A Survey of Impact Evaluations of Infrastructure Projects, Programs and Policies

Antonio Estache
ECARES, Université Libre de Bruxelles

ECARES working paper 2010-005

ECARES
ULB - CP 114
50, F.D. Roosevelt Ave., B-1050 Brussels BELGIUM
www.ecares.org

A survey of impact evaluations of infrastructure projects, programs and policies

Antonio Estache [1]

February 2010 (revised March 2010)

Abstract

This paper surveys the main lessons from impact evaluations of infrastructure projects, programs and policies relevant to policymakers. In the process, however, it also reveals some major research gaps of relevance to academics. After a brief discussion of the motivation for the explosion of the demand for impact evaluations, it starts with some sometimes underestimated lessons for infrastructure specialists from theoretical experiment. The main focus of the paper is, however, the impact evaluations derived from experimental and quasi-experimental techniques. It covers energy, water and sanitation as well as the various transport subsectors (ports, railways, rural roads and highways). The main value added of the survey may be to show that despite the relatively modest number of published evaluations concluded in infrastructure as compared to health or education for instance, there is a growing coverage of the sector in evaluation efforts. The survey also offers an opportunity to get a sense of the creativity of researchers conducting these evaluations. It summarizes the main questions asked, the main techniques used and when available the results available. It concludes with a discussion of some of the limitations of evaluations in the context of infrastructure interventions.

1. Introduction [2]

Since the mid-2000s, the interest in analytically robust evaluations of the impact of projects, programs or policies has exploded among development academics and field workers. There is a particularly keen interest in evaluations based on randomized field experiments and quasi-experiments.[3] These approaches are now viewed by many as the most effective technique yet to measure the effectiveness of efforts to target the poor in major development projects and to ensure that the poor get their fair share or more of the economic and social benefits achieved through these projects. In depth experimental or quasi-experimental impact evaluations are now mainstreamed in

[1] The author is at ULB - Université Libre de Bruxelles, ECARES - European Centre for Advanced Research in Economics and Statistics , CEPR and ECORE, the association between the CORE and ECARES

[2] This paper builds on work conducted for the Asian Development Bank. Comments and discussions with Francois Bourguignon, Jesus Felipe, Frida Johansen, Martin Rossi, Gael Raballand and Luc Savard. A special thanks to Daniel Camos, Marisela Montoliu and Francesc Trillas who provided really detailed comments and suggestions. Any misinterpretation of results or any mistake is mine and mine only!

[3] Just as a reminder, in a randomized experiment, a study sample is divided into two groups: one will benefit from an intervention (the treatment group) and the other will not (the control group). For instance, we may want to study the health effects of a subsidy to clean water in a particular region. The region is randomly divided into treatment and control groups. The outcome measured –health effects– can then be compared in the two groups. The difference in health will provide an evaluation of the effectiveness of the intervention. Note that randomization in this context means that the evaluator ensures that no pattern exists between the assignment of families into groups and any characteristics of those subjects. Every family is as likely as any other to be assigned to the treatment (or control) group. When the selection of beneficiaries cannot be done randomly, evaluations can rely on groups similar to those benefiting from a treatment. The data are then used using econometric techniques such as regression discontinuity, matching techniques, difference in differences and regression discontinuities which best approximate the results of an randomized approach.

many health and education activities in which international development agencies are involved. They are also increasingly adopted for infrastructures interventions as the length of the reference list of this survey indicates.[4] Impact evaluation has now essentially become a sub-field of development economics.

The large size of some of the infrastructure projects and some other factors discussed later may limit the possibility to assume randomness. But there are many standard interventions in the sector that are appropriate for the latest evaluations techniques based on randomized field experiments or quasi-experiments. Moreover, when the experimental and quasi-experimental approaches are not feasible for some technical reason, there are other approaches such as general equilibrium modeling that can deliver useful and robust evaluations as well.

The paper takes stock of the lessons of recent impact evaluations in infrastructure. It emphasizes those based on experimental and quasi-experimental techniques but also discusses substitutes for infrastructure activities for which these techniques cannot be used.[5] The focus is on the last 5-10 years for developing countries, although, when useful, it also reports studies conducted for developed countries. It is aimed at infrastructure specialists interested in an evaluation to signal a commitment to quantitative evidence of impact through their projects.[6] Hopefully, it is also useful to academics interested in knowledge gaps in the field that could also become interesting research areas.

The paper is organized as follows. It first provides some background on the sources of the increase in demand for evaluations. Next, it very briefly summarizes key results relevant to infrastructure projects, programs or policies from theoretical ("economic laboratory") experiments which have preceded many of the field experiments conducted across sectors. It then summarizes the relevant results from field evaluations on each of the main infrastructure subsectors. It concludes with some comments on very pragmatic challenges evaluators may face. [7]

2. Why an evaluation explosion?

The explosion in the number of evaluations conducted across sectors in developing countries has two main sources. The first is the evolution of expectations on evidence for effectiveness of infrastructure interventions among the key development stakeholders. The second is a change in opportunities to work on evaluations in the academic community.

---

[4] See Bourguignon and Pereira da Silva (2003) for an early exhaustive presentation of the various techniques and Duflo et al (2008) or Ravallion (2008) for more recent updated overviews.

[5] We do not deal with dams or irrigation although both have been subject to recent high profile evaluations. See Duflo and Pande (2007) who end up with a controversial mixed review of the distributional effects of a dam in India and Sawada et al (2008) for an irrigation project in Sri Lanka showing that irrigation reduces chronic poverty by enhancing permanent income through higher agricultural productivity and to some extent better credit availability.

[6] Most donor agencies have documented their own experience in the last 2 years and disseminated it through NONIE (the Network of Networks on Impact Evaluations). With the exception of DFID and JBIC, none has covered in much details their infrastructure evaluations.

[7] Because the paper is aimed at practitioners, it does not deal with the critically important statistical, econometric and modeling issues. For instance, it does not deal with the crucial debate on the choice of instruments to be used in the econometric estimations. These are typically not the responsibility of the officers responsible for the implementation of the projects.

Modern evaluation techniques respond to the increasing demand expressed by most bilateral and multilateral aid agencies--and some NGOs--for reliable quantitative evidence of an explicit causal link between interventions and observed effects.

Many important donors have put significant political pressure to catalyze a more systematic use of these techniques by key development agencies. A simple look at the web site of the British, Spanish or Norwegian bilateral agencies reveals the importance of the topic for their decisions to allocate their development resources. Most work on evaluations on their own. But they mainly contribute financially to collective efforts to scale up the use of evaluation, including through financing the major multilaterals and independent networks of academic researchers.[8]

From an academic viewpoint, the explosion of interest stems from the increase in data availability and from our improved collective capacity to process the data. The academic evaluation field has been able to grow so fast in recent years thanks to: (i) significant improvement in the volume and the quality of household income and consumption surveys conducted in developing countries; (ii) major improvements in microeconometrics and general equilibrium modeling techniques in the last 15 years or so, (iii) the increased ability of personal computers to process large data sets quite fast.[9]

Given the convergence of their interests, the development stakeholders and the academic communities have quite rapidly converged on the implementation of the evaluation agenda. It should thus not be a surprise that academics have been quite effective at delivering what the development stakeholders want since sound academic research in the field with direct policy implications is now just as easy if not easier than in many developed countries. The financial efforts of key donors (e.g. France, Japan, Norway, Japan, Sweden, Spain and the UK, to name the most generous) are often acknowledged by the academics leading the evaluations.

Note that the more quantitative and formal techniques surveyed here for infrastructure have already been popular, and hence tested, in public economics and many other areas of public policy in developed countries for almost 20 years. The large volume of experimental literature in labor, health and education economics assessing the effectiveness of targeted efforts to help the poor in rich countries through unemployment benefits, school vouchers or health subsidies has provided a very strong basis for the academic agenda on development evaluations. The designers of infrastructure projects have also benefited from over 30 years of theoretical experiments in economic laboratories.[10]

[8] See the sources of financing for NONIE, 3ie, the MIT poverty lab or the World Bank Dime initiative for instance.

[9] See Angrist and Pischke (2009) for an outstanding user friendly introduction to the theory and practice of the techniques available to establish a causal relationship with illustrations from a wide range of policy fields relying on experiments.

[10] For instance, Gruber (2007), one of the most popular textbooks on public finance and public policy, includes a full discussion of the econometrics of randomized trials and quasi-experiments as part of his discussion of the standard empirical tool for public finance economists. Many of the ideas on how to use randomized and quasi-randomized experiments in public economics including the optimal decision to deliver public goods (such as many infrastructure projects) were to a large already identified by Rivlin (1974), 35 years ago as discussed in the recent survey by Kling (2007)

The explosion in the use of these evaluation techniques in development economics has not left all academics equally happy. It has its critics, arguing quite aggressively sometimes with its promoters.[11] The two sides disagree on complex technical issues, on the risks associated with poor choices in the implementation of these evaluations and on the scope of policy issues that can be evaluated. But they also agree that the techniques are useful when accountability demands the measurement of payoffs of policy interventions, in particular to avoid recurring white elephants, a risk often associated with large infrastructure projects (the "roads to nowhere" syndrome).

The critics have not tamed the excitement of "evaluation academics' for the field, and hence their support to the concerns of development stakeholders. The evaluation field has, to some extent, now taken over the main economic development journals. This is because they tend to report on interesting, innovative and technically sound work conducted by key development actors. Evaluations of the poverty reduction effects of development actions are widely seen as the source of some of the most important new policy relevant knowledge in development economics. This is also why impact evaluations naturally fit in the sort of things these journals want to publish. In 2009, a journal specializing on impact evaluation was even inaugurated![12]

The upshot most relevant to this survey is that the academic enthusiasm for the field means that since the mid 2000s, the number of policy interventions that has enjoyed top quality control by top econometricians, modelers and development specialists has become quite large. It also means that in less than five years, development economists and policymakers have managed to collect a valuable body of evidence on how well standard development policies, programs and projects actually work in practice and why. Those conducted for infrastructure interventions are the focus of this note.[13]

3. What infrastructure evaluations learned from theoretical experiments [14]

It is easy to forget that long before these field experiments and quasi-experiments became popular, academics used experimental techniques to try to assess the effectiveness and the drivers of the sustainability of policy interventions. Some of these early theoretical experiments already offered some useful lessons for the design of infrastructure interventions and the drivers of their effectiveness.

The efficiency and equity of interventions were among the main concerns addressed by these experimental approaches. But they were not the only ones. There were also efforts to assess the effectiveness of various financing designs to recover the costs of these interventions. These are the

[11] See for instance the very harsh criticism by Deaton (2009), Heckman and Urzua (2009)and the very technical and systematic reaction by Imbens (2009)

[12] The Journal of Development Effectiveness, with a No1, Vol. 1 dated from March 2009 and launched by the International Initiative on Impact Evaluation (see http://www.3ieimpact.org/)

[13] There are other recent efforts to document the experience with impact evaluations in some of the infrastructure subsectors (water, rural roads, rural electricity and slum upgrading). This survey builds on some of the earlier survey, adding more recent information when relevant and covering a wider range of subsectors. For a useful complement to this paper for source of information, see the report commissioned by DFID--Jones et al. (2009). It offers an overview of how IE are produced, used to guide policy and to communicate with key stakeholders and cover all sectors quite thoroughly, including infrastructure but focuses on processes rather than policy concerns. Similarly, a report Ito et al. (2008) offers an overview of JBIC's experience with impact evaluation in the infrastructure sectors but is restricted to the lessons of the Japanese experience in the field.

[14] For details and a useful complement, see Alm and Jacobson (2007).

studies that may have generated some of the most interesting results from the viewpoint of infrastructure specialists and inspired the design of some of the field experiments currently underway in infrastructure public interventions.

