

# **Impact Evaluation of Rural Road Projects**

**September 2008**



## Acknowledgement

This paper was written by Dominique van de Walle<sup>1</sup> in 2008. The work was task managed by Emmanuel Skoufias. The paper benefited greatly from comments received by Niall Keleher, Gayatri Koolwal, Ren Mu, Ambar Narayan, Martin Ravallion and Howard White. The support of the PREM Evaluation Unit is gratefully acknowledged. This work program was financed by a Norwegian Trust Fund for Impact Evaluation in Infrastructure and by the Bank-Netherlands Partnership Programs.

---

<sup>1</sup> World Bank, 1818 H ST, NW, Washington, DC 20433. tel: 202-473-7935, fax: 202-522 1154, [dvandewalle@worldbank.org](mailto:dvandewalle@worldbank.org) The opinions reflected in this paper are the opinions of the author and not opinions of their institutions.

## TABLE OF CONTENTS

<b>INTRODUCTION.....</b>	<b>1</b>
<b>I. GENERIC ISSUES IN ASSESSING THE IMPACTS OF RURAL ROADS.....</b>	<b>3</b>
A. DERIVED AND CONDITIONAL BENEFITS.....	3
B. SOURCES OF ENDOGENEITY .....	4
C. DISPERSED EFFECTS ON MULTIPLE OUTCOMES .....	5
D. DISTRIBUTIONAL IMPACTS.....	6
E. IMPACTS MAY TAKE A LONG TIME TO EMERGE.....	6
<b>II. METHODS AND DATA .....</b>	<b>7</b>
F. EVALUATION METHODOLOGY.....	7
G. IMPLICATIONS FOR DATA .....	11
<b>III. STEPS IN SETTING UP AN IMPACT EVALUATION .....</b>	<b>11</b>
A. ASSESSING THE FEASIBILITY OF DOING AN IMPACT EVALUATION .....	12
B. LEARNING FROM THE PROJECT AND ITS EX-ANTE EVALUATION .....	13
C. THE EVALUATION DESIGN: DECIDING WHAT DATA ARE NEEDED.....	13
D. EVALUATION DESIGN: COLLECTING THE DATA .....	19
E. A NOTE ON RISKS AND PITFALLS .....	25
<b>IV. SELECTED FINDINGS FROM RECENT RURAL ROAD IMPACT EVALUATIONS .....</b>	<b>26</b>
<b>V. OUTSTANDING ISSUES FOR RURAL ROAD IMPACT EVALUATIONS.....</b>	<b>28</b>
A. SPECIFIC AND GENERIC ISSUES WHERE MORE NEEDS TO BE KNOWN .....	28
B. STRUCTURAL METHODS .....	29
<b>CONCLUSIONS.....</b>	<b>30</b>
<b>BIBLIOGRAPHY .....</b>	<b>35</b>

## Introduction

Given limited redistributive instruments, governments and aid donors look to basic health, education and infrastructure interventions to help achieve distributional objectives. Some of these interventions are also believed to have efficiency benefits, by relieving the constraints people face in taking up new economic opportunities.

In recent years, rural roads have been extensively championed as poverty alleviation instruments by the World Bank and donor institutions. It is argued that rural roads are key to raising living standards in poor rural areas (for example see Gannon and Liu, 1997). By reducing transport costs, roads are expected to generate market activity, affect input and output prices, and foster economic linkages that enhance agricultural production, alter land use, crop intensity and other production decisions, stimulate off-farm diversification and other income-earning opportunities, and encourage migration. Claims have also been made that by facilitating access to social service facilities, better roads enhance social outcomes. Yet despite a general consensus on the importance of rural roads for development and living standards, there is surprisingly little hard evidence on the size and nature of their benefits, or their distributional impacts. Indeed, there have been relatively few rigorous and credible impact evaluations of rural roads.

For the selection and appraisal of rural roads projects, the transport sector has typically measured the benefits based on the vehicle operating cost savings, and time savings.<sup>2</sup> Champions of rural roads have rightly observed that benefits are likely to be much broader. One response among donor agencies has been to devise special selection and appraisal criteria for rural roads that simply assume important social benefits, despite a general lack of rigorous empirical evidence. These are used as justification for abandoning economic analysis when, as is the case in many rural areas of developing countries, traffic levels are too low for conventional consumer surplus measures to make sense (van de Walle, 2002). Impact evaluations can help provide better information on the non-pecuniary benefits of rural roads.

While there have been a number of recent attempts at assessing rural road impacts using impact evaluation methods, research has newly underlined the enormous difficulties inherent in estimating the magnitude of the effects attributable to infrastructure. Problems arise due to the endogeneity of much infrastructural development and the many other factors that are at work (Binswanger et al., 1993; Jalan and Ravallion, 1998). Not allowing for initial conditions — both time-invariant and time-varying — and the way in which infrastructure is allocated to specific regions will tend to bias impact estimates, often downwards in poor areas. The most common criticisms of past impact evaluations of roads are that they did not have appropriate comparison groups; that the results were not likely to be robust to unobserved factors influencing both program placement, subsequent growth, and outcomes; and that they did not follow

---

<sup>2</sup> This is typically also the methodology used by developed country Transport Planning Agencies (Weisbrod, 2000).

projects long enough to capture full impacts. Doing better requires a combination of better evaluation design, methods and data.

Although little is known about the magnitude of the benefits from rural roads, plenty is known about their costs. These are relatively expensive investments. An aid donor faced with a limited budget needs to know more about how benefits compare to costs in order to make informed decisions between a road and, say, an education intervention — also believed to have considerable impacts on well-being.

As a first step, much more needs to be known about the impacts of roads. Further knowledge and understanding about the conditions under which those impacts will be maximized would also be extremely useful. For example, a key policy issue is whether, and in what way, initial conditions at village and household level interact with roads to influence outcomes. Are impacts higher or lower in places with higher levels of initial development in terms of market and other institutions? More such information based on careful, rigorous impact evaluations will help improve road project appraisals, design, and selection methodologies.

This paper seeks to discuss the various issues that arise in setting up and conducting *ex-post* impact evaluations of rural road improvements — new road construction as well as road rehabilitation — and to give practical guidance on those issues. I have tried to write the sort of paper that would have been very useful to me when I embarked on an evaluation of rural roads. The paper is intended to guide someone who wants to evaluate a specific rural roads project. There is also a literature on the impacts of roads or other infrastructure, which can be thought of as impacts of a great many past projects (Binswanger et al., 1993; Fan et al., 2000; Fan and Chan-Kang, 2005; Mogues et al., 2007). While one might use the latter to provide insights into the impact of a new project (e.g. Gibson and Rozelle, 2003), this paper focuses on methods for evaluating new projects.

The paper focuses on *ex-post* impact evaluations, meaning that one is evaluating impacts after an intervention has happened and had time to have impacts. By ‘impact evaluation’ I mean that an effort is made to establish causality by exploiting the fact that the interventions are assigned to specific units, namely households, firms, or locations in the case of roads, such that non assigned units exist that provide clues to the counterfactual situation of what would have happened in the absence of the intervention.

‘Rural roads’ are defined here as small local roads or even paths and tracks in rural areas that have low or no motorized traffic volumes. They typically link up villages with other villages or with the road network and are expected to have local impacts. Therefore, highways and long distance trunk roads joining up regions, that tend to have macroeconomic effects and are not assigned to specific communities — so that a counterfactual can not be identified based on observable data, and classic evaluation tools are useless — are not considered in depth here. However, a short discussion of these cases is included.

The paper assumes an understanding of the most basic principles of impact evaluation. It will refer to various impact evaluation and econometric methods but does not provide an in depth explanation or review of these as there are other sources that already do this well.<sup>3</sup> Likewise, a reading of this paper should be complemented with the information provided in World Bank (2006) concerning generic and practical issues in linking impact evaluation and the project cycle. Throughout the paper, examples are taken from a few recent rural road evaluations and particularly one that was conducted by the author in Vietnam.

In order to set the stage for the discussion of design and data, the paper first discusses in Section 2 evaluation issues that are specific to rural road impact evaluation. Section 3 then turns to the implications for evaluation methods and data requirements. Next, Section 4 examines various steps in the evaluation and issues specific to road evaluations that need to be addressed at different stages, as well as some risks and difficulties that may be faced as the evaluation proceeds. Section 5 briefly notes some general findings from recent rural road impact evaluations, while Section 6 turns to some of the outstanding generic questions that rural road evaluations could help throw light on. The section also briefly introduces structural methods to assess rural road impacts as an alternative to standard impact evaluation methods.

## **I. Generic issues in assessing the impacts of rural roads**

It is useful to start by thinking about the nature of roads and the ways in which they may differ from other policy investments. Their distinguishing features suggest a number of researchable evaluation questions about which relatively little is known. They also have important implications for evaluation design and methods, as well as for data requirements.

### ***A. Derived and conditional benefits***

The first distinguishing characteristic of roads is that their prospective benefits — whether positive or negative — are derived and conditional. One does not obtain utility directly from a road, but indirectly via the access to opportunities for extra consumption that it allows. The nature and extent of impacts are also likely to be heavily dependent on interactions with other investments, other social and physical infrastructure, and the geographical, community and household characteristics of where they are located. For a new road to enhance mobility, people must have access to private (e.g. bicycles, motorcycles) or public (e.g. buses, taxis) means of transport along the road. For impacts to be felt on input and output prices there must be some means of freight transport and local production for which external demand exists. For the road to have impacts on

---

<sup>3</sup> For excellent introductions to econometrics and impact evaluation methods see respectively Wooldridge (2003) and Ravallion (2001). A more in depth review of impact evaluation issues is given in Ravallion (2008).

social sector outcomes there must be functioning health and education facilities in its vicinity. In the absence of such complementary factors, new or improved roads may simply facilitate and hasten population out migration. That still constitutes a benefit as it generates new opportunities but it is a qualitatively different benefit and not (typically) the intended one from a rural roads project.

Thus a large number of factors can be expected to interact with a road intervention in determining what impacts it has on its surroundings, as well as who its beneficiaries are.<sup>4</sup> In evaluating impacts it is clearly important to control for such conditioning factors and their heterogeneity across road projects.

## ***B. Sources of endogeneity***

Roads are clearly not randomly placed, and it is likely that the factors that matter to road placement will also affect outcomes. Roads are built (or rehabilitated) in certain locations and not in others, for reasons that tend to have a lot to do with the attributes of those locations. For example, policy makers might assign a new road to a specific region because it is deemed to have economic potential or because it has a strong political constituency, clamoring (for example) to get their produce out to markets. To give another example, improved school outcomes associated with a road improvement might stem from the fact that committed parents were demanding better access to schools. The evaluation problem is then to work out how much of the change in outcomes is attributable to the road as distinct from the prior demand for schooling.

Many practitioners are now well aware of the potential for selection bias — also known as the problem of ‘endogenous road placement.’ There are three distinct ways in which endogeneity can arise in this context. Table 1 provides a summary of the generic types of endogeneity. In each case there is some third relevant variable X that is jointly correlated with the road’s placement and the outcome variable.<sup>5</sup> Failure to properly control for X can generate large biases in estimates of the impacts of rural roads (an illustrative example is given in van de Walle 2002).

