REVISED FINAL REPORT

Impact Evaluation of the Transportation Project in Honduras

Millennium Challenge Corporation Contract MCC-10-0133-CON-20 TO01

DECEMBER 20, 2013

PRESENTED TO:
Millennium Challenge Corporation
875 15th Street NW
Washington, DC 20005

PRESENTED BY:
NORC at the University of Chicago
55 East Monroe Street, 30th Floor
Chicago, IL 60603
(312)759-4000 (312)759-4004

NORC at the UNIVERSITY of CHICAGO

TABLE OF CONTENTS

List of Tables

List of Boxes

List of Exhibits

EXECUTIVE SUMMARY

Transportation Project Background

This document is the final report for the impact evaluation of the Transportation Project funded by the Millennium Challenge Account – Honduras and the Millennium Challenge Corporation (MCC) in Honduras over the period 2007- 2102. The project was implemented by the Millennium Challenge Account - Honduras (MCA-H) under a Compact between the governments of Honduras and the United States of America.

The Goal of the Compact in Honduras, which ended on September 30, 2010, was to stimulate economic growth and poverty reduction. To accomplish this goal, the MCA - Honduras Program aimed to achieve the following objectives:

- Increase the productivity and business skills of farmers who operate small and medium sized farms and their employees (the “Agricultural Objective”); and
- Reduce transportation costs between targeted production centers and national, regional, and global markets (the “Transportation Objective”).

Over the course of the Compact, two projects were implemented by MCA - Honduras to achieve these Objectives:

(1) The Rural Development Project, which was comprised of four activities: (I) farmer training and development, (ii) facilitation of access to credit by farmers, (iii) upgrading of farm to market roads (resurfacing of 20 rural roads) and (iv) provision of an agriculture public grants facility.

(2) The Transportation Project, which upgraded two major sections of the CA-5 Logistical Corridor, and paved approximately 65 km of secondary roads.

Under the NORC–MCA - Honduras contract (May 2007 to September 30, 2010) and the follow-on contract between NORC and MCC (October 1, 2010 to December 31, 2013), NORC undertook rigorous impact evaluations of two MCA - Honduras Program activities: the Farmer Training and Development Activity (FTDA), and the Transportation Project. This report discusses and presents the findings of the Transportation Project impact evaluation. A separate report presents the findings of the Farmer Training and Development Activity (FTDA) impact evaluation.

The Transportation Project aimed to reduce transportation costs between targeted production centers and key national markets and ports. The initial scope called for rehabilitating two major sections of Highway CA-5 (totaling 106.9 km in length), upgrading and paving at least 70 km of secondary roads, and developing a vehicle weight control system. Under the Rural Development Project, MCA - Honduras sought to upgrade and pave at 600 km of rural roads (farm-to-market roads). Given that it is part of the national road network, for the purpose of this evaluation we consider the evaluation of the rural roads improvement within the framework of the Transportation Project. Due to increases in costs and a partial re-scoping of the road rehabilitation component of the project, only 65 km of secondary roads and 495 km of rural roads were ultimately upgraded

The Transportation Project sought to improve conditions of the national road network by constructing and upgrading a number of secondary and rural roads, two segments of the CA-5 Highway and implementing a national weight control system. Improved conditions throughout the road network are expected to:

- Lower transport costs and travel time for businesses, including farm households;
- Provide better access to a wider range of job opportunities for individuals (labor market effects);
- Lower price of consumables and inputs by increasing competition and reducing barriers to entry posed by poor transport infrastructure;
- Improve access to health establishments and schools.

The overall expected result of these changes is higher incomes and employment at the household level. We also hypothesize a possible increase in use of health facilities (improved health-seeking behavior) and school attendance.

Based on these hypotheses, we examine changes over time in a number of income and expenditures indicators, as well as travel time to key points of interest including schools and health centers.

Key Features of the Evaluation Design

To comprehensively evaluate the impact of the MCA - Honduras Transportation project, we used a model-based approach, in which the treatment effect is represented by change in travel time, and the program impact is represented as a function of change in travel time caused by the program intervention. The model relies heavily on geographic information system (GIS) for several purposes, including the estimation of changes in travel times.

Box 1: Key Features of the Evaluation Design

- Estimation of multiple benefit streams – captures widespread benefits of the road projects
- A single integrated network model for all three roads projects – captures network interrelationships
- Measures of project treatment on a continuous scale (instead of a simple binary treatment (treated / untreated) model); estimation of conditional impacts
- Analytical survey design matched to impact estimation goals
- Use of variables from external models to improve accuracy and scope of inference

The evaluation design used for the Transportation Project significantly expanded the scope of the design outlined in the MCA - Honduras 2007 *M&E Plan*. The original design proposed the estimation of a single before-after benefit stream that accrues in the form of decreased vehicle operating costs and decreased travel time for CA-5 and secondary roads, and a separate estimation of changes in income for those households within a specified zone of influence (or “buffer zone”) around the rural roads, compared to a comparison group of households outside the zone of influence. The evaluation design that we used for the impact estimation was considerably different and has several important features, which we describe below, along with their associated limitations and benefits.

