Impact Evaluation
in Transport

Impact Evaluation
in Transport

Contents

Introduction

The direct benefits of transport projects are well known; paving roads reduces vehicular operating costs, mass transport systems reduce commuting times and road maintenance prolongs their operating lifespan. Impact evaluation reveals a range of ways in which transport projects improve lives. The empirical evidence shows that paved roads encourage preventive healthcare and increase employment opportunities, mass transport systems improve air quality and reduce crime, and satisfaction of road maintenance increases when services are provided by local communities.

Impact evaluations generate valuable information for both executing agencies and stakeholders. Their results provide rigorous measurements of the effect of a project, who benefits the most and why. As a result, the evidence generated by well-conducted impact evaluations is increasingly used to attract resources to effective investments and to guide institutional choices.

Impact Evaluation is Complementary to Other Evaluation Methods

Impact evaluation is a great complement to monitoring and cost-benefit analysis. Monitoring tracks performance of projects, both in terms of costs and outputs, against their expected values. Cost-benefit analysis compares the monetary costs and benefits of a project. Impact evaluation makes a causal link between a project and a set of results. Its most distinctive feature is the comparison of the world as observed against what would have happened had the project not taken place. This hypothetical result in the absence of the project is called the *counterfactual*. While in a cost-benefit analysis this "without project" scenario is simulated based on certain assumptions, in an impact evaluation it is empirically measured. The results from well-designed impact evaluations can be used to calculate the actual cost-benefit

analysis of the projects. In addition, they can be a very useful input for the ex-ante cost-benefit analyses of similar projects.

Causality: The Main Challenge of Impact Evaluation

Impact evaluation seeks to answer: What is the *causal effect* of a project on an outcome of interest? For instance: Does access to public transport expand job opportunities? How does public transport affect air quality? By how much does a transit law reduce car accidents? What is the effect of paving residential roads on housing prices? The main challenge in answering these questions is capturing the effect directly attributable to the transport project, which requires eliminating any other factors that may also affect the observed results but are not related to the project.

Suppose you want to measure the impact of paving residential streets on households' well-being. Simply observing that consumption increases after improving a road is not sufficient to establish causality. Indicators of well-being may have improved because of other factors unrelated to transport, such as economic growth, an unusual economic boom, a government cash transfer, etc. Thanks to its ability to obtain a counterfactual,

an impact evaluation can separate the changes generated by improving the streets from the ones generated by other events occurring during the same period.

The Unobserved Counterfactual Approximated by a Comparison Group

The idea of the counterfactual is very useful, but it is impossible to observe what would have happened in the absence of a transport project or any other project. Suppose we want to measure the impact of public transport services on employment. We either observe individuals with or without access to public transport, but never the same individual in both conditions at the same time.

Impact evaluation solves this issue by comparing outcomes in a *treatment group* that benefits from a project with outcomes in a *comparison group* that does not benefit from the project. A good comparison group has three properties:

- It is similar on average to the treatment group.
- Remains unaffected by the project during the evaluation.
- The impact of a project on the comparison group would be expected to be identical to the impact observed on the treatment group.

The mechanisms selecting the treatment and comparison groups are crucial in impact evaluation. The main concern is the presence of a *selection bias*. There is selection bias when the treatment and comparison groups differ in ways that are related to the outcome evaluated. If there is selection bias it will be difficult if not impossible to attribute the observed outcome changes to a project rather than to the underlying differences between the two groups. One of the major objectives and challenges for any impact evaluation is ensuring that the selection of the control and comparison group is free of bias. Different ways of dealing with potential selection bias give rise to different evaluation methods: Experimental, quasi-experimental and non-experimental.