The first type of endogeneity is when there is an initial condition that is correlated with road placement and is also correlated with the level of the outcome variable. For example, if places with agricultural cash crop potential are selected to get better roads, then a bias can be expected when assessing road impacts on agricultural incomes, given that the measured values of those incomes will partly reflect the differences in initial potential. In the second type, the initial condition that influences placement also influences the changes in outcomes as would be the case if roads were targeted to poor areas with attributes that determined both their poverty and their subsequent growth path. This is likely to be an important source of endogeneity in the rural road context (Jalan

---

<sup>4</sup> Health and school facilities are similar in that the buildings without the personnel and equipment are useless in promoting health and education outcomes. But this seems to be better internalized by policy makers than the potential need for complementary investments with road projects.

<sup>5</sup> A distinction is sometimes made between endogeneity arising through reverse causation whereby placement and outcomes are jointly determined, and through the presence of a common third variable that determines both (also known as omitted variable bias). However, the first implies the second and the distinction has no bearing on the estimation, so here the focus is on the more general second case.



and Ravallion, 1998). Finally, the endogeneity problem can also arise when changes in program placement are a function of time-varying factors, such as when road expansions accord with changing economic or political conditions that are themselves correlated with the changes in outcomes. For example, if new roads are built in places that are growing faster, there will clearly be a problem in disentangling the road's impact on incomes. Note that in all three cases the problematic initial condition or time-varying factor may be observed or unobserved; this difference has important implications for the methods used to try to address the endogeneity problem, as discussed in Section 3.

It is sometimes argued that endogeneity is not a concern when one is studying the impacts of roads at the micro (household or firm, e.g., the users) level, since assignment is to communities or geographic areas. However, there could still be a problem in this case because there are likely to be individual characteristics that are unobserved but geographically determined — or geographically correlated (such as through location choice) — and are also correlated with the things that influence program placement.

Understanding the potential sources of endogeneity is critical to collecting the right data and choosing the right methodology for estimating unbiased impacts.

### ***C. Dispersed effects on multiple outcomes***

A third issue is that roads and road networks are widely expected to have geographically dispersed effects on numerous outcome variables. The sorts of rural roads considered here are not expected to have economy-wide effects such as a major new trunk road or highway might have. However, through various linkages, an improved rural road could conceivably have community-wide impacts on livelihoods and socio-economic development and impacts that extend beyond the community the road links up to. This has a number of implications. First, traditional evaluation methods that use a single cross-section cannot be relied upon. Observable local area attributes — that would typically be considered as potential controls for program placement and used to ensure conditional exogeneity in a single difference model (see Section 3) — may have themselves been determined or contaminated by the road investment project. Even with a rich dataset, much of the data may plausibly be determined by the road. This leads me to be pessimistic about conducting a worthwhile rural road evaluation without adequate baseline (pre-intervention) data.<sup>6</sup>

The fact that roads may have dispersed effects also has implications for how one defines the road's zone of influence. Say an intervention passes through community A which is identified as the zone of influence for data collection purposes. If there are externalities from the project to neighboring community B, the evaluation would miss these impacts as they are outside the defined zone of influence. It also has implications for identifying appropriate comparison areas that are close enough to the intervention areas — and hence similar in various geo-climatic and other respects — and yet sufficiently distant to avoid being contaminated by spillover effects. The latter might

---

<sup>6</sup> Of course, there may be ways of recreating the key necessary baseline variables through recall or using existing data unconnected to the road project evaluation.

happen if community B above were selected as a comparison area leading one to misapprehend impacts. (Section 4 further discusses these issues in the context of defining the zone of influence and comparison groups.) A final implication is that there may well be numerous potential outcome indicators of interest since a priori many factors could be affected directly and indirectly by a road improvement.

#### ***D. Distributional impacts***

Little is known about the distributional impacts of rural road investments. There are likely to be both vertical impacts (between ‘rich’ and ‘poor’) and horizontal impacts (at a given level of pre-intervention welfare). There is a real possibility that in addition to gainers there will be losers, as there almost always are with policy change. On average, benefits may well be positive but it is essential to understand who the losers are if one is to understand distributional impacts, and differences in impacts at given levels of living. For example, if new roads lead to higher land values there may be a heightened tendency towards land concentration and landlessness (Howe, 1984). Those with greater initial land, education, wealth or influence will be better able to take advantage of the changes. The distribution of current income and future income earning opportunities may widen. There may be a reduction in common property resources which may hurt the poor the most. As cheaper goods are brought in, traditional job displacement may ensue.

In short, it is important to adopt a sampling design and collect data that allows one to explore whether and how impacts vary across groups.

#### ***E. Impacts may take a long time to emerge***

Little is known about the time it takes for the full welfare impacts of an improved road to emerge, how quickly intermediate impacts emerge and how short term impacts on these outcomes may differ from longer term impacts. In contrast to interventions with relatively rapid impacts (such as transfer programs), the welfare impacts of road infrastructure are generally expected to take some time to appear. This creates problems for an impact evaluation. An evaluation that doesn’t allow sufficient time for the linkages to play themselves out could vastly under- or over-estimate the net impacts on living standards. However, long-term evaluations are hard to set up and sustain. The more time one allows for impacts to emerge, the more the evaluation needs to contend with potential attrition in the data (whereby some sampled units drop out), confounding exogenous shocks and spillover effects such as related or unrelated other investment programs in the control or treatment areas that make it harder to assess the impacts of the original road project.<sup>7</sup>

Of course, some less comprehensive outcome indicators — such as migration or school attendance to a pre-existing school — may respond rapidly, while others take longer. Additionally, initial impacts could differ from medium to longer term impacts for the same indicator. Some first-round mean impacts may well be volatile and alter as time

---

<sup>7</sup> A road project evaluation in Bangladesh reports losing two of four initial comparison villages due to new paved roads being built near them post baseline (Khandker et al., forthcoming).

passes. For example, some facilities and services may relocate away from the village center in the immediate aftermath of a road access improvement. Small traders, who had stocked locally produced goods and ones hauled back long distances on foot, may be driven out of business now that better stores are more accessible by bicycle and motorcycle. In due course, such a store may set up in the village but for an initial period the road will have caused the village food stall to disappear. Distributional impacts may also differ between the short and longer term with even initial losers eventually benefiting. These are all testable propositions on which existing rural road evaluations have had little to say. The data must clearly follow the experience of participants a sufficiently long time after project completion to capture impacts. How long it needs to do so will ultimately depend on the specific outcome indicator and the evaluation objective.

## II. Methods and Data

What are the implications of the foregoing discussion of road intervention features for evaluation design and data collection?

### *F. Evaluation methodology*

Important generic sources of endogeneity bias were identified above in the context of rural road interventions. Initial conditions, such as poverty levels or pre-existing infrastructure, are likely to determine road placement, as well as to influence the levels of relevant outcome variables and also the subsequent growth paths and prospects of the communities, firms and households within them. A credible evaluation methodology needs to correct for these potential sources of selection bias through a careful construction of the counterfactual. Furthermore, given that roads may have dispersed effects on many factors, information on conditions prior to the intervention must naturally be available. Table 1 provides a summary of specific solutions associated with the types of endogeneity discussed in Section 2.2.

Single difference comparisons can be either reflexive (before and after) comparisons that track gains solely in project areas, or with and without comparisons that take single differences in mean outcomes between participants and non-participants using cross-sectional data. The reflexive comparison is unable to separate the road impact from the general economic and other changes that would have happened without the roads project. The with and without comparison can only identify the impact of the roads if one can deal with the selection bias.

**Double difference.** If it is believed that the data do not include some of the relevant X variables that jointly influence program placement and the level of the outcome measure then one might consider a double difference (DD) (or ‘difference-in-difference’) design where a first difference is taken between outcomes in the project areas after the program and before it, minus the (second) corresponding difference in comparison non-project areas. This can be obtained from a simple OLS regression where the dependent variable is the change in the measured outcome variable between the pre-

intervention and the post-project data and the only explanatory variable is a dummy for whether the areas participated in the project.<sup>8</sup> Used alone, this method assumes that placement is a function of initial conditions (observed or unobserved) that take the form of an additive time-invariant error term (type 1 endogeneity, see Section 2.2 and Table 1). Once this is differenced out, any remaining differences in outcomes between the treated and non-treated can be attributed to the project.

**Double difference combined with propensity score matching.** A conventional DD does not allow for initial conditions (type 2, Table 1) or time-varying factors (type 3, Table 1) that also influence subsequent changes in outcomes over time. To deal with these sources of endogeneity bias, one needs to control for the initial conditions and any time variant factors that simultaneously influence the placement of roads and subsequent growth rates. One recently popular approach for correcting for these biases combines double differencing with propensity score matching (PSM) to select ideal comparison areas from among the sampled non-project areas (Ravallion and Chen 2005; Lokshin and Yemtsov, 2005; Mu and van de Walle, 2007 and 2008; Chen et al. 2007).<sup>9</sup>

Matching methods — whereby non-project units are ‘matched’ to project units to minimize observable heterogeneity — are one way to construct comparison groups that are as similar as possible except for the intervention. PSM matches on the basis of the predicted probability of participating in the road project (the propensity score).<sup>10</sup> The propensity scores are estimated using a logit (or probit) on the pooled pre-intervention treatment and non-treatment area data. The logit model should control for all initial conditions and time-varying factors that may explain placement and/or subsequently affect changes in the areas. Exact matching on these attributes implies that the resulting matched subjects have the same distribution of the covariates. PSM also allows one to trim the sample of comparison groups with propensity scores that do not overlap with those for the treatment groups.

Once the matched comparison group is constructed, there are various possible estimators. For example, the mean impact estimate can be calculated as the mean difference in the outcome indicator between the project and the matched comparison areas. Other estimators use the propensity scores to weight the comparison areas (following Hirano et al., 2003). PSM thus eliminates bias in estimated treatment effects due to initial observed heterogeneity and time variant selection bias.

These methods can only control for the differences in observable variables jointly influencing outcomes (or their changes) and program placement. The estimates will still be biased if there are unobservables that affect both placement and outcome changes. Typically missing from even good data bases and often mentioned as possibly important

---

<sup>8</sup> An alternative that does not require a balanced panel is to use a regression with an interaction term as follows:  $y = a + a_1 project + a_2 time_1 + a_3 project * time_1$  where  $a_3$  gives the DD.

<sup>9</sup> For the theory of propensity score matching and weighting, see Rosenbaum and Rubin (1983) and Hirano et al. (2003), respectively.

<sup>10</sup> Since road project assignment is typically at the community level, matching is implemented — and propensity scores estimated — for communities not households, even when the analysis of impacts is conducted at the household level. See Box 3.

in this respect are factors such as political clout, local leadership, social capital and empowerment. There has been progress at measuring such conditions in recent years and an effort to measure them or proxies may be crucial to the evaluation in certain settings. The better we measure initial conditions and relevant time variant attributes, the less concern there will be that latent factors might influence changes both in road placement and outcomes over time. However, one can never completely rule out the possibility of omitted initial conditions or changes over time in those variables that are correlated with placement or outcomes.