Consider one the main challenges in the delivery of infrastructure services: the need to ensure enough financing to reimburse the loans obtained by the poorest countries to finance the investments needed to deliver those services. The challenge results from the fact that many of these infrastructure services have some characteristics of public or club goods. Unless a toll is imposed to access a road, no potential users can be excluded from its use. Also, unless the engineers grossly underestimated the traffic on that road, there is no rivalry in the use of that road. It is unlikely that most rural roads will be subject to tolls and to recurring major traffic jams. A similar logic applies to many of the urban services often financed by development agencies.

Another dimension of the financing challenge in infrastructure relates to the externalities associated with some of the services. These externalities are such that in theory, unless the government intervenes, the willingness to pay of the users is likely to reflect the private benefits of the consumption of these services but not the social costs. The provision of clean water delivers private benefits to its users but the increased sanitation costs associated with the increase in water consumption are much more difficult to recover—largely because may users do not relate some of their health problem to the poor management of their waste water.

This free rider concern identified in most public economics textbook has long been part of the discussions of the financing of infrastructure services. The increased case for, at least, a partial financing through user fees or specific taxes has initially been validated, some would argue *stimulated*, by theoretical experiments studying quantitatively the institutional drivers of the risks of financing failures for infrastructure. One of the messages of that research is that, for some public goods including some infrastructure goods and services, tax financing can be less than 100% because users will be willing to cover the difference. Experimental economics provided many of the insights currently used in the design of evaluations. It helped understand when and how the risks of market failures can be minimized, with obviously strong implications for infrastructure.

From a very pragmatic perspective, theoretical experiments have shown that there is no reason to reject any financing mechanism for public goods, at least in the short run, *as* long as the institutional environment in which the projects are taking place has been carefully assessed (including rules for the punishment of free riders). The concern for free riding is real but may sometimes be overestimated in textbooks and by some practitioners in some contexts.

In real life infrastructure activities, voluntary contributions are actually higher than the zero voluntary contribution implied in traditional public economics textbooks. These voluntary contributions include cash but also the willingness to offer free labor so commonly observed in slums and rural areas for the maintenance and rehabilitation of infrastructure investments. This result has been validated in field experiments as discussed later.

The really important contribution of these theoretical experiments, which can easily be underestimated, is how and how much the institutional environment can influence the willingness to volunteer contributions. For instance, the forms of communication between providers and users or

among users matter. Peer pressure, face to face communication, group sizes for instance have been found to be relevant dimensions in the odds of obtaining voluntary contribution in financing public goods or dealing with undesirable externalities. This result has now also been validated by field experiments covered later in the survey.

There is however no reason for overconfidence either. Laboratory experiments are unlikely to be always as precise as field experiments. For instance, Laury and Taylor (2008) found that individuals' contributions to a public good in the laboratory were not always able to predict their actual willingness to contribute to a public good in field experiments. Not totally surprisingly, these experiments also sometimes overestimate the real willingness to contribute. As always, the devil is in the details, in particular institutional details so important to pick up correctly in the design of the evaluations.

Finally, the experimental research validates the intuition that suggests that the willingness to contribute varies across users and may evolve over time. In practice, this means that it is essential to design and regularly reassess financing mechanisms to ensure they maintain the incentive to contribute. Indeed, high initial levels of contributions could diminish over time (observed when experimental subjects are examined over repeated rounds). This implies that experienced users are less willing to contribute than inexperienced users, a crucial concern for the many long lived assets constructed through infrastructure projects.

Ultimately, these experiments conducted in "economic laboratories" have thus been crucial in thinking through the design of field experiments and quasi-experiments. They have been particularly useful in organizing learning through the design of questionnaires and projects as will be obvious in the surveys discussed later.

The insights from experimental economics on the necessity to understand the context and the institutions to assess fully the effectiveness of development actions were validated by researchers working on field experiments and quasi-experiments early on. For instance, Jalan and Ravallion (2003) found that in rural India the prevalence and duration of diarrhea among children under five are significantly less on average for families with piped water than for families without it as somewhat expected. But, just as importantly, they also found that these health gains largely bypass children in poor families where educational levels tend to be lower because the less educated mothers capture fewer of the health benefits of these interventions. These results point to the necessity to consider combining, in the design of water projects, investments with effective public action to promote basic health education and income poverty reduction. It may not always help, but then it may sometimes help as in the example presented here.

Some of these insights may now seem relatively standard in the design of infrastructure projects. But those of us who have been working in the field for over 15 years know that most of our earlier projects did not really address these concerns systematically or carefully.

4. How to learn from the field evaluation experience so far?[15]

There are two broad ways to learn from the experience with evaluations so far. The first deals with the nature of the interactions between evaluators and infrastructure specialists. The second deals with some of the quantitative information that can be extracted from published evaluations.[16]

Learning from interactions with evaluators

From the perspective of infrastructure specialists, the key is probably to focus on the dimensions from experience for which their role is essential. They understand how the sector works. They know better than anyone the incentive and other institutional problems faced routinely in the sector. It is essential that this knowledge be built into the evaluation to ensure that the right interventions and the right range of outcomes are considered.

We now have collectively a lot of information on how the multiple inputs and the multiple outcomes of infrastructure interventions have been dealt with. The proper identification of the key inputs and outputs are the result of a close collaboration between infrastructure specialists and experienced evaluators. Learning from these collaborations should thus be helpful to any infrastructure specialist interested in generating robust evidence of causality between an intervention and its outcomes. These earlier experience of close collaboration should also show infrastructure specialists how they should influence evaluators.

For instance, infrastructure specialists could probably teach a thing or two to anyone about patience, since the payoffs to many infrastructure interventions tend to be slow to show up. This is quite clear in the case of transport projects as discussed later. Any evaluator expecting quick evidence in some common transport interventions is likely to be disappointed. It would be disastrous for the poor to conclude that these projects have no impact, simply because this impact does not show up fast enough. Ensuring that this dimension is built in the design of the evaluation seems to be basic but it is essential to avoid rejecting an effective intervention.

Another issue on which the close collaboration between the infrastructure specialist and the evaluator is crucial is the assessment of the area of influence of an intervention. This is easy for very local projects, but in the case of large projects it is essential to figure out how far from the project zone the evaluators should collect data on possible outcomes. Some energy and transport projects can generate benefits quite far from the zone in which the intervention is taking place.

Finally, the distributional impacts that need to be identified may arise in fairly complex indirect ways. In this case, evaluators with enough field experience in the sector should be able to build that into a design of the evaluations. This can only be done well if the collaboration with the

[15] See Jones et al. (2009), for a helpful complement to this paper offering an evaluation of evaluations across sectors, including infrastructure based on a questionnaire sent to practitioners. A key point, from the viewpoint of this survey, these authors made is that most infrastructure evaluations tended to be qualitative in their sample.

[16] There is a large number of ongoing infrastructure evaluations in progress at IFPRI (sometimes on behalf of the World Bank for its own projects) and the IDB for instance, but there is too little public information available to do them full justice in this survey.

infrastructure specialist works well. Evaluators typically tend to rely on the infrastructure specialist to identify the full range of consequences of an intervention.

Overall, figuring out *how long* the evaluations should take, *where* it should take place and *how wide* the range of its distributional implications should be is central to the collaboration between evaluators and infrastructure specialists. The answer to these questions drive the economic benefits (and sometimes the economic costs) of an intervention. Learning from field practice in these areas should thus help ensure that the causality between the infrastructure projects, programs or policies and the promised outcomes is established properly.

Learning from published evaluations

Although the number of evaluations for infrastructure interventions is small than for some other sectoral interventions, there are enough to draw some useful lessons to ensure sound quality standards in future evaluations. In a nutshell, the main relevant lessons for infrastructure specialists to learn from (and hence to cover in this survey) are those that answer the following three questions:

1. *What are the main questions addressed by the evaluation?*
2. *What are the main inputs, outputs, outcomes and impact being monitored??*
3. *What are the main conclusions to be drawn on the main project, programs of policy design and the associated institutional variables of the projects?*

The obvious start to any survey of evaluations aimed at infrastructure specialists should cover how evaluators have ensured the explicit recognition of the causal relationship in any PPP. In practice, this is achieved by spelling out a clear question or set of questions the evaluation is supposed to answer. It is quite essential to recognize that the policy questions that the experiment tries to address are simply the formal expression of a concern with the effectiveness of one or several interventions in terms of its expected quantitative effects, mainly on poverty related outcomes but also sometimes in terms of other dimensions such as trade level or diversification as in the case of transport projects.

These questions can reflect the wide range of expectations that can be assigned to an intervention. For instance, these questions range from: Is one type of subsidy more cost effective than another? To: Is one form of decision making process to share the responsibility to maintain a local investment more cost effective than another? Or: Is giving people the choice of quality of service more effective at getting them to use a service with a potentially high poverty reduction effect than setting a standardized quality?

It is thus useful to review these questions for each sub-sector to get a sense of how sound evaluations have dealt with relatively common policy concerns across subsectors. Obviously, some questions will differ across intervention and across sectors, since each intervention has to be tailored to the specifics of a country, a region or a community. But there are common elements that can be picked up in each sector, and we will try to identify them for each subsector.

In terms of the measurement of the key variables required by modern evaluation techniques, the challenges are also easy to underestimate. Most evaluators identify simple indicators for the

widely expected poverty reduction outcomes. But best-practice evaluations are showing that it is also quite relevant to identify the effects on the key channels to achieve poverty reduction such as job creation, educational achievements and overall health conditions as well to get a sense of the outcome and impact indicators one would look for in the design of the experiments.

The set of indicators typically covered by infrastructure evaluations can be grouped as follows. The allocation of funds to finance an infrastructure is the basic input. The infrastructure built is the main output (a road, a generator, a water pump, …). But these are only intermediate indicators that do not say much about impact. The outcomes are typically indicators that reflect in a very concrete way how life is changing for the communities involved (improvements in the quantity and/or quality of a service such as time to go school to a clinic or to get water). The impact are changes in the levels of health, educational, employment, income or poverty levels, as discussed later, due to the specific intervention or set of interventions analyzed (an investment with or without associated policy or institutional changes).

The outcome and the impact indicators can be quite challenging to identify when the focus is on the long term. As mentioned earlier, one of the main challenges results from the fact that the main payoffs of many infrastructure interventions are really slow to come. For instance, road investments are based on traffic forecasts to 20-30 years horizon to ensure long term cost minimization. This makes the causal relationship harder to establish from short term experiments. This is essentially why the policy questions to be addressed by the experiment are much more challenging to define as discussed below. It is also why transport evaluations based on field experiments are much rarer than water or even energy evaluations for instance where short terms payoff abound.

Ultimately, the set of experiments reviewed here suggests that the outcome indicators are at the most visible part of the measurement exercise. They are thus crucial from the accountability viewpoint and, for that reason also, to compare across experiments. This is what many in bilateral aid agencies pushing for increased evidence of aid effectiveness will be doing. There is a wide range of relatively standard outcome indicators influenced by the set of data most commonly available from regularly scheduled surveys in countries. These are often easy to complement through additional field survey work. They also often appeal to the evaluators as they are often close enough to the sort of concerns addressed in the context of a sector and hence facilitate the communications between the evaluator and the infrastructure specialist. For instance, health related outcome indicators are generally common for water evaluations. Time of travel to work, school or clinics are just as common in the evaluation of roads for instance. They are all easy to include in the design of the monitoring surveys conducted as part of the experiments to monitor progress and impact.

