existed in the sample villages at baseline, data on these variables before women became members is unavailable, posing a challenge for ensuring proper matching.
  - To reduce endogeneity bias, the study uses SHG membership at midline and matches on predetermined, exogenous variables measured at baseline, including women's characteristics, household traits, time spent on household tasks, and village, district, and state characteristics.



## The power of the collective empowers women: Evidence from self-help groups in India.

- **Results:**
  - SHG participation has significant impacts on the overall empowerment score and the empowerment gap between spouses (both pro-WEAI and A-WEAI), with a weak positive impact on men's empowerment (only A-WEAI).
  - The results suggest that closing the intrahousehold empowerment gap did not disempower men, countering the potential unintended consequence of programs aimed solely at women.
  - Strong positive impacts are seen in women's credit access, decision-making, and control over income use, though the impacts are weaker in areas like decision-making on production and asset ownership, due to deep-seated gender norms.



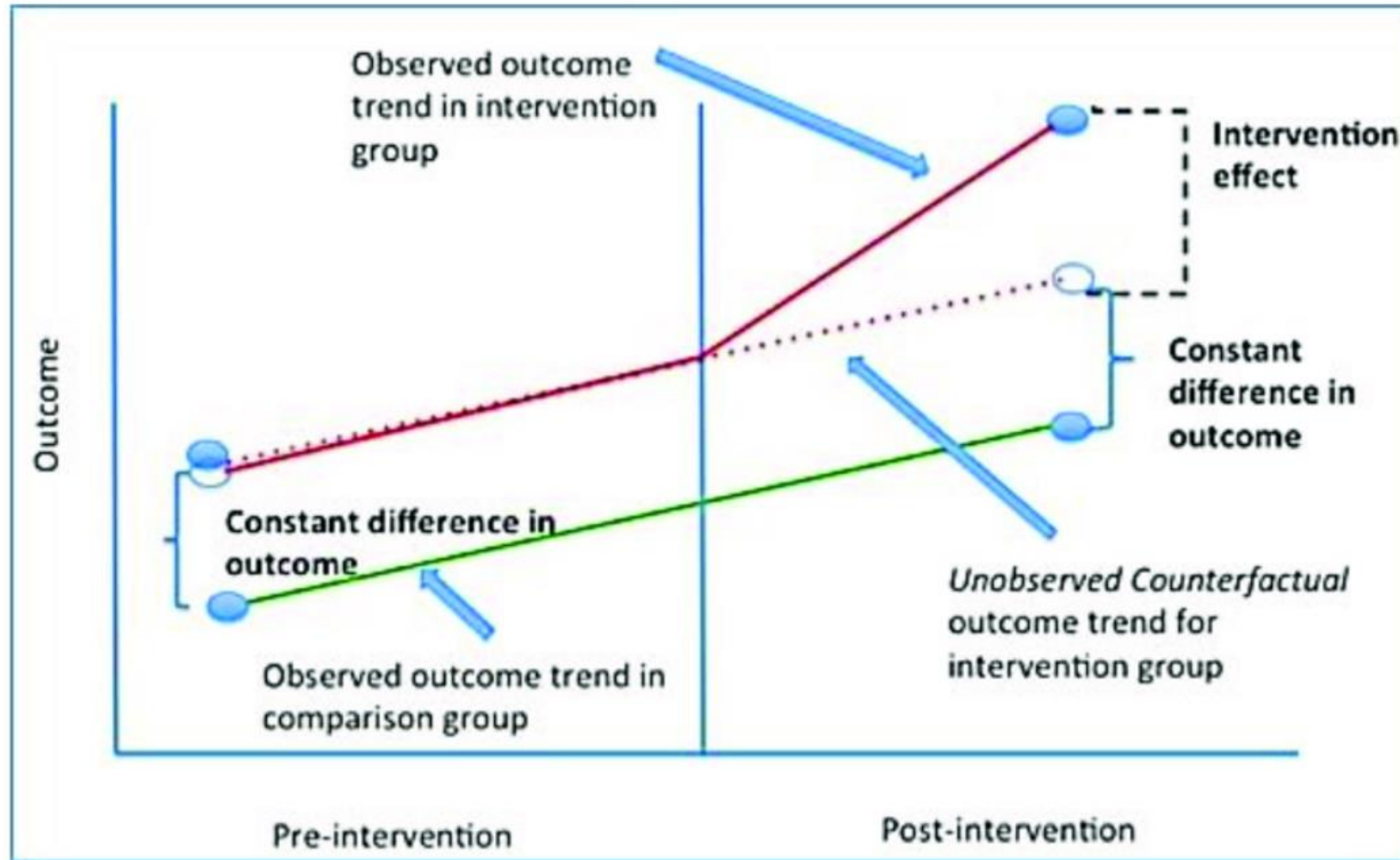


# Quasi-experimental methods : Differences-in-Differences

- Difference-in-differences **compares changes in outcomes over time** between units enrolled in a program (**treatment group**) and those that are not (**comparison group**)
    - Instead of comparing outcomes between the treatment and comparison groups after the intervention, the difference-in-differences method compares trends between the two groups over time
    - This approach allows for **correcting any differences between** the treatment and comparison groups **that remain constant over time**
- ➔
- Key **assumption** to test: **existence of parallel trends in the outcome of interest before the intervention**