Evaluation scope. From a technical viewpoint, experimental designs based on randomized selection of units from an eligible population and randomized assignment of those units to treatment levels are generally preferred as the basis for evaluation studies. Randomized assignment to treatment assures that the distributions of variables (other than treatment) related to outcome are the same for the treatment and control groups. Randomization is usually not feasible for large infrastructure projects because infrastructure improvements, such as road rehabilitation, are typically targeted for certain locations and not others for a host of reasons, such as economic potential and/or political considerations. Similarly, the unique nature of the Honduran Highway CA-5, which serves as the main north-south Central American trunk expressway, precluded the possibility of random selection at a project level. In its initial conception, in 2007, the evaluation design did call for using randomization in the selection of rural roads to be upgraded. However, in the implementation of the Transportation Project, MCA - Honduras determined that it was not feasible to employ randomization in the selection of the rural roads to be improved (treated), and added a requirement to the selection criteria for eligible roads that municipalities provide matching funding and/or in-kind contributions towards the road improvements. In summary, the process used to select project roads did not allow for randomized selection from a population of eligible roads. Given the lack of randomization in assignment to treatment, the scope of the evaluation project was to assess the impact of the *particular* road-improvement activities comprising the Transportation Project, not to attempt to estimate the impact of a conceptually infinite population of similar roads projects in other locations at other times.

A single, integrated network model that has greater efficiency and validity. The evaluation model we used recognizes that the Honduran road system functions as a *single, integrated road network*, thereby allowing for network effects to be taken into account. In other words, it takes into account the fact that improvements to a single road section is likely to have impacts that are felt across the entire road network, not just locally, and these impacts may differ depending on where in the road network the improvement section is located, and the degree to which the section serves as a key access point between different sections of the overall network. This new model, which represents the physical road network as an integrated computer/mathematical network (through the GIS), recognizes that in reality, rural households are likely to benefit not only from rural-road improvements, but also from improvements to secondary (or even primary) roads. For example, for a farmer who must travel to a distant location to obtain fertilizer, improvements to his local rural roads may not reduce his travel-time cost nearly as much as secondary-road improvements might. The integrated model captures synergies and interaction effects between improvements made to different parts of the total road network, thereby allowing us to assess the combined impact of different road improvements. Furthermore, this unified approach enables assessment of the nationwide impacts of road improvements.

The integrated network model had the additional benefit of optimizing project resources: all resources that were planned for the three separate roads-project evaluation efforts were allocated to the development of a single model. This not only led to a more valid representation of road-related phenomena (e.g., the interaction of road segments), but also to a substantial increase in the precision of model estimates and the statistical power of tests of hypotheses.

On the cost side, the survey sample size required to develop this integrated model for the three roads projects amounted to approximately the same as that for a sample survey to develop a model for one sub-component project, such as rural roads.

Continuous treatment variables and conditional impact. The design comprehensively evaluates the impact of road improvements by measuring changes in benefit stream variables (such as income and employment) for samples of households, relative to incremental changes in travel time or travel cost (accessibility).

The approach makes use of a mathematical model that describes outcomes of interest (such as income) as continuous functions of travel times to places of interest. This approach embodies the fact that households are affected in varying degrees by road improvements. It enables the analysis to make full use of the power of the GIS to reflect the physical effect of road improvements, and permits the estimation of impact conditional on assumptions about road improvements (through the relationship to travel time).

The continuous (incremental-treatment) approach is a more accurate representation than a binary (dichotomous, zero-one, treatment/no-treatment) approach because it explicitly addresses the fact that the impact of road improvements varies over space as a continuum (as a function of variation in travel-time accessibility to roads and markets).

It should be noted that with respect to treatment units (road segments), the evaluation design is a pretest-posttest-group design – each road segment in the country is either treated (improved) or not treated. With respect to the unit of analysis – the household – however, it is a continuous-treatment-variable design, in which the intensity of treatment is reflected in changes in travel times from the household to points of interest. From this perspective, each experimental unit (household) serves in effect both as a treatment unit and comparison unit.

Analytical model to estimate impact. As a simple conceptual representation, the relationship may be represented as:

> *Vehicle speed = f(project intervention (road improvement), road characteristics (primary, secondary, rural; elevation variation), given vehicle type, season, day of week, time of day, weather)*
>
> *Travel time (for pickup truck) = g(road characteristics, mean vehicle speed given road characteristics)*
>
> *Outcome measure = h(travel-time variables and other variables ("covariates"))*
>
> *Impact = Expected value (mean) of outcome measure conditional on completion and maintenance of road-improvement project and on covariate means for both survey rounds – Expected value of outcome measure at beginning of project, conditional on covariate means for both survey rounds*

where f(.), g(.), and h(.) denote continuous functional relationships. A more general, unconditional, measure of impact would have been the difference in unconditional means, "*Expected value of outcome measure at end of project, conditional on covariate means for both survey rounds – Expected value of outcome measure at beginning of project, conditional on covariate means for both survey rounds*." It was not useful to estimate the unconditional measure since not all project activities were completed at the end of the evaluation project, and in many instances road improvements did not last because of a lack of maintenance.