Having the benefits of a wide range of standard indicators to pick from does not mean that there has to be a competition for improvements in every outcome. Not all projects should be expected to have the same impact on every outcome. Moreover, not all outcomes are that easy to pick up from simple indicators. When it comes to the growth, trade or productivity effects of an intervention, simulations through general equilibrium models or econometrics based on past observations may be more appropriate to generate more robust approximations of relevant outcome indicators.

These insights on how best to design an intervention to maximize its effectiveness are not the only ones offered by the experience of field work so far. There are also useful technical aspects to learn from in the infrastructure evaluations conducted in the last 5-10 years. For instance, why and how the econometric techniques have been selected and implemented—how to deal with the possible relevance of the initial conditions, how to deal with the non-randomness (the recurring endogeneity concern voiced by the evaluations technicians) of the selection of some projects (typically roads), how to pick instruments to ensure reliable econometric results, how the surveys have been designed to generate the information needed to establish causality between intervention and impact, how much consultation took place to ensure the support for the exercise and maximize its payoffs.

Although these technical lessons are essential, they are typically dealt with by technical experts (mainly econometricians) who want to ensure there is no systematic upward or downward bias in the assessment of the impact of a infrastructure project, program or policy on monitored outcomes, in particular in the benefits of the poorest. These lessons are all important but go beyond the scope of this note.

The rest of this paper focuses on these three broad sets of lessons directly relevant to infrastructure specialists—*i.e. main questions, main indicators and main conclusions*. I focus first on water and sanitation, the sector the best covered by evaluations based on randomized experiments. I then discuss the very hybrid evaluations experiences offered by the transport sectors and its various subsectors. I conclude with a review of the few evaluations that have taken place for electricity.

5. Lessons from Water and Sanitation evaluations

As summarized by Poulos et al. (2006), most WSS interventions have at least one of the following three expected outcomes:

1. *efficient access to safe drinking water and/or basic sanitation services;*
2. *sustainable access to safe drinking water and/or basic sanitation services; and*
3. *equitable access to safe drinking water and/or basic sanitation services.*

These are quite well known goals which have motivated a wide range of intervention for the longest time. The real contribution of modern evaluation techniques has been the linkage of these goals with its welfare impacts.

Most evaluations focus on the impact of the interventions on health outcomes such as infant and child mortality, nutrition (weight at birth and later), and childhood diseases like diarrhea (e.g three or more loose stools over 24 hour period/ in last two days or two weeks among children under 5 years) but also acute respiratory infection – e.g. incidence of cough and cold in last two weeks among children under 5). Some look at the impact on education (whether children attend school regularly – e.g. number of days children in sample attended school in the last month). Others use indicators such as time savings, social inclusion, social capital, political participation, women privacy of water use, safety of water use perception. Finally, many focus on synthetic or representative indicators of welfare effects. These include income or consumption levels such as household per capita income

and consumption, expenditures on household water treatment, water storage containers or expenditures on medical treatment.

These impacts are the expected product of six groups of interventions typically found in this sector:

- *water supply coverage and operations improvements* (e.g. investment and maintenance planning, commercialization)—which is the standard impact engineers and economists working on the sector have been focusing since evaluations started to be published
- *water treatment improvements* (clarity, odor, taste, and control for bacteriological and chemical contaminants, point of use of water treatment and other source water treatment)—which have also been studied for a while
- *sanitation improvements* (e.g. disposal and management of community, industrial, agricultural and household liquid and solid wastes through drains, garbage dumps and wastewater treatment facilities, hygiene improvements (e.g. hand-washing with soap as discussed in the review by Peterson-Zwane and Kremer (2007)),
- *education* (e.g information campaigns—see Jalan and Sommanathan (2004) in India on the importance of education but also Peterson-Zwane and Kremer (2007)) for a useful survey of evidence of the relative impact of this intervention in relation to other interventions in the sector)
- *policy changes (e.g. tariff reforms, subsdies programs, support to technological innovations as discussed for instance in Estache, Foster and Wodon (2002) for the impact of tariff or subsidy structure changes or Kremer et al (2009) for the impact of various price based strategies to increase use of dilute chlorine tested in rural Kenya)*
- *institutional changes* (e.g. privatization—Galiani et al. (2007), decentralization (Rawlings et al (2004) or social norms (Kremer et al. (2009)).

The published evidence on the extent to which these goals have been achieved and the resulting impact obtained is starting to become really impressive. This large volume also allows researchers and field workers to benefit from surveys and meta-analysis conducted from the viewpoint of different groups of stakeholders (health specialists, health economists and development economists). At least 6 recent reviews of water and sanitation interventions have been produced over the last 4 years or so. The World Bank Dime initiative published one (Poulos et al. (2006)), the World Bank Evaluation department, published its own (IEG(2008)), four sets of academic researchers also published one (Fewtrell et al (2005), Clasen et al. (2007) and Peterson Zwane and Kremer (2007)) and a new think tank focusing on impact evaluations is working on its own and has released early results (Waddington et alt. (2009).

The Waddington et alt (2009) work, also the most recent survey, identified 110 studies evaluating water related activities and almost 300 papers in English, French, Chinese or Spanish. This is a testimony to the importance of the concern for impact evaluation in the water and sanitation sectors—even if Waddington and his colleagues focused on "only" 84 of these studies in their meta-analysis to draw their conclusion, their sample still covered 34 low and middle income countries.

According to the Waddington survey, studies vary from 6 to 19 months in duration of collection of water related disease data, with their average sample sizes varying from 327 for point of use treatment to almost 6000 for water supply. All studies found some impact for each intervention type but there was a significant diversity of efforts to be precise across studies.

Unfortunately, their survey also suggests that the sustainability of gains tends to be limited—although more results are needed to be able to validate this concern.

To get to assess the impacts, the most recurring questions are the following:

- *Does improved water supply, sanitation or hygiene improve health?*
- *Does improved water supply, sanitation or hygiene improve educational outcomes?*
- *Does improved water supply, sanitation or hygiene improve income?*
- *What is the welfare impact of costs of providing water/sanitation/hygiene?*
- *Are the cost savings from improved productivity passed on to consumers?*
- *Does improved water supply increase business activity (SME or other)?*
- *What is the impact of water sources (e.g. spring) protection on child health and anthropometrics?*
- *How do improvements in basic hygiene education impact health outcomes?*
- *How does the increased involvement of local communities in water and sanitation-related decisions impact health, education and income outcomes?*
- *What are the behavioral changes (e.g. in terms of transportation frequency, boiling or storage) of improvements in water access or water source protection?*

Considered jointly, the quantitative and analytical answers to these questions offered by studies validate many of the common perceptions regarding the desirability of meeting the Millenium Development Goals (MDGs) soon. Water and sanitation can easily be connected with desirable health, education, nutritional, employment and income outcomes. Not much news to report on that front. But these studies also reveal some variance in the effectiveness of the interventions aimed at reaching the MDGs. For instance, unless all connections come from piped water, water supply interventions tend to be less effective in terms of health (although they can help save time) than water treatment at point of use interventions or than many sanitation and hygiene interventions. These differences in effectiveness across intervention types and within groups of interventions yield the most interesting lessons from these evaluations, explaining in particular why interventions effectiveness varies across locations.

So, why are interventions not equally effective across locations? One of the reasons is that not all interventions are comparable. Indeed, one lesson of the recent more technical evaluation experience is the recognition of the need to be precise in the assessments of impact from physical interventions. Assessments need to reflect quality of water and quality of service and not just the quantity resulting from the intervention. In practice, interventions deal with a variety of interventions (including multiple quantity and quality options) which tend to address multiple goals (e.g. residential, agricultural, industrial, medical, educational, …) and do so relying on multiple forms of interventions (investment, policy, education, institutions, …). In sum, evaluations need to recognize the multidimentionality of objectives, interventions or outcomes as suggested by Poulos et al. (2006). Impact evaluations dealing with multiple concerns are simply somewhat more complex but deliver just as important results as simpler evaluations (see for instance Labonne and Chase (2008) in their evaluation of CDDs).

Second, evaluations also show the need to be quite specific about technology whether the focus is on water supply, treatment or sanitation. As already mentioned, if there was no budget constraint, piped water and systematic collection and treatment of waste water would be perfect and technology choice would not be an issue. But budget constraints are pressing, water pollution is increasing fast, and mortality rates continue to be high in villages not investing in low cost technologies simply

because they keep being told that within the next two years they will get piped water. These next two years never deliver the promised technology and this is why too many children continue to die, many more miss school and reduce their future earning opportunities while adults. This is also why, often, women continue to walk long hours to get water, losing the opportunities to earn income as well.

Third, evaluations document the importance of social norms in maximizing the efforts to improve hygiene and in ensuring the cooperation needed to guarantee the sustainability of interventions in the sector. Kremer and his various co-authors (2008, 2009) provide ample evidence of this importance in terms of hygiene, and Labonne and Chase (2008) illustrate this message in the context of improvements in social and political capital in a case study of the Philippines.

Fourth, the assessment of the impact of the policy and institutional changes has provided some of the most surprising and debated results on the importance of the policy and institutional context in which the evaluation is conducted. The most widely quoted evaluation concerns the impact of privatization on health in Argentina (Galiani et al. (2007)) which shows that, for that specific case study and the specific impact assessed, this type of reform was positive. It had health payoffs typically ignored in standard assessment of privatizations which tend to focus on their fiscal and employments effects. This result is in contrast with the conclusions reached on a quasi experimental assessment of water privatization in Argentina, Bolivia and Brazil conducted by Clarke et al (2004) focusing on other dimensions. More specifically, they find no differences in increases in access rates to water under public and private options—and which of course does not say much about the health impacts of interest to Galiani and his co-authors.

Fifth, the experience with evaluations suggests that fair assessments of impact depend on the full recognition of spillover effects. Positive externalities such as reduction in the spread of some health problems resulting from local efforts to deal with water quality are often quoted. But negative externalities associated with interventions can also be important. This is particularly true in the context of water and sanitation since an improvement of access to one group of users may hurt access to another (e.g. the recurring disputes for scarce water resources between farmers and non-farmers and the recurring concerns with increased water source pollution from intensified used of fertilizers or from fast industrialization which creates jobs and hence helps reduce poverty).

Finally, evaluations have been quite effective at showing that educating water users can have high payoffs as well but that the form of education matters a lot more than many field workers sometimes recognize. The effects can be very different if the knowledge comes from peers or if it comes from common formal training for instance. There is however no clear ranking of approaches since many other factors need to be taken into account. (Jalan and Somanathan (2004) or Peterson-Zwane and Kremer (2007) for instance).

Given the broad array of possible interventions, it is hardly surprising that there are so many studies and that there is such a large diversity of assessed outcomes around a relatively stable mean positive answer to the efforts made to reach the MDGs in this sector. Not all is perfect however since many types of common interventions in the sector could do with a stronger evidence base to inform policy (e.g. which water delivery and treatment technologies have the highest payoffs in terms of health outcomes, wasted time, income, sustainability?)

6. Lessons from transport evaluations

The transport sector typifies the challenges facing evaluators of infrastructure interventions. While many small scale or rural transport projects can be evaluated using real or quasi trials, not all transport projects lend themselves to these techniques. Randomization is not obvious for large scale transport project that capture a large share of international resources (highways, ports, airports, railways).

Why is it so hard to find randomized experiments in transport? To perform a purely randomized experimental approach, we would need two similar areas in terms of their geography and economic situation. For instance, in one area, a new highway should be constructed while in the other no highway would be constructed and the allocation of the highway would be based on a lottery. This is not a common decision mechanisms in this sector.

