Understanding Quasi-Experimental Designs: Techniques for Measuring Impact without Randomization

When randomization is not possible, quasi-experimental designs offer rigorous alternatives for evaluating impact. This brief explores three widely used methods: Regression Discontinuity Design (RDD), Interrupted Time Series (ITS), and Propensity Score Matching (PSM).
You could refer to the first guiding brief titled: *Foundations of Impact Evaluation* for more details and background.

This brief is part of a series of five brief guides produced by the Queen Rania Foundation, based on materials developed by Prof. Howard White (presenting the Research and Evaluation Center (REC)) for the training course titled "Impact Evaluation: Methods, Advocacy, and Scalability". The training was funded by the Education Endowment Foundation (EEF), and the BHP Foundation, and the Queen Rania Foundation.

Regression Discontinuity Design (RDD)

Definition:
RDD estimates causal effects based on a clear eligibility threshold in program eligibility. Participants just above and below the threshold serve as natural comparison groups, minimizing selection bias.

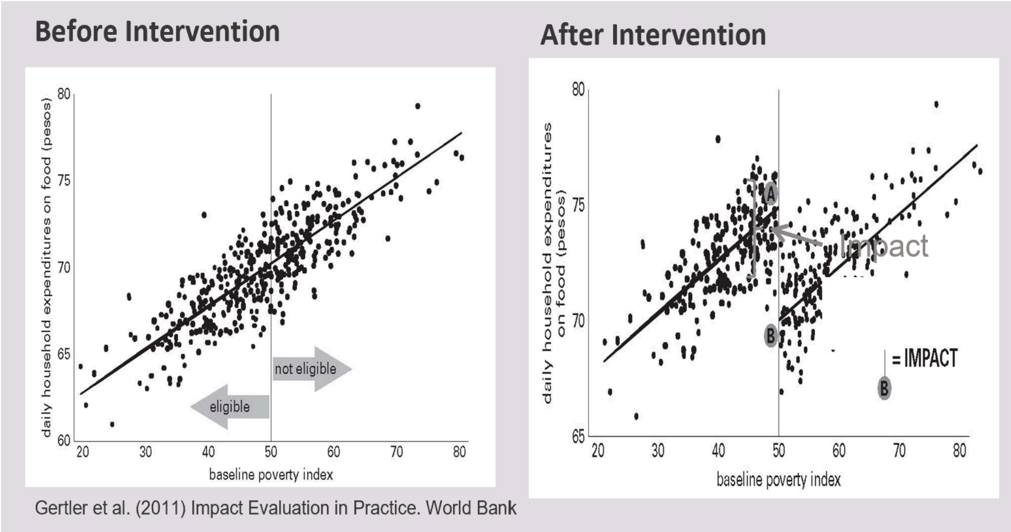
Example: Poverty Index & Conditional Cash Transfers

- Evaluated the impact of cash transfers based on a predetermined poverty score.
- Households above the threshold were excluded, creating a natural control group.
- Findings: Significant increases in school enrollment and improved child nutrition outcomes, particularly among girls.

The graph below plots household food expenditure against a baseline poverty index, where higher values indicate less poverty. The cutoff defines eligibility. After the intervention, households just below the threshold show improved food expenditure, revealing a discontinuity that estimates the program's impact.

The size of the structural break - calculated by estimating a regression line - is the intervention impact.

Fig.1: Basic Characteristics: RDD before and after intervention.



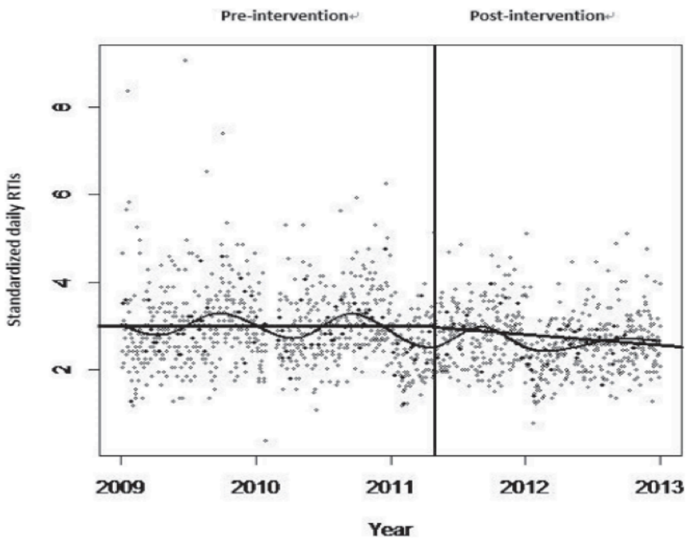
Interrupted Time Series (ITS)

Definition: ITS examines trends before and after an intervention to assess its effect over time.

Example: Criminalization of Drunk Driving in China

- Evaluated the effect of a new law on road traffic incidents.
- Findings: The intervention did not cause an immediate drop in accidents but led to a gradual decline, suggesting behavioral shifts driven by enforcement and public awareness.

Fig.2: Criminalization of drunk driving in China: example illustrates the impact of a change



Source: Zhao, A., Chen, R., Qi, Y., Chen, A., Chen, X., Liang, Z., ... & Kan, H. (2016). Evaluating the impact of criminalizing drunk driving on road-traffic injuries in Guangzhou, China: a time-series study. *Journal of epidemiology*, *26*(8), 433-439.

Propensity Score Matching (PSM)

When random assignment isn't possible, PSM builds a comparison group by matching participants and non-participants with similar observable characteristics.