**Double difference combined with Instrumental Variables.** The idea here is to find a variable — known as the Instrumental Variable (IV) — that determines the road placement (or changes in placement) but does not determine outcomes conditional on placement. In other words, the IV only affects outcomes through the program placement, thus allowing identification of impacts robustly to selection bias. In a first step the IV is used to explain the road access variable. The predicted value from that equation is then used instead of the road variable in the DD outcomes regression. A good IV purges the evaluation of endogeneity bias. It is the only option for dealing with an unobserved X that is thought to be correlated with placement or changes in placement and changes in outcomes. (The latter requires an IV that is correlated with the changes in placement but only affects outcomes via changes in placement.)

In practice, it is often difficult to find a valid IV. One example comes from an assessment of the impact of roads on consumption in rural Papua New Guinea (Gibson and Rozelle, 2003).<sup>11</sup> The authors argue that the year that a district was linked to the national highway system was not a function of a region's wealth or productivity potential and hence, provides a good instrument for subsequent access to rural roads. The linkage led to feeder roads being built and yet does not affect consumption independently of its effect on feeder roads.

**Dynamic panel data models:** In a situation where longitudinal data on households or communities is available for at least three periods, it may be possible to implement a dynamic panel data growth model to estimate the impacts of roads on income or consumption growth. Jalan and Ravallion (2002) use a micro model of consumption growth with Chinese household panel data covering 5 years and find that the density of roads has a positive impact on subsequent consumption growth. Dercon et al. (2007) try something similar to assess the impacts of changes in road access on consumption in Ethiopia. It is more difficult to explore impacts of policy changes as it clearly raises the further concern that the changes in program placement are endogenous to outcome changes even when it is plausible that the initial assignment is exogenous (conditional on observable controls). Moreover, it is not common to find longitudinal data with at least three points in time and information on road access for a particular road project area that one wants to evaluate.

---

<sup>11</sup> Using cross-sectional data, this study examines the impact of road access as opposed to changes in access such as in a road project evaluation context. It then uses the estimates to simulate the impact of a change in access.

**Why not use randomization?** The above discussion has assumed that roads are placed non-randomly and has thus focused on quasi-experimental evaluation methods. Could the allocation of road sub-projects be randomized?

The potential for randomization to evaluate roads and deal with selection bias seems limited. One can imagine a situation whereby a government could be convinced to randomize road interventions through a phase in of the road project — phasing in the sub-project interventions randomly and using the initially randomly delayed areas as the comparison group. However this common approach to randomizing when it is politically infeasible to exclude some eligible areas does not appear promising given that the full impacts of a road intervention may take a long time to emerge. The second phase might have to be delayed beyond what would ever be politically acceptable. Experience suggests that road project phases do not typically get implemented in a short period of time but often over the course of a few years as different road links take longer to get implemented than others. For the randomly assigned and non-assigned units to be observationally identical and randomization to work, sample sizes must be sufficiently large. Again, one would have to delay the second round of sub-projects even longer in order to get a sufficiently large treatment sample with enough lead time for impacts to emerge. Finally, if the randomized out communities get wind of the future road improvement, contamination is likely, as households or firms try to set themselves up to maximize their ability to take advantage of the road improvements. These arguments suggest that it will be necessary to control for the heterogeneity of factors that jointly influence outcomes and road placement using the non-experimental methods discussed above.

**Situations where all these methods fail.** What happens if the project is perfectly targeted on observables? Then, there are no areas with those same observable covariates that do not receive the intervention. For example, a project might target road improvements to all communities with a poverty headcount of 30% or higher. In such a case, it is impossible to identify an average treatment effect on the treated since there is no possible counterfactual. However, one might still identify the marginal impact by a discontinuity design which compares outcomes for communities with a poverty incidence of just above 30 percent with those for communities just below 30%. This identifies the local impact in a region of the 30% poverty rate under the assumption that they should be very similar in the absence of the road project (see Hahn et al., 2001). One can then identify the marginal impacts for the group of communities found around the project eligibility cutoff point. Impacts are then estimated only for a subgroup of participants. However, road projects that are perfectly targeted on observables appear to be extremely unlikely in practice. I am not aware of this method being applied to rural roads.

The outlined methods also fail in the case of larger, non-assigned rural road interventions where one does not strictly observe non-participants. This paper is primarily concerned with the simple binary case where some units are assigned the road and others are not. The treatment variable is then a dummy which takes the value 1 for being in the project area and 0 for non-project areas, from which an appropriate comparison group can be drawn. This is reasonable in the case of small rural roads where the benefits go to a self-contained area and one can identify non-beneficiaries.

Thus, benefits go to zero at some well-defined distance from the intervention. In the case of bigger roads, the approach is no longer reasonable as there will be important spillover effects and one doesn't strictly observe non-beneficiaries. Benefits are more continuous and diffuse.

In such cases, one approach would be to use the distance to the road as one's treatment variable under the assumption that accessibility to the road as measured by distance is still an important aspect of the benefits. This approach still rules out the case of huge road projects such as major highways that require very different methods since they have general equilibrium implications and all prices and wages in the economy could be affected.

Regressing outcomes on distance to the intervention, one must still worry about the same endogeneity problems as before. For example, the possibility that poor people live in the more inaccessible places at further distance from the road, where land values are also lower (Jacoby, 2000). One must now control for initial conditions in a regression context since matching is no longer feasible.

### ***G. Implications for data***

The foregoing discussion suggests that a credible rural roads impact evaluation will require panel data, including pre-intervention baseline data. This will need to be available for both project and non-project areas that resemble the project areas to enable creation of an appropriate counterfactual. In order to further control — whether by means of a regression or a matching model — for initial conditions that may have led to placement as well as outcomes, one will need detailed information of baseline attributes that potentially influenced selection into the project as well as appropriate controls for exogenous time varying factors. The data should also contain a set of outcome indicators of interest, controls for heterogeneity, and data that allow one to differentiate between welfare groups. There are tremendous gains from data that is specifically designed and collected for the evaluation.

When using PSM, balancing tests should be used to assure that initial conditions (including initial outcomes) are similar between treatment and comparison samples (see for example Chen et al., 2007; Mu and van de Walle, 2007). The post-project follow-up data collection should allow for a sufficiently long time interval for impacts to emerge. Finally, detailed information on the project — including when and where the road works were implemented and what was done — will also be needed in order to determine its impact.

The data issues in practice are discussed in greater detail in the next section.

## **II. Steps in setting up an impact evaluation**

The following discussion is organized around the key recommended steps to be taken in implementing a rural roads impact evaluation. Box 2 outlines all the steps. This section will not cover in detail those steps that are shared wholly with evaluations in

general, but focuses on issues that arise specifically in road evaluations and need to be carefully thought about in practice. However, it is not feasible to mention all potential solutions and options available at each juncture since these will, in the end, be very context and project specific. The discussion is meant to help prospective evaluators think through these matters rather than to provide a generic blue print. One needs always to remain flexible and to work within the constraints and specifics of one's own evaluation problem, including the degree of in-country capacity.

#### ***A. Assessing the feasibility of doing an impact evaluation***

Before making the decision to launch an impact evaluation, it is important to ensure that a number of key factors that could make or break the evaluation are in place. First, one needs to secure support and the promise of continued full cooperation and access to information on project details and its progress from the government and other key counterparts in the country, such as the project management unit (PMU), the World Bank (or other donor agency's) project task team and the latter's management. Throughout the course of the evaluation, the evaluation team will need to work closely with those responsible both in the country and in the donor agency, most notably to be kept abreast of information on sub-project selection and design, timing, and changes. For example, such information will be crucial to getting the baseline in on time (before the project starts) and to sampling enough areas that will and will not get the project. This is typically a difficult period when the evaluation team needs to move quickly with sampling design and fielding the baseline, while the project team is revising and finalizing its plans subject to its own constraints. Clearly, cooperation and coordination from project counterparts can be of fundamental importance to implementing the evaluation as desired. If there is strong opposition to conducting an evaluation from these players, it may well be better to abandon the idea since the quality of the evaluation can be severely compromised without sufficient cooperation of all relevant parties.

Second, rural road evaluations typically require considerable funding and sustained commitment, particularly when panel data are to be collected. Therefore one also needs to determine what sources of funding will be available and whether adequate financial backing for conducting the evaluation rigorously, independently and over a sufficiently long time horizon can be lined up.

Thirdly, one must figure out early on whether, given the project's objectives, design and reach, there is potential for constructing a credible counterfactual. As noted earlier, this should be possible in the context of small rural road improvements that are expected to have local impacts. At the other extreme, a nationally rolled out road rehabilitation scheme offers little or no hope of finding an adequate comparison group.

A fourth issue is data: to ascertain potential data sources and assess in-country capacity for data collection. If there are no existing or planned surveys that can be used or adapted for the evaluation, is quality data collection a feasible undertaking given the country's statistical capacity? Can the national statistics office be relied on in terms of implementation capacity and assuring data quality? What are the private sector (including NGO) capabilities?



## ***B. Learning from the project and its ex-ante evaluation***

Much can be learnt from the project itself. Thinking about the evaluation from the project's inception can help to understand biases in the ex-post evaluation and to define an appropriate counterfactual. In particular, understanding the selection process of the regions chosen for project placement as well as the specific roads is key to dealing appropriately with the endogeneity of placement issue. The 'appraisal' stage used to select the road links to be rehabilitated can be thought of as 'ex-ante evaluation.'

One should ask the following questions of the government (and donor) project staff:

- How are intervention areas picked? Why are certain provinces (states) included, and within them certain districts and finally certain communities? Are poor areas targeted? On what basis? Are there political influences?
- How are the specific road links selected within the chosen areas? Factors that often play a role in road link selection include population density, the cost of the proposed investment and poverty incidence. But numerous other factors, such as political connectedness or the existence of a charismatic local leader may also come into play and be much more difficult to pinpoint.

This information will best be learnt from the project's administrative unit and the transport ministry (or equivalent) that is responsible for making decisions. At times this attempt to understand program placement may be frustrating. In the Vietnam rural roads case for example, the Project Appraisal Document (PAD) said that participating provinces were chosen based on poverty indices, but the evidence did not appear to support this claim. It may be a good idea to check the claims made by the PAD (or equivalent). The understanding one gets about placement should help greatly in identifying specific information that will need to be in the evaluation data base.

But this also means understanding well the locations — including their history — and again, has important implications for the data to be collected in the baseline. One would like to have a recent history of the chosen project areas including information on prior public spending and other interventions. The more one knows in advance about the characteristics of selected intervention areas (relative to non-selected areas), the better for the evaluation. This information will be crucial for adequately controlling for initial conditions but equally importantly, for controlling for the initial trajectories of the participating locations as compared to comparison areas. It can also help guide the choice of appropriate comparison areas.

## ***C. The evaluation design: Deciding what data are needed***

**Outcome variables:** As noted, these can be numerous. How many outcome variables an evaluation focuses on depends in large part on the objectives of the evaluation, which will be setting and project specific.