To deal with the need for randomness, one could think of a sectoral strategy as an opportunity to monitor quasi-experiments, in which some cities or regions get "treated" with a transport infrastructure investment aiming at helping the poor while others are not "treated". We could then infer effectiveness of an intervention based on that comparison between the sub-samples as one can do for quasi random allocations of water interventions. This is an approach unlikely to be implementable. It can be done for some rural roads interventions and some bus projects that may require public financing when service obligations are too costly to be passed on to the users of the services and projects need to be delivered in sequence. But these sorts of interventions are not the norm in the sector. Moreover, even in the case of a possible randomization of the implementation of activities associates with service obligations (i.e. the obligation to have a stop with all the associated infrastructure in some cities or neighborhoods), there is often no reason to expect the poor to be the main expected beneficiaries. For instance, they may reflect the concern to service large groups of workers, rich and poor, that need cost effective mobility from the suburbs where they live to the center of the city where they work.

An additional characteristic of transport projects that makes them tough candidates for impact evaluations reflects some common ways in which their benefits and costs need to be internalized early on in the design of the projects. On the benefit side, investments are based on demand forecasts with 20-30 years lead time. This is because it is a lot more expensive to widen a road when traffic has increased enough than to build-in the high traffic levels in the initial design of the road. These sorts of projects only have extended release payoffs, starting slow and peaking late since most of the bulky investments associated with transport infrastructure are based on their long term needs. On the cost side, trial oriented evaluations tend to have a hard time dealing with the uncertainty due to structural variation in the infrastructure over time (e.g. uncertainty about road degradation and variation in future capital road works strategies that may be needed). It requires an evaluation approach capable of separating the physical and the policy components of the project over the long run.

Being able to assign proper credit or blame to a specific investment is actually even much more challenging than those implied by simple economic concerns! Ports, railways or roads are often built for other complex reasons that influence the impact of the investment and which the

investment itself influences. For instance, opaque political negotiations and coalitions between regional governments trading support for each other investment projects that require national financial support are quite common. Similarly, concerns for military risks associated with investment otherwise aimed at international integration have often resulted in the rejection of economically and socially sound roads or railways.[17] It is hard to argue in those cases that investment programs have been allocated by "pulling a yes or a no" out of a hat to ensure a random allocation of investment resources. It is also hard to establish causality between the investment and the poverty alleviation outcome one may be hoping to document.

There are feasible approximations to obtain an evaluation (propensity scores) but they are not simple either. The most common practice is to rely on general equilibrium and other structural models. These can be used to run before and after simulations of traffics and their impacts on jobs, productivity, education or poverty based on demand and supply elasticity to transport prices (and time) observed in other similar projects than on natural experiments or quasi-experiments.[18] These allow various types of simulations ex ante to get a range of possible outcomes and ex post to test and compare various mechanisms through which poverty related impacts can be obtained. An alternative is rely on detailed economic cost benefit analysis accounting for distributional weights explicitly (both for ex ante and for ex-post evaluations). This is the most common approach in urban transport but it goes beyond the scope of this survey.[19] A third approach is to rely on spatial analysis techniques which, thanks to improved access to data, are providing new insights on the impact of transport interventions. [20]

In sum, while a transport infrastructure investment may indeed contribute to reduce poverty by cutting the costs to get to a job, a school or a medical center, it may achieve this objective through many possible mechanisms which are all location and context specific. Evaluations based on randomized experiments may help learn and anticipate social outcomes as seen later in the review of the sector experiences. However, more often than not, evaluation such as those of large scale transport infrastructure projects may be best conducted through macro models with detailed modeling of household consumption patterns and income sources for various quintiles or deciles. The rest of this section provides an overview of the best practice approaches that have used for the main sub sectors benefiting from development financial support: rural roads, highways, railroads and ports.

[17] Some obvious cross national roads opportunities have often not been implemented because there was and still is a concern that it would make military invasion easier for a neighbor. Many transport experts working in the Southern cone of Latin America argue for instance one of the reason why a rail or road tunnel in the mountains between Chile and Argentina near Mendoza was never built.

[18] It is not surprising to see Deaton use a railways project to illustrate many of the econometric challenges faced by an evaluator.

[19] For a recent review which shows how little real impact evaluation of poverty effects using modern evaluation techniques is taking place in urban transport, see Boarnet (2007). Most of the work somewhat relevant to this survey consist in survey of the travel patterns of users of public transportation per income classes. But none really assess the distributional impact of various types of interventions systematically in a statistically robust way. Good examples of this type of work are Barone and Rebello (2003) on Sao Paulo and Baker et al. (2005) on Mumbai

[20] See for instance Dorosh et al. (2008) or Lall et al. (2009)

Rural Roads

The most honest start to this overview is to refer the reader to another author! Indeed, Dominique van de Walle (2009) offers a very thorough overview of the technical dimensions of impact evaluations of rural road projects. The very first observation of her paper is that very few of the many aid-financed rural road projects in developing countries have been subject evaluations. She could have added that the same applies to highways or small road networks. The reason is simple: they are simply hard to do using (quasi-)randomized evaluation techniques.

The most challenging characteristic of road projects in terms of the techniques approximating random trials is that they have no natural comparison group. It is indeed hard to find two similar regions in all the relevant characteristics such as the initial conditions in the composition and level of production activities, composition and levels of skills of workers, the number of users, access to other transport modes, access to schools or any variable that may influence the evolution of the derived demand for the road and hence the comparability of the evolution of regions with and without the road project. In addition, evaluators also have a hard time addressing all relevant spillover effects as well as time dimensions associated with many road PPPs. This is why it is still common to see assessments of the impact of rural roads interventions conducted through general equilibrium modeling.[21]

Despite the challenges, there are a few well known paper delivering top quality evaluations: Jacoby (2001) on Nepal, Jalal and Ravallion (2002) on India, van de Walle and Cratty (2002) and van de Walle and Mu (2007) on Vietnam, Gibson and Rozelle (2003) on Papua New-Guinea), Lokshin and Yemtsov (2005) on Georgia,), Escobal and Ponce (2004) and Escobal (2005) on Peru, Khandaker et al. (2006) on Bangladesh and Dercon et al. (2007) on Ethiopia.[22]

An overview of these evaluations suggests that the main questions asked can be summarized as follows:

- *Do rural roads reduce poverty?*
- *Are locations or people with favorable initial conditions in terms of development and access to better institutions enjoying stronger or weaker impacts?*
- *What is the mechanism by which roads improvements can be translated into improved living conditions and lowered inequality?*
- *Do time savings from improved roads used improve the living condition of women?*
- *Do rural roads ease labor mobility?*
- *Do rural roads lead to switches in economic activity (agricultural vs non agricultural?)*
- *Do rural roads improve educational outcomes?*
- *How do the local initial conditions affect impacts of road investments?*
- *What investments (i.e. complimentary infrastructure, productive activities) and services (i.e. transport options) interact with roads to improve development outcomes?*
- *What management schemes are most effective for maintaining good-quality roads?*

[21] See Fan et al (2005a, 2005b) on the cases of China and India.

[22] Each of these papers shows how the choice of instruments in the econometrics part of the paper are driven by local data availability.

- *Do rural roads influence the modal choice (bicycles, cars, buses, motorbikes, animal drawn vehicles, …)?*
- *Do rural roads improve the development of small firms?*

The indicators used to assess outcomes are quite standard. They include traffic, transit time, transport costs, accidents, local growth, local income level, household access rates to various social services, academic results, food availability, disease data, mortality, job creation, and number of SME creation.

The overall policy message of these papers is quite robust. Rural roads provide substantial benefits to households in low-income countries, especially the poorest. But not all roads beneficiaries get the same benefits. There is indeed a wide range of outcomes, including situations in which a specific outcome is present in one project and not in another one within the same country. Moreover, they also show that rural roads are not a panacea for poverty alleviation and the mechanics of poverty alleviation can vary quite a lot across projects.

The ability to decompose what component of infrastructure spending generates which benefits under what kind of circumstances is one of the most helpful policy contributions of these studies. It may be useful, for example, to know the extent to which the composition of spending on road investments makes a difference to outcomes. Is it more important to pay the road workers well or to improve significantly the quality of the gravel used?

A powerful illustration of the way in which crucial institutional dimensions in the sector can be analyzed is offered by Olken (2007). He studied how to reduce theft and graft in public works projects by doing controlled field experiments in 608 Indonesian villages. Some village leaders involved in the building of roads were told that, upon completion of the project, they would be visited by government auditors. Other villages were chosen to participate in grassroots "accountability meetings," during which project coordinators would publicly account for the use of government funds in a town-hall-like venue. Villagers would be offered anonymous forms to report graft.

Olken concludes that audits cut missing expenditures, as measured by discrepancies between official project costs and an independent engineer's estimate of costs—by 8%. Increasing grassroots participation in monitoring had little average impact! The overall policy implication is that in some settings, traditional top-down monitoring can play an important role in reducing corruption and hence improve the use of resources in road maintenance. A strong message when deciding how best to ensure the long run sustainability of commitments made to maintain a new or a rehabilitated road!

Highways.

The large and distant spillovers of highways render the use of experimental approaches even less attractive in the case of highways and other large traffic roads. To perform them properly we would need two similar areas in terms of their geography and economic situation. In one a new highway should be constructed while in the other no highway were constructed. This comparison is very seldom possible. But other evaluations techniques can help.

A fair number of studies have been relying on quasi experimental approaches to assess the impact of highways in the US. The set of questions asked are not very different from those identified in the case of rural roads. They also use very similar indicators. The outcomes analyzed are transport costs, jobs, income, tax revenue, productivity of firms, diversification of firms and changes in the types of vehicles used. There are however interesting differences across the few studies available that deserve a place in this survey.

The first paper credited with a quasi experimental approach to the evaluation of the impact of highways is by Rephann and Isserman (1994). They assess the impact of highways construction on counties, comparing the outcomes for those who got an access to that in counties who did not get one. Using a creative (3 steps) experimental approach, they conclude that in the US, the benefits of the interstate system are concentrated on the areas close to large cities or with a high degree of urbanization in the pre-treatment period. Isolated rural areas and areas close to the interstate network do not receive any benefit.

Chandra and Thompson (2000), using a different experimental approach analyze the relationship between interstate highway construction and the level of economic activity. Dealing with a crucial technical concern (the endogeneity of the location of new roads), they find that in the case of metropolitan areas, roads are related to the past economic performance but that it is unrelated for non-metropolitan areas. They use this exogeneity assumption to justify a quasi-experiment in which the treatment group is the non-metropolitan US counties that received an interstate highway and the control group is the non-metropolitan counties that never had an interstate. They also study the effect of a new highway on areas that are close to the counties that received the highway but the infrastructure does not cross their territory. They show that non-metropolitan counties that received a highway experienced an increase in earnings compared to counties where the highway did not cross through. But counties that where adjacent to highway counties suffered a reduction in retail trade and government earning.

Holl (2004) analyzed if the municipalities that are closer to a highway attract more new business than the ones further away. He also argued that highway construction can be assumed to be exogenous to changes at the municipality level because the decision about the route of the highway is taken at a higher governmental level. Using this exogeneity assumption, he explains the number of new manufacturing establishments on proxies for intra and inter-regional demand accessibility, supplier accessibility and distance to the closest highway. He finds that highways affect the spatial distribution of new manufacturing establishments increasing their number in municipalities close to highways. He also finds sectoral differences in the attractiveness of municipalities close to highways.

Michaels (2008) considers the possibility that political or economic conditions may have affected the specific placement of highways in contrast with the original design. Since changes in trade openness are typically confounded with other factors, it has been difficult to identify the labor market consequences of increased international trade. The advent of the United States Interstate Highway System provides a unique policy experiment, which he uses to identify the effect of reducing trade barriers on the relative demand for skilled labor. The Interstate Highway System was designed to connect major metropolitan areas, to serve national defence and to connect the United States to Canada and Mexico. As a result, many rural counties were also connected to the highway

system. The outcome was an increase in trade-related activities, such as trucking and retail sales, by 7-10 percentage points per capita. Most significantly, by increasing trade the highways raised the relative demand for skilled manufacturing workers in counties with a high endowment of human capital and reduced it elsewhere.