  - **We need data for at least 3 points in time**: before the intervention, when the intervention started and after the intervention



# Quasi-experimental methods : Differences-in-Differences





- Motivation/Problem:
- Intervention:
- Methodology:

Bryan, E., & Mekonnen, D. (2023). Does small-scale irrigation provide a pathway to women's empowerment? Lessons from Northern Ghana. *Journal of Rural Studies*, 97, 474-484.



## Quasi-experimental methods : Regression Discontinuity Design

- Regression Discontinuity Design (RDD) is a methodology suited for interventions that use a continuous eligibility index (or variable) with a clear cutoff score to determine participant eligibility.
- Key considerations:
  - The index must be continuous and smoothly ranked, like poverty scores or test scores, but not categorical variables like employment status or education level.
  - A clearly defined cutoff score must determine eligibility, such as a poverty index below 50 or an age threshold for pensions.
  - The cutoff must be unique to the program being evaluated, ensuring no other programs use the same eligibility threshold.
  - Scores must not be manipulable by enumerators, beneficiaries, or administrators to maintain the integrity of the evaluation.



## Quasi-experimental methods : Regression Discontinuity Design

- Consider an agriculture program that aims to improve total rice yields by subsidizing farmers' purchase of fertilizer.
  - The program targets small and medium-size farms, which it classifies as farms with fewer than 50 hectares of land.
  - Farms are eligible for fertilizer subsidies if they have less than 50 hectares, while those with 50 or more are not.
    - Farms just below the cutoff (e.g., 49.9 ha) are similar to those just above (e.g., 50.1 ha), except for program participation.



This allows for comparison to assess the causal effect of the subsidy



# Quasi-experimental methods : Regression Discontinuity Design

Figure 6.1 Rice Yield, Smaller Farms versus Larger Farms (Baseline)