One option is to focus on a single welfare or poverty indicator such as household consumption or income, assumed to encompass and reflect the net effect of myriad factors. Collection of such data can be difficult to do well and needs a carefully crafted household survey and well-trained interviewers. As a rule, consumption is more reliably measured than income and is arguably generally the better welfare indicator.<sup>12</sup> Working with a tested consumption or income module from an existing household survey questionnaire and relying on interview teams with experience in such data collection can vastly facilitate the task as well as ensure data quality. Using an indicator of welfare as the outcome indicator has the advantages that it allows a distributional analysis of welfare impacts, and facilitates analysis since the focus is on one all-encompassing indicator only. It may require that more time elapse between project completion and the post-project data collection so that impacts have had a chance to be reflected in the welfare measures.

An alternative is to focus on a large variety of outcome indicators without attempting to aggregate into a single welfare measure. This allows an analysis and understanding of the underlying linkages and potential heterogeneity across impacts. Although the distinction is not always clear, one may want to include “intermediate” or “output” outcomes — such as access, costs and quality of transport options — and more “final” outcome indicators that one expects to be affected by improved mobility. Outcome indicators could also include indicators of potential deleterious effects such as environmental damage, deforestation and traffic accidents that are unlikely to be reflected in standard one dimensional aggregates such as consumption.

Evaluations in the transport sector may naturally want to focus on impacts on beneficiaries’ time allocation and travel behavior, and potential time and transport cost savings. Others will be interested in transport modal shifts and traffic density on the improved road. The degree to which local market prices of inputs and outputs change in the project areas as well as in neighboring areas will also be of interest. These may all be considered outcome variables in their own right but they can also be of considerable use in understanding other outcomes including broader welfare measures. How changes in access affect the daily activities and mobility behavior of individual beneficiaries eventually also has bearing on more general socio-economic local area development. Migration-related outcomes are another case in point. The degree to which the road improvement facilitates and augments migration in search of work and the flow of remittances will have wide-ranging implications for living standards in the community. Similarly, how access improvements affect the type and frequency of traffic along the road has implications for what other factors will be affected and how; how prices of inputs and outputs are changing may help in understanding how markets and trade patterns are developing, as well as potential externalities of the project.

Given how little is known about rural road impacts, this type of analysis done well has large public good potential. However, the more numerous the outcome indicators, the more complex the interpretation and presentation of results becomes (for

---

<sup>12</sup> For further discussion see Deaton (1997).

an example see Mu and van de Walle, 2007). For example, different indicators may go in opposite directions and show contradictory results, making interpretation more difficult without a means of aggregating the multiple outcome variables to define the relevant tradeoffs.

There are intermediate approaches between these extremes. One can focus on a few key sets of indicators more directly linked to project objectives even if these are not final welfare outcomes. Common project objectives include market development, agricultural productivity enhancement and schooling and/or health outcome improvements. One may choose to pick a few good indicators linked to one or more such objectives (for example, Mu and van de Walle 2008). These indicators can be measured at household or community level or both.

Another approach relies on another survey, unrelated to the evaluation. A careful weighing of the constraints faced by the survey instrument and the severe difficulties involved in collecting reliable welfare information, may lead to a decision not to measure household income or consumption expenditures. However, if a nationally representative household level consumption survey exists for a date that is close to the evaluation's survey dates, it may be possible to combine it with the data collected for the evaluation to predict consumption expenditures for the evaluation survey's households. The latter survey instrument must then be designed with this objective in mind. Specifically, this means collecting information on household characteristics that can be expected to be highly correlated with household consumption and are available, similarly defined, from the national consumption survey. (These correlates of consumption may or may not be considered directly relevant outcome variables for the road evaluation. The key thing is that they are good predictors of consumption.) A consumption model can then be estimated on the national survey and the estimated coefficients used to predict consumption expenditures for the sampled households in the evaluation survey using the corresponding variables in that survey. Ideally, this can be done for the baseline as well as for the post-project follow-up round. The latter will be more credible if it is based on a model re-estimated on a national survey implemented close to the date of the follow-up round.

The results from this procedure allow a ranking of households into welfare groups and a distributional analysis without having to collect data on the numerous components of consumption. For example, impacts can be disaggregated by poor non-poor households (or more groups depending on sample size), as well as by poor non-poor communities using the household level predictions to create community means. Predicted consumption expenditures can also be used as an outcome variable as long as the impact estimation strategy controls for observables in a non-linear way (such as with PSM).<sup>13</sup>

**Control variables:** Regardless of the option chosen with respect to outcome indicators, additional data will be required for conditioning on observables, controlling

---

<sup>13</sup> Since these data are predicted based on household and community level characteristics they can not be used as outcome indicators in a regression context where one controls linearly for the same characteristics.

for changing circumstances over time, exploring heterogeneity and allowing for exogenous shocks (e.g. natural disasters, weather shocks). It can also be important to collect data on other development projects and investments — including road infrastructure — that vary over time across communities. These may be influenced by the roads project through spillover effects or complementarities (such as when new roads induce telephone lines). Focusing on a single outcome variable does not necessarily reduce the amount of data that has to be collected. There may also be overlap between variables defined as outcome indicators and their use as controls in their baseline value. Covariates can also be used to disaggregate impacts in interesting ways such as by income, education, or gender.

Data on initial conditions and historical factors that affect placement and subsequent area trajectories will need to be quite extensive to deal adequately with selection on observables. Leads on what covariates of participation to collect will hopefully have emerged from a close study of the project and ex-ante appraisal conducted in Step 2 (Section 4.2). These and the factors that may affect future trends typically include population characteristics (e.g. demographics, human capital, ethnicity), infrastructure facilities (e.g. road conditions, rural road density, non-road forms of transport, social and banking facilities) and services (e.g. passenger and freight transport services), household vehicle ownership, geo-climatic attributes, agricultural and other economic potential, sector of employment shares and so on.<sup>14</sup> As noted earlier, a number of area characteristics such as governance, culture and social capital have typically been unobserved in data, and cited by critics as potentially biasing the evaluation results. Efforts should be made to measure these directly.

Such variables are necessary to implement a matching estimator but also to explore the heterogeneity of impacts or to control for changing circumstances over time. Covariates can also be used to disaggregate impacts in interesting ways such as according to the average level of income or education, or by gender. Data on exogenous shocks such as natural disasters, severe weather shocks, and other development projects and investments that vary over time across communities should also be included in the data base as they could affect impacts. Past (pre-intervention) values of the outcome variables can also be very useful as controls for assessing impacts on the current values of those variables.

**Project data base.** Impacts will vary according to road attributes, the change in road conditions, and the method of project implementation. Sub-projects will differ widely within a specific project. Some road links may only require spot improvements, while others need a bridge or some serious engineering to be passable. Some may be paved and others may remain gravel and dirt based. Some may connect up to small villages and others to large towns. The impacts from rehabilitating a road can be expected to vary according to these various characteristics.

---

<sup>14</sup> None of the initial conditions included in the logit for PSM can be policy variables that are determined by the levels of government that are also responsible for picking the project participants — for example, whether a community received outside funding for rehabilitation in recent years — since matching on variables that are jointly determined with project participation may lead to mismatching.

Therefore, it will be important to set up a project level database for each of the project areas surveyed with information relevant to the evaluation including the condition of the roads before and post rehabilitation, kilometers built or rehabilitated, employment days of local labor, costs, start and end dates, and precise location. Another important factor to keep track of is whether funding and institutional arrangements for routine maintenance are provided for and whether maintenance is being carried out post intervention. Its lack can significantly reduce the impacts of newly improved roads as well as the sustainability of impacts. One might also collect measures of how well the post project road condition is holding up. Finally, collecting data on the size and attributes of the places the road connects up with may also be important.

One way to capture the heterogeneity in road works in the analysis might be to report impacts disaggregated by the level of road rehabilitation, its cost or whether a post-project maintenance program was adopted and enforced. At any rate, it will be important to collect a detailed project database.

**Defining the road's zone of influence:** An issue arises in defining the zone of a road's influence — hereafter the project area (for definitions see Box 3) — and the data collection domain. Where do we look for impacts? There is surprisingly little explicit discussion of this issue in recent attempts to implement rural road impact evaluations. Most studies typically test for impacts in well-defined communities such as villages serviced by an improved road or located in its vicinity.

In practice, this decision is likely to be influenced by the specifics of the road project under study and the practicalities of the evaluation and data collection. For example, the survey design for the Vietnam rural roads evaluation takes the commune — a local administrative geographical unit which typically comprises of a number of small villages — as the project's zone of influence. This is justified first by the project objectives — namely to link commune centers (where key social, economic and administrative facilities are located) with road and market networks. The second important justification is that the commune is an enumeration level at which data is commonly and more easily collected in Vietnam. It usefully provides a sample frame. For these reasons, the project area is defined as the commune that the road passes through. This approach will not always be ideal, but may often make sense given the constraints faced.

An alternative that is sometimes followed is to set a maximum distance on either side of the road link — studies typically use anywhere from 2 to 15 km — and confine the search for impacts to this area. A Global Positioning System (GPS) receiver (discussed shortly) can help identify the spatial coordinates of the area desired (Landry and Shen, 2005). Identifying the project area in this manner will typically require that a listing be made of all communities and households located in the area so that a sample of primary sampling units and households within them can be randomly selected (see Box 3 and section on sampling).

The definition of the project area can be important. Obviously, data collected within the road project's defined zone of influence do not allow an investigation of

impacts that may spillover to neighboring areas. One can imagine a situation where better roads displace local development to the now more accessible closest external commercial center. As a result, negative impacts may be found for the project areas while total impacts are underestimated since they fall partly outside the delineated sample. One should endeavor to assure that the chosen zone of influence is appropriate to capturing the likely benefits in each setting, based on a priori information about the project setting.

**Potential beneficiaries and the unit of analysis for measuring benefits.** Beneficiaries are simply the individuals, households, firms and communities found within the project road's delineated zone of influence. Impacts can be measured for any one or more of these beneficiary groups. The choice depends a lot on the objectives of the road project and of the evaluation, and on the constraints on data availability. For example, data and analysis can focus on community level variables if one is solely interested in assessing whether improved roads lead to better transport services at community level. Alternatively, if the focus is on whether the road improvements result in a reduction of the time women spend on chores, then household or individual level data will also be required.

In most cases, one will probably need to collect both types of data — individual (household or firm), and community level. Since project placement is at the community or geographic level, one will typically need to collect community level information for the PSM and for understanding the heterogeneity of impacts. Note that while some area level variables can be created by aggregating household level information that is unlikely to be possible for all the things one would need to know at community level. Hence, surveys will be needed at both levels.

**Choosing non-project comparison areas:** An issue that is related to the delineation of the project area is that of identifying appropriate non-project comparison areas.<sup>15</sup> Ideally, comparison areas differ from the treatment areas in terms of mean outcomes only in so far as they do not receive an intervention. Sometimes, pre-existing survey data — perhaps a recent agricultural or population census — can be used to do a pre-matching of project and other locations on the basis of geographic and other community level characteristics. However, the requisite data to make informed choices on ideal comparison areas is often not available when decisions need to be made. Thus, a preliminary set of non-project comparison areas will frequently need to be selected in an ad hoc way and later modified using the data collected for the impact evaluation to form a better comparison group. Sample design should anticipate that such trimming of the original sample of comparison areas will be required.