Experimental Methods

The Gold Standard of Impact Evaluation
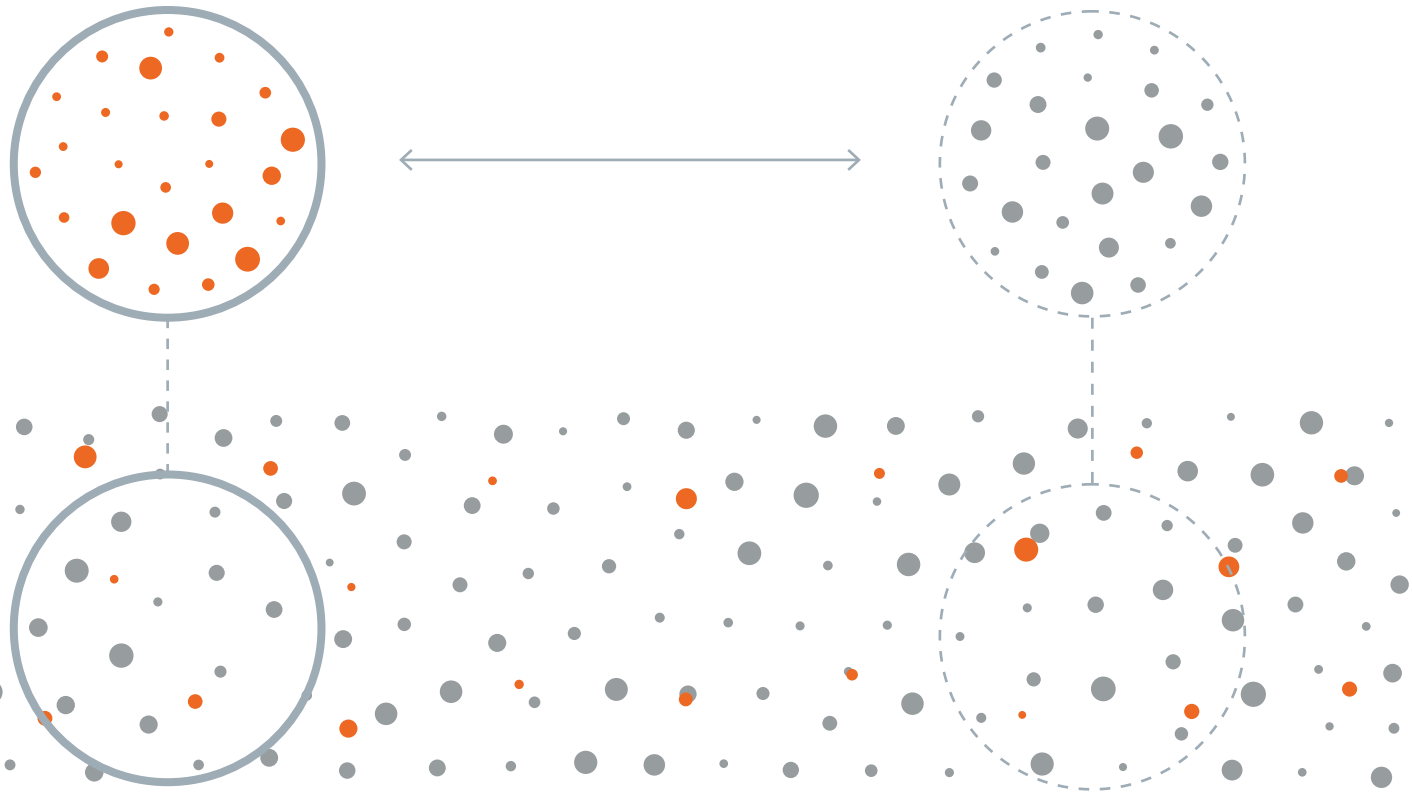
Experimental methods use lotteries and other random selection devices to create a comparison group, called the *control group*. The key feature of random selection is that all potential beneficiaries face the same probabilities ex-ante of being part of a project. An impact evaluation using this approach is called a Randomized Control Trial (RCT).

Well conducted evaluations using experimental methods are considered the gold standard of impact evaluation because they require less assumptions to establish the comparability between treatment and control groups. When the sample is sufficiently large, random selection creates a control group that is similar, on average, in all characteristics to the treatment group. RCTs ensure that all characteristics that could be related to the outcome of interest are equally distributed among the control and treatment groups. That is, implying that there is no selection bias.

Randomizing Infrastructure Projects: *The Impact of Paving Residential Streets on Welfare*

ACAYUACÁN, MEXICO

The local government of Acayucán had a long list of residential roads that needed to be paved. Resources were scarce and it was impossible to pave all the streets at once.

A computer-generated lottery was used to select a group of streets that would be paved out of a long list of suitable candidates. The streets selected to receive the project formed the treatment group, while the streets that remained unpaved served as the control group. The procedure used to select the two groups guaranteed that the two groups of streets were comparable.

Comparing outcomes between the treatment and the control groups two years after the project, the evaluation showed that paving residential streets increases the value of nearby properties and that households whose streets were paved were able to transform their increased property wealth into larger rates of vehicle ownership, increased ownership of household appliances and investment on home improvements.

In contexts of excess demand or when projects are developed gradually over time, lotteries and other random selection devices can be seen as fair tools to allocate treatment among equally deserving beneficiaries, as well as offering the opportunity to conduct rigorous impact evaluations.

Randomizing Service Access: *The Impact of a Transport Subsidy on Job Search*

WASHINGTON, D.C., THE USA

Relatively expensive transport services may constrain job search in urban areas, disproportionately affecting poor populations. This is the case in Washington D.C., where job-seekers often live far from their potential employers.

