United Nations University
World Institute for Development Economics Research

Session 9A.1:
Ex-Post Evaluation of TEs II – Concepts & Methods

Amina Ebrahim

6-8 December 2023 | Dar es Salaam, Tanzania

ATI Regional Follow-up Technical Meeting on Tax Expenditures: East Africa

UNU WIDER

Monitoring & Evaluation Fundamentals

Monitoring	Evaluation
Ongoing data collection	Time-specific
Internal	Can be internal or external
Informs ‘real-time’ decision making	Is the policy effective, impactful, efficient, relevant?
	Informs current and future policy-making

UNU WIDER

Impact evaluation

1. Studies the existence and *magnitude* of impact
2. Aims to identify the **causal** relation
3. Uses scientific methods

- Impact evaluation is an essential tool for evidence-based policy-making
- Evidence-based policymaking means using empirical knowledge and research-supported facts in planning and implementing policies
- Help tracing the **accountability**, **efficiency** and **impact** of policies
- Results can be useful for redesigning, comparing and improving policies

UNU WIDER

Components of impact evaluation:

1. Questions to be answered

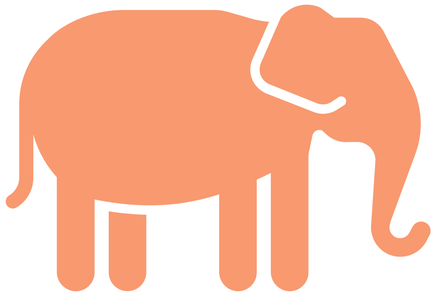
- Depend on the policy design and information needs.
- Examples:
  - Is the policy change increasing tax revenue?
  - What are the distributional impacts of the tax schedule?
  - Is the tax deduction impacting investment?

UNU WIDER

Components of impact evaluation

2. Outcomes of interest

- Outcomes we are interested in investigating → Impact indicators
- Examples:
  - Number of registered taxpayers
  - Amount of tax revenue
  - Sales → business activity
  - Number of formal employees



UNU WIDER

Components of impact evaluation

3. Evaluation design

- Impact evaluation uses scientific methods (Statistics and Econometrics) to estimate average impacts.

UNU WIDER

What is an indicator?

To evaluate policy = to measure change.

We need indicators

Indicators are signals of change

- What are we measuring

Examples: Taxpayer registration campaign

- More taxpayers → increase in taxpayer registration
- More taxpayers → increase in revenue collection

An indicator is a quantitative or qualitative factor or variable

- provides a simple and reliable means to measure achievement or to reflect the changes
- In principle, an indicator is anything that you can measure, and that reflects a change

Quantitative indicators

Measures of quantity common in research conducted with tax data

Expressed as numbers (units, proportions, rates of change, %, prices etc)

Information that can be counted and aggregated

UNU WIDER

Commonly used methods: An example

Consider the following situation:

- Program: Introduction of a tax incentive to all firms
- Objective: Increase the number of young people employed
- Question to be answered: How many additional jobs have resulted from the Tax Incentive?

Some commonly used methods to answer the question “What was the impact of the TE on the number of jobs?” are:

- Compare beneficiaries before and after the start;
- Compare beneficiaries and non-beneficiaries.

Are these methods capable of estimating the impact of the program on productivity?

Consider possibility 1: Compare beneficiaries before and after the program.

Suppose that:

- Before: Average number of employees of beneficiaries was 10;
- After: Average number of employees of beneficiaries was 12;

We can say that there was an **increase** in employees among beneficiaries over time, but we cannot guarantee that this increase was exclusively due to the TE:

- Other simultaneous factors may have contributed to this observed productivity increase (e.g. reduced electricity cuts, etc).

UNU WIDER

Consider possibility 2: Compare beneficiaries to non-beneficiaries.