To ensure similarity as well as for practical and logistical reasons, a common strategy is to pick potential comparison areas in the vicinity of the treatment areas (though outside the catchment area of the project's impact). The comparison areas chosen this way should share many of the same characteristics as the project areas. However, we

---

<sup>15</sup> Since project placement is at the geographic level, comparison units must be defined at the same geographic level. Analysis of impacts can then be conducted among households located within that geographic area.

cannot be confident that they are a good comparison group on a priori grounds. A more rigorous matching using propensity scores and based on the data collected for the evaluation should be used to test the selection of comparison groups. Any areas from the set of selected non-project areas with unusual attributes (relative to the treatment areas) and hence extreme propensity scores (far lower than for the project areas) can be deleted from the comparison group. The sample size needs to allow for this possibility.

The same principles apply to the case where the project zone of influence is defined as the area within a specified distance from the road project. Comparison units will have to be identified in a similarly defined catchment area around a comparable road with no planned improvements in the case of rehabilitation projects, and a similar but roadless area for new road projects.

#### ***D. Evaluation design: Collecting the data***

**Data sources.** Various pre-existing data sources (population, agricultural and community level censuses, geographic, administrative and financial data bases, detailed road and transport network inventories, and household surveys) can be useful to the evaluation. For example, they can help to implement a pre-matching of prospective comparison areas as discussed, or as a source of information on the history of community trajectories to control for selection bias. In some cases it may be possible to use a recent community or household survey, or piggyback on one being fielded prior to the project start. In this case, it will generally be necessary to get the surveyors to agree to over-sample in project areas so that there are sufficient observations on beneficiaries to evaluate impacts. The rest of the data can then provide comparison groups. Alternatively, if pre-identification of potential comparison areas has been possible, over-sampling households in those areas may also be desirable.

But the timing, sampling properties, or included information of surveys not connected to the evaluation will seldom be ideal or sufficiently flexible for the purposes of the evaluation. Furthermore, ensuring that the survey can be replicated in panel form in a future follow-up round may also create difficulties. Another consideration is that the data for project and non-project areas should ideally come from the same source. A number of past studies have demonstrated that impact estimates can be highly questionable when survey design, implementation, and processing are not identical between treatment and comparison units (see, for example, Heckman et al., 1999). Most commonly then, one or more purposive surveys will need to be implemented by the evaluation.

When a baseline does not exist it may in some cases be possible to recreate it ex-post. One way is to rely on existing data as discussed above. The problem with this approach will be that a survey planned and implemented independently of the project evaluation will rarely have sufficient geographical overlap or observations on beneficiaries. Alternatively, it may be feasible to create a baseline through recall. The degree of measurement error associated with recall data depends on the recall time period and what's being measured. The expectation is that recall is more reliable for recent events (Deaton 1997). So, there's a clear trade-off between longer recall periods and

more information, versus higher recall error. Furthermore, some events and information can be expected to be more easily and correctly recalled than others. One is likely to remember buying a bicycle a year ago and even how much it cost much more readily than food consumption back then. In sum, using recall may be a good option for baseline information on area characteristics used in matching or as regression controls but less reliable for outcome measures such as of welfare.

Household level questionnaires should whenever possible be based on an existing national household expenditure survey's consumption and other relevant modules. Questionnaires that have been used for the impact evaluation of rural road interventions in other countries at both household and community level can be useful but must be carefully adapted to the specific country and project context (notably its objectives and placement).

Much of the necessary information for the impact evaluation will have to be collected by way of quantitative, structured sample surveys of communities and households. Grosh and Glewwe (2000) provides an exhaustive treatment of how to go about implementing such surveys. The community level questionnaires can include modules of interest that require different methods of collection. A price and goods availability module would need to be filled out with visits to the main market. Specific modules may require specific respondents such as the school master or local transport operators. Naturally, in some settings, there will be no transport services (or other outcomes of interest) in the baseline and it is of interest to assess whether the road project has an impact on their availability. In other cases, transport service providers will be based elsewhere and not easily interviewed.

Collecting qualitative information could be useful in such cases. For example, focus group discussions with transport operators may be one way of getting at perspectives on how the road improvement has altered the situation from the point of view of service providers. Qualitative data from unstructured interviews can be of considerable use in interpreting and understanding the underlying and intermediate processes that explain final impacts, as well as to inform the design and content of the quantitative survey instruments. However, such data can not on its own properly identify the counterfactual, establish causality, or identify generalizable impacts with known statistical properties. See Rao and Woolcock (2005) for a discussion of how to best combine qualitative and quantitative data and data collection methods in the context of impact evaluations.

Data on time allocation, travel behavior and traffic volume require specialized survey instruments. Time use and travel behavior information require individual level questionnaires and can substantially raise the costs of data collection. It is usually recommended that they be administered to all household members above a certain age (ten years, say). To enhance reliability, they may need to be collected over the course of more than just a single day. Both time allocation and travel behavior can be collected in the form of diaries. However, this typically requires adequate literacy. Time use surveys measure how individuals allocate their time across different activities over a typical day.



Travel diaries collect information on all of an individual's trips small and large, the mode of travel, time spent, distance, cost and purpose. Baker and Denning (2005) provide an example of a traffic diary module.

We would expect the road intervention to have impacts on the time it takes to get to (fixed) key work and non-work destinations. Measures of travel time savings can be calculated (DFID, 2005). It has often been noted that survey respondents' estimates of the time it takes to get to places and their distance are unreliable and inter-personally and cross-spatially not comparable (Escobal and Laszlo, 2007; Gibson and McKenzie, 2007; Roberts et al., 2006). Roberts et al. (2006) use pedometers to assess subjective reports of times and distances in rural Albania and Tanzania. They conclude that self-reports are more reliable in places such as Albania where respondents are more steeped in a culture of using timepieces. But, in general, time taken to destinations was found to be less error prone than estimated distances to those same destinations.

In the context of Peru, Escobal and Laszlo (2007) compare self-reported times and distances to markets and social infrastructure with actual respondent travel times as measured by an accompanying surveyor for part of their sample. They find substantial error in self-reported times and distances that is correlated with socio-economic characteristics (including whether the respondent has a watch which is itself related to income), with implications for using such data for policy analysis. In settings where people walk a lot it may be possible to provide them with pedometers that measure distance (by counting steps) and time to places. These could be used as part of a travel or time use diary survey instrument.<sup>16</sup> Interview teams may also be able to measure time and distances consistently at the individual and household level as part of the survey work, as done by Escobal and Laszlo (2007). However, this will be costly in time and resources and may only be possible for a small sample. Alternatively, Escobal and Laszlo (2007) suggest and describe using Global Positioning System (GPS) devices together with a Geographic Information system (GIS) to measure time to destinations that allows for terrain and natural barriers to mobility for larger samples.

Given recent technological advances and falling costs, GPS receivers — that mark the precise location of people and places — used in conjunction with a GIS — that maps, combines and displays various spatial information — can be of great help in transport studies. Gibson and McKenzie (2007) summarize the various uses and benefits of using GPS, GIS and geo-referenced data for various analyses and provide many useful references. Recent surveys have successfully experimented with hand held computers called Personal Digital Assistants (PDA). These can be equipped with global positioning satellite cards so that the geographical coordinates of a location can be recorded as part of an interview. It is recommended, for example, that the household survey record each sampled household's coordinates, and its distance to the road and to key facilities such as schools, health posts, and markets, some of which may not be in the surveyed project area. Community level data should also record GPS coordinates of social and economic infrastructure. Having before and after maps of project areas and their surroundings that

---

<sup>16</sup> A detailed description of the use of pedometers in a survey context is given in Annex 5 of Roberts et al. (2006).

identify social and economic infrastructure would help understand the processes that underlie impact measures as well as potential spillover and displacement effects.

One issue is whether GPS — relatively straightforward and easy to use — can be recommended for measuring traveling time and distances without being combined with GIS — which is clearly more complicated and demanding of skills and resources. With two sets of GPS coordinates, one can calculate a straight line distance between two locations. For some applications, that measurement will be preferable to self-reported distances where measurement errors may be correlated with outcomes of interest. However, travel times based on GPS do not account for topography and existing transport and travel routes and can therefore be misleading in some situations. Gibson and McKenzie (2007) discuss this issue and conclude that setting and topography will be the decisive factors. Road or travel distances will clearly be preferable to straight line distances in locations that are isolated, mountainous or dotted with natural obstacles such as lakes or rivers.

Traffic volume surveys manually count traffic on a given road for a given period of time — for example over daylight hours during the course of one week. Typically, details are collected on numbers of different types of vehicles — including bicycles, motorcycles and animal drawn carts (depending on the setting) — and their passenger and freight capacity. Pedestrian traffic and cases of head-loading can also be recorded. Gender specific counts may be of interest. See Grootaert (2003) for an example of a traffic density module. If a way can be found to stop traffic operators and informally ask them questions about destination and distance traveled, this may be a good way of combining qualitative and quantitative information.

A good source for the necessary project data is the monitoring data base that is set up for the specific project and its sub-projects. This typically assembles data reflecting project inputs, costs and outputs (e.g. road conditions pre and post, start and end dates, exact location details, kilometers of road and type, number of beneficiaries for projects completed, number of people employed) and transport characteristics before and after the project (e.g. average daily traffic, freight rates, passenger transport frequency and rates). It will often be necessary to add to this data base a number of additional project related variables deemed useful for the evaluation such as whether a maintenance plan has been put in place and is being successfully implemented post-project. These data may be collected and maintained both at the local level and/or centrally.

An evaluation team member who is independent from the project will need to be in charge of gathering this data, and making sure that it is correct and complete. This task often requires care and close supervision to check that the information in the monitoring data base actually accords with what is going on at the sub-project level. Infrastructure projects are notorious for being subject to corruption and it is easy to doctor the project monitoring data base. Honest mistakes are also easily made. In the Vietnam evaluation, after we queried the central PMU about project data assembled for the evaluation at the provincial level, we noted large disparities with the data received from the center. Following a bout of intense detective work and visits to each province,

we uncovered a number of inconsistencies in how the data were being recorded. To give just one example, some provinces recorded the date at which contracts were signed as the start date while others and the central PMU recorded the start of the actual work which was sometimes as long as a year later. The reliability of these data is key to being able to conduct the analysis of project impacts.

**Sampling and sample size:** The precision of impact estimates (their standard errors) will depend on both the sample size and the design of the sampling. Two stage sampling is the norm. In the first stage one samples geographic areas such as communities — called the primary sampling units (PSU) — within the overall project area defined as the area within which the road project is expected to have impacts (Box 3). In the second stage one samples households or firms within each PSU; this second stage is often done in the field (notably when one does not have an up-to-date listing of all households or firms from which to draw the sample). A sampling expert may be needed to help determine how many communities in project and non-project areas should be surveyed as well as deciding on the number of households to be surveyed and what sampling strategy to pursue inside the selected communities. The classic reference on the design of sample surveys is Kish (1965).