The case studies reviewed suggests that all empirical evaluations have been conducted for the US. Many have, but for an unknown reason, many have been replicated for Spain and Spanish researchers have also contributed new approaches. Recently, Garcia-Mila and Montalvo (2008) have relied on a quasi experimental technique to evaluate upgrades of a road to a highway. The basic idea is to compare the number of new establishments in the catchment area of the roads transformed into highways/dual carriageways with respect to the roads not upgraded. They rely on a quasi-experimental approach because the transformation may not be independent of the characteristics of the areas transformed. They divide the Spanish national roads/highways system into 20-km long segments. Then, they use the GIS location of each new firm to assign it to the catchment area of one of these segments. They next compare the number of new firms per squared kilometer in the transformed and untransformed segments. They find that the upgrade is economically important but statistically insignificant since it did not have an additional attraction effect of firms.

This range of applications and the recognition that they have been replicated is good news. Their main residual drawback is that they are only helpful to conduct ex-post evaluations. For evaluations that require both an ex-ante and an ex-post assessment, general equilibrium or engineering models with added social modules are the norm for highways. They focus on the same questions and on the same outcomes indicators as any other evaluation, but they rely on modeling of the structure of the economy in which the demand for transport and transport costs are explicitly addressed for each economic agent of interest, including the various income classes if needed.

Most large general equilibrium studies assessing roads, highways or road networks consist of multi-sector static-comparative regional studies with a focus on spatial aspects of road and highway projects.[23] The latest generation consists of macro-level intertemporal simulation studies (Rioja 1999, 2003a, 2003b and Kim, Hewings & Hong 2004) The static-comparative simulation studies, typically, evaluate specific highway projects, while the intertemporal simulation studies, typically, aims at computing optimal levels of road and highway infrastructure as well as the optimal levels of commitments to maintain them. Suboptimality (and hence evaluation) is measured in terms of fiscal, jobs and income losses. In general, simulation-based evaluation studies in the economics tradition focus on measuring short and long term outcomes in comparative statics. Some more recent models make an explicit effort to calculated of lifetime benefits, the way the engineering simulation-based evaluation do (e.g. Archondo-Callao (2009)).

While useful to get a sense of the global macroeconomic impact of road project or programs, these studies do not have much to say about the transmission mechanism of the impact estimates. This is why there is now a new generation of model trying to achieve this unbundling, equivalent to the one obtained in randomized experiments. Jensen (2009) introduces an integrated economy-wide simulation framework for evaluation of capital road works for existing roads and highways, which formalizes transmission mechanisms and captures important static and dynamic spillover effects, and thereby allows for measurement of lifetime net benefits. It is tested on Ghana's roads investment

[23] See Jensen (2009) for a really useful overview of all the modeling approaches.

program. It documents the transmission mechanisms between capital road works, road degradation, vehicle operating costs, transport sector productivity, and socioeconomic outcome variables. Its outcome indicators are the usual ones as well except that most are measured in changes in net present value. They include real GDP, household welfare and poverty gap, as well as the increase in employment years and the reduction in poverty years. For Ghana, the author finds that the planned future sequence of capital road works is crucial for (evaluating) the impact of current capital road works. He estimates that dynamic effects may account for over 95 percent of the total GDP impact. Socioeconomic spillover effects may account for half of the total GDP impact.

Overall, three main messages emerge from this overview of highways evaluations. First, "clinical trials" are hard to use in this sector and dynamic CGEs and probably evaluations based on techniques used more commonly in spatial analysis are often going to be more useful. Second, from a policy viewpoint, one of the most recurring messages to come from the last generation of ex-ante and ex-post evaluations of highways is that the expected outcome depends a lot more than traditionally recognized on the long terms commitments to maintain the investments. Maintenance is good for users, but also good for the workers, of poor since most maintenance activities do not require significant qualifications. The third message is only implicit but important. Roads and highways projects tend to have distributional effects that are too often ignored. There are regional winners and losers for many common roads or rail interventions.[24] Moreover, underfunding maintenance favors current generations over future generations while over building to meet the need to future generations get current users and taxpayers to pay more than they need to meet the current needs—unless these roads are financing with long term loans or bonds to be paid back by future generations.

Railways

There are even fewer full economic evaluations of rail projects than there are of roads, all type of methodologies considered. Most evaluations are costs-benefit analysis ignoring distributional concerns. They essentially offer rates of return assessment, and too often only private rates rather than social returns. An an increasing number of studies try to explain the difference between ex-ante and ex-post returns also. Their goal is often to get a sense of the unprogrammed cost to users and taxpayers. [25] There have also been a few interesting studies conducting before and after comparisons focusing on property prices as proxies for outcome. Before-and-after comparisons of these prices have frequently been applied to assess the impact of intra-urban train transit, particularly in studies on U.S. cities (Bajic (1983); Dewees (1976); McDonald and Osuji (1995); McMillen and McDonald (2004) for the US and Gibbons and Machin (2005) for the UK.

A recent paper by Ahlfeldt (2009) focuses on access to inter-city connections. Instead of distinguishing between treatment and control groups, he refer to areas that are positively or negatively affected by an intervention, defined on the basis of whether they experience an increase or decrease in accessibility. He also focuses on property prices but relies on a range of techniques, including differences-in-differences to get to his conclusions.

[24] This is one of the main policy contributions of a recent paper by Polasek et al. (2009) studying the regional effects of new railways infrastructure in Austria for instance.
[25] See for instance Premius et al (2008).

The most innovative evaluations available for now are three recent studies focusing on the railways systems in developing countries using experimental approaches. They deserve a somewhat longer discussion since they are more directly related to the audience of this paper.

The first is an assessment of the benefits of the Indian railways and a discussion of their explanation by Donaldson (2009) as part of his PhD thesis at LSE. From archival data from colonial India, he estimates the impact of India's vast railroad network. He looks at fairly common concerns in the analysis of trade in large countries or trade across smaller countries. He covers the impact of rail on trade costs, on interregional price differences, on international and interregional trade volumes, on the interactions between productivity shock and local prices, on income levels and their volatility and on overall welfare levels. In a nutshell, he looks at impact in terms of trade competitiveness. To conduct his analysis, he follows an approach also used to evaluate ex-post (and ex ante) road projects. He develops a multi-region, multi-commodity, a (Ricardian) trade model, where trade occurs at a cost. Because of geographical heterogeneity, regions have different productivity levels across commodities. This creates incentives to trade to exploit comparative advantage. A new railroad link between two districts lowers their bilateral trade cost, allowing consumers to buy goods from the cheapest district, and producers to sell more of what they are best at producing. He finds that that railroads in India: (1) decreased trade costs and interregional price gaps; (2) increased interregional and international trade; (3) eliminated the responsiveness of local prices to local productivity shocks (but increased the transmission of these shocks between regions); (4) increased the level of real income (but harmed neighboring regions without railroad access); (5) decreased the volatility of real income. All could be used to make a case to consider railways rehabilitation in many countries. The case could potentially be even stronger if one considers some of the environmental consequences of a modal switch to rail in many settings.

The second interesting paper is also part of a PhD thesis (Harvard) and also focuses on the impact of railways on trade performances, adding assessment of the impact on firm productivity and on growth. Sandra Sequeira's quasi-experimental study actually focuses not only on the impact of railways but also on the mechanisms through which transport infrastructure affects firm behavior, trade, investment and growth. To measure the causal impact of transport costs on the performance and geography of business, she relies on a quasi-experiment based on the building of a railway connecting the industrial and agricultural heartland of South Africa to the Port of Maputo in Mozambique. Adopting a difference-in-differences strategy to compare economic outcomes in regions close to the railway to those farther away, before and after the investments in infrastructure took place, she focuses primarily on indicators of firm and industry-level productivity, export behavior and growth, collected over a three-year period for 1,600 firms that started operations long before the railway began to be built. This is work in progress, but it provides already a great model of the sort of questions one wants to deal with in these evaluations and which indicators to focus on.

The third is a preliminary paper by Banerjee, Duflo and Qian (2009) estimating the effect of access to transportation networks on regional demographic and economic outcomes across counties in China during 1986-2003. They go beyond the trade related impacts of the two other papers. They also try to assess the effects of greater factor mobility, better access to education, health care and finance, and other effects of diffusion of ideas, technologies, etc. Their results are still preliminary and somewhat surprising. They do not find a significant effect on GDP levels, population, or the composition of population. However, with a few important caveats, they find a distributional impact

across space from distance to railways. On average, increasing distance from railroads by 1% decreases annual GDP growth by 0.12-0.28% across sectors. In other words, a policy which "randomly" places transportation infrastructure will have a positive economic effect on those areas. Their data does however not allow them to say much more about the social or private return on investing in transportation infrastructure because we have no idea of the relevant costs. The result is however provocative enough to fuel new ideas in the debates on the necessity to relaunch railways on behalf of environmental concerns.

Once more, one of the benefits of this review of research is practice is to identify the sort of questions of interest to evaluators. The current bias towards trade integration and trade growth associated with railways project may be biasing the current selection, but it is representative of the motivation that tends to drive railways projejcts in developing countries. The main questions identified can be summarized as follows:

- *Do investments in railways (or any transport infrastructure) reduce transport cost to firms?*
- *To what extent and how do transport costs influence the prices, productivity, employment generation and export propensity of firms?*
- *Does increased competition between transport corridors increase efficiency in the transport system and reduce transport costs for firms?*
- *Do reduced transport costs improve regional growth?*
- *Are investments in rail and road complement or substitutes?*
- *What is the nature of spillovers or network effects across transport modes or corridors ?*
- *Do investment in infrastructure lead to uniform reductions in unit transport costs across firms and commodities?*
- *Is there variation in the firms' reaction to reductions in transport cost?*
- *Do changes in transport costs affect firm dynamics regarding the pattern of investments, exports, factor productivity, entry and exit from the market?*
- *What are the mechanisms though which reductions in transport costs influences firm's productivity, investment and growth?*
- *What are the distributional effects of transport infrastructure interventions?*

Clearly, if these questions can be addressed in the context of railways projects, they should also be relevant in the context of ports and roads. Since demand for these transport infrastructure is often a derived demand, it is not surprising that many of the evaluation questions will try to look at the extent to which they improve competitiveness by cutting costs and from there impact the poor through increases in demand for labor. It is more surprising to see the extent to which the distributional of railways may actually be important in some cases than the static and dynamic efficiency much of the literature has tended to focus on.

Ports

Ports generate cargo and produce tradeable services that attract ships and shippers. This creates jobs, including jobs requiring minimum skills. The low skilled jobs help the poor. That's what port economists have been arguing for the longest time. The link with poverty is only indirect and this may be why there does not seem to be much interest in the port community to conduct impact evaluation of their industry that address the distributional impacts. They are more interested in

understanding what could be done to cut their costs in comparison to the costs of their competitors. Most evaluations of impact on costs and from there on growth have been conducted through structural models simulating causality or partial equilibrium models which under today's econometric standards would often be criticized for weak instrumentalizations.