Suppose that:

- Beneficiaries: Average number of employees is 15;
- Non-beneficiaries: Average number of employees is 15;

We can say that non-beneficiaries are the same, but we cannot guarantee that the program had *zero or negative impact*;

- Non-beneficiaries may be different from beneficiaries in relevant ways (e.g. different firm sizes, different sectors, etc).

Commonly used methods: Limitations!

Possibility 1. Comparing beneficiaries before and after will only give us the impact if productivity was not affected by any *other factors* besides the intervention.

- This is rare in practice.

Possibility 2. Comparing beneficiaries to non-beneficiaries will only give us the program impact if non-beneficiaries are equal to beneficiaries except for intervention participation.

- We must have information about beneficiaries and non-beneficiaries to investigate whether this comparison is adequate.
- It is not always possible to guarantee that these two groups are similar in all relevant characteristics, as some might not be observable and will be correlated to both program participation and outcomes (e.g. firm's motivation, firm size, firm sector).

What is causality?

Impact evaluations identify ***causal relationships.***

- Main question: what is the impact (causal effect) of a TE on the outcome of interest?

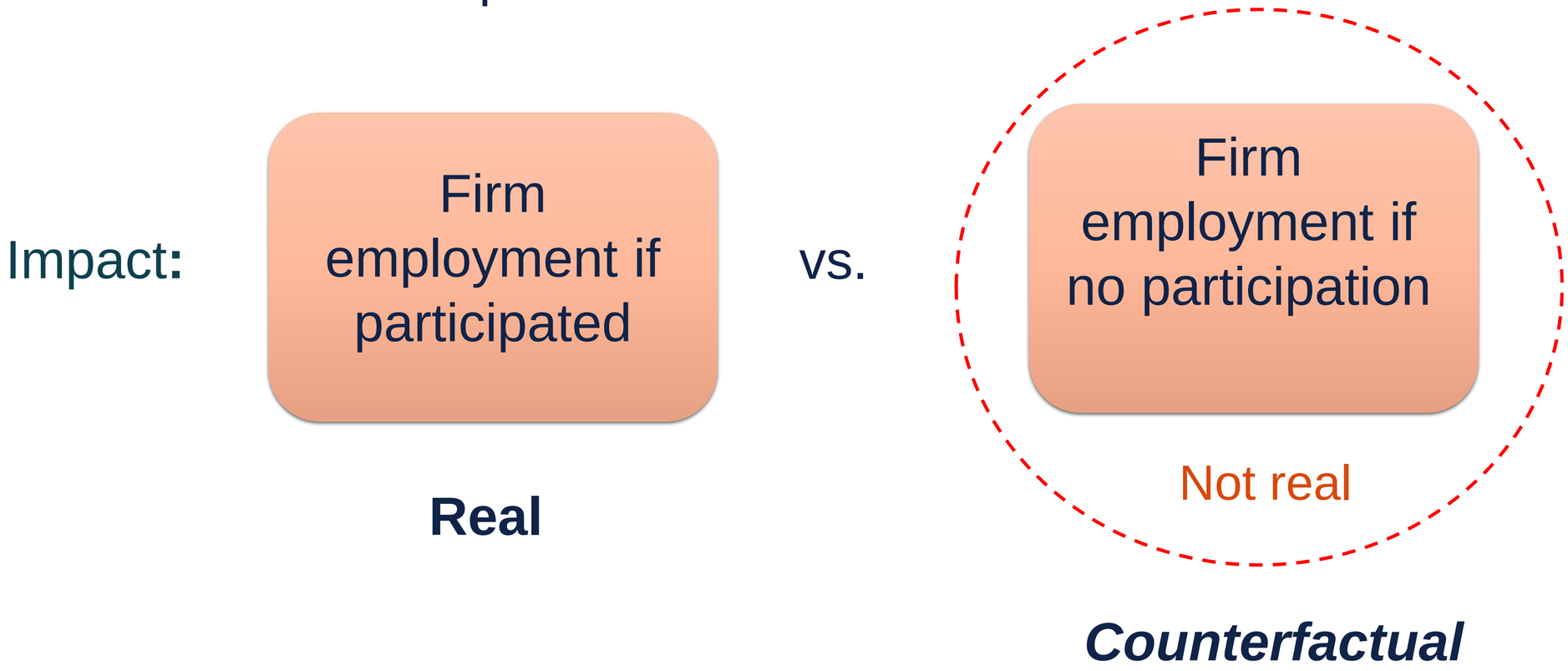
***Causality* implies that:**

- We can attribute observed changes in the outcome of interest to the TE.
- In the absence of the TE, such changes in the outcome of interest would not have been observed.

UNU WIDER

How do we establish causality?

