Loyola

Impact Evaluation

Impact Evaluation

Impact evaluation aims to quantify and measure the effects or impacts of a program, policy, or intervention on the population of interest. By conducting an impact evaluation, evaluators seek to determine the extent to which the intervention has achieved its intended outcomes and assess the overall impact on the target population.

The evaluation process involves collecting and analyzing data to assess the changes that can be attributed to the intervention. This requires comparing the outcomes of the intervention group, which received the program or intervention, with a control group or a counterfactual scenario that did not receive the intervention. By comparing these two groups, evaluators can isolate the effects of the intervention and estimate its impact.

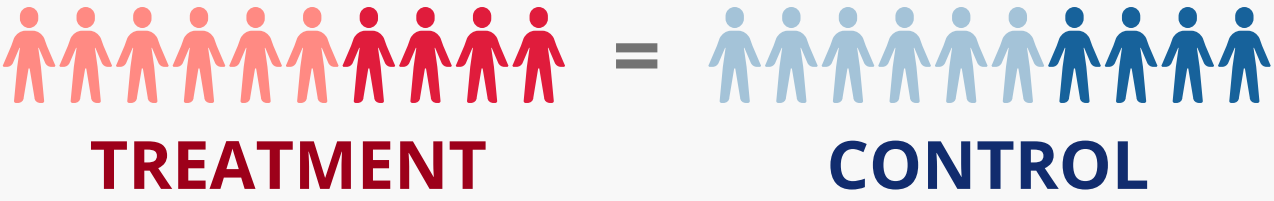
The ultimate goal of impact evaluation is to provide evidence-based insights into the effectiveness and efficiency of interventions, allowing policymakers, program managers, and stakeholders to make informed decisions. By isolating the effects of the intervention, impact evaluation provides evidence-based insights into its effectiveness and efficiency, helping decision-makers make informed choices. The evaluation also considers unintended consequences and assesses the overall impact on the target population.

In every impact evaluation, there are **three essential elements:**

1. **Intervention:** This refers to a specific action or policy implemented to bring about a change. The objective of the intervention is to address and improve an unsatisfactory problem or situation. We define a problem as a situation that raises concern when an indicator shows that the value of a variable is above or below what is considered acceptable.
2. **Outcome variables:** The indicators that raise alarms and lead us to promote interventions are the same ones we must use to measure the intervention. These variables determine the success of the program.
3. **Target population:** It is the defined population for which the intervention is designed and intended to benefit. The target population refers to the specific group of individuals, households, or entities that a program, policy, or intervention aims to reach or impact. The characteristics and criteria used to define the target population may vary depending on the objectives and nature of the intervention. These criteria could include demographic factors (age, gender, income), socio-economic status, geographic location, specific needs or vulnerabilities, or other relevant factors.

Impact evaluation goes beyond a simple measurement of the program's effect before and after implementation. It involves a counterfactual analysis, which means comparing the outcomes achieved by the group that received the program (treated group) with the outcomes of a group that did not receive the intervention (control group). This comparison helps determine the program's causal effect by creating a hypothetical scenario that simulates what would have happened to the beneficiaries if they had not received the program. This approach helps isolate the program's contribution and assess its effectiveness.

Therefore, the core idea in impact evaluation is to compare the outcomes of a group of individuals who receive the program with another group that does not receive it but is on average similar to the treated group. Obtaining a group of "twins" or similar individuals is not always easy, and therefore, various statistical and econometric techniques are used to simulate this control group. The aim is to ensure that the treated and control groups are identical in both observable and unobservable characteristics.



Beneficiaries selection

As mentioned before, to properly evaluate a program, we need a control group, meaning a group of households or individuals that are similar to the treatment group.