Quasi experimental methods could actually provide some insights on what to do to cut costs and enjoy all the desirable derived payoffs. Cutting corruption is one of them. Sequeira and Djankov (2008) and Sequeira, Djankov and Mullaninathan (2009) investigate how bureaucrats set bribes and whether these payments impose significant economic costs based on an original dataset on bribe payments at ports in Southern Africa. They find that bribes can represent up to a 14% increase in total shipping costs for a standard 20ft container and a 600% increase in the monthly salary of a port official. Bribes are paid primarily to evade tariffs, protect cargo on the docks and avoid costly storage. They find that various distortions due to corruption. The most relevant here is a "diversion effect" where firms go the long way around to avoid the most corrupt port. This is one of the key consequences that could have consequences on the demand for labor and hence on the poor. Overall, the evaluation rejects the idea that corruption has no allocative effects and that it is simply a transfer a rent between economic actors. Unfortunately, the paper does not go very far in the discussion of possible interventions and of their effectiveness in reducing these costs. But this does not diminish the fact that the distributional implications of the lack of accountability in the sector also has efficiency implications which hurt jobs and hence the poor. In other words, an intervention that reduces corruption should be pro-poor.

7. Lessons from Electricity

As in the case of water and roads, impact evaluations tend to focus a lot more on rural populations. As in the case of rail or ports, there are very few publications on the impact of electricity interventions! There is a fair amount of ongoing evaluations (e.g. Afghanistan, Bangladesh, El Salvador, Ethiopia, Mozambique, Pakistan, Peru, Tanzania, Vietnam …) but it is too early to draw major conclusions from these projects. There is however enough information on the sort of questions being asked by the evaluations, the outcome and impact indicators of interest to the evaluators.

Why so few evaluations in this sector? Once more, excessive technical challenges may be blamed for the relatively modest volume of energy interventions. Some projects are quite large and require some creativity to avoid technical difficulties in the design of the evaluation. Many of the projects are interconnected and hence leave little room for randomness. This may be why most of the projects being evaluated tend to be rural where small and local are more the norm.

Few projects try to have a wide encompassing view of the impact of the sector. This is what the three oldest relatively robust impact evaluations of energy projects in developing countries tried to deliver. They are less than 7 years old and unfortunately share a significant drawback from the viewpoint of an evaluator. All relied on subnational data comparisons to draw some conclusions from differences in access to electricity. Fan et al. (2002) used Chinese data from 1970-97 to argue that for every 10,000 yuan spent on electricity development, 2.3 persons are brought out of poverty. Balisacan and Pernia (2002) used Filipino data from 1985-1997 to argue that the rich tend to benefit more from increases access to electricity. Balisacan et al. (2002) did the same for Indonesia in 1990

and concluded that a 10% improvement in access to a composite technology measure (including electricity in a village) raised the income of the poor by roughly 2 percent. More recently, Taylor (2005) and Escobal and Torero (2005) conducted similar assessments for Guatemala and Peru and drew very similar positive conclusions on the gains from electrification but suffered from some of the main econometric problems as the older studies.

As explained quite lucidly by Lipscomb et alt. (2008), the problem with these papers is that they do not fully address the causality between the intervention and the impact. This is because they do not account for the fact that electricity is often installed first in areas with the greatest potential for economic growth.

There are other problems which tend to be relevant to electricity impact evaluations also mentioned by Lipscomb et alt (2008). For instance, when drawing conclusions from contrasting areas with and without electricity with cross-sectional data, it is essential to address all the ways in which they may differ. Unfortunately, they sometimes differ in unobservable ways. Being able to rely on panel data may help to compare changes in outcomes over time in 'treatment' (with electricity) and 'control' (without) areas. The main problem there is that treatment areas have not been randomly selected.

The upshot is that this is a sector in which the little evaluation evidence available up to now had somewhat been tainted from the viewpoint of evaluators. There are, however, a few recent papers that address some of the main concerns. None of the evaluations in progress provides enough information to be reported here except for ongoing work on Peru by Torero (2009).[26]

The finalized papers look at the impact of electricity interventions in terms of their impact on growth and population density (Lipscomb et al (2008) health (Gonzalez-Eiras and Rossi (2007), female labor participation and cooking technologies (Dinkelman (2008)). The interventions assessed are the decision on the location of power generation plants (Lipscomb et al.), the impact of privatization (Gonzalez-Eiras and Rossi (2007) or rural electrification roll out (Dinkelman). I include in this overview ongoing Torero's (2009) work on policies surrounding electrification in Ethiopia because it illustrates how far the current evaluation technique can go to help understand some of the basic policy options commonly considered in the context of rural electrifications or in urban slums.

Lipscomb et alt. (2008) shows how the evaluation techniques serve to demonstrate the impact of electrification interventions on major macroeconomic concerns, including growth and the concentrations of populations. They exploit quasi-random variation in hydro-power generation and transmission in Brazil in order to isolate the causal effects of electricity grid expansion on changes in population density and GDP. Since hydro-power generation requires intercepting water at high velocity, there is a random component to households' access to electricity in a country that relies heavily on hydro-power. Indeed, access depends on the household's proximity to rivers with a gradient suitable for hydro-electricity generation. They predict hydropower plant placement based on exogenous geologic characteristics (river gradient and water flow) of locations throughout Brazil, and then develop a cost-minimizing "engineering model" to predict the expansion of the transmission lines from each of those predicted hypothetical stations every decade. This allows them

[26] The web site of all major donors lists the evaluations they are supporting. Unfortunately, they provide not much more than basic information on each project. Too little to deserve a discussion in this survey.

to examine the impact of electricity availability. They find that grid expansions strongly induce people to move to areas with electricity. This effect is much larger than a simple correlation between electricity availability and population density would suggest. Providing electricity can more than double the population of a location over a decade. Electricity is also found to increase GDP per capita, probably due to a causal effect of electricity on some aspect of productivity.

Dinkelman (2008) focuses on rural electrification in a region of South Africa. Her paper is very similar in spirit to the Lipscomb et al. (2008) paper but provides complementary insights on the impact of electrification efforts. The impact it is interested in is cooking technologies and employment. These effects are identified by exploiting variation in electricity project placement and timing from South Africa's mass roll-out of rural household electricity. Dinkelman finds that within five years, treated areas substitute sharply towards electricity in cooking. She also finds an significant 13.5% increase in women employment but no effect on male employment. This employment effect is driven by the switch to electricity from cooking wood that is usually collected by women. The importance of this paper for infrastructure specialists is to show that roll-overs in network industries can be modeled as quasi experiments. It also shows that there are employment, gender and technological payoffs that are worth anticipating as factors driving possible income and growth payoffs of infrastructure expansions.

But as often in the case of new services, it is important to also anticipate the reluctance of some potential users to connect to a grid. This can be addressed in an evaluation. Torero (2009) presents a good illustration. It is based on an ongoing rural electrification project in Ethiopia. The country is implementing an ambitious electrification plan to support the high demand growth and to extend access to the rural population at a very accelerated pace. The objective of this plan is to electrify rural towns, as well as villages, through extension of the grid. These rural towns/villages range in size from 70 to 4000 households. Beneficiaries include residential, commercial and industrial customers (such as flour mills, irrigation centers, water pumping, and telecommunications). About 50% of rural towns are to be electrified over five years and 100% in 10 years. The project component would connect about 382 towns to the grid (through substation expansions and installation of sub-transmission lines (typically 33kV)), install efficient public street lighting in towns electrified; and distribute CFLs (compact fluorescent lamps) to rural customers to foster energy efficiency and therefore affordability.

This project clearly begs for simple evaluation questions: (i) what is the impact evaluation of this program on the welfare of the population, and (ii) are some specific interventions more effective than others in achieving positive outcomes?

The first question is pretty standard across ongoing and past evaluations. The second one is the one that opens new areas of understanding of how policy options work in this sector. To figure out the best policy instrument, the evaluators are assessing what drives the incentive to connect from the mainline of the town of each of the households and business. Their randomization instrument is a discount voucher to pay the last mile connection. This is an extremely useful proxy to get a sense of how much prices matter to the decision to connect.

Between 10-50% of eligible survey respondents will be randomly assign vouchers for a discount charge to cover the connection of the household or business. Varying amounts of the offer

will be distributed in equal proportions among voucher recipients. These vouchers will have a period of duration (4 months). In a subsample of towns randomly selected no vouchers will be distributed.[27] No results are available yet, but clearly, this evaluation should tell us a lot about how and how much prices drive the willingness to connect and hence to contribute to the overall welfare gains for the electrification plan.

Another recent paper also offers insight on the impact of an important policy option. Gonzalez-Eiras and Rossi (2007) evaluate the impact of privatization of electricity distribution companies on health. They use provincial-level data for Argentina and examines the impact of privatization on two output measures: incidence of low birth weight and child mortality rates caused by food poisoning. They find that privatization improves service coverage which, through the use of refrigerators, may improve nutritional intake. Privatization also results in a reduction in the frequency of power outages, and thus may reduce the likelihood of food poisoning. They conclude however with a very sobering statement suggesting that their evidence are not strong enough to inform the policy debate with respect to the benefits of privatization for the welfare of the poor. But they do provide interesting insights on how welfare gains can be achieved.

To conclude this overview of electricity impact evaluations, it may be useful to mention a recent study (World Bank (2008b) by the evaluation department (IEG) of the World Bank. IEG conducted a quasi-meta evaluation of the evidence on the impact of rural electrification in 12 countries focusing on the economic analysis of the causal chain from the provision of electricity to the various benefits which it is claimed to bring.

The main results of this evaluation are that: (i) World Bank projects have been generally successful in establishing electricity infrastructure but weak in strengthening supplier institutions; (ii) the largest share of benefits from rural electrification is captured by the non-poor; (iii) high connection fees and community selection criteria that emphasize economic returns continue to be barriers to reaching the very poor; (iv) consumer education and promotion of productive uses would enhance the benefits of electrification; and (v) properly calculating willingness to pay can demonstrate good rates of return on rural electrification projects. All these conclusions are going to have to be tested in forthcoming validations relying on more robust evaluation techniques.

From a practical viewpoint, it is interesting to see the range of indicators considered of relevance by all these evaluations. Evaluators indeed monitor relatively easy to follow indicators to establish the impact of electrification. For household without electricity for instance, they cover the types and number of nonelectric lighting appliances household are using (for example, candle, wick lamp, hurricane lantern, pressurized kerosene lamp). They consider the number of hours and minutes each lamp is used during the last 24 hours and the quantity of fuels used for lighting. They consider the type of fuel used for cooking and the monetary and non monetary costs of the various types of fuel. The most common indicator is however total spending for lighting and cooking fuel. For households connected, they consider the types, number and wattage of electric lighting appliances household are using (for example, incandescent lamp, fluorescent tube, and compact fluorescent. They also monitor the number of hours and minutes each light is used during the last 24 hours and the total kWh used last month. Of course, the total spending for electricity last month is also recorded.

[27] From a technical viewpoint, these vouchers will serve as an instrumental variable for intensity of electricity access.

The range of impact indicators used in these studies is almost just as large. It includes employment, educational achievements, changes in income, changes in consumption patterns (types of house appliances associated with improvement in welfare), safety, indoor air pollution and when possible distinguish the benefits on the basis of gender. All these indicators are needed to answer the wide range of questions found in evaluations in this sector. The most common include:

- What is the impact of expanded access and use of electricity on household welfare and local economic development?
- What is the impact on educational and health outcomes?
- What is the impact of introducing energy efficient technology on uses of electricity?
- What is the impact of alternative financing mechanisms on household connection rates
- How do households respond to renewable energy technologies?
- What is the appropriate level of subsidies for promoting a sustainable market for renewable energy usage?
- What is the impact of electrification on household and firm-level income outcomes?
- What marketing methods are most effective at increasing usage of renewable energy technologies?
- How do impacts compare between on-grid and off-grid electrification?
- How does electrification impact the cooking technology choices?
- What do households do with the time saved thanks to electrification?
- What is the optimal subsidy to CFLs to maximize CFL usage per dollar spent?
- What is the impact of alternative modes of delivery of CFLs (public distribution, private, vouchers)?