Example: to estimate the impact of a tax incentive on firm employment we would like to compare beneficiary firms in two different states after the intervention has taken place:



UNU WIDER

However, *it is impossible to observe two different states simultaneously*.

Selection of a comparison group

To estimate the counterfactual, we will use a ***comparison group*** that mimics what would have happened to participants without the tax incentive.

An adequate comparison group is a key element for any impact evaluation!

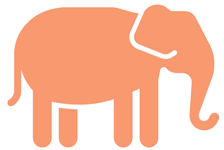
UNU WIDER

Difference-in-Differences: When to Use It

- In many cases, the assignment rules of the program are not as clear or as well defined as discussed in the previous approaches.

- If that is the case, approaches such as experimental evaluation, regression discontinuity design, and instrumental variables will not be feasible.

- If baseline **data** is available, one possibility is to use *difference-in-differences*.



Source: Gertler et al (2011).

UNU WIDER

Difference-in-Differences: When to Use It

The ***difference-in-differences*** approach requires stronger assumptions and is only feasible if baseline data is available for both participants and non-participants.

- Baseline data must be from a pre-intervention period.

This method does not require specifying treatment assignment rules.

In this method, we compare the changes in outcomes (impact indicators) over time (before and after) between a treatment group (participants of the intervention) and a control group (non-participants of the intervention).

- This is the reason why the method is called *difference-in-differences*.