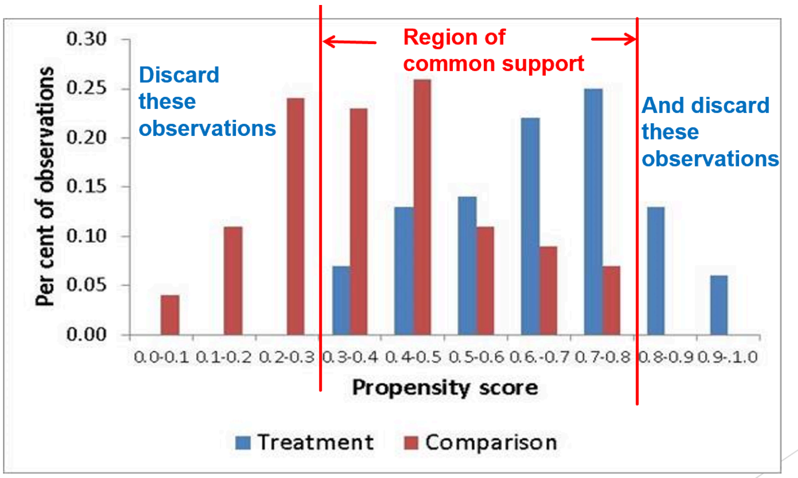
The goal is simple: to ensure that we compare individuals who are statistically similar (meaning that they have the same average observed characteristics) , except for receiving the intervention.

Consider the example of evaluating a scholarship program that provides financial aid to students from low-income backgrounds. If we simply compare the performance of scholarship recipients to those who didn't receive the aid, the results may be misleading because students who apply for scholarships may already be more motivated, higher attaining students than those who don't. PSM helps mitigate this bias by calculating the probability (propensity score) that a student would receive the scholarship based on observable factors like family income, prior academic performance, and parental education. By matching students with similar scores, we create a balanced comparison group that allows us to isolate the impact of the scholarship itself.

Key Steps in PSM Implementation:

1. Define the intervention and determine eligibility criteria.
2. Identify baseline variables that influence participation.
3. Estimate propensity scores using logistic or probit regression.
4. Define the region of common support and exclude observations without suitable matches.
5. Perform nearest neighbor, caliper, or kernel matching.
6. Compare outcomes between matched groups.
7. Conduct sensitivity analysis to test robustness of findings.

Fig.3: Graphical representation of the region of common support



While PSM is powerful, it has limitations. PSM assumes all relevant confounding variables are observed and included. If key variables are missing, bias remains.

It only accounts for observable characteristics, meaning that unmeasured differences—like personal motivation—may still bias results. Therefore, combining PSM with other methods, like Difference-in-Differences (DiD) or Instrumental Variable (IV) analysis, can further strengthen the credibility of findings.

Table.1: Comparison of Quasi-Experimental Designs: Strengths and Limitations

Design	Strengths	Limitations
RDD	Strong internal validity, mimics randomization	Limited external validity beyond the cutoff region. Obtaining sufficient sample size may be a challenge, especially that it only applies to those around the threshold.
ITS	Accounts for long-term trends, requires no control group	Susceptible to confounding factors if other policies change concurrently
PSM	Creates a balanced comparison group using observable characteristics, mitigates selection bias	Only accounts for observable characteristics, unmeasured differences may still bias results May be more suitable for data rich environments.

At a Glance: When to Use These Methods?

- **RDD: When a strict eligibility threshold exists.**
- **ITS: When you have time-series data around a policy or program change.**
- **PSM: When randomization isn't possible, but strong pre-treatment data exists.**

For those seeking to apply or deepen their understanding of these methods, the following resources offer accessible and technical insights.

Further Reading

Baird, S., Ferreira, F.H.G., Özler, B. and Woolcock, M. (2013), Relative Effectiveness of Conditional and Unconditional Cash Transfers for Schooling Outcomes in Developing Countries: A Systematic Review. Campbell Systematic Reviews, 9: 1-124. https://doi.org/10.4073/csr.2013.8

Carvalho, S., & White, H. (2004). Theory-based evaluation: the case of social funds. *American journal of evaluation*, *25*(2), 141-160.

Davies R, 2018, Representing Theories of Change: A Technical Challenge with Evaluation Consequences, CEDIL Inception Paper 15: London

Vajja, A., & White, H. (2008). Can the World Bank build social capital? The experience of social funds in Malawi and Zambia. *The Journal of Development Studies*, *44*(8), 1145-1168.

Dickinson-Craig, E., Turbat, T., Hemming, K., Pope, F. D., Bartington, S. E., Anjaa, S., ... & Manaseki-Holland, S. (2025). Improving Air Quality and Childhood Respiratory Health in Mongolia: The Impact of the Raw Coal Ban and COVID-19 Restrictions—An Interrupted Time-Series Analysis. *Atmosphere*, *16*(1), 46.

Zhao, A., Chen, R., Qi, Y., Chen, A., Chen, X., Liang, Z., ... & Kan, H. (2016). Evaluating the impact of criminalizing drunk driving on road-traffic injuries in Guangzhou, China: a time-series study. *Journal of epidemiology*, *26*(8), 433-439.

White, H., & S. Sabarwal (2014). Quasi-experimental Design and Methods, Methodological Briefs: Impact Evaluation 8, UNICEF Office of Research, Florence. https://www.betterevaluation.org/sites/default/files/Quasi-Experimental_Design_and_Methods_ENG.pdf

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

Ikira, M., & Ezzrari, A. (2021). Evaluating the impact of conditional cash transfer programs: Evidence from Morocco. *American Journal of Educational Research*, *9*(5), 320-329. https://users.nber.org/~rdehejia/!@$DRM/Lecture%2001/papers%20to%20discuss%20-%20Reserach%20vs%20Evlaution/CCTs%20education-9-5-10.pdf