The diversity of these questions mirrors the diversity of local concerns to be addressed in evaluations in this sector as it does in others. And yet it is probably incomplete. One could have easily added to this list, questions that appear in evaluations of the health impacts of indoor pollution since some of the solutions to that problem are directly related to the choice of fuel and the policies implemented to reduce the use of wood for instance. But most of these evaluations is also work in progress for now.[28]

8. Concluding comments

The main messages of this survey may be summarized as follows.

First not all infrastructure interventions are suitable for impact evaluations based on experiments or quasi experiments. For those unsuitable cases, there are other ways of generating robust quantitative evaluation of the effectiveness of infrastructure interventions which can be just as effective in increasing accountability for intervention selection, implementation and sustainability.

Second, whatever the form of evaluation, research and practice of the last 5-10 years has provided many insights on why not all apparently comparable interventions have sometimes generated dissimilar impacts across locations. Differences in institutions, in legal or social incentives

[28] See Duflo, Greenstone and Hanna (2008) for a recent survey of the evidence on indoor air pollution and its policy solutions

and norms, in access to and sources of financial resources, in technological preferences and choices or in initials conditions can all explain quite convincingly differences in impact.

Third, modern evaluation techniques are delivering on their promise to identify poverty related and distributional issues with many of the interventions considered in infrastructure activities, whether projects, programs or policies.

Fourth, some of these differences can be internalized in the design of interventions and this may be the most important lesson to be learned from the fast growing number of experiences. Learning from mistakes and omissions is the minimum what this research can contribute to best practice and it is already a big step towards better and more systematic evaluations. The real challenge now is to make sure that practioners can internalize these lessons and are willing to contribute to our collective knowledge of "what works and what does not when" by building-in evaluation components in their own projects.

To be fair, this is not a simple matter of will. It is a matter of incentives and more importantly of financial incentives. Many infrastructure interventions are financed by organizations currently unwilling to pay for the cost of these evaluations. Most have not been budgeted to do so. This is quite surprising since most international organizations are also quite keen to generate knowledge public goods. It turns out that evaluations generate a part of that public good. In the short run, evaluations are benefiting from the subsidy of some large donors. But will it happen in the long run?

Evaluations will only survive if all development stakeholders accept to include them in the costs of the intervention (i.e. loan preparation and monitoring costs) as a matter of routine. This is a small price to generate a public good and, maybe more importantly to some actors, a small price to increase accountability for aid effectiveness.

Unless this global social norm changes, evaluations run the risk a being another development fad and remain in the class rooms rather than used as evidence of aid effectiveness at the board of major donors and other key development stakeholders. This would be a pity given the many lessons documented in this survey for a sector delivery many basic goods.

Bibliography

1. Ahlfeldt, Gabriel M. (2009), "The train has left the station: Do markets value intra-city access to inter-city rail connections?", MPRA Paper No. 13900, March

2. Alm, J. and S. Jacobson (2007), "Using laboratory experiments in public economics", *National Tax Journal*, March, Vol, LX, No1, 129-152

3. Angrist, J.D. and J.-S. Pischke (2009), *Mostly harmless econometrics: an Empiricist's Companion,* Princeton University Press

4. Archondo-Callao, R., 2009, "RONET Road Network Evaluation Tools, User Guide, Version 2.0." SSATP Working Paper No. 89-A, World Bank.

5. Bajic, V. (1983), "The Effects of a New Subway Line on Housing Prices in Metropolitan Toronto", *Urban Studies*, *20*(2), 147-158.

6. Balisacan, A., A. Banerjee, S. Cole and E. Pernia (2002), "Probing beneath Cross-national averages: poverty, inequality and growth in the Philippines", ERD Working Paper Series No7, Asian Development Bank

7. Balisacan, A., E. Pernia amd A Asra (2002), "Revisting growth and poverty reduction in Indonesia: What do Sub-National Data show?, ERD Working Paper Series No25, Asian Development Bank

8. Banerjee, A., E. Duflo, and N. Qian (2009), "On the road access to transportation infrastructure and economic growth", mimeo, MIT, available at http://igov.berkeley.edu/China09papers/On_the_Road_Access_to_Transportation_Infrastructure_and_Economic_growth.pdf

9. Baum-Snow, N., and M.E. Kahn (2000) , "The effects of new public projects to expand urban rail transit", *Journal of Public Economics*, *77*(2), 241.

10. Baker, J., R. Basu, M. Cropper, S. Lall and A. Takeuchi (2005), , Urban Transport and Poverty: The Case of Mumbai", mimeo, the World Bank, available at http://www-wds.worldbank.org/external/default/main?pagePK=64193027&piPK=64187937&theSitePK=523679&menuPK=64187510&searchMenuPK=64187283&theSitePK=523679&entityID=000016406_20050824160629&searchMenuPK=64187283&theSitePK=523679

11. Bamberger, M. and E. Hewitt (1986), "Monitoring and Evaluating Urban Development Programs: A Handbook for Program Managers and Researchers." World Bank Technical Paper No. 53. Washington, D.C.

12. Barone, M. and J. Rebelo (2003), "Potential Impact of Metro's Line 4 on Poverty in Sao Paulo Metropolitan Region (SPMR)", mimeo, the World Bank, available at http://siteresources.worldbank.org/INTURBANTRANSPORT/Resources/poverty_14_sp.pdf

13. Binswanger, H., S. Khander and M. Rosenzweig (1993) ,"How infrastructure and financial institutions affect agricultural output and investment in India", *Journal of Development Economics*, 41 (2) 337-366

14. Boarnet, M.G. (2007), Conducting Impact Evaluations in Urban Transport, Doing Impact Evaluation #5, The World Bank

15. Bourguignon, F. and L. Pereira da Silva, ed. (2003), *The Impact of Economic Policies on Poverty and Income Distribution: Evaluation Techniques and Tools*, Oxford University Press, New York

16. Buckley and Kalarickal (2005), "Housing policy in developing countries: Conjectures and refutations" *The World Bank Research Observer*, 20, 2:233-257.

17. Chandra, A. and E. Thompson (2000), "Does public infrastructure affect economic activity? Evidence from the rural interstate highway system," *Regional Science and Urban Economics*, 30, 457-490.

18. Chong, A., V. Galdo and M. Torero (2008), Access to Telehone Services and Household Income in Poor Rural Areas Using a Quasi-natural Experiemmnt for Peru", Economica, 1-26
19. Clarke, G., K. Kosec and S. Wallsten (2004), *"Has Private Participation in Water and Sewerage Improved Coverage? Empirical Evidence from Latin America"* , World Bank Policy Research Working Paper No. 3445

20. Clasen, T.,W. Schmidt, T. Rable, I Roberts and S. Craincross (2007). "Interventions to improve water quality for preventing diarrhoea: Systematic Review and Meta-Analysis", *British Medical Journal*, 334: 782-91

21. Dammert, A.C., J. C. Galdo, V. Galdo and T. Sanz (2009), "*Can ICT Prevent Malaria? Evidence from a Randomized Experiment in the Border Between Ecuador and Peru***"** Proposal prepared for Inter-American Development Bank, Research Network Project "Information and Communication Technologies (ICT) and Societal Empowerment

22. Deaton, A., (2009), "Instruments of Development: Randomization in the Tropics, and the Search for the Elusive Keys to Economic Development," *NBER Working Paper* #14690.

23. Dercon, S., D.O. Gilligan, J. Hoddinott and T. Woldehanna (2007), "The impact of roads and agricultural extension on consumption growth and poverty in fifteen Ethiopian villages", Oxford University, CSAE WPS/2007-01

24. Dewees, D. N. (1976). The Effect of a Subway on Residential Property Values in Toronto. *Journal of Urban Economics, 3*(4), 357.

25. Dinkelman, T. (2008), "The Effects of Rural Electrification on Employment: New Evidence from South Africa." PSC Research Report No. 08-653. August 2008

26. Donaldson, D. (200) "Railroads of the Raj: Estimating the Economic Impact of Transportation Infrastructure", mimeo, LSE

27. Dorosh, P., H.-G. Wang, L. You and E. Schmidt (2008), "Crop Production and road Connectivity in SSA: A Spatial Analysis, mimeo, The World Bank

28. Duflo, E., R. Glennerster, and M. Kremer, (2008), "Using Randomization in Development Economics Research: A Toolkit," in Schultz, T.P. and J. Strauss eds, *Handbook of Development Economics*, 3895-3962.

29. Duflo, E., M. Greenstone and R. Hanna (2008), Indoor Air Pollution, Heatlh and Economic Well-being, memo, MIT, February, available at http://econ-www.mit.edu/files/2375

30. Duflo, E. and R. Pande (2007), "Dams", *Quarterly Journal of Economics*, 122(2):601-646

31. Escobal, J., (2005). "The Role of Public Infrastructure in Market Development in Rural Peru," MPRA Paper 727, University Library of Munich,

32. Escobal, G. and M. Torero (2005), "Measuring the impact of asset complementarity: the Case of Rural Peru, *Cuadernos de Economia*, 42 (May), 137-164

33. Escobal, J. and C. Ponce (2004), "The benefits of rural roads: enhancing income opportunities for the rural poor", GRADE working paper 40, Lima, Peru

34. Esrey S. A. *et al.* (1985), "Interventions for the control of diarrheal diseases among young children: Improving water supplies and excreta disposal facilities," *Bulletin of the World Health Organization* 63(4), 757-72.

35. Esrey S. A. and J.-P. Habicht (1986), "Epidemiologic evidence for health benefits from improved water and sanitation in developing countries," *Epidemiology Review* 8, 117-128.

36. Estache, A., V. Foster and Q. Wodon (2002), *Accounting for Poverty in Infrastructure Reform – Learning from Latin America's Experience*, World Bank Institute Publications, Studies in Development Series

37. Fan, S. and C. Chan-Kang (2005), "Road Development, Economic Growth, and Poverty Reduction in China", Research Report 138, Washington, D.C.: IFPRI.

38. Fan, S., C. Chan-Kang, and A. Mukherjee (2005), "Rural and Urban Dynamics and Poverty: Evidence from China and India), Washington, D.C.: IFPRI

39. Fan, S. and C. Chan-Kang (2002), "Growth, Inequality, Poverty Reduction in Rural China", Research Report 125, Washington, D.C.: IFPRI.

40. Fewtrell, L., R. Kaufmann, D. Kay, W. Enanoria, L. Haller, and J. M. Colford. 2005. "Water, Sanitation, and Hygiene Interventions to Reduce Diarrhoea in Less Developed Countries: A Systematic Review and Meta-Analysis." *Lancet Infectious Diseases* 5: 42–52.

41. Field, E. and M. Kremer (2006) "Impact evaluation for slump upgrading interventions", *Doing Impact Evaluation* #3, The World Bank

42. Fjelsoe, B., A. Marshall and Y. Miller (2009) "Behavior change interventions delivered by mobile phone short message service" *American Journal of Preventive Medicine*, 36(2):165-73

43. Galiani, S., P. Gertler and E. Schargrodsky (2005), "Water for life: The impact of the privatization of water supply on child mortality. *Journal of Political Economy*, Volume 113, 2005, pp. 83-120

44. Garcia-Mila,T. and J. G.Montalvo (2007), "The impact of new highways on business location: new evidence from Spain", mimeo, Universitat Pompeu Fabra, February

45. Ghazala M. and V. Rao (2004), "Community Based (and Driven) Development: A Critical Review," *The World Bank Research Observer*, vol. 19, no. 1,. 1-39.