A computer-generated lottery randomly selected job-seekers in segregated neighborhoods to receive a transport subsidy. This created a treatment group who had access to city buses and trains for a reduced price. The control group consisted of job-seekers who were not selected in the lottery and had to look for work without the transport subsidy.

The differences in outcomes between the treatment and the control group after three months showed that subsidized access to urban transport speeds up the job search process, which could translate into shorter unemployment durations. This result highlights the important social and economic impacts that transport prices can have on vulnerable populations.

Infrastructure components normally do not lend themselves easily to randomized control trials, but some complementary projects like transport subsidies may. The knowledge generated by this type of experimental evaluations could be highly valuable for policymakers.

Experimental Evaluation May Be "The One", But Not the Only One

"If you can do a Randomized Control Trial, by all means, do it." [1] This said, quasi-experimental and non-experimental methods offer alternatives for rigorous impact evaluation. The choice among these methods depends on the specific characteristics of the project to be evaluated and the availability of data.

Quasi-experimental Methods:

"Things Happen",
Sometimes at Random

Quasi-experimental methods exploit unexpected events or ad hoc rules that govern the roll-out of projects or selection of beneficiaries to measure their impact. They are called quasi-experimental because the factors that determine who benefits from a project or who benefits first are considered "as good as random." The most common quasi-experimental methods are Regression Discontinuity and Natural Experiments.

Regression Discontinuity:
Using External Project Rules to Estimate Their Impact

Regression discontinuity methods are useful when administrative rules or project guidelines, represented by a threshold or a cutoff, determine if certain locations or people become beneficiaries of a transport project or not. For example, selection criteria based on population size that establish which communities benefit from a project, or a minimum length in kilometers of a road as inclusion criterion to be rehabilitated.

Ad hoc administrative rules sometimes create discontinuities in the probability of becoming a project beneficiary. These discontinuities can be exploited to evaluate the impact. Those immediately above and immediately below the cutoff are often statistically comparable and equally deserving. Being just above or just below the cutoff can be considered almost random. The only difference is that one group benefits from a project while the other does not.

The key element of evaluations under regression discontinuity is that the rule that determines the implementation of a project cannot be precisely manipulated either by program administrators or potential beneficiaries. This condition increases the comparability between the treatment and comparison groups in all respects.
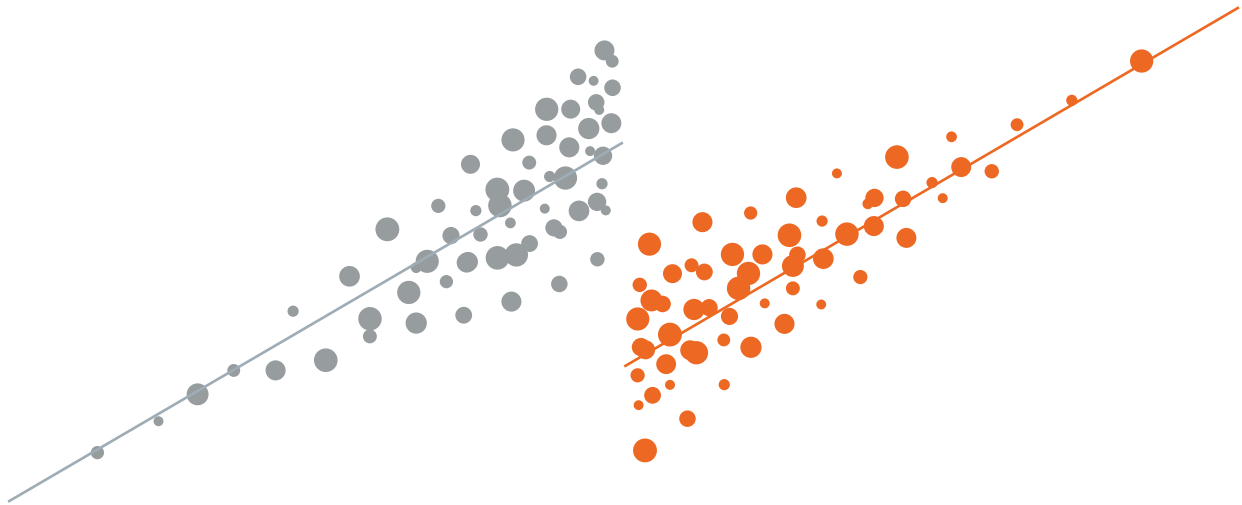
## *Village Size: A Cutoff Rule to Evaluate the Impact of Connecting Roads on Healthcare and Labor*

INDIA

The government of India mandated that all villages with more than 500 inhabitants should have at least one road connecting to the national transport network.

The cutoff rule in this transport program acted almost as a random selection mechanism for villages close to the threshold. Villages with a few more than 500 inhabitants benefited from new connecting roads, creating a treatment group. Villages with a few less than 500 inhabitants are a good comparison group because no roads were built even though they should be similar to the treatment group.