Finding a control group is not always easy. There are different techniques to do it and depends on how beneficiaries are selected. We are going to mention the most common:

- **Random Selection:** In this method of allocation, individuals or groups are selected randomly to form either the treatment group or the control group. The aim is to ensure that any differences between the groups are due to chance, providing a strong basis for causal inference.
- **Scoring:** In this method of allocation, specific criteria or scores are used to determine which individuals or groups should be included in the treatment group. The idea is to identify those who are most likely to benefit from the intervention based on certain characteristics or conditions.
- **Phases:** Interventions in phases refer to situations where a program or policy is implemented gradually over time, targeting different subsets of the target population in different stages or phases. Instead of applying the program to the entire target population at once, it is rolled out incrementally to specific groups or areas. The first group or phase to receive the treatment will serve as the treatment group, and the subsequent groups or phases that receive the treatment later will serve as the control group until they receive the intervention.
- **Open:** In open allocation, the treatment and control groups are not assigned or determined by any specific criteria or selection process. Instead, individuals or groups self-select into either the treatment or control group based on their preferences or circumstances.

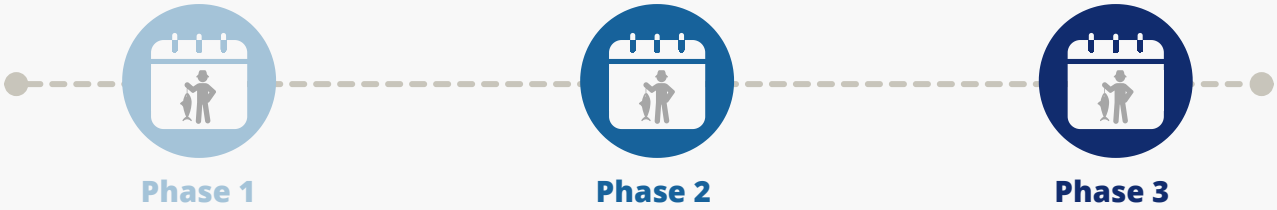
Let's imagine an impact evaluation to assess the effectiveness of using specialized nets that reduce bycatch in helping to minimize overfishing and protect sea life.

Random Allocation:
For the random allocation approach, researchers or policy makers would randomly select a representative sample of fishing vessels or fishing crews from a larger population of fishermen who currently use regular fishing gear. The selected vessels or crews would be provided with the specialized nets as the treatment group, while the remaining vessels or crews continue using their regular fishing gear and form the control group. The random allocation ensures that any differences between the two groups are due to chance, providing a solid basis for causal inference.

Scoring allocation:
Researchers could use the scoring method to select fishing vessels or crews based on specific criteria. The vessels or crews with the highest scores would receive the specialized nets as the treatment group, while the others continue with regular gear as the control group.

While fishermen who score higher on the assessment clearly differ from those at the bottom, the situation becomes more nuanced near the cutoff for selection. For example, the fishermen who just made the cutoff (positions 901 to 1000) may not significantly differ from those on the waiting list (positions 1001 to 1100). Random luck may have played a critical role in their selection, making it plausible that some on the waiting list possess the same capabilities. Thus, we assume equivalence between subjects in the ranges of 901-1000 and 1001-1100, these are the groups we will use as control and treatment. This methodology is referred to as regression discontinuity (RD) in impact evaluation.

Allocation in Phases:
In the impact evaluation for specialized nets to reduce bycatch and protect sea life, the allocation in phases is chosen due to limited funds available for the intervention's implementation. By adopting a phased approach, different subsets of fishing vessels can be selected over time to receive the specialized nets, allowing for a more manageable and cost-effective implementation. In each phase, a subset of fishing vessels will be randomly selected to receive the specialized nets (treatment group), while the remaining vessels continue using regular gear as the control group.



Open allocation:
In this open allocation approach, all fishermen in the community would be informed about the availability of specialized nets designed to reduce bycatch and promote sustainable fishing practices. Fishermen interested in participating would have the option to voluntarily adopt and use the specialized nets, becoming part of the treatment group. Others who choose not to use the nets would form the control group.

The random allocation scenario is the easiest to evaluate. When an external mechanism, such as a lottery, is used to assign beneficiaries and non-beneficiaries, there are no distinguishing household (or individual) characteristics that differentiate the treated group from the non-treated group. This means that the treatment and control groups are identical before the intervention occurs. In this situation, we have an immediate "twin" or comparison group, allowing us to employ the most rigorous impact evaluation technique: randomized controlled trials (RCTs).

