# Impact Evaluation Glossary

**Alternative hypothesis:** In the context of impact evaluation, the alternative hypothesis is usually that the null hypothesis is false; in other words, that the intervention has had an impact on outcomes of interest.

**Attribution:** The extent to which the observed change in outcomes is the result of the intervention, taking into account all other factors which may also affect the outcomes of interest.

**Attrition:** Attrition occurs when some units drop from the evaluation sample between data collection exercises. Attrition is a case of unit non-response and can create bias in impact evaluations if it is correlated with treatment status.

**Average Treatment Effect (ATE):** The average value of the impact on the beneficiary group (i.e., treatment group). In an experiment, the ATE is the average difference between the value of the outcomes for the treated group and the value of the outcomes for the control group. See also intention to treat (ITT) and treatment effect on the treated (TOT).

**Baseline:** Pre-intervention, ex-ante. The situation prior to an intervention, against which impact can be evaluated or comparisons can be made. Baseline data are collected before a program or policy is implemented to assess the "before" state.

**Before-and-after comparison:** Also known as "pre-post comparison" or "reflexive comparison," a before-and-after comparison attempts to establish the impact of a program by tracking changes in outcomes for program beneficiaries over time, using measurements before and after the program or policy is implemented. It is not possible to establish a causal effect based on this strategy since many factors other than the program could have affected the change in outcomes before and after implementation.

**Bias:** The bias of an estimator is the difference between an estimator's expectation and the true value of the parameter being estimated.

**Cluster:** A cluster is a group of units that are similar in one way or another. For example, in a sampling of school children, children who attend the same school would belong to a cluster because they share the same school facilities, teachers, and curriculum.

**Contamination:** When units in the control group receive or are exposed to the treatment.

**Comparison group:** Also known as a "control group." A valid comparison group will have, on average, the same characteristics as the group of beneficiaries of the program (treatment group), except that the units in the comparison group are not exposed to, or do not benefit from, the program or policy. Comparison groups are used to estimate the counterfactual.

**Confounding factors:** Factors (i.e., variables) other than the program or policy which affect the outcome of interest.

**Cost-benefit analysis (CBA):** A comparison of all the costs and benefits of the intervention, in which these costs and benefits are all assigned monetary value. The advantage of CBA over analysis of cost effectiveness, is that it can consider multiple outcomes and allows comparison in the return to spending in different sectors, helping the efficient allocation of development resources as a result.

**Cost-effectiveness:** Determining cost-effectiveness entails comparing similar interventions based on cost and impact. For example, impact evaluations of various education programs allow policy makers to make more informed decisions about which intervention may achieve the desired objectives with more efficient use of resources, given their particular context and constraints.

**Counterfactual:** What would have happened without the program or policy intervention. The counterfactual is an estimate of what the outcome would have been for a program beneficiary in the absence of the program. By definition, the counterfactual cannot be directly observed. Therefore, it must be estimated using comparison groups. The best counterfactual is the one estimated using random assignment.

**Dependent Variable:** A variable believed to be predicted by or caused by one or more other variables (independent variables). The term is commonly used in regression analysis. In the context of impact evaluation, the dependent variables are also called "outcomes."

**Difference-in-differences (DiD):** Also known as “double difference” or “DiD.” Difference-in-differences estimates the counterfactual for the change in an outcome for the treatment group by subtracting the change in the same outcome for the comparison group. This method allows us to take into account any differences between the treatment and comparison groups that are constant over time. The two differences are thus before and after, and between the treatment and comparison groups.

**External validity:** To have external validity means that the causal impact discovered in the impact evaluation can be generalized to the universe of all eligible units. For an evaluation to be externally valid, it is necessary that the evaluation sample be a representative sample of all of eligible units.

**Follow-up survey:** Also known as “post-intervention” or “ex-post” survey. A survey that is fielded after the program or policy intervention has started, once the beneficiaries have benefited from it for some time. An impact evaluation can include several follow-up surveys.

**Hypothesis:** A hypothesis is a proposed explanation for an observable phenomenon. See also null hypothesis and alternative hypothesis.

**Impact:** The difference between what happened with the program or policy intervention and what would have happened in its absence.

**Impact evaluation:** An impact evaluation is a study that tries to make a causal link between a program or policy intervention and a set of outcomes. An impact evaluation tries to answer the question of whether an intervention is responsible for changes in the outcomes of interest. Contrast with process evaluation.

**Independent Variable:** A variable believed to cause changes in the dependent variable, usually applied in regression analysis.

**Intention-to-treat estimator (ITT):** The ITT estimator is the difference in outcomes for the group randomly assigned to receive the treatment and the group randomly assigned to the control group, regardless of take-up of the treatment or compliance with random assignment. Contrast with treatment-on-the-treated.

**Internal validity:** To say that an impact evaluation has internal validity means that it uses a valid comparison group, that is, a comparison group that is a valid estimate of the counterfactual.

**Matching:** Matching is a non-experimental evaluation method that uses large data sets and heavy statistical techniques to construct the best possible comparison group for a given treatment group. The comparison groups are constructed based on observed characteristics, so the danger that unobserved characteristics might bias the results is always present.

**Minimum desired effect:** The minimum change in outcomes that would justify the investment that has been made in an intervention, counting not only the cost of the program and the benefits that it provides, but also the opportunity cost of not investing funds in an alternative intervention. The minimum desired effect is an input for power calculations; that is, evaluation samples need to be large enough to detect at least the minimum desired effect with sufficient power.

**Null hypothesis:** A null hypothesis is a hypothesis that might be falsified on the basis of observed data. The null hypothesis typically proposes a general or default position. In impact evaluation, the default position is usually that there is no difference between the treatment and control groups, or in other words, that the intervention has no impact on outcomes.