A comparison of outcomes of villages just above and just below the cutoff (with and without new roads) showed that the project improved household welfare. Preventive healthcare increased among women and men obtained access to better-paid jobs in cities nearby.
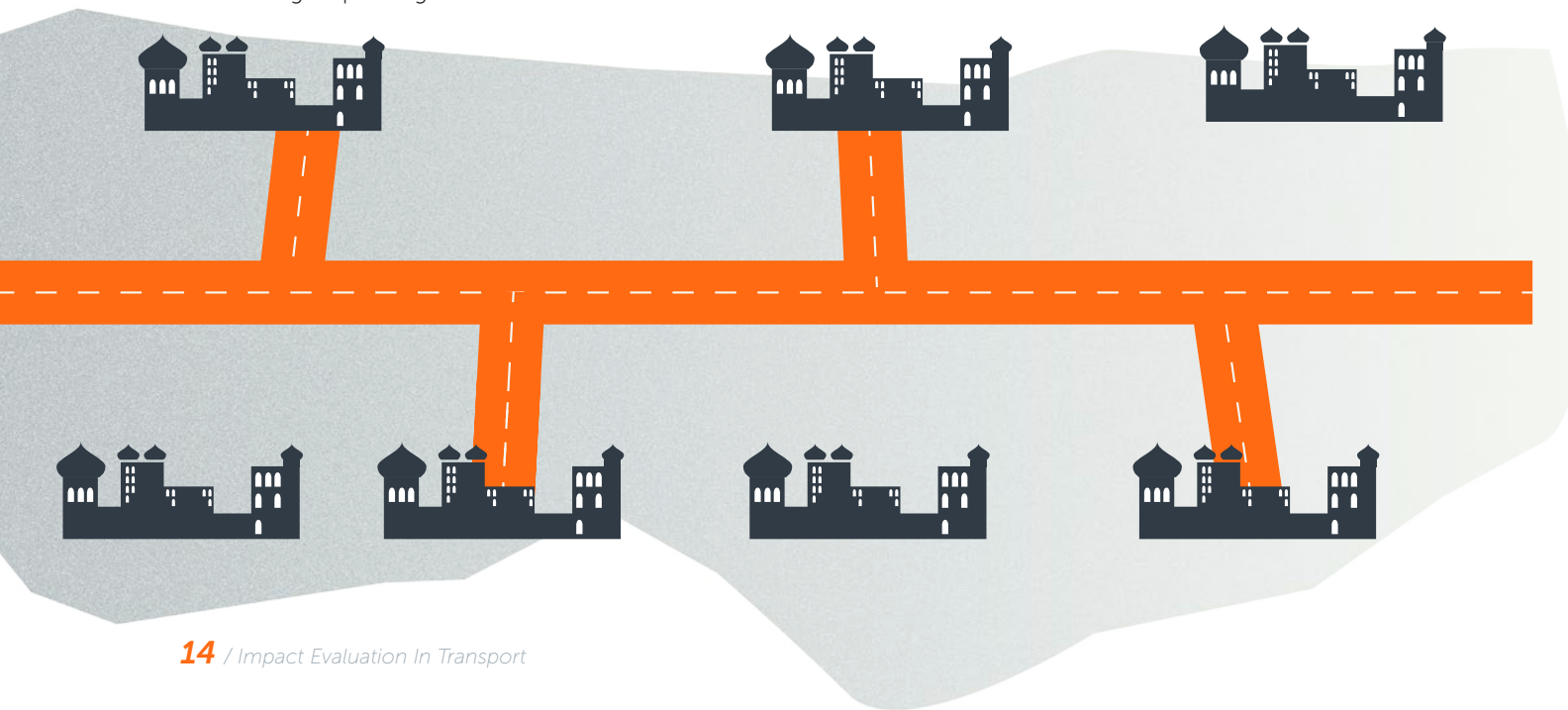
Natural Experiments: *The Unexpected Chance for Evaluating Impact*

Natural Experiments are external events that create treatment and comparison groups by chance. In natural experiments a comparison group may form due to project delays or because part of a project is cancelled due to external factors that have nothing to do with its expected impact or outcomes.