There is a particular case where interventions naturally create twins without any explicit mechanism. Let's consider a coral reef preservation intervention organized in phases due to limited funds. The goal is to protect coral reefs in different locations along the coast. In Phase 1, certain coral reef areas are selected for active preservation efforts, including implementing regulations to prevent destructive fishing practices and promoting sustainable tourism. These chosen areas constitute the treatment group.

Conversely, in Phase 2, other coral reef sites are temporarily excluded from the intervention due to budget constraints, forming the control group. The allocation of sites to each phase is based on logistical and financial considerations, ensuring that the selected sites in both phases have similar ecological characteristics.

Since the decision to include coral reef sites in either phase is not influenced by the specific attributes of each site, we can assume that all the coral reefs in the evaluation are similar in their baseline conditions. As a result, this phased implementation creates a control and treatment group similar to a randomized controlled trial (RCT) without explicit randomization.

Finally, there are interventions where the criteria for allocating treatments differ from the previous cases, making it more difficult to identify a suitable twin for the treatment-control comparison. In such situations, the strategy is to find a twin that closely resembles the treated group as much as possible. To achieve this, statistical techniques are employed to identify households (or individuals) with characteristics that closely match those who have received the intervention. This technique is called propensity score matching (PSM).

Impact evaluation techniques

1. Randomized Controlled Trials (RCTs) are a type of experimental study design used in scientific research to evaluate the effectiveness of an intervention or treatment. RCTs are considered the gold standard for establishing causal relationships between an intervention and its outcomes. They are widely used in various fields, including medicine, public health, social sciences, and education.

Key Features of Randomized Controlled Trials:

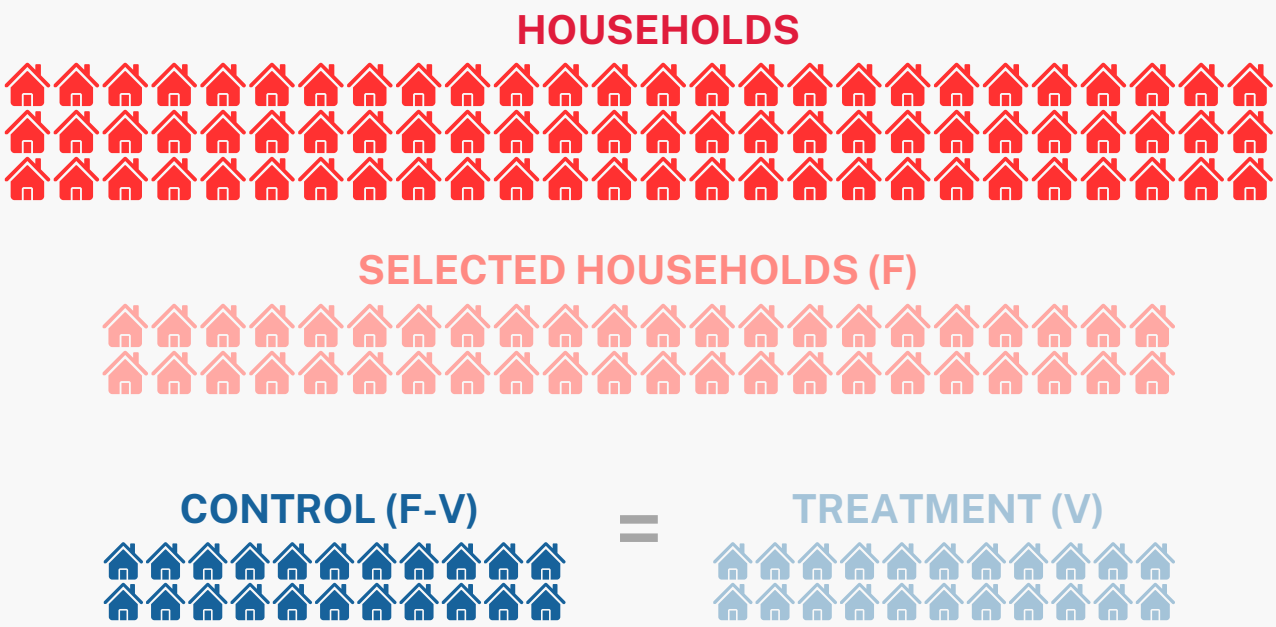
- Randomization: In an RCT, participants or subjects are randomly assigned to either the treatment or control groups. Randomization ensures that each participant has an equal chance of being assigned to either group, which helps create comparable groups and minimizes selection bias. The random allocation helps ensure that any differences in outcomes between the groups are more likely to be due to the intervention being tested rather than other confounding factors.
- Treatment Group: The treatment group receives the intervention being studied. It is the group that receives the new program or any other intervention being tested for its effectiveness.
- Control Group: The control group serves as a comparison to the treatment group. It does not receive the intervention but is treated similarly in all other respects. The control group helps establish a baseline for comparison, allowing researchers to determine whether the intervention has a significant impact compared to the absence of the intervention.