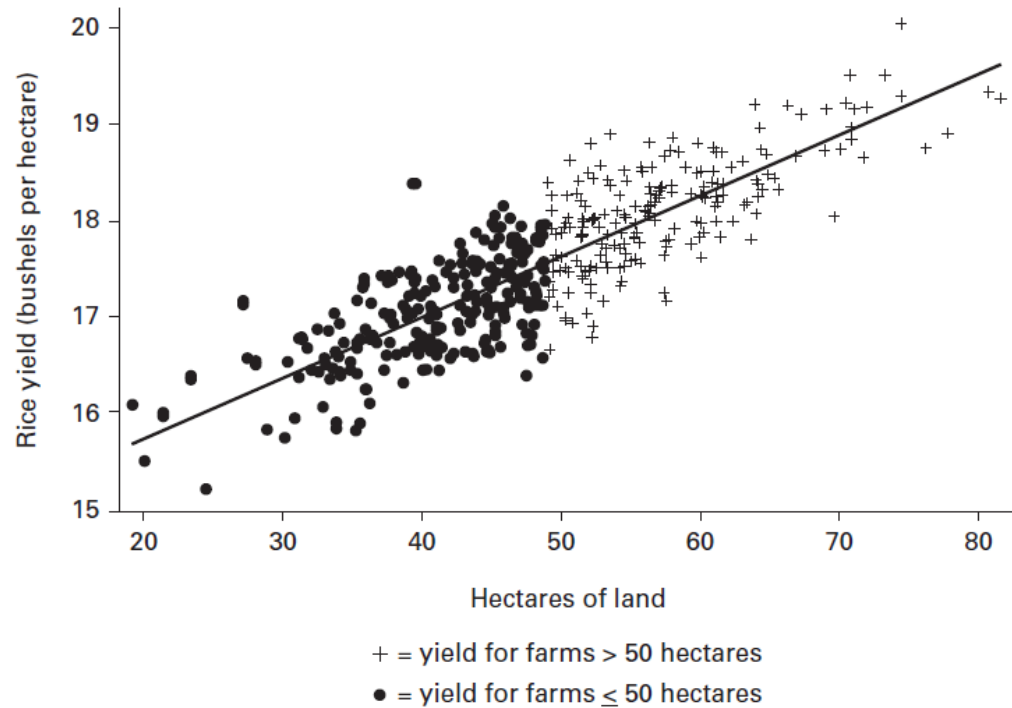
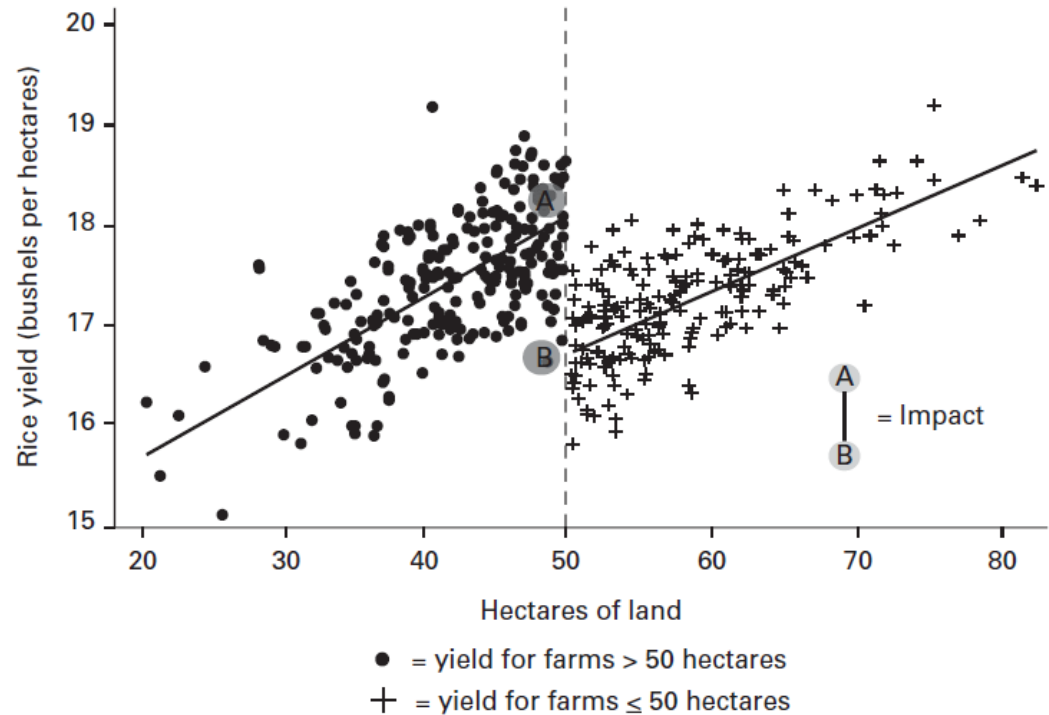