Natural experiments do not need to be "natural." The fundamental condition of a natural experiment is that the implementation or expansion of a program is disrupted by factors outside the control of potential beneficiaries.
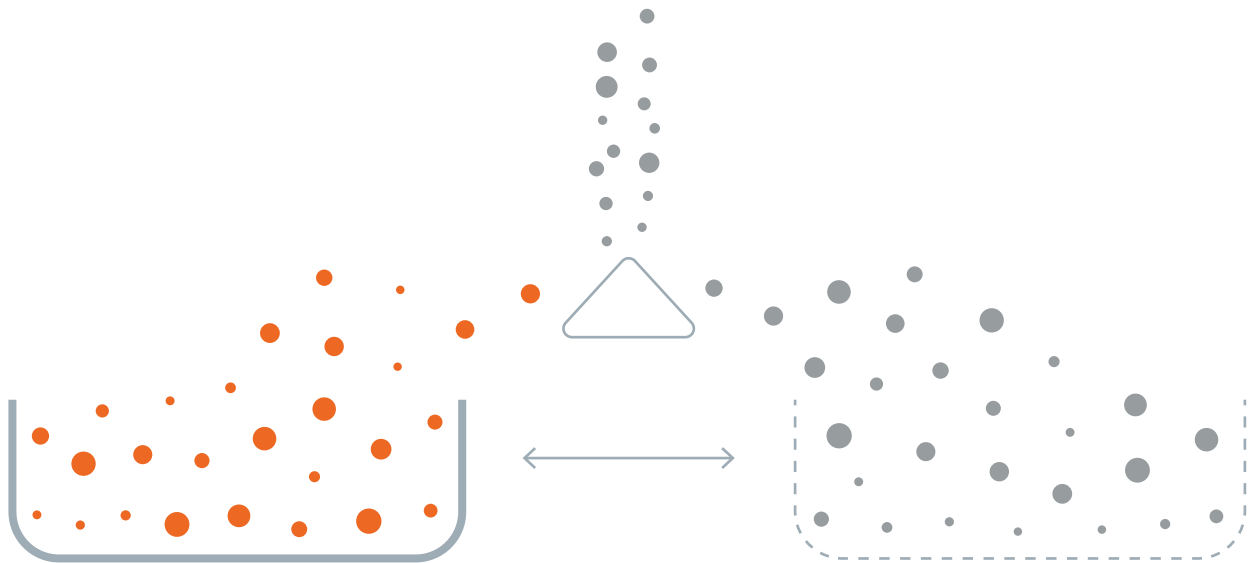
Olympic Games: A "Natural" Opportunity to Evaluate the Impact of Emissions on Health

BEIJING, CHINA

In preparation for hosting the 2008 Olympic Games, the city of Beijing introduced several measures to improve the city's air quality. The measures included heightened vehicular emissions standards and restricting the use of vehicles by license plate number. The bundle of measures related to transport in combination with other emission cuts effectively reduced air pollution during the Olympic Games period. After a few weeks, things returned to "normal", creating a sharply defined window of time of relatively good air quality.

This sudden reduction in emissions during a certain number of months allowed measuring the impact of air pollution on neonatal health. Infants born during the Olympics were significantly heavier than infants born during the same period the year before and the year after the event . This impact evaluation suggests that even a short-term decrease in air pollution favors fetal development in late pregnancy, which in turn could have long-lasting effects on health.

New Metro Lines: A "Natural" Opportunity to Evaluate the Impact of Mass Transport on Air Pollution

DELHI, INDIA

A natural experiment was created by the phasing in of several metro rail extensions in Delhi. This opportunity was used to evaluate the impact of a clean mass transport system on air pollution.

The increase in ridership observed some weeks after the opening of the extensions created an opportunity to compare air pollution levels in the surrounding areas to the new stations, right before and right after the metro service entered in operation.

Comparing air quality in areas of a city that have transport access with respect to areas of the city that do not have transport access could lead to misleading conclusions. Both types of areas are likely to be different in many other aspects aside from access to public transport. For example, metro lines are usually located in areas of high economic activity where people ride automobiles and contamination is prevalent. Areas without metro lines are typically more residential and less polluted.

The short-term effects show that the increase in metro ridership during the window of analysis resulted in an important decrease in pollutants that can have adverse effects on health.

Non-experimental Methods

If no Counterfactual Exists, Let's Find One

Non-experimental methods rely on the richness of data to estimate what would have happened in the absence of a project. Matching method use available data prior to the treatment to construct a comparison group that "looks like" the beneficiaries of a project. The Difference-in-Differences method uses data from treated and comparison groups before and after a project to measure its effect. These two methods are often combined.

Propensity Score Matching

In general, matching methods use large data sets and statistical techniques to construct the best possible comparison group for a treatment group, based on observed characteristics (prior to the treatment) of neighborhoods, roads or households that could potentially influence the project impact.