Let's use the example of a "Zero Waste" program, where a country aims to promote waste reduction and environmental sustainability by encouraging households to adopt zero waste practices. The program involves providing incentives to a select number of households to implement zero-waste strategies, and a randomized control trial will be used to evaluate its impact.

To conduct a randomized control trial, a pool of households from a specific region or community would be identified. These households should represent a diverse range of demographics, waste generation habits, and attitudes toward waste management.

From the pool of identified households, a random selection would be made to participate in the "Zero Waste" program. The selected households (called F) would then be divided into two groups: the treatment group and the control group.

- Treatment Group: A certain number of households (let's call it V) would be randomly assigned to the treatment group. These households would receive incentives and support from the "Zero Waste" program to implement zero-waste practices in their daily lives.
- Control Group: The remaining households (F - V) would be part of the control group. They would not receive any special incentives or support related to the "Zero Waste" program and would continue with their usual waste



2. Regression Discontinuity (RD) is a research design used to estimate the causal impact of an intervention or treatment when the assignment to the treatment group is determined by a cutoff point or threshold on a continuous variable. In RD, individuals or units just above or below the cutoff are compared to each other, assuming that they are similar except for their proximity to the threshold. This design allows researchers to draw causal inferences in situations where random treatment assignment is not feasible or ethical.

Key Features of Regression Discontinuity:

- Cutoff Point: The key element of RD is the existence of a cutoff point or threshold on a continuous variable. Individuals or units with a score or value just above or below the cutoff are subject to different treatment conditions.
- Treatment and Control Groups: In an RD design, individuals or units just above the cutoff point are assigned to the treatment group, while those just below the cutoff point form the control group. The assumption is that individuals close to the cutoff are comparable in all other aspects except for their treatment status.

In the "Zero Waste" program will be evaluated using a Regression Discontinuity (RD) design to assess its impact on waste reduction and sustainable practices.

Let's assume that households will be eligible to participate in the program if they achieve a certain score on a "Zero Waste Index" that measures their current waste reduction practices and environmental sustainability efforts. Households with a Zero Waste Index score just above the cutoff point will be assigned to the treatment group, while those with a score just below the cutoff point will be assigned to the control group.

3. Interventions in phases refer to situations where a program or policy is implemented gradually over time, targeting different subsets of the population in different stages or phases. Instead of applying the program to the entire population at once, it is rolled out incrementally to specific groups or areas.

There are two common ways in which interventions in phases are carried out:

- **Random Allocation:** In this approach, the timing of the program implementation is determined randomly. The program might randomly select certain regions or communities to receive the treatment or intervention in the first phase, then different regions in the second phase, and so on. By randomly selecting the groups to receive the intervention at each phase, researchers can create comparable treatment and control groups, making it easier to measure the program's impact.

- **Events or Fortuitous Circumstances:** In some cases, the decision of which population or area receives the program in each phase is driven by an external event or a fortuitous circumstance. In this case, the occurrence of the event determines the timing of the intervention, and the affected area serves as the treatment group while other unaffected areas serve as the control group.