Here I will only focus on some specific issues of importance in sampling for the purposes of doing an impact evaluation. The best design will depend in part on the purposes of the evaluation. The standard recommendation from sampling text books may need modification in the light of the objectives of the study, notably its intention to draw useful lessons for policy.

If the usefulness for policy purposes depends on one's ability to draw reasonably precise estimates for certain sub-groups (such as communities in a specific region, or poor households) then one will of course need a larger sample or a sample that is appropriately stratified along the relevant dimensions (or available proxies for them). When designing household panel surveys, such as often used in estimating difference-in-difference estimates, the sampling must allow for potential household attrition, as units will no doubt drop out of the subsequent survey rounds.

Given budget constraints, one typically faces a choice between sampling a larger number of households within each PSU versus sampling a larger number of PSUs. Large sample sizes within PSUs, with correspondingly fewer PSUs sampled, will create a high clustering effect on overall standard errors given that sample points are not independent within a PSU.<sup>17</sup> Thus, in conventional sample designs, household sample sizes collected within communities (PSUs) will often be small to enhance the precision of the overall sample estimates. Against that, if one wants to draw reasonably precise inferences about heterogeneity in local level impacts (at the PSU level or possibly higher) then one may be drawn to choosing higher sample sizes at the PSU level since even random samples within each PSU may give high standard errors for impact estimates or other relevant

---

<sup>17</sup> Sample clustering needs to be taken into account when estimating the standard errors. If as expected, intra-cluster correlations are high, standard errors will also be high. A larger sample of PSUs will then be needed to ensure precise average impact estimates.

variables for a given PSU. Again, for a given budget available for the evaluation, this will mean a lower number of sampled PSUs. The design effect (whereby the households sampled within a PSU cannot be treated as independent of each other) will then tend to increase the standard errors of the overall impact estimates. Thus there can be a tradeoff between the precision of overall impact estimates and the ability to understand the heterogeneity in impacts; in the context of impact evaluation, this trade off is discussed further in Ravallion (2008).

In choosing a sample of the treatment areas one normally wants a representative sample of the population of project participants (which constitutes the “sample frame”). However, in sampling comparison areas it should be kept in mind that the objective is not to obtain a representative sample of the non-participants, but rather to obtain a good comparison group for the participants. Simply drawing a random sample of the non-project areas would almost certainly not be efficient, since one will probably end up throwing away data (outside the region of common support). This can add complexity to the design of the comparison group sample. Stratification will be needed to permit deliberately over-sampling types of non-participants who are likely to have similar characteristics to the participants.

It is important that the household level sampling is representative of the relevant sample frame. Drawing a random sample assures that the household sample is representative, but that will only be true of the final sample data if all those actually sampled agree to be interviewed. Rich households appear often to be unwilling to participate in surveys. Sometimes, the interviewers themselves might want to avoid certain groups, possibly amongst the poorest and richest, or certain ethnic groups. Such sources of non-sampling errors can create biases when one is concerned with distributional impacts.<sup>18</sup> Extra effort may be needed to ensure that sampled households came from different positions in the distribution of income; for further discussion see Groves and Couper (1998).

In many countries, some kind of welfare ranking is available within communities. It may then be possible to conduct a form of stratified sampling whereby, say, a third of the sample households are chosen from each of three lists, containing the poorest, middle and richest thirds of all households in the commune. Clearly, these rankings are to some extent subjective, but stratified sampling on this basis should assure a sample that is reasonably representative of each community’s main socio-economic groups. This was done successfully in Vietnam. The Ethiopian survey also relied on village household listings (Dercon et al., 2007).

**Timing:** The evaluation needs to start as soon as possible after the project has been identified and it will need to extend at least a few years after the end of project implementation to convincingly capture its impacts. Much of the evaluation-related work will be concentrated in a few months at the beginning and end of the period. The data

---

<sup>18</sup> See Korinek et al (2006) for a discussion of the implications of selective compliance and a correction method.

collection will also take place at the beginning and end of the period, with a possibility of an additional survey being fielded sometime earlier in the post-intervention period.

The evaluation team should be identified and hired as soon as possible and should begin work on the evaluation, sampling and questionnaire design. Timing of the baseline data collection is critical. It needs to be fielded prior to the project start and preferably before communities and households have discovered that they are to benefit from road improvements. Clearly one does not want the baseline to be done too far in advance of the project either. There are risks and uncertainties associated with project timing which complicate things. Often, project sites will not be fully approved or even selected by the time the baseline needs to be prepared and sampling implemented. In the Vietnam case, for instance, many of the sampled project areas did not have a completed road link project as intended pre-baseline by the time of the first follow up round. One small advantage of this delay is that the areas still awaiting projects may be used as comparison communes for analysis of early impacts.

Follow-up (post project) rounds need to mimic as closely as possible the implementation process followed for previous rounds of data collection. Geographic coverage timed across months should coincide as much as possible with the schedule followed during the baseline. It is a good idea to have the same in-country survey team and supervisors for all survey rounds. Even then, retraining before each round is likely to be required.

Another timing issue has to do with the potential seasonality of outcome and control variables. Depending on the setting, many variables of interest — including transport costs, time allocation and agricultural production — may vary by season. As already noted, one should ensure that baseline and follow-up survey rounds are fielded at the same time of the year for the sake of comparability across time. It may also be important to collect data on some variables for different times of the year, for example by rainy and dry season in case impacts are also seasonal.

### ***E. A note on risks and pitfalls***

There are many risks to conducting a rural roads evaluation, though some are also shared with other types of evaluations. The paper has already alluded to some of the problems that can arise. Others that it may be useful to be aware of include the following:

- The evaluation team may have little control over project timing. Sub-projects may be delayed or changed en route so that far fewer road links are completed by the time the post-project survey round is slated to be fielded. This extends the amount of time needed to have an adequate sample size and complete the evaluation.
- Governments in many countries have a habit of periodically splitting administrative units such as may be chosen as project areas. This was the case for some of the communes in the Vietnam case. The baseline data for a particular

commune may then no longer be representative of either of the new administrative units formed and so the commune must be dropped from the study, affecting sample size.

- Things don't stand still for the evaluation. There may also be household attrition for various reasons. Other development projects, including road projects may be implemented in the survey areas, complicating the evaluation design.
- The weather can present severe problems when one is collecting panel data and trying to time data collection similarly across years and regions in a specific year.
- Finally, it is not easy to maintain outside interest and enthusiasm for a long lasting study of this kind, where a great deal of work must be done before there is much to show for it. Patience and considerable staying power are required.

#### **IV. Selected findings from recent rural road impact evaluations**

In recent years a number of studies have assessed the impacts of rural roads rigorously using impact evaluation methods that expressly deal with selection. These studies show mixed results, some finding substantial impacts and others more muted impacts. They have examined disparate outcome variables, in diverse circumstances, using various techniques — some of which can be questioned. Sources of ambiguity in impacts can also be expected due to heterogeneity. It thus remains difficult to draw definitive conclusions concerning the impacts of rural roads. The following discussion briefly summarizes the main findings from these evaluations.

Escobal and Ponce (2004) use PSM to find that rural road rehabilitation enhanced non-agricultural, particularly wage-based, income earning opportunities in poor areas of rural Peru. Higher incomes did not result in higher consumption but were reflected in higher savings in the form of livestock. Unfortunately, the study suffers from a lack of panel data. Instead it uses a cross-section of households living near a rehabilitated road and compares their outcomes with those for households not living near the road matched using propensity score methods. It is not clear how well this can deal with selection bias. It is also unclear why the outcome indicators would be affected by the road, while the characteristics included in the matching exercise can be considered exogenous.

Khandker et al. (2009) use a panel fixed effects regression method, with regression controls for initial conditions. They find that road investments in Bangladesh reduced poverty significantly by raising agricultural production, wages and output prices and lowering input and transport costs. Their results also suggest that schooling outcomes for both boys and girls improved as a result of the road improvements. Impacts were found to be proportionately higher for the poor relative to the non-poor.

A study of the impacts of road rehabilitation in Georgia using a double difference with propensity score matching also finds that impacts vary between poor and non-poor households (Lokshin and Yemtsov, 2005). It estimates that in the aggregate, opportunities for off-farm and female wage employment were significantly increased in project versus control villages. However, further disaggregation indicates that off-farm employment improved solely for non-poor households and female wage employment increased for poor women only. In contrast, the study finds no impact of road quality on agricultural product sales.

Also combining a DD with PSM and focusing instead on the community level impacts of road rehabilitation on local market development, Mu and van de Walle (2008) also find generally higher impacts in poorer areas of rural Vietnam. There are significant average impacts on the presence and frequency of markets, the availability of various services, household diversification out of agriculture into off-farm service based activities and on primary school completion rates. But the study also documents substantial impact heterogeneity, with a tendency for poorer communes to have higher impacts due to lower levels of initial market development. Moreover those impacts are tempered by various other attributes of poor communes — such as high shares of ethnic minorities and high illiteracy rates — that tend to reduce the benefits of road improvements. These findings have important implications for project design.

Using a dynamic growth model approach, both Jalan and Ravallion (2002) and Dercon et al. (2007) estimate that better rural roads promoted higher rates of growth; for example, Dercon et al. estimate that changes in access to quality roads increased consumption growth in rural Ethiopia by 16 percent and reduced poverty by 7 percent. Jalan and Ravallion assume that initial road placement is conditionally exogenous, arguing that they have sufficiently good controls for this assumption to be justified. Dercon et al. use household fixed effects but do not control for initial conditions that may affect changes in placement and subsequent growth rates, raising the concern that they have not adequately dealt with endogeneity biases.

These rural road studies differ in a number of ways that make comparisons and generalizations difficult. One difference is between the use of regression controls versus PSM. The difference between these alternative ways of conditioning on covariates is not yet well understood in the literature. It appears to be thought by some researchers that controlling for initial conditions in a linear regression (as done by Khandker et al. 2009 and Jalan and Ravallion 1998, 2002) is similar to doing so non-parametrically with PSM (as done by Ravallion and Chen 2005; Lokshin and Yemtsov 2005; and Mu and van de Walle 2008). However, the methods make a number of rather different assumptions, including on functional form and the use of the full sample versus trimmed or re-weighted samples and may not give similar results. In particular, the linearity conditions may not be innocuous. Matching or re-weighting the data using propensity scores can claim to be the more flexible way of controlling for observable heterogeneity. Comparing the two approaches, van de Walle and Mu (2007) find quite striking differences in the impact estimates despite controlling for the same initial conditions.

## V. Outstanding issues for rural road impact evaluations

This section points to a number of issues that lie beyond the main scope of this paper, but may arise in practice. Some of these issues are not as yet well understood, but future methodological research may well change that.