Propensity score matching involves two steps. First, constructing an index to measure "project fitness" called *propensity score*. If two units have similar propensity scores, it means that they are both equally likely to receive treatment. Second, each unit that benefits from the project is matched to a comparison group, a set of units with similar propensity scores but that are not affected by the project in practice. The key idea behind this method is that the comparison group is "as good as random." This is true if the propensity score truly captures the likelihood of becoming a project beneficiary. This is a strong assumption that may or may not hold, depending on the context and the project.
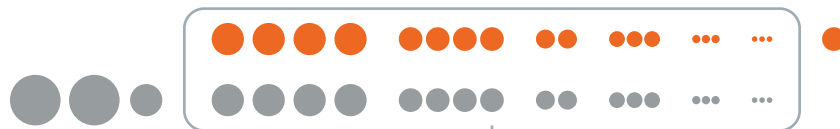
Diff-in-Diff

The difference-in-differences method also called Diff-in-Diff or double differences, exploits data before and after a project in a treatment and a comparison group. The first differences refer to the changes over time within the treatment and within the comparison group measured separately to obtain trends over time within each group. The second difference compares the trend over time in the treatment group with the trend over time in the comparison group. The Diff-in-Diff method assumes that the second difference between the two groups can be attributed to the project. The evolution over time of the comparison group approximates the trend that the treatment group would have followed in the absence of the program.

The most important requirement for the validity of the Diff-in-Diff method is that treatment and comparison groups follow similar trends prior to the project. As with matching, this can be a strong assumption, that may or may not hold, depending on the context and the project. This method requires historical data to convince evaluators that its main underlying assumption holds.
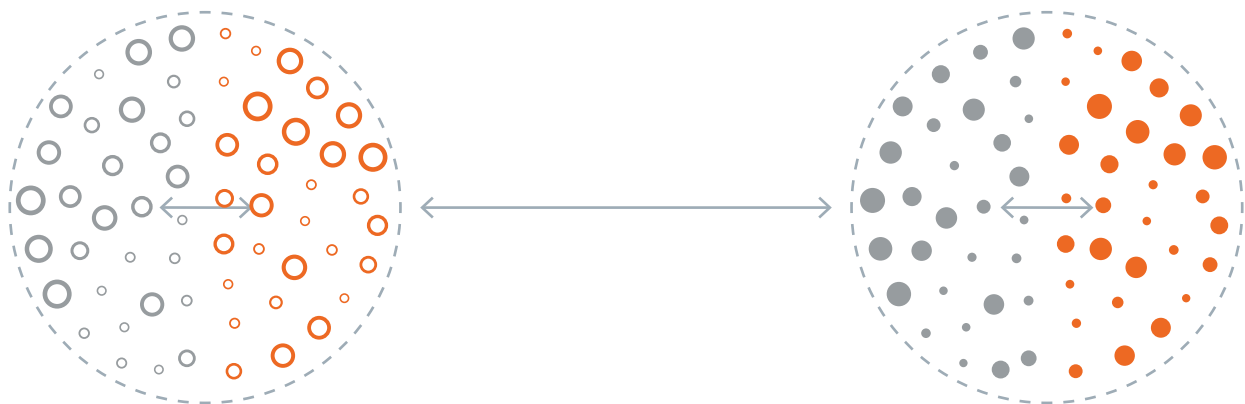
## *Metrocable: Estimating the Impact of Urban Transport Systems on Violence*

MEDELLÍN, COLOMBIA

With the objective of decreasing social conflict and promoting urban development, the local government of Medellín constructed a cable-propelled transit system known as Metrocable. A fleet of ski-like gondolas connected impoverished neighborhoods in its mountainous periphery with the city center.

The choice of location for this transport investment was certainly not random and many other local initiatives were in place to reduce violence in disadvantaged neighborhoods. The evaluation used propensity scores to match areas which benefited from this innovative transport system with similar areas in the city which did not benefit but that were as similar as possible based on the propensity score. Both treated and comparison neighborhoods were under a widespread territorial plan to promote urban and rural development, but only the treated group benefited from the new transport system.