Figure 6.2 Rice Yield, Smaller Farms versus Larger Farms (Follow-Up)





## Cash Transfers and Women's Agency: Evidence from Pakistan's BISP Program

- **Motivation/Problem:** Cash transfer programs in developing countries aim to reduce poverty, support consumption, and improve human capital, often targeting women to enhance empowerment. This study examines the impact of Pakistan's Benazir Income Support Program (BISP) on women's agency after two and five years.
- **Intervention:** In 2010, Pakistan introduced a Proxy Means Test (PMT) to target BISP transfers to the poorest, replacing selection by local parliamentarians. Households scoring below 16.17 in the 2010–11 Poverty Scorecard survey qualified, with ever-married women holding a valid identity card eligible to receive benefits.
- **Methodology:**
  - The analysis combines household survey data from 2011, 2013, and 2016, specifically collected for the BISP evaluation. And administrative records on eligibility and payments

Ambler, Kate; and de Brauw, Alan. Cash transfers and women's agency: Evidence from Pakistan's BISP program. *Economic Development and Cultural Change* 72(3).



## Cash Transfers and Women's Agency: Evidence from Pakistan's BISP Program

- **Methodology:**
  - The way BISP was targeted and designed allows for a causal identification through RDD, with households just above and below the eligibility threshold serving as comparison groups.
  - The key assumption is that eligible and ineligible households near the cutoff only differ through BISP eligibility, with no other programs targeting beneficiaries using the same poverty score.
  - RDD identifies program effects for households near the cutoff, using eligibility as an instrument for beneficiary status, with a fuzzy RDD estimate accounting for imperfect compliance.
  - The analysis combines household survey data from 2011, 2013, and 2016, specifically collected for the BISP evaluation. And administrative records on eligibility and payments

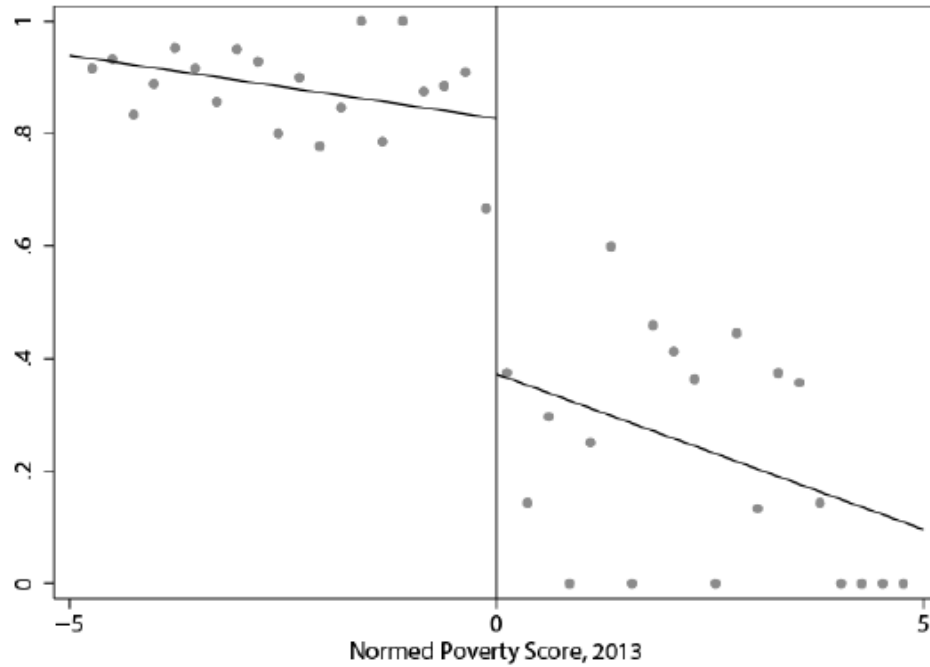
Ambler, Kate; and de Brauw, Alan. Cash transfers and women's agency: Evidence from Pakistan's BISP program. *Economic Development and Cultural Change* 72(3).