UNU WIDER

Difference-in-Differences: Example

Consider the following situation

Program: Introducing a new Income Tax rule

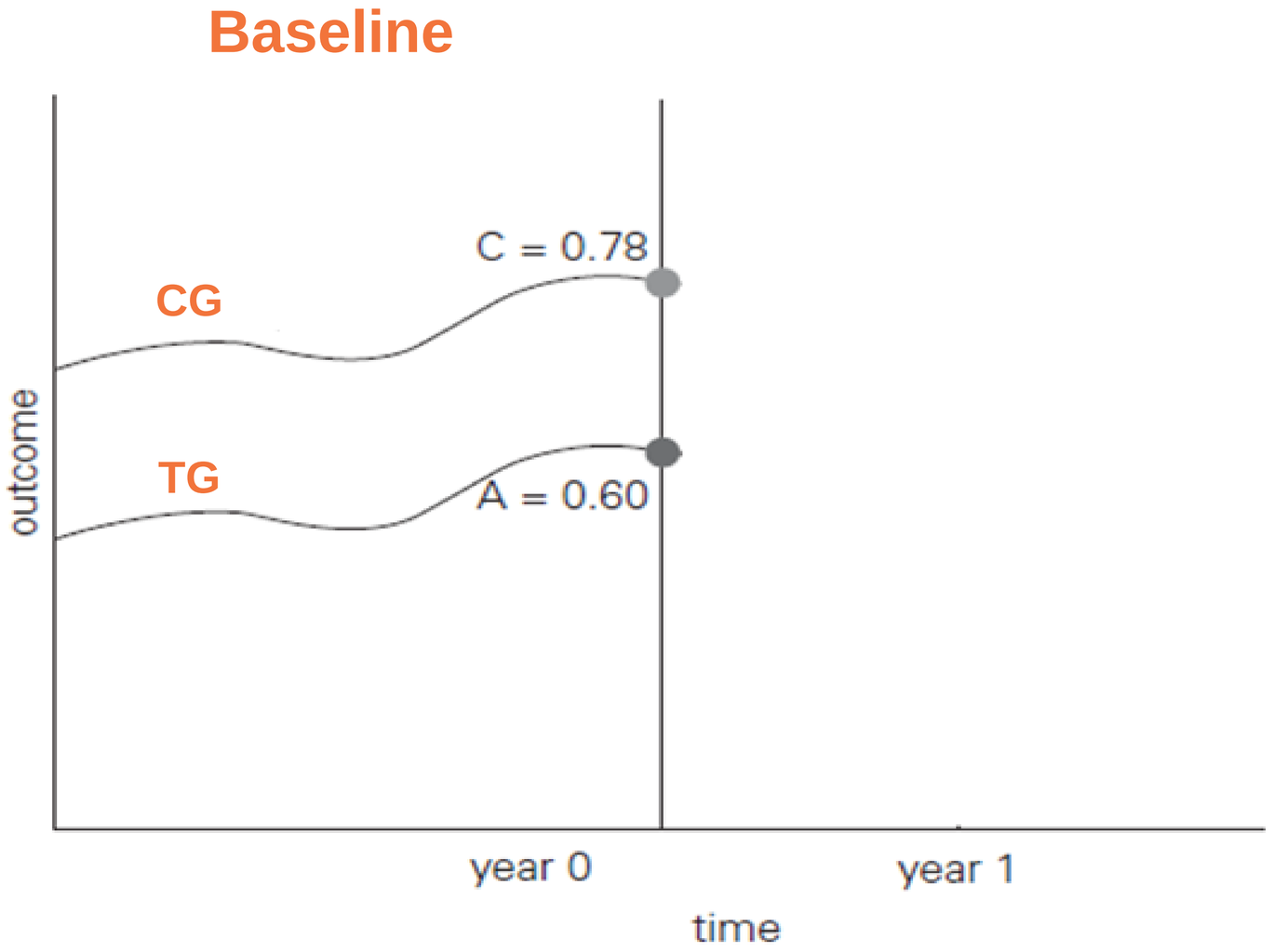
Objective: increasing government revenue by increasing tax rates for richer individuals

Eligible units: Individuals with income different from zero

Selection to participate: Income higher than a specific threshold

Impact evaluation objective: estimating the impact of the program on government revenues