Results suggest that areas with Metrocable experienced a reduction in crime when compared to similar areas without this system.
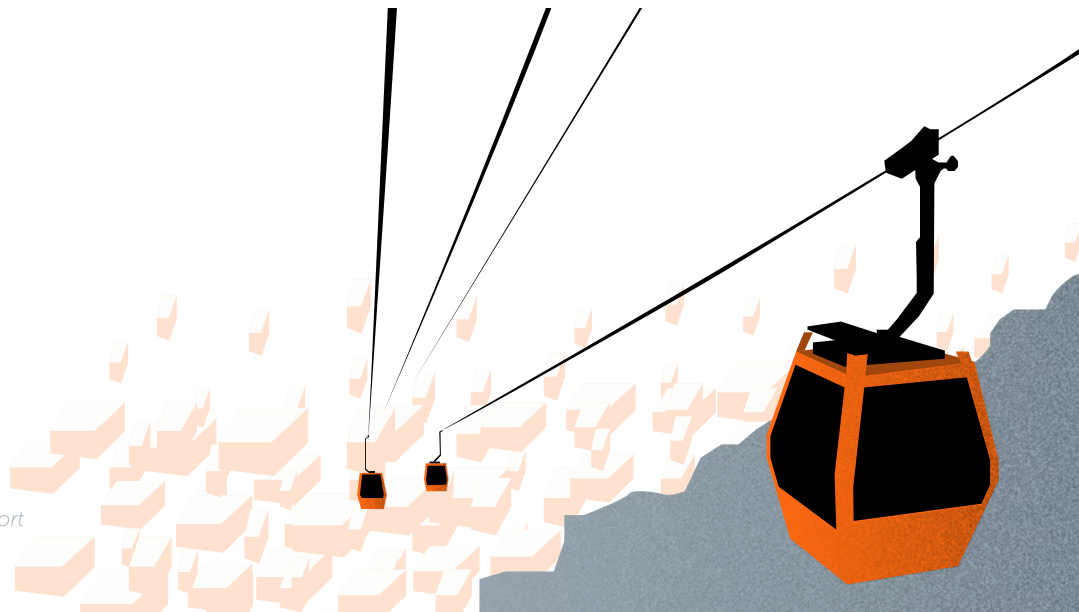
Zero Tolerance Law: Evaluating the Impact of a New BAC Regulation on Traffic Accidents

CHILE

The Chilean government passed a law decreasing the permissible Blood Alcohol Content threshold (BAC) for driving and implemented severe punishments for offenders. The impact of this new regulation was uncertain and controversial, given the limited evidence for developing countries.

The evaluation of this regulatory change exploited local administrative records on traffic accidents and driving to measure the effect of the new law on fatal accidents or drunk driving. A difference-in-differences approach was used to compare the evolution of alcohol-related accidents and infractions before and after the new law, with respect to the evolution of other types of accidents and driving offenses over the same period. Alcohol-related accidents and infractions are considered here the treatment group, as they should be probably affected by the law. Other types of crashes and driving offenses are considered the comparison group as they should probably not be affected by the law. These double differences (over time and over types of accidents) allow measuring the impact of the new law and separate it from other confounding factors that affect alcohol-related traffic accidents, such as festivities, holidays, weather conditions or traffic controls.

The new law succeeded in reducing the overall traffic accident rate related to alcohol consumption, while the number of serious injuries also related to alcohol remained stable.
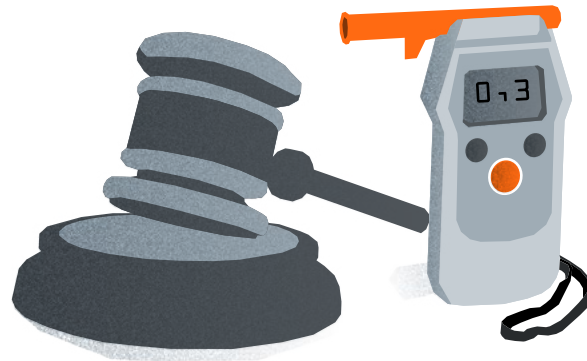
Road Maintenance: Estimating the Impact of Institutional Innovations on Transport Services

PERÚ

The efficiency in the maintenance of the road network is a serious challenge for developing economies. The Peruvian rural roads program implemented an institutional innovation by contracting local private firms for the maintenance of existing rural roads. This Public-Private Partnership (PPP) at small scale aimed to improve the efficiency of the maintenance services through a community monitoring system.

This institutional innovation was evaluated using a combination of the difference-in-differences method and the matching approach. First, treated roads were matched to comparison roads with similar features. Then, a comparison of outcomes before and after the maintenance contracts showed that households near the roads in charge of the local private firms received better maintenance service than the comparison roads which were subject to standard maintenance provided directly by the governments or other institutions.

The decentralization of maintenance services improved road quality, reduced travel time and resulted in general satisfaction with the rehabilitation service.

Conclusions

The quality of an impact evaluation can only be as accurate as the data used. "The inferences that one can logically draw are determined by the available data and the assumptions that one brings to bear."[2] When interpreting the results of an impact evaluation one should keep in mind that statistically significant effects may or may not be economically meaningful depending on their size.

The data sources for impact evaluation vary, and with them, the cost of the evaluations. While some impact evaluations require the collection of primary data, secondary data such as censuses, existing surveys, or administrative records may be excellent sources of information. Furthermore, the digital revolution has widened the spectrum of possibilities. Satellite images, call detail records and other innovations continue expanding the data available for impact evaluation.