### A. *Specific and generic issues where more needs to be known*

The classic impact evaluation of an assigned intervention can reveal disappointingly little about the precise ways in which the intervention does or does not have impact. To better understand welfare impacts, evaluations need to focus on understanding some of the key supply-side and other behavioral responses to road improvements. There are a large number of specific questions on which negligible evidence exists: How do freight and passenger transport services respond? Do facilities and services relocate? Is there evidence that better teachers, doctors and other professionals are more easily recruited? Do improved roads act as a magnet for other public and private development projects in the vicinity? What factors or conditions determine whether such responses occur, and how much of the total impact on welfare depends on such supply responses? For example, do better roads lead to higher school enrolments and health care facility utilization as is often claimed by advocates? How much of the increase is attributable to the direct effect of better access, versus the indirect effects of better school quality, a new bus service, or higher incomes, all indirectly attributable to the road improvement? Many of these questions will call for data on what can be thought of as “intermediate indicators,” i.e. variables that are not of interest as outcomes per se but rather as variables that transmit impacts. (In doing so, one takes a step in the direction of the “structural methods” discussed later in this section.

The importance of heterogeneity has long been recognized in the literature in principle, but empirical work has often ignored it. This has risked seriously misinforming policy and choices. Heterogeneity of impacts can be expected to arise according to the characteristics of the community where the road is placed and across households within a community according to their characteristics. Of considerable interest is to better understand which contingent factors influence welfare impacts. One objective here concerns better understanding what complementary public policies or investments would enhance the returns to rural road investments. Future work needs to strive to understand the conditions and factors that influence mean outcomes and their distribution, including interaction effects with other development initiatives. There are obviously important implications for data collection — to assure that one collects data on factors likely to be dimensions of heterogeneous or explanatory factors.

As has already been discussed, little is known about the distributional impacts of rural road investments. Some have argued that rural roads lead to greater disparity between rich and poor and hence, to rising inequality (Cook et al. 2005; Hettige 2006; Howe 1997). Short and longer-term impacts may differ for different groups. For example, the poor may lose out in the short term and yet experience significantly improved livelihood opportunities and living standards over time as a result of the



dynamic effects resulting from the road investment. Key questions for impact evaluations concern what factors at community and household level attenuate and augment distributional impacts, and how they alter with the amount of time elapsed following a road improvement. Are there other policies that would enhance impacts for the poor if implemented concurrently?

Further issues on which there is little evidence are aid fungibility and spillover effects in the context of transport projects. Does the externally funded road financing actually improve roads? The government may simply cut its own funding of roads in response to the external aid. Alternatively, the local authorities may decide that for the sake of fairness, non project communities should get a higher share of local spending on infrastructure since they are not benefiting from the donor financed project. Equally well, the road project may itself attract other investments. Clearly, such displacement or induced spending have important implications for data and assessing impact. van de Walle and Mu (2007) found that road rehabilitation project resources in Vietnam were diverted to road building, but stuck to the sector. Similarly see Chen et al. (2007) on how local government spending (including on roads) responded to aid from a World Bank poor area development project.

Finally, an important objective of the analysis of ex-post impacts should be to draw implications for the ex-ante evaluation of rural road investments. Current methods for the selection and appraisal of rural road projects leave a lot to be desired (van de Walle, 2002, Huenemann, 2008). Impact evaluations can help provide better information on the non-pecuniary benefits of rural roads.

Similarly, evaluations can help draw useful lessons for the choice of road project monitoring indicators. As noted in Section 4, Bank infrastructure projects typically include data collection for monitoring and performance evaluation purposes. These measures tell us little about impacts on living standards per se. Some project evaluations have tried to get at this more directly by also collecting information on pre- and post-project indicators such as agricultural production, yields and costs. But, without the same data for control groups, these data may be uninformative and even misleading about causal relationships. A common (economy-wide) effect could mean that similar changes in living standards are observed in areas not covered by the project. Good impact evaluations can provide a better basis for making decisions about appropriate monitoring indicators when a full-fledged ex-post impact evaluation is not feasible.

## ***B. Structural methods***

The impact evaluation context discussed in this paper is largely a-theoretical and reduced form. It often remains unclear how exactly roads have their impacts. To some extent, the careful choice of ‘outcome’ variables can help identify the channels of impact, by focusing on intermediate outcomes. However, there are clear limits to how far this is possible in practice.

Alternative approaches identify impacts on the basis of economic assumptions about how the world works. These approaches rely heavily on a structural model of

behavior where clear economic assumptions propel the specification. This helps make up for important missing data. The downside is that the assumptions on which the model is built may not be evidently plausible or empirically testable. The upside is the gain in terms of what we can learn, as long as the assumptions are valid.

Structural methods can use either partial or general equilibrium models. There are a number of examples of partial equilibrium work estimating the impacts of reduced transport costs through road provision. Two good examples are Jacoby (2000) and Jacoby and Minten (2007). The latter develops an agricultural household model to derive the willingness-to-pay for a reduction in transport costs in a poor, isolated, rural region of Madagascar where transport costs vary commensurately with distance to the market town. This is used to estimate the benefits of a hypothetical road project under the assumption that transport cost disparities can be treated as exogenous. The paper concludes that a road that eliminated transport costs would raise the incomes of the most isolated households by nearly half, primarily by raising non-farm earnings.

Jacoby and Minten carefully model and estimate an agricultural household model, in the expectation that the key benefits will be from the large reduction in the costs of transporting goods in and out of remote areas where households are overwhelmingly farmers. Off-farm incomes are not explicitly modeled but added in an ad hoc way once it is discovered that they account for a substantial share of income for households with access to the urban area.

The latter study illustrates two broader lessons. First, some road impacts may be difficult to model and estimate in structural form. Second, road infrastructure may have unexpected impacts not factored into structural models.<sup>19</sup> This suggests that reduced form approaches such as ex-post impact evaluations can help provide insights to structural models that attempt to estimate road impacts, and vice versa. The two approaches — classic reduced form impact evaluation and structural form models — should be seen as complements.

## Conclusions

Despite their popularity, very few aid-financed rural road projects in developing countries have been the subject of rigorous impact evaluations. Knowledge about their impacts and the heterogeneity in those impacts continues to be limited. And yet more information based on careful, rigorous impact evaluations could substantially improve aid effectiveness generally, as well as road project appraisals, design, and selection methodologies specifically.

---

<sup>19</sup> Other studies caution that road investments are not sufficient to resolve the barriers to commercialization. Eskola (2005) examines the impact of commercialization on the consumption of rural Tanzanian farmers, and the role of roads and transport in inducing market participation. She concludes that higher participation would help raise incomes significantly. However, other market rigidities, household behavior and structural constraints to trade also militate against farmers' commercialization in Tanzania.

Yet it must be acknowledged that rural roads pose challenges for evaluation. The benefits of rural roads are indirect and conditional on interactions with the geographic, community and household characteristics of their location. Road locations are typically determined by those same characteristics confounding inferences based on comparisons of places with roads versus without them. Additionally, impacts may be distributional, felt across multiple outcomes and take a long time to emerge. These features of rural roads have implications for evaluation design and data collection.

A credible rural roads impact evaluation requires panel (with pre-intervention) data for project and appropriate non-project areas; detailed information on outcome indicators, baseline attributes and controls for heterogeneity and exogenous time varying factors. Geo-referenced data can also be extremely useful. Such a database allows for an evaluation design that combines a double difference with controls for initial conditions either through propensity score matching, regression controls or an IV. With the appropriate data and techniques a large number of interesting and far-reaching questions about rural road impacts can begin to be answered and discussed with large public good benefits.

**Table 1: Types of endogeneity of program placement associated with the existence of third variables (X)**

Source of endogeneity bias		Appropriate estimation method	
Xs are:		Xs are observed	Xs are unobserved
Type 1	Initial condition correlated with the level of the outcome variable	Single difference with controls	DD
Type 2	Initial condition correlated with changes in the outcome variable	DD with controls for levels of X	IV
Type 3	Time varying factors correlated with changes in the outcome variable	DD with controls for time-varying Xs	IV

Note: “Controls” can be either PSM or regression controls. DD: double difference; IV: Instrumental Variable; PSM: propensity score matching.

**Box 1: Deceptive assessments of the gains from rural roads**

	Mean incomes in villages with and without a road (\$/day/person)		
	Without road (n = 56)	With road (n = 44)	% increase (t-test)
Case 1: Road yields 20% income gain	1.287	2.413	87% (2.29)
Case 2: Road yields no income gain	1.287	1.976	54% (2.00)

The table shows mean incomes for a group of villages that do not have road access and a group that does. Mean income is much higher for the villages with roads. From such statistics the conclusion is sometimes drawn that the roads generated these large gains — 87% increase in mean income for one group of villages and 54% for the other in this particular case.

These numbers were created, however, by a model in which roads generated an income gain of only 20% for case 1, and no gain for case 2. The model's pre-intervention incomes were drawn randomly from log normal distributions. Road placement was determined endogenously, as a function of village income (with 25% weight) and a second independent log-normal random variable (75%). The latter could represent population size, ethnicity, likely votes, historical accident, or any other variable influencing road location by the government. Thus, roads are distributed across villages in terms of a latent variable  $z = 0.25y + 0.75x$  where  $y$  is log income and  $x$  is the other determinant of road placement. The model gave a road only to villages with positive values of  $z$ .

Of course, the evaluator does not know the true impact of the roads and is tempted to base an estimate on the observed differences in mean incomes between villages with a road and those without. This yields a large overestimate.

Source: Reproduced from van de Walle (2002).

## **Box 2: Steps in the evaluation of a rural roads project**

### ***Step 1: Deciding whether to implement an impact evaluation***

Is there sufficient support and cooperation:

- From the government?
- From the bank project team and bank management?
- From funding sources?

Is a credible evaluation feasible:

- Is there in-country capacity (data collection, supervision)?
- Are there existing or planned surveys that can be used or questionnaires that can be adapted?
- Is there a potential sampling frame in the prospective zone of influence?
- Is there time to prepare and field a baseline before the project begins?
- Can a counterfactual be identified under seemingly plausible assumptions?

### ***Step 2: Learn from the ex-ante evaluation***

Understanding program placement: to understand biases in the ex-post evaluation and define an appropriate counterfactual

### ***Step 3: Set up the evaluation team***

- Finding a stable in-country home for the evaluation.
- Choosing an evaluation team that is reasonably independent of executing agency yet can work with that agency as need be: local counterpart, interviewers, data processors.

### ***Step 4: The evaluation design: Deciding what data are needed***

Outcome variables (distributional impacts, traffic counts, time use, travel diaries)

Control variables

Project data

Choice and definition of:

- Zone of influence
- Beneficiaries: communities, households, firms, individuals?
- Comparison areas

### ***Step 5: The evaluation design: Collecting the data:***

Identify data sources and data collection methods

Sampling and sample size

Designing survey instruments

Deciding on timing of baseline and follow-up rounds

### ***Step 6: Analysis and writing up***

Plan adequate time for data processing: entry, cleaning, lessons for follow-up; analysis of baseline

Plan for follow-up survey(s).

### **Box 3: Some definitions**

The project area for a rehabilitated or new road can be defined as its “zone of influence”, or “catchment area,” given by the area around the road where project impacts are expected. This could be a local government area or community serviced by the road, or might consist of a number of communities in its vicinity. The set of all such areas defines the sampling frame from which one selects a random sample of primary sampling units (PSU) and within these, a random sample of beneficiaries. The beneficiaries of the project can be defined as the entire project area or the communities, firms, households or individuals located within the area.