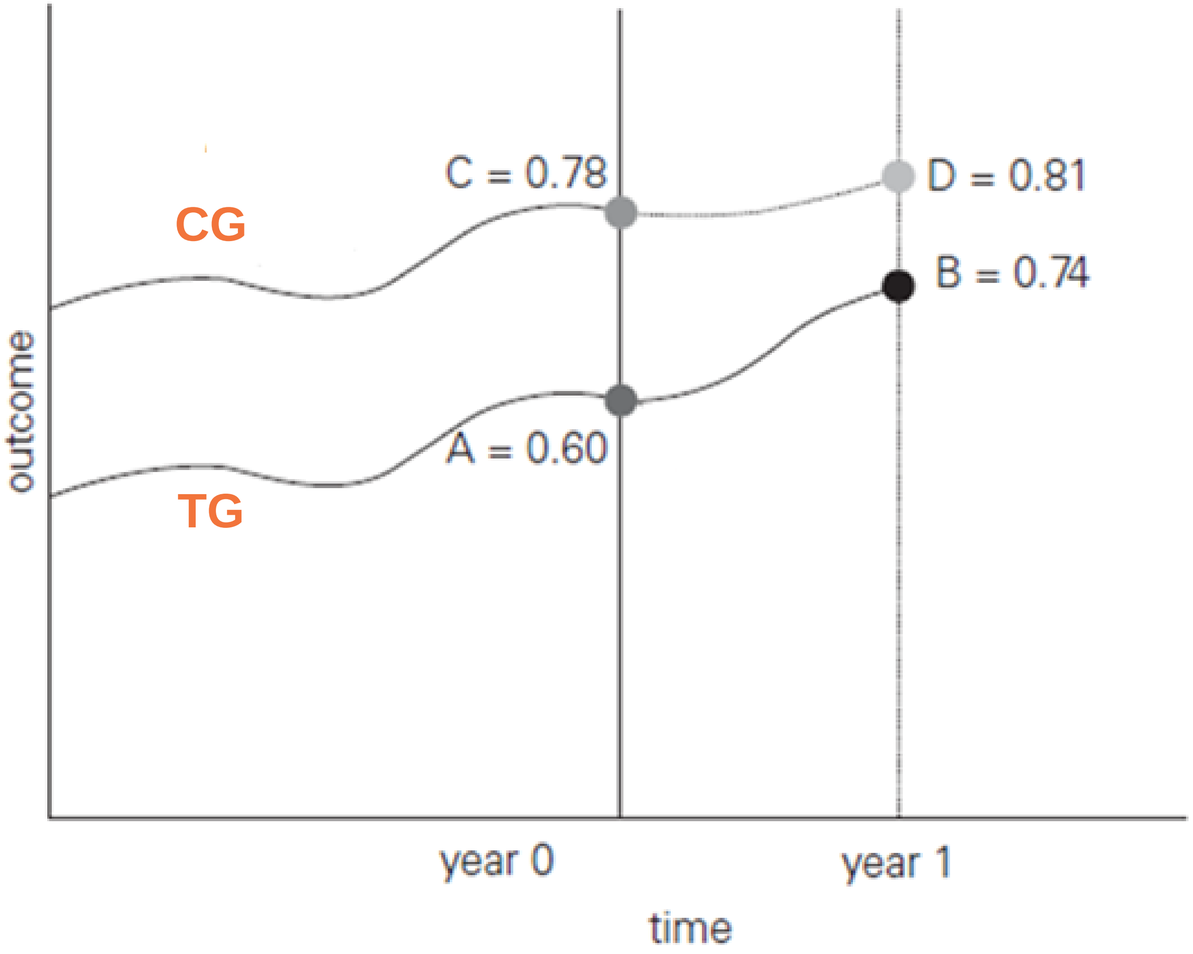
Difference-in-Differences: Example



Source: adapted from Gertler et al (2011).