All the evaluation methods presented here can provide valid approximations of what would have happened in the absence of a project. Nonetheless, each method has specific requirements to ensure its validity. Experimental evaluations require true random selection of project beneficiaries; regression discontinuity assumes program rules are consistently applied; natural experiments must effectively determine who benefits from a project in response to external conditions; matching methods imply that all determinants of receiving a project that affect outcomes are observed; and difference-in-differences must comply with the assumption that beneficiaries and non-beneficiaries evolve over time in the same way prior to the project. Which method is best suited to evaluate a project depends on which of these requirements are appropriate for the context. A deep understanding of the operational rules and the context of a project can help select the most appropriate evaluation method.

The evidence generated by impact evaluations can serve as basis for conducting cost-benefit analysis as well as for building knowledge on how infrastructure and transport projects improve wellbeing. While impact evaluation is a strong policy tool, it is important to acknowledge its limits. The extrapolation of an impact evaluation effect outside the evaluated project may be inappropriate. All impact evaluations, included those conducted using RCTs, are valid for the project under analysis if the assumptions of the method used hold. But nothing guarantees that the results apply in other contexts or time periods. The uniqueness of some transport projects implies that their impact evaluation results may not be generalizable to other contexts or time periods.

References

General References

Gertler, Paul J.; Martinez, Sebastian; Premand, Patrick; Rawlings, Laura B.; Vermeersch, Christel M. J. (2011), *Impact Evaluation in Practice*, Second Edition. World Bank. (goo.gl/Ir58Hr)

Manski, Charles (2004), *Measuring Expectations, Econometrica*, 72, issue 5, p. 1329-1376. (goo.gl/gJB5YU)

William Shadish (2010), *Quasi-Experimental Workshop, Part II, Workshops and training*, Northwestern University, Institute for Policy Research. Aug, 2010. (goo.gl/j45jS3)

RCTs

Marco Gonzalez-Navarro and Climent Quintana-Domeque (2016), *Paving Streets for the Poor: Experimental Analysis of Infrastructure Effects*, Review of Economics and Statistics, May 2016, Vol. 98, No. 2 , Pages: 254-267. (goo.gl/EmMVjh)

David C. Phillips (2014), *Getting to work: Experimental evidence on job search and transportation costs*, Labour Economics, Elsevier, vol. 29(C), pages 72-82. (goo.gl/V14eUc)

RD

Asher, S. and Novosad, P. (2016), *Market Access and Structural Transformation: Evidence from Rural Roads in India*. Working Paper. (goo.gl/rCPv7d)

Banerjee, Rakesh and Ashish Sachdeva (2015): *Pathways to Preventive Health, Evidence from India's Rural Road Program*. USC Dornsife Institute for New Economic Thinking, Working Paper No. 15-19. (goo.gl/ER6O8d)

Natural Experiments

Goel, Deepti & Gupta, Sonam (2015), *The effect of metro expansions on air pollution in Delhi*, Centre for development Economics, Department of Economics, Delhi School of Economics, Working Paper No. 229. (goo.gl/h3Cr1u)

Rich DQ, Liu K, Zhang J, Thurston SW, Stevens TP, Pan Y, Kane C, Weinberger B, Ohman-Strickland P, Woodruff TJ, Duan X, Assibey-Mensah V, Zhang J. (2015), *Differences in birth weight associated with the 2008 Beijing Olympics air pollution reduction: results from a natural experiment*, Environmental Health Perspectives 123:880–887. (goo.gl/dhJXmP)

Matching

Cerdá, M; Morenoff, JD; Hansen, BB; Tessari Hicks, KJ; Duque, LF; Restrepo, A; Diez-Roux, AV (2012), *Reducing violence by transforming neighborhoods: a natural experiment in Medellín, Colombia*. American Journal of. Epidemiology. 175: 1045–53. (goo.gl/tS1i8c)

Diff-in-Diff

Otero, Sebastián and Rau, Tomás (2015), *New Evidence of the Effects of Zero Tolerance Laws on Drinking and Driving. The Case of Chile*. Working-Paper. (goo.gl/mDiCVq)

Valdivia, Martin (2010), *Contracting the Road to Development: Early Impacts of a Rural Roads Program, Partnership for economic policy- PMMA 2010-18*, Working Paper. (goo.gl/O8BAmS)

Notes

1 William Shadish (2010), [Quasi-Experimental Workshop, Part II, Workshops and training](), Northwestern University, Institute for Policy Research. Aug, 2010. ([goo.gl/j45jS3]())

2 Manski, Charles (2004), [Measuring Expectations, Econometrica](), 72, issue 5, p. 1329-1376. ([goo.gl/gJB5YU]())

Authors:

René Alejandro Cortés

María Adelaida Lopera

Oscar Alberto Mitnik

Patricia Yañez

IDB
Inter-American
Development Bank