**Power calculations:** Power calculations indicate the sample size required for an evaluation to detect a given minimum desired effect. Power calculations depend on parameters such as power (or the likelihood of type II error), significance level, variance, and intra-cluster correlation of the outcome of interest.

**Process evaluation:** A process evaluation is an evaluation that tries to establish the level of quality or success of the processes of a program; for example, adequacy of the administrative processes, acceptability of the program benefits, clarity of the information campaign, internal dynamics of implementing organizations, their policy instruments, their service delivery mechanisms, their management practices, and the linkages among these. Contrast with impact evaluation.

**Randomized assignment:** Randomized assignment is considered the most robust method for estimating counterfactuals and is often referred to as the "gold standard" of impact evaluation. With this method, beneficiaries are randomly selected to receive an intervention, and each has an equal chance of receiving the program. With large-enough sample sizes, the process of random assignment ensures equivalence, in both observed and unobserved characteristics, between the treatment and control groups, thereby addressing any selection bias. This is possible because the selection mechanism depends on a random process and not on any observed or unobserved characteristic of the eligible individuals.

**Regression:** In statistics, regression analysis includes techniques for analyzing several variables,

when the focus is on the relationship between a dependent variable and one or more independent variables. In impact evaluation, regression analysis helps us understand how the typical value of the outcome indicator (dependent variable) changes when the assignment to treatment or comparison group (independent variable) varies, while characteristics of the beneficiaries (other independent variables) are held constant.

**Regression discontinuity (RDD):** Some programs or policy interventions use continuous measures to determine who is eligible to receive it and who is not. RDD is a method used to evaluate the impact of interventions that have a continuous eligibility measure and a clearly defined eligibility "cutoff" or threshold (i.e., age, poverty index, etc.). Units in close proximity to the eligibility cutoff on either side (treated or non-treated) are assumed to be as-if randomly assigned and, therefore, allow for estimation of casual effects of the intervention.

**Sample:** In statistics, a sample is a subset of a population. Typically, the population is very large, making a census or a complete enumeration of all the units in the population impractical or impossible. Instead, researchers can select a representative subset of the population (using a sampling frame) and collect statistics on the sample; these may be used to make inferences or to extrapolate to the population. This process is referred to as sampling.

**Sampling error:** The error which occurs as estimates are used making data from a sample rather than the whole population.

**Selection bias:** Selection bias occurs when the reasons for which an individual participates in a program are correlated with outcomes. This bias commonly occurs when the comparison group is ineligible or self-selects out of treatment.

**Significance level:** The significance level is usually denoted by the Greek symbol, α (alpha). Popular levels of significance are 5 percent (0.05), 1 percent (0.01), and 0.1 percent (0.001). If a test of significance gives a p value lower than the α level, the null hypothesis is rejected. Such results are informally referred to as "statistically significant." The lower the significance level, the stronger the evidence required. Choosing the level of significance is an arbitrary task, but for many applications, a level of 5 percent is chosen for no better reason than that it is conventional.

**Single difference:** Either, the comparison in the outcome for the treatment group after the intervention to its baseline value (also called before versus after), or an ex post comparison in the outcome between the treatment and control groups. Compare to double difference.

**Spillover effect:** Also known as contamination of the comparison group. A spillover effect occurs when the comparison group is affected by the treatment administered to the treatment group, even though the treatment is not administered directly to the comparison group. If the spillover effect on the comparison group is negative (that is,if they suffer because of the program), then the straight difference between outcomes in the treatment and comparison groups will yield an overestimation of the program impact. By contrast, if the spillover effect on the comparison group is positive (that is, they benefit), then it will yield an underestimation of the program impact.

**Stratified sample:** Obtained by dividing the population of interest (sampling frame) into groups (for example, male and female), and then drawing a random sample within each group. A

stratified sample is a probabilistic sample: every unit in each group (or stratum) has a well-defined probability of being drawn.

**Theory of change:** Laying out the underlying causal chain linking inputs, activities, outputs, and outcomes, and identifying the assumptions required to hold if the program or policy intervention is to be successful. A theory of change is the starting point for theory-based impact evaluation.

**Treatment group:** Also known as the treated group or the intervention group. The treatment group is the group of units that benefits from an intervention, versus the comparison group that does not.

**Treatment-on-the-treated (TOT):** The effect of TOT is the impact of the treatment on those units that have actually received or benefited from the treatment. Contrast with intention-to-treat.

**Type I error:** Error committed when rejecting a null hypothesis even though the null hypothesis actually holds. In the context of an impact evaluation, a type I error is made when an evaluation concludes that a program has had an impact (that is, the null hypothesis of no impact is rejected), even though in reality the program had no impact (that is, the null hypothesis holds). The significance level determines the probability of committing a type I error.

**Type II error:** Error committed when accepting (not rejecting) the null hypothesis even though the null hypothesis does not hold. In the context of an impact evaluation, a type II error is made when concluding that a program has no impact (that is, the null hypothesis of no impact is not rejected) even though the program did have an impact (that is, the null hypothesis does not hold). The probability of committing a type II error is 1 minus the power level.

**Unobservables:** Characteristics which cannot be observed or measured. The presence of unobservables can cause selection bias in quasi-experimental designs, if these unobservables are correlated with both participation in the program and the outcome(s) of interest.

**Variable:** A variable is an attribute that describes an individual, place, or idea. In statistical terminology, the value of a variable can change (i.e., vary) from one unit to another.

Based and adapted from:
*3ie* (2011). *Impact Evaluation Glossary. New Delhi, India.*
Inter-American Development Bank (2016). Impact Evaluation Portal: Glossary.