Note that road projects typically select road links or segments, not geographic areas. However, these road segments are not independent of their project areas; by selecting a road segment one automatically selects a project area as defined above. Selection of a road segment can thus be treated as the (implicit) selection of a project area.

The comparison units must then be selected from within the sub-set of the non-project areas that appear to best represent the counterfactual of what would have happened in the project areas in the absence of the project. Matching to ensure a sub-set of non-project areas that best represents the counterfactual should be done at the same geographic level of aggregation (e.g. local government area or community level) used in defining the PSUs.

The unit of analysis for measuring benefits will often be at a level below the project area or PSU. For example, one might look at outcomes for households or firms within the project area, recognizing that certain units may benefit more than others. The fact that these units of analysis are geographically clustered must be allowed for in calculating the standard errors; the households or firms within the PSUs cannot be treated as being sampled independently, since they are all linked by their location. This has implications for sample design as discussed in the text.

## Bibliography

- Abramson, Daniel Benjamin. 1997. "Marketization and Institutions in Chinese Inner-city Redevelopment: a Commentary of Lu Junhua's Beijing's Old and Dilapidated Housing Renewal." *Cities* 14 (2): pp 71-75.
- Baker, Judy and William Denning (2005), "Development of a Transport Module for Multi-topic Household Surveys," Transport Papers No. 5, Transport Sector Board, World Bank, Washington, D.C.
- Binswanger, Hans, Shahidur Khandker and Mark Rosenzweig (1993), "How Infrastructure and Financial Institutions Affect Agricultural Output and Investment in India," *Journal of Development Economics*, 41(2):337-366.
- Chen, Shaohua, Ren Mu and Martin Ravallion, 2007, "Are there Lasting Impacts of Aid to Poor Areas? Evidence from Rural China," Policy Research Working Paper 4084, World Bank, Washington DC.
- Cook, Cynthia, Tyrrell Duncan, Somchai Jitsuchon, Anil Sharma and Wu Guobao (2005), "Assessing the Impact of Transport and Energy Infrastructure on Poverty Reduction," Manila: Asian Development Bank.
- Deaton, Angus (1997), *The Analysis of Household Surveys: A Microeconomic Approach to Development Policy* Johns Hopkins University Press, Baltimore, Maryland.
- Department for International Development (DFID) (2005), "How to Manual: The Valuation of Rural Travel Time Savings in Least Developed Countries," I.T. Transport Limited, Ardington, United Kingdom.
- Dercon, Stefan, Daniel Gilligan, John Hoddinott, and Tassew Woldehanna (2007), "The Impact of Roads and Agricultural Extension on Consumption Growth and Poverty in Fifteen Ethiopian Villages," CSAE WPS 2007-01, University of Oxford, UK.
- Escobal, Javier and Sonia Laszlo (2007) "Measurement Error in Access to Markets Data," *Oxford Bulletin of Economics and Statistics*, forthcoming.
- Escobal, Javier and Carmen Ponce (2004), "The Benefits of Rural Roads: Enhancing Income Opportunities for the Rural Poor," GRADE Working Paper 40, Lima Peru.
- Eskola, Elina (2005), "Commercialization and Poverty in Tanzania: Household -level Analysis," University of Copenhagen Discussion Paper 05-27, Dept. of Economics, Copenhagen, Denmark.
- Ettema, D, H Timmermans, and L. Veghel, 1996 "Effects of Data collection methods in travel and activity research", European Institute of Retailing and services studies.

- Fan, Shenggen and Connie Chan-Kang (2005), *Road Development, Economic Growth, and Poverty Reduction in China*, International Food Policy Research Institute, Washington, D.C.
- Fan, Shenggen, Peter Hazel, and Sukhadeo Thorat (2000), "Government Spending, Growth and Poverty in Rural India," *American Journal of Agricultural Economics* 82(4): 1038-1051.
- Galiani, Sebastian (2007), "Notes for the Impact Evaluation of the Construction and Rehabilitation of Rural Roads," mimeo, DIME, World Bank, Washington, DC.
- Gannon, Colin and Zhi Liu (1997), "Poverty and Transport." TWU discussion papers, TWU-30, World Bank, Washington, DC.
- Gibson, John and David McKenzie (2007), "Using the Global Positioning System in Household Surveys for Better Economics and Better Policy," *The World Bank Research Observer*, 22(2): 217-241.
- Gibson, John and Scott Rozelle (2003), "Poverty and Access to Roads in Papua New Guinea," *Economic Development and Cultural Change*, 52: 159-185.
- Grootaert, Christiaan (2001), "Socioeconomic Impact Assessment of Rural Roads: Methodology and Questionnaires," mimeo, INFTD, World Bank.
- Grosh, Margaret and Paul Glewwe (2000). *Designing Household Survey Questionnaires for Developing Countries: Lessons from 15 Years of the Living Standards Measurement Study*. Volumes 1, 2, and 3. The World Bank.
- Groves, Robert E., and Mick P. Couper (1998), *Nonresponse in Household Interview Surveys*. New York: John Wiley and Sons.
- Hahn, Jinyong, Petra Todd and Wilbert Van der Klauw (2001), "Identification and Estimation of Treatment Effects with a Regression-Discontinuity Design," *Econometrica* 69(1): 201-209.
- Heckman, James, H. Ichimura, and Petra Todd (1997), "Matching as an Econometric Evaluation Estimator: Evidence from Evaluating a Job Training Programme", *Review of Economic Studies*, 64, 605-654.
- Heckman, James, Robert Lalonde and James Smith (1999), "The Economics and Econometrics of Active Labor Market Programs," *Handbook of Labor Economics, Volume 3*, Ashenfelter, A. and D. Card, eds., Amsterdam: Elsevier Science.
- Hettige, Hemamala (2006), "When Do Rural Roads Benefit the Poor and How?" Operations Evaluation Department, Asian Development Bank, Manila, Philippines



- Hirano, Keisuke, Guido Imbens and Geert Ridder (2003), "Efficient Estimation of Average Treatment Effects Using the Estimated Propensity Score," *Econometrica*, 71, 1161-1189.
- Howe, John (1984), "The Impact of Rural Roads on Poverty Alleviation: a Review of the Literature," in John Howe and Peter Richards (eds.) "*Rural Roads and Poverty Alleviation*," International Labour Organization, Intermediate Technology Publications Ltd., London.
- Howe, John (1997), "The Impact of Rural Roads on Poverty Alleviation," IHE WP T&RE-18, International Institute for Infrastructural, Hydraulic and Environmental Engineering, Delft, The Netherlands.
- Huenemann, Ralph, (2008). "The Continuing Quest for a Pro-Poor Methodology for Project Appraisal: The Case of Rural Roads in Western China," unpublished, Faculty of business, University of Victoria, Victoria, BC, Canada.
- Jacoby, Hanan (2000), "Access to Markets and Benefits of Rural Roads," *The Economic Journal*, 110: 713-737.
- Jacoby, Hanan and Bart Minten (2008), "On Measuring the Benefits of Lower Transport Costs," Policy Research Working Paper 4484, World Bank, Washington, DC.
- Jalan, Jyotsna and Martin Ravallion (1998), "Are There Dynamic Gains from a Poor-area Development Program?" *Journal of Public Economics* 67, 65-86.
- Jalan, Jyotsna and Martin Ravallion (2002), "Geographic Poverty Traps? A Micro Model of Consumption Growth in Rural China," *Journal of Applied Econometrics* Vol.17(4), pp. 329-346
- Khandker, Shahidur, Zaid Bakht and Gayatri Koolwal (2009), "The Poverty Impact of Rural Roads: Evidence from Bangladesh," *Economic Development and Cultural Change* forthcoming.
- Kish, Leslie (1965), *Survey Sampling*, New York: John Wiley.
- Korinek, Anton, Johan Mistiaen and Martin Ravallion (2006), "Survey Nonresponse and the Distribution of Income," *Journal of Economic Inequality*, 4(2), pp. 33-55.
- Landry, Pierre and Mingming Shen (2005), "Reaching Migrants in Survey Research: The Use of Global Positioning System to Reduce Coverage Bias in China," *Political Analysis* 13: 1-22.
- Lokshin, Michael and Ruslan Yemtsov (2005), "Has Rural Infrastructure Rehabilitation In Georgia Helped the Poor?" *The World Bank Economic Review* 19(2), 311-333.

- Mogues, Tewodaj, Gezahegn Ayele and Zelekawork Paulos (2007), "The Bang for the Birr: Public Expenditures and Rural Welfare in Ethiopia," IFPRI Discussion Paper No. 702, International Food Policy Research Institute, Washington, DC.
- Mu, Ren and Dominique van de Walle (2007), "Rural Roads and Poor Area Development in Vietnam," Policy Research Working Paper No. 4340, Development Research Group, World Bank, Washington, DC, August.
- Mu, Ren and Dominique van de Walle (2008), "Rural Roads and Market Development in Vietnam," mimeo, PRMGE, World Bank, Washington DC.
- Rao, Vijayendra and Michael Woolcock (2003), "Integrating Qualitative and Quantitative Approaches in Program Evaluation," in Luiz A. Pereira da Silva and François Bourguignon, eds., *The Impact of Economic Policies on Poverty and Income Distribution: Evaluation Techniques and Tools*. World Bank and Oxford University Press.
- Ravallion, Martin (2001), "The Mystery of the Vanishing Benefits: An Introduction to Impact Evaluation," *World Bank Economic Review*, 15(1): 115-140.
- Ravallion, Martin (2008), "Evaluating Anti-Poverty Programs," in *Handbook of Development Economics Volume 5*, edited by T. Paul Schultz and John Strauss, Amsterdam: North-Holland, forthcoming.
- Ravallion, Martin and Shaohua Chen (2005), "Hidden Impact: Household Saving in Response to a Poor-Area Development Project," *Journal of Public Economics*, 89: 2183-2204.
- Roberts, Peter, KC Shyam, and Cordula Rastogi (2006), "Rural Access Index: A Key Development Indicator," Transport Papers No 10, Transport Sector Board, World Bank, Washington DC.
- Rosenbaum, Paul and Donald Rubin (1983), "The Central Role of the Propensity Score in Observational Studies for Causal Effects," *Biometrika*, 70, 41-55.
- van de Walle, Dominique (2002), "Choosing Rural Road Investments to Help Reduce Poverty," *World Development*, 30(4):575-589.
- van de Walle, Dominique and Ren Mu (2007), "Fungibility and the Flypaper Effect of Project Aid: Micro-evidence for Vietnam," *Journal of Development Economics*, 84(2): 667-684.
- Weisbrod, Glen (2000), "Current Practices for Assessing Economic Development Impacts from Transportation Investments: A Synthesis of Highway Practice," Washington, D.C.: National Academy Press.

Wooldridge, Jeffrey (2003), *Introductory Econometrics: A Modern Approach*, Ohio: Thomson South Western.

World Bank (2006), "Impact Evaluation and the Project Cycle," Doing Impact Evaluation Series No. 1, Thematic Group on Poverty Analysis, Monitoring and Impact Evaluation, PREM.