46. Gibbons, S., and S. Machin, (2005). Valuing rail access using transport innovations. *Journal of Urban Economics, 57*(1), 148-169.

47. Gibson, J and S. Rozelle (2003)"Poverty and access to roads in Papua New Guinea", *Economic development and cultural change*, 52, 159-185

48. Gonzalez-Eyras, M. and M. Rossi (2007), "The impact of electricity sector privatization on public health", Inter-American Developmetn Bank, Research Network Initiative, Working Paper #R-524 available at http://www.iadb.org/res/pub_desc.cfm?pub_id=r-524

49. Gruber, J. (2007), *Public Finance and Public Policy*, Worth Publishers, 2nd edition.

50. Heckman, J., and S. Urzua., (2009), "Comparing IV With Structural Models: What Simple IV Can and Cannot Identify," NBER *Working Paper*, # 14706.

51. Holl, A. (2004), "Manufacturing location and the impacts of road transport infrastructure: empirical evidence from Spain," *Regional Science and Urban Economics*, 34, 341-363.

52. Imbens, G. (2009), "Better LATE than Nothing: Some Comments on Deaton (2009) and Heckman and Urzua (2009)", *NBER Working Paper, #* 14896.

53. Ito, S. N. Kobayashi and Y. Wada (2008), "Leearning to evaluate the impact of aid", Nonie *Working Paper, #* 6, January

54. Jacoby, H. (2001), " Access to Markets and the Benefits of Rural Roads', the Economic Journal, 110, 713-737

55. Jacoby, H. and Minken (2007), "On the benefits of lower transport costs with an application to rural Madagascar", mimeo, The World Bank

56. Jalan, J. and Sommanathan (2004), "The importance of being informed: Experimental evidence on the demand for environmental quality", Indian Statistical Institute, Delhi Planning Unit, Discussion Paper 04-08

57. Jalan, J. and M. Ravallion (2003a), "Does piped water reduce diarrhea for children in rural India?," *Journal of Econometrics,* vol. 112(1), 153-173, January.

58. Jalan, J. and M. Ravallion (2003b), "Estimating the Benefit Incidence of an Antipoverty Program by Propensity Score Matching", *Journal of Business and Economic Statistics*, 19-30

59. Jensen, T. H. (2009), "General equilibrium impact evaluation of road sector investment projects in Ghana", Paper prepared for presentation at the Twelfth GTAP conference in Santiago (10-12 June 2009) and the ECOMOD 2009 conference in Ottawa (24-26 June 2009), see http://www.ecomod.org/files/papers/1045.pdf

60. Jones, N., H. Jones, L. Steer and A. Datta (2009), "Improving Impact Evaluation Production and Use", Working Paper 300, March, London, Overseas Development Institute.

61. Khandker, S.R.., Z. Bakht, and Koolwal, K.B. Gayatri (2006), "The poverty impact of rural roads : evidence from Bangladesh",

62. Karlan, D. (2009), Cairo Evaluation Clinic: "*Thoughts on Randomized Trials For Evaluation of Development"*, Economic Growth Center, Yale University, Center Discussion Paper, No. 973

63. Kim, E., G. J. D. Hewings and C. Hong (2004), "An application of an integrated transport network-multiregional CGE model: A framework of economic analysis of highway projects." Economic Systems Research Vol. 16(3): pp. 235-258.

64. Kling, Jeffrey R (2007), "Methodological Frontiers of Public Finance Field Experiments." *National Tax Journal, Vol. LX, No.1,* , March, 109-127.

65. Kremer, M., J. Leina, E. Miguel and A. Peterson Zwane (2007), "A randomized evaluation of source water quality improvement" mimeo, Kennedy School, Harvard University

66. Kremer, M., J. Leina, E. Miguel and A. Peterson Zwane (2009), "Spring Cleaning- Rural Water Impacts, Valuation and Institutions:, mimeo, Kennedy School, Harvard University

67. Labonne, J. and R. S. Chase (2008), "Do Community Driven Development Projects Enhance Social Capital?", The World Bank, Policy Research Working Paper, #46778

68. Labonne, J. and R.S. Chase (2008), "Do Community-Driven Development Projects Enhance Social Capital?", The World Bank, Policy Research Working Paper, No. 4678, July

69. Lall, S.V., H. Wang and T. Munthali (2009), Explaining High Transport Costs within Malawi, The World Bank, Policy Research Working paper No. 5133

70. Laury, S.K. and L.O. Taylor (2008) "Altruism Spillovers: Are Behaviors in Context-free Experiments Predictive of Altruism Toward a Naturally Occurring Public Good?" *Journal of Economic Behavior and Organization*, 65:9-29

71. Lipscomb, M., A.M. Mobarak and T. Barham (2008) ,"Returns to Electricity: Evidence from the Quasi-Random Placement of Hydropower Plants in Brazil", mimeo, University of Colorado

72. Lokshin,M. and R. Yesimov (2005), "Evaluating the Impact of Infrastructure Rehabilitation Interventions on Household Welfare in Georgia.", *World Bank Economic Review*, Vol. 19(2): 311-333

73. Garcia, M. and J.G. Montalvo (2008), "Do new highways attract new businesses? A new microeconometric methodology", Martinez-Vazquez, J. and F. Vaillancourt, Public Policies for Regional Development, Routledge: 180-196

74. McCord and van Seventer (2004), "The Economy-wide impacts of the Labour Intensification of Infrastructure Expenditure in South Africa", CSSR Working Paper No. 93, Centre for Social Science Research, University of Cape Town.

75. McDonald, J. F., and C.I. Osuji (1995), "The effect of anticipated transportation improvement on residential land values", *Regional Science & Urban Economics, 25*(3), 261.

76. McMillen, D. P. and J.F. McDonald (2004). Reaction of House Prices to a New Rapid Transit Line: Chicago's Midway Line, 1983-1999. *Real Estate Economics, 32*(3), 463-486.

77. Michaels, G. (2008). The Effect of Trade on the Demand for Skill: Evidence from the Interstate Highway System. *Review of Economics and Statistics, 90*(4), 683-701.

78. OECD (2002), "Impact of Transport Infrastructure Investment on Regional Development", OECD programme of Research on Road Transport and Intermodal Linkages (RTR), OECD, Paris

79. Olken, B. A. (2007) "Monitoring Corruption: Evidence from a field Experiment in Indonesia", *Journal of Political Economy*, 115:2.

80. Pereira A. M. & O. R. Sagales (2007), "Public infrastructures and regional asymmetries in Spain." College of William & Mary Department of Economics Working paper No. 46.

81. Peterson Zwane, A. and M. Kremer (2007), "What works in fighting diarrheal diseases in developing countries? A critical review", NBER working paper, w12987, March

82. Polasek, W., W. Schwarzbauer and R. Sellner (2009), "aggreage and Regional Economic Effects of New Railway Infrastructure", Institute for Advanced Studies, Economics Series, No. 242, June

83. Poulos, C., S.K. Pattanaya, and K. Jones (2006), "A Guide to Water and Sanitation Sector Impact Evaluations , *Doing Impact Evaluation* #4, The World Bank

84. Priemus, H., B. Flyvbjerg, and B. van Wee, eds (2008), *Decision-Making On Mega-Projects: Cost-benefit Analysis, Planning, and Innovation,* Cheltenham, UK and Northampton, MA, USA: Edward Elgar

85. Ravallion, Martin (2008) "Evaluating anti-Poverty Programs," in P. Schultz and J. Strauss, ed. *Handbook of Development Economics*, Volume 4, Amsterdam: North-Holland: 3787-3846.

86. Rawlings, L.B., L. Sherbune-Benz and J. Van Dommelen (2004)," Evaluating Social Funds A cross country analysis of community investments", mimeo, The World Bank

87. Rephann, T. and A. Isserman (1994), "New highways as economic development tools: an evaluation using a quasi-experimental matching method," *Regional Science and Urban Economics*, 24, 723-751.

88. Rioja, F. K. (2003a), "Filling potholes: Macroeconomic effects of maintenance versus new investments in public infrastructure." *Journal of Public Economic* Vol. 87(9-10): 2281-2304.

89. Rioja, F. K., (2003b), "The penalties of inefficient infrastructure." *Review of Development Economics* Vol. 7(1): 127-137.

90. Rioja, F. K. (1999), "Productiveness and welfare implications of public infrastructure: A dynamic two-sector general equilibrium analysis." *Journal of Development Economics* Vol. 58(2): pp. 387-404.

91. Rivlin, A.(1974), "How Can Experiments Be More Useful?" American Economic Review 64 No. 2, May: 346-354.

92. Rodrik, Dani (2008) "The New Development Economics: We Shall Experiment, But How Shall we Learn?" Brookings Global Economy and Development Conference. May.

93. Röller, L. and L. Waverman (2001) "Telecommunications Infrastructure and Economic Development: A Simultaneous Approach" *American Economic Review*, 91(4): 909-923.

94. Sawada, Y., M. Shoji, S. Sugarawa and Shinya Sugawara (2008), "The role of infrastructure in reducing chronic and transient poverty: The case of JBIC supported irrigation project in Sri Lanka", JBICI Discussion paper, No. 16, August

95. Sequeira, S.(2009), "Railroad to Growth: a Quasi-experimental Study of the Impact of Railroads on Trade, Firm Performance and Growth", mimeo, Harvard University, Economics Department

96. Sequeira, S. and S. Djankov (2008), "On the Waterfront: An Empirical Study of Corruption in Ports", mimeo, The World Bank, available at http://www.doingbusiness.org/documents/Corruption_in_Ports.pdf

97. Sequeira, S., S. Djankov (2009), “The long way around: the real consequences of corruption in ports”, Harvard University, mimeo,

98. Taylor, M. (2005), “Electrifying Rural Guatemala: Central Policy and Rural Reality”, *Environment and Planning C: Government and Policy, 2*3, 173-189

99. Torero, M. (2009), “Estimating impact of rural electrification programs in Ethiopia”, PPT presentation, July 20, IFPRI

100. Torero, M. and J. von Braun (2005) *Information and Communications Technology for Development and Poverty Reduction*, Johns Hopkins University Press.

101. Van de Walle, D. (2009), “Impact evaluation of rural road projects”, *Journal of Development Effectiveness,* Vol. 1, No.1, March, 15-36

102. Van de Walle, D. & D. Cratty (2002), “Impact Evaluation of a Rural Road Rehabilitation Project”, mimeo. See: http://siteresources.worldbank.org/INTISPMA/Resources/383704-1153333441931/11258_Dwalle-Impact-Rural_Roads.pdf

103. Van de Walle, D. and R. Mu (2007), “Fungibility and the flypaper effect of project aid: Micro-evidence for Vietnam.” *Journal of Development Economics* Vol. 84(2): pp. 667-685.

104. Waddington, H., B. Snilstveit, H. White and L. Fewtrell (2009). “Water and sanitation inteverntions for better child health: evidence from a synthetic review”, mimeo, Powerpoint presentation, 3ie

105. White, H. (2009), “The welfare impact of rural electrification”, *mimeo,* Powerpoint Presentation, The World Bank

106. World Bank (2008a), “Evaluating the macroeconomic and distributional impacts of lowering transportation costs.” Report No. 40020-BR, Brazil country management unit, PREM sector management unit, Latin America and the Caribbean Region, World Bank, Washington DC.

107. World Bank (2008b), *The Welfare impact of rural electrification : A reassessment of the costs and benefits,* An IEG study, The World Bank

108. Wright, J., S. Gundry, and R. Conroy (2004), Household drinking water in developing countries: A systematic review of microbiological contamination between source and point-of-use,” *Tropical Medicine and International Health* 9(1), 106-17.