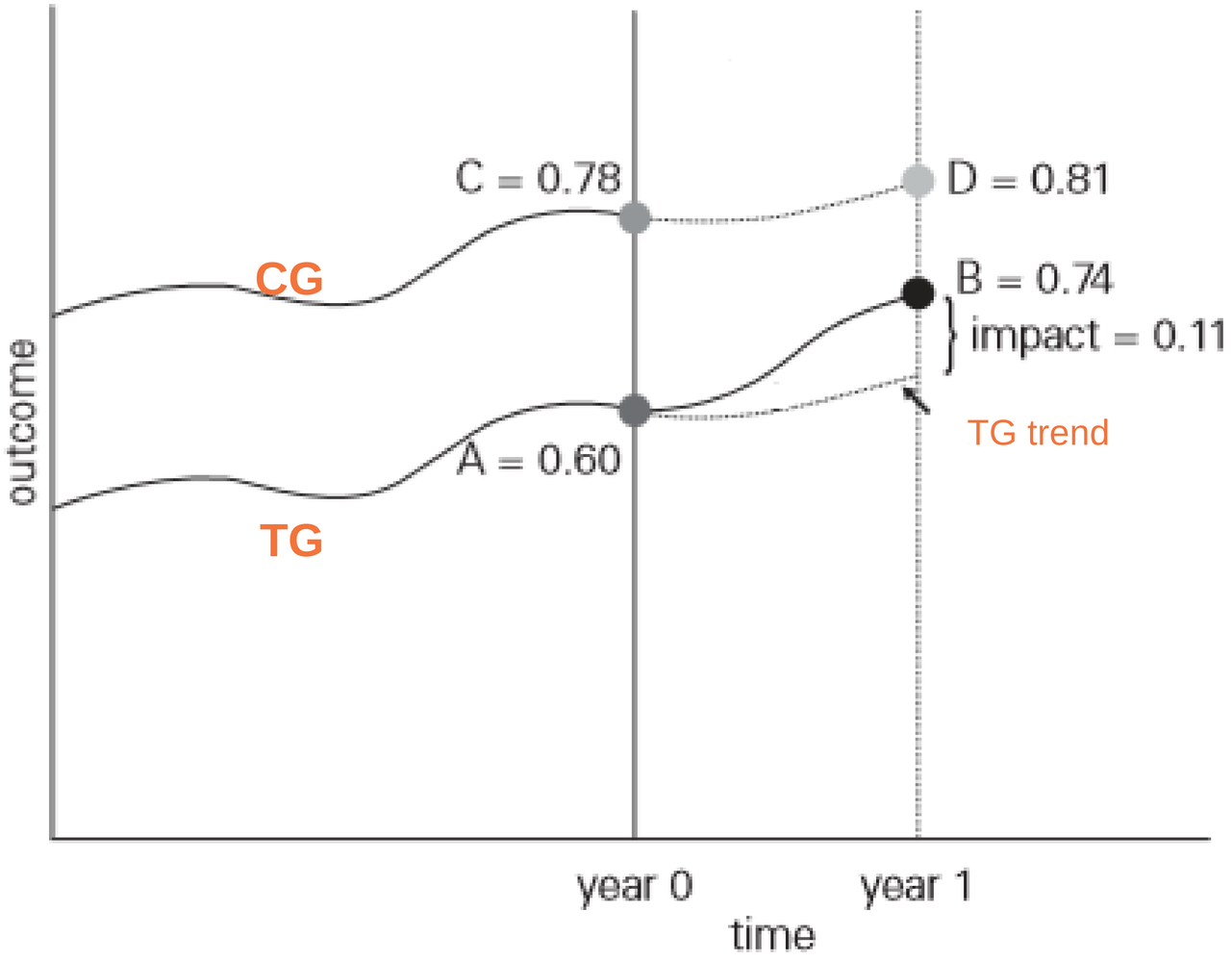
Difference-in-Differences: Example

Baseline and Endline



Source: adapted from Gertler et al (2011).

Difference in Difference: Example calculation



Impact calculation

Impact = (D-B) – (C-A)
= (0.81-0.74) – (0.78-0.60)
= 0.14 – 0.03
= 0.11

Positive Effect!

UNU WIDER

Difference-in-Differences: Assumptions

1

In the absence of the intervention, outcomes of treatment and control groups would display equal trends

If true, changes in outcomes over time for the control group represent the changes in outcomes that would have been observed for the treatment group in the absence of the intervention.

2

Treatment and control groups should not be affected in different ways by other factors at the same time as the intervention.

If this does not hold, impact estimation may be biased.

3

Group composition (treatment and control) should not change significantly over time (before and after the intervention).

UNU WIDER

Difference-in-Differences: Limitations

- The equal trends assumption must hold for the estimated impact to be unbiased. It is a strong assumption and its validity cannot be tested.
- If other simultaneous factors are present and affect the difference in trends between treatment and control groups, impact estimation using this approach will be invalid or biased (Gertler et al, 2010).

Recap

- Need to be clear about the indicators/outcomes
- Difference in Differences can offer a rigorous evaluation method
- Important to consider the comparison group carefully.

UNU WIDER

Thank you!

Amina Ebrahim
Research Fellow | UNU-WIDER
amina@wider.unu.edu

UNU WIDER