The goal of interventions in phases is often to manage logistical challenges, budget constraints, or external factors that make it difficult to implement the program across the entire population simultaneously. By implementing the program in stages, policymakers and researchers can better manage resources, assess the program's effectiveness, and adjust the approach based on early findings before expanding it to other groups or areas.

In the "Zero Waste" program due to budget constraints or logistical challenges, it is not feasible to implement the program for all households in the country simultaneously. Therefore, the program is carried out in phases, targeting different subsets of households at different stages. The first group or phase to receive the treatment will serve as the treatment group, and the subsequent groups or phases that receive the treatment later will serve as the control group until they receive the intervention.

4. **Propensity Score Matching (PSM)** is a statistical technique used in observational studies and program evaluations to estimate the causal effect of a treatment or intervention. It helps address the issue of selection bias by creating comparable treatment and control groups, even when random assignment is not possible.

The basic idea behind PSM is to estimate a propensity score for each individual in the dataset. The propensity score represents the probability of receiving the treatment (being in the treatment group) given a set of observed characteristics (covariates) of the individual. These observed characteristics could include demographics, socioeconomic status, pre-treatment outcomes, or any other variables that may influence both treatment assignment and the outcome of interest.

Key Features in Propensity Score Matching:

- Propensity Score Estimation: Using logistic regression or other statistical methods, the propensity score is estimated for each individual in the dataset. The logistic regression model predicts the probability of being in the treatment group (receiving the intervention) based on the observed covariates.
- Matching: Once the propensity scores are estimated, individuals in the treatment group are matched with those in the control group with similar or identical propensity scores. The goal is to create pairs (or groups) of treated and control subjects with similar probabilities of receiving the treatment, thereby creating comparable groups.
- Assessing Balance: After matching, it is essential to assess the balance between the treatment and control groups in terms of the covariates. The matched groups should have similar distributions of observed characteristics, indicating that they are now comparable in terms of these variables.

In the "Zero Waste" program, some households participate in the program (treatment group) and receive incentives to adopt zero waste practices, while others do not participate (control group). The challenge is that households self-select into the program based on their interest and willingness to adopt zero-waste practices, which could lead to biased results.

To address this issue, we can use Propensity Score Matching:

We estimate the propensity score for each household, which represents the probability of a household being part of the treatment group (participating in the "Zero Waste" program) based on observable characteristics such as household size, income level, education, location, etc. Next, we match households in the treatment group with households in the control group that have similar propensity scores. After matching, we assess whether the treatment and control groups have similar distributions of observable characteristics. If the matched groups are balanced, it means we have created comparable groups with similar characteristics.

In short, by using Propensity Score Matching, we reduce the bias introduced by households self-selecting into the program. It helps us create more comparable treatment and control groups, making the evaluation of the "Zero Waste" program's effectiveness more reliable and trustworthy.

Further issues

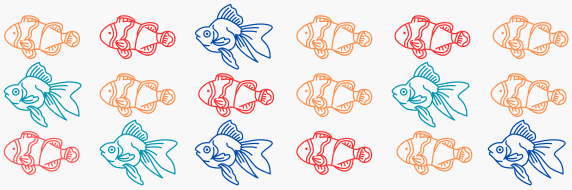
Firstly, it is common for interventions to impact a large population, and the required sample size for impact evaluation is typically smaller than the population of beneficiaries. However, there may be cases where an intervention is small-scale, resulting in a small number of beneficiaries. In such cases, even the entire population of beneficiaries may not be sufficient to achieve the required sample size. This can lead to statistical issues due to low power; therefore, extreme caution must be exercised when drawing conclusions.

Secondly, it is crucial to consider that the optimal sample size mentioned earlier assumes that the random selection unit is the individual or household. However, there are situations where the selection level is not the household but rather clusters or groups.