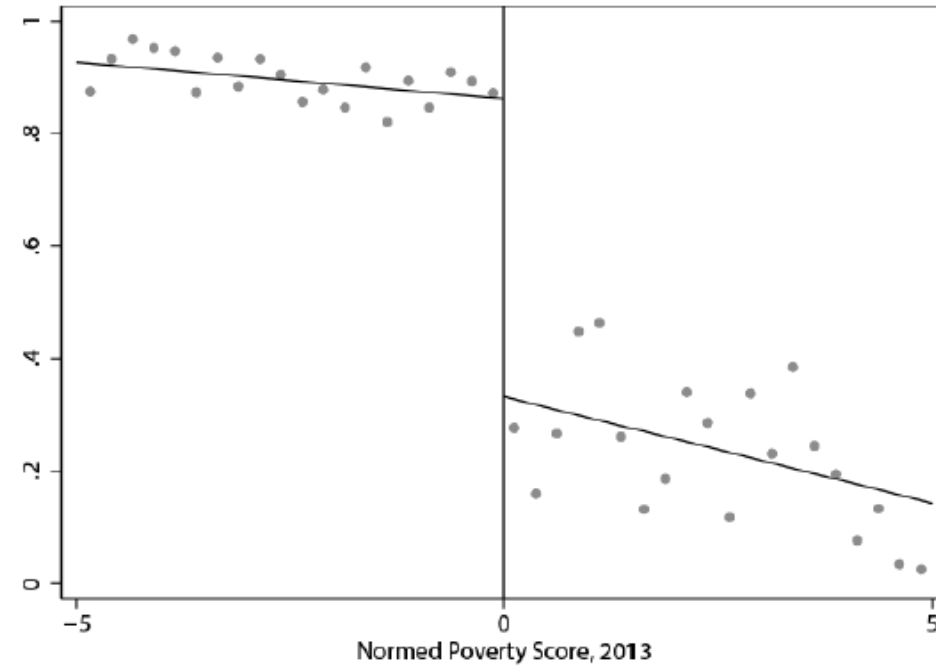


Relationship between poverty score and probability of receiving BISP transfers (\*score normalized).