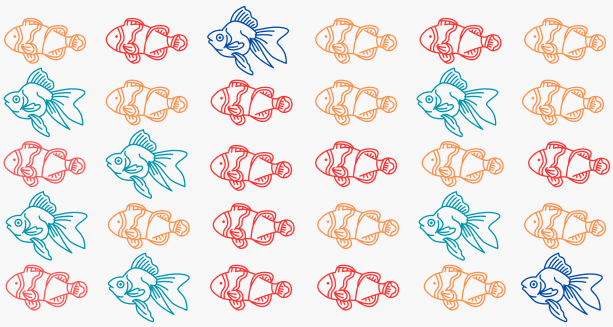
a) **Heterogeneous Effects:** In some cases, treated individuals may not fully benefit from the intervention due to environmental or contextual factors. This can lead to heterogeneous effects among the treated individuals, where the impact of the intervention varies across the treated group. It is important to account for these variations and understand the potential limitations in evaluating the overall impact.

In an impact evaluation of a marine life conservation program, researchers implement a strategy to protect different marine species from overfishing and habitat degradation. The intervention includes creating marine protected areas (MPAs) and implementing stricter fishing regulations, including a fishing ban during the fall season, to protect marine species from overfishing. The ban aligns with the reproductive season of "Species A," a fish species that reproduces during the fall.

BEFORE



AFTER



Due to the fishing ban during their reproductive season, the population of "Species A" shows a remarkable recovery within the marine protected areas (MPAs). The ban allows them to spawn and reproduce without disturbance, leading to an increase in their numbers.

On the other hand, "Species B" reproduces during the summer, and the fishing ban during the fall does not directly benefit their reproductive cycle. As a result, the impact of the conservation program on "Species B" is not as significant as it is for "Species A." While the stricter regulations might still provide some benefits to "Species B" by reducing overall fishing pressure, it does not specifically target their reproductive season.

	Species	Before	After	Increase of population
	A	12	22	83%
	B	6	8	33%

This difference in the impact of the conservation program on the two fish species exemplifies heterogeneous effects. The intervention has varying outcomes for different marine species based on their unique life cycles and environmental dependencies. Understanding these variations is crucial in designing effective conservation strategies that consider the specific needs and vulnerabilities of each species. Policymakers might consider additional measures or targeted interventions to ensure the preservation of all marine species and maintain the ecological balance of the marine ecosystem.

b) Spillover Effects or Contamination: There may be situations where untreated individuals who are connected or interact with treated individuals also experience some effects of the intervention. This is known as a spillover effect. It is important to consider the possibility of such effects and assess their impact on the evaluation results.

The existence of spillover effects is a paradoxical issue. On one hand, if there are spillover effects, it means that more people benefit from the intervention, making it a clear success in terms of public policy (since more people benefit from the intervention at the same cost). However, on the other hand, it poses a challenge for impact evaluation because it makes the evaluation more complicated. The control group will no longer be purely untreated but rather partially treated due to the spillover effects. In some cases, the evaluation will become inconclusive, meaning that it fails to show a statistically significant difference between the treatment and control groups.

In the impact evaluation of the marine life conservation program, the researchers chose a control group located a few kilometers away from the marine protected areas (MPAs) where the intervention was implemented. The control group was intended to represent a similar marine environment but without the strict fishing regulations and fishing ban during the fall season.

However, as a spill-over effect, the recovery of "Species A" and "Species B" within the MPAs has been so successful that some individuals from these species have migrated to the control group area. The improved conditions in the MPAs, such as increased food availability and reduced fishing pressure, have attracted fish from neighboring areas, including the control group location.

As a result, the control group area has become contaminated with fish from the MPAs, leading to changes in the ecosystem dynamics and fish populations. The presence of "Species A" and "Species B" in the control group area may lead to increased competition for resources or changes in predator-prey relationships, affecting the abundance and distribution of other marine species in the control area.

Understanding and addressing spill-over effects is crucial in impact evaluations, as it helps researchers draw more accurate conclusions about the effectiveness of conservation programs and the broader implications for marine ecosystem management.

When to collect data?

The effects of interventions may take a long time to manifest, and in such cases, one year would not be sufficient. Once the control group has been identified, an important question is when and how often to collect information. The most appropriate way to conduct a proper impact evaluation is to have at least two measurements: one before the intervention and another after the intervention.