**A** 2013 survey, urban sample



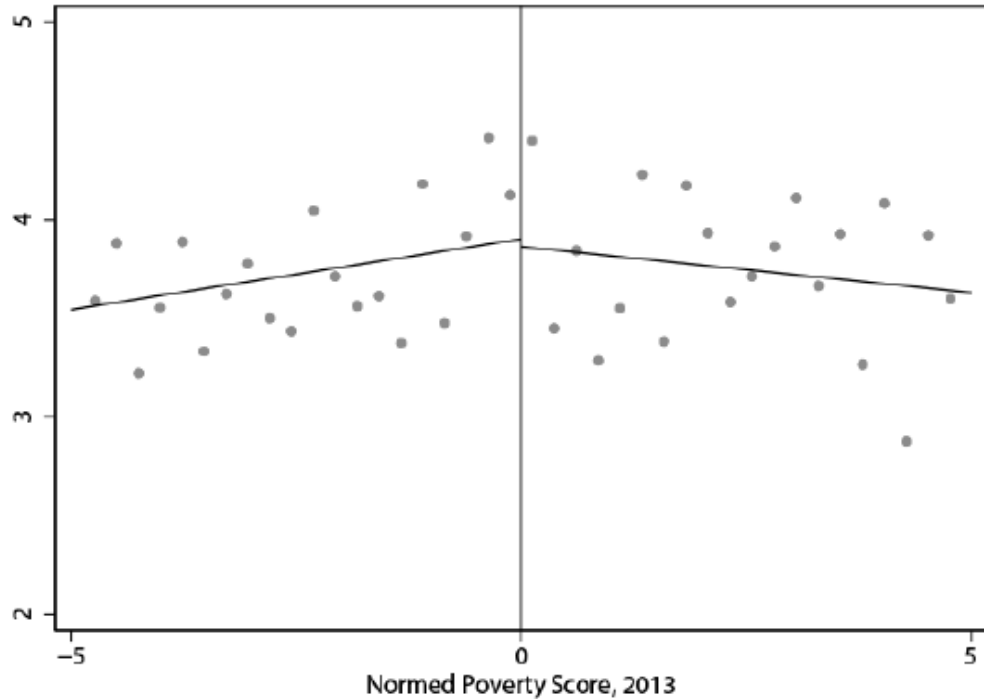
**B** 2013 survey, rural sample



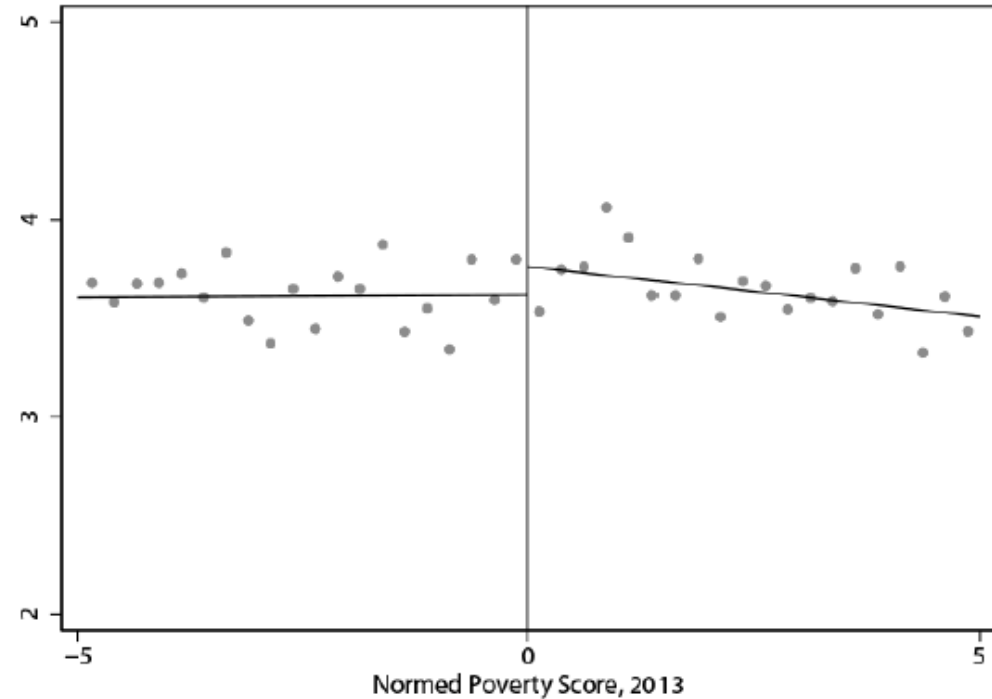


# Relationship between gender-norms index and poverty score (\*score normalized).

## A 2013 survey, urban sample



## B 2013 survey, rural sample





## References

- Gertler, Paul J.; Martinez, Sebastian; Premand, Patrick; Rawlings, Laura B.; Vermeersch, Christel M. J.. 2016. Impact Evaluation in Practice, Second Edition. © Washington, DC: Inter-American Development Bank and World Bank. <http://hdl.handle.net/10986/25030>
- The elements of a randomized evaluation, JPAL  
<https://www.povertyactionlab.org/resource/elements-randomized-evaluation>
- Randomized Control Trials, DIME-World Bank  
[https://dimewiki.worldbank.org/Randomized Control Trials](https://dimewiki.worldbank.org/Randomized_Control_Trials)
- Kumar, N., Raghunathan, K., Arrieta, A., Jilani, A., & Pandey, S. (2021). The power of the collective empowers women: Evidence from self-help groups in India. World Development, 146, 105579.
- Ambler, Kate; and de Brauw, Alan. Cash transfers and women's agency: Evidence from Pakistan's BISP program. Economic Development and Cultural Change 72(3).



## Useful resources

- [Power size calculations-DIME-World Bank](#)
- [Power size calculations-JPAL](#)