When conducting an impact evaluation, it is crucial to collect data before the intervention takes place. However, the timing of data collection is not simply a matter of capturing information prior to the start of the program. It is also important to ensure that the individuals being evaluated are unaware of their potential eligibility for the intervention.

This consideration is significant because individuals who are aware of their eligibility may exhibit certain biases that can affect the evaluation results. For example, eligible individuals who believe they will be selected may have high expectations and anticipate positive outcomes, leading them to provide overly optimistic data. On the other hand, those who are eligible but not ultimately chosen may feel a sense of unfair treatment, which could negatively influence their responses.

To address these biases, it is essential to choose a data collection time that not only precedes the implementation of the intervention but also occurs before the selection or pre-selection of beneficiaries.

The delayed effects of interventions can be attributed to various factors, outcomes may require a significant amount of time to manifest. By conducting a third round of data collection, several years after the intervention, evaluators can capture the longer-term effects and assess whether the desired outcomes have been sustained over time. This additional evaluation allows for a more comprehensive understanding of the intervention's impact and helps determine its long-term effectiveness.

Therefore, when evaluating interventions, it is crucial to consider the potential time lag between the completion of the intervention and the appearance of significant effects. Multiple rounds of data collection over an extended period allow for a more accurate assessment of the intervention's impact and its sustainability in the long run.

How to compute the effect?

The difference-in-differences (DiD) method is a statistical technique that compares changes in outcomes over time between two groups: a treatment group that receives an intervention or treatment, and a control group that is similar to the treatment group in all relevant characteristics except for the treatment.

The DiD method aims to estimate the causal effect of the intervention or treatment by controlling for other factors that may be affecting the outcome. This is achieved by comparing the changes in outcomes over time in the treatment group to those in the control group.

To implement the DID method, researchers typically collect data on both the treatment and control groups before and after the intervention. They then calculate the difference in outcomes between the pre-and post-intervention periods for each group and subtract the difference in outcomes in the control group from the difference in outcomes in the treatment group. This difference in differences provides an estimate of the causal effect of the intervention or treatment, net of other factors that may be influencing the outcome.

As said, in a DiD analysis, the value of the outcome variable (y) is taken before (pre) and after (post) the treatment. This measure is taken for both the treated or treatment group (T) and the untreated or control group (C), resulting in:

- Difference before-after for the treatment group: $Dif_1 = y^t_{post} - y^t_{pre}$
- Difference before-after for the control group: $Dif_2 = y^c_{post} - y^c_{pre}$

The difference between Dif_1 and Dif_2 is, therefore, the effect of the program.

$$DiD = Dif_1 - Dif_2$$

Therefore, this DiD measures how much the treatment group has changed compared to the control group. If the difference is positive (DiD > 0), it means that the impact of the intervention on the studied variable has been greater in the treatment group than in the control group. This indicates that the intervention has been truly effective in the group it was intended to target.

Number of observations

A key question in any evaluation is determining the sample size. We know that small samples can lead to erroneous conclusions, but we also understand that large samples come with high costs. Therefore, we need a sample size that is large enough to provide informative results while being affordable. In other words, we want to avoid spending more money than necessary.

In an impact evaluation, our goal is to test whether the effect we observe from the program is zero or different from zero (either positive or negative). Using various statistical techniques, we compare:

- The hypothesis that there is no effect (**null hypothesis**) versus
- The hypothesis that the effect is different from zero (**alternative hypothesis**).

If we reject the null hypothesis, we can conclude that there is an effect. Conversely, if we fail to reject the null hypothesis, we can conclude that there is no effect. The effect will be a positive or negative number which we will call μ.

However, tests are not perfect and can generate errors, in other words, they may fail to detect the effect we are looking for. There are two possible errors: Type I and Type II Errors.

- **Type I error**, also known as a false positive, happens when the test incorrectly indicates the presence of an effect that doesn't exist. It is conventionally set at a 5% significance level ($\alpha = 0.05$), meaning there is a 5% chance of falsely detecting an effect.

- **Type II error**, also known as a false negative, occurs when the test fails to detect an actual effect. The acceptable percentage for Type II error can vary between 10% and 20% ($\beta = 0.10\text{-}0.20$). This error can be reduced by increasing the sample size or redefining the desired effect size (μ).

In summary, in hypothesis testing, Type I error refers to falsely detecting an effect that doesn't exist, with a 5% significance level. Type II error refers to failing to detect an actual effect, with an acceptable range typically between 10% and 20%. The sample size and the desired effect size (μ) play important roles in controlling these errors. Larger effects are easier to detect, while smaller effects require larger sample sizes. Increasing the sample size helps reduce Type II error, but it is inversely related to the actual effect size.

Overall, when choosing a sample size, it is crucial to consider the actual effect size we want to detect (μ) and the acceptable level of Type II error we can tolerate. If we conduct impact evaluations to calculate effects, it means that we do not know the size of the effects. Therefore, the key question is what to do when we do not know the true effect of a project and we conduct an impact evaluation to estimate it.

In such cases, we employ a technique called power calculations. These calculations involve assuming an acceptable Type II error rate and analyzing the minimum detectable effect (MDE) across various sample sizes. The objective is to determine the smallest effect size that can be reliably detected.

The minimum detectable effect is quantified in standard deviations (SD) and signifies the minimal change anticipated between the treated group and the control group. Statistical power is defined as the probability of correctly identifying a program's effect when it actually exists, whereas Type II error (β) refers to the likelihood of erroneously stating that the effect is null when it is not. Consequently, statistical power is equal to 1 minus the Type II error rate ($1-\beta$). For instance, a 20% Type II error would correspond to a statistical power of 80%.

Within the impact evaluation literature, it is customary to set the statistical power (1-β) between 80% and 95%. However, it is essential to not only determine the desired power but also specify the magnitude of the impact being sought.

The subsequent table integrates minimum detectable effects (MDE) and varying levels of acceptable statistical power. It provides information on the necessary number of observations to conduct the impact evaluation, for both control and treatment groups.

MDE	Statistical power		
	80%	90%	95%
0.1 SD	3200	4350	5200
0.2 SD	790	1055	1305
0.3 SD	360	470	580
0.4 SD	200	270	330
0.5 SD	130	170	210

Aiming for the lowest MDE and high statistical power in an impact evaluation study requires a sample size of 5,200 participants or more. Similarly, impact evaluations conducted with small samples (for instance < 500) would be useful only for "large" expected impact (or would suffer enormous type II errors)

To conclude

Impact evaluations are a valuable method for assessing the effectiveness of programs or interventions. They aim to measure the causal impact or effect that a specific program has on its beneficiaries, compared to a group of individuals who did not receive the program.
These evaluations are typically conducted after the program has been implemented (ex-post), and they employ various statistical techniques to estimate the program's impact. One common approach is the DiD method, which compares the changes in outcomes between the treatment group and the control group.

To create a control group that mimics the absence of the program, evaluators often select individuals who are similar in characteristics to the beneficiaries. This process involves creating a "twin" or control group that represents what would have happened to the beneficiaries in the absence of the program. By comparing the outcomes of the treatment group with those of the control group, the impact of the program can be estimated.
It is important to note that every methodology has its limitations, and as we move away from random assignment, finding a twin or control group becomes increasingly challenging. This results in lower precision and consistency of the estimator of the program's effect.

Randomized Controlled Trials are considered the gold standard for impact evaluations because they use random assignment to create treatment and control groups. This ensures that any differences observed between the groups can be attributed to the program. However, in practice, it may not always be feasible or ethical to conduct RCTs.

As the resemblance between the treatment and control groups diminishes, the precision and consistency of the program's effect estimator decrease. This introduces uncertainty into the impact evaluation, making it more challenging to draw definitive conclusions about the program's effectiveness.

Despite these limitations, impact evaluations remain a valuable tool for assessing program impacts and informing decision-making. Researchers and evaluators strive to employ rigorous methodologies and statistical techniques to minimize biases and improve the reliability of the estimates.

Overall, impact evaluations provide valuable insights into the effectiveness and impact of programs, helping policymakers and practitioners make informed decisions and improve program design and implementation.