The travel times are "direct" or "intermediate" results of the project intervention. Impact is measured in terms of indirect, or "ultimate" outcomes of interest, such as income, employment, and access to health and education facilities.

The vehicle speed estimates are obtained from traffic surveys on project roads and comparison roads. The travel-time estimates are estimated using a geographic information system road network model, using the speed estimates and taking into account road characteristics. An impact estimate is obtained by combining the travel-time estimates with data from a nationwide household sample survey.

Note that some project activities were terminated early in the project. The objective of the evaluation is to estimate the impact of the road-improvement activities that were not terminated early.

An advantage of the preceding model is that it expresses outcome as a function of travel time. Because of this, it is possible to estimate conditional impacts, given different values of travel times. This feature can be exploited to estimate the conditional impact of an individual project activity (such as improvement to a particular road segment), given completion and maintenance of the activity. This ability to estimate impact as a function of explanatory variables is not available from the binary-treatment-variable model.

Analytical survey design matched to impact estimation goals. The estimation of impact was based on collection of data from two sources – a household survey and a number of traffic surveys. The household survey data may be used to estimate higher-level outcomes of interest, such as household income and employment, as well as intermediate (direct) effects such as changes in travel times to places of interest. Traffic-survey data may be used only for the latter purpose. More project resources were allocated to the household survey than to the traffic surveys. These surveys were conducted near the beginning and end of the Transportation Project. The purpose of the household survey was to enable household-level estimation of direct and indirect outcomes of interest, such as changes in travel time and travel behavior and income. The household survey was a national survey stratified by travel times from project roads and other variables believed to be related to outcomes of interest, and the traffic surveys were conducted on project roads, or "treatment" roads, and a matched sample of non-project (non-treatment) roads.

The sample size for the household survey was large, and effective use was made of sample survey design techniques and statistical power analysis in the survey design. The household survey design was an "analytical" survey design that was configured to provide an efficient return of precision and power for estimating impacts of interest and making tests of hypotheses about them. To achieve high precision and power, the design was marginally stratified on variables believed to have a significant effect on outcomes of interest, such as travel times and anticipated changes in travel times associated with the project. The marginal stratification, which was implemented by setting variable probabilities of selection for sample units, assured adequate variation (balance, spread) in explanatory variables believed to have a significant effect on outcomes of interest.

The purpose of the traffic surveys was to provide data to enable estimation of travel times from the GIS model, for use in the statistical analysis of the household-survey data. The primary result of the traffic surveys was a table that showed mean speed of a pickup truck (the most common vehicle in Honduras) over Honduran roads, as a function of road type (primary, secondary, rural), elevation variation, and treatment status (improved / non-improved). Elevation variation was selected as the conditioning road characteristic because this could be determined (in the GIS) for all Honduran roads. To achieve high precision, this table was estimated under controlled conditions (season, day of week, time of day, weather). The survey was conducted for all project roads and a matched sample of comparison roads. The matching was done to reduce the effect of extraneous variables on the conditional estimates of the mean-speed table.

Use of geographic information system. The evaluation effort made use of geographic information system (GIS) data and data processing capabilities, both in the design and analysis phases of the project. The GIS served as the repository of a detailed digital, geo-spatial road network database (which includes detailed and up-to-date data on primary, secondary and rural road networks, and extensive physiographic data (elevation, land cover)). It was used to estimate travel times between any two points in the road network under alternative treatment conditions. The travel-time estimates were used to assist the design of the household sample survey, by enabling stratification on travel-time-related variables (and other variables), including the estimated reduction in travel time caused by the MCA - Honduras road improvements. The travel-time estimates constructed after the traffic survey data were available were used to support the construction of estimates of project impact conditional on project completion and maintenance.

Estimates of Project Impact

Table ES1 presents estimates of the total treatment effect for the complete set of response indicators analyzed in detail. Impact estimates that are statistically significant (two-sided or one-sided, as appropriate, .05 significance level) are marked with an asterisk (*). Note that the components of IncEmp (IncEmpAg and IncEmpNonAg) do not sum to IncEmp, because they are estimated independently.

Table ES1: Ordinary-Least-Squares (OLS) Regression Estimate of Average Treatment Effect (ATE) based on Continuous Treatment Variables, for Selected Outcome Measures

Table ES1. Ordinary-Least-Squares (OLS) Regression Estimate of Average Treatment Effect (ATE) based on Continuous Treatment Variables, for Selected Outcome Measures

Outcome Variable (with definition)		Estimate of Impact	Standard Error of Estimate
Household Income and Expenditure [1]			
IncEmp	Income from labor-market employment (monthly)	5.00	41.5
IncEmpAg	Income from labor-market employment in the agricultural sector (monthly)	71.9*	17.9
IncEmpNonAg	Income from labor-market employment in non-ag sectors (monthly)	-109*	30.1
TotHHExp	Total household expenditures (monthly)	-25.2	23.0
NetHHInc	Net household income (annualized)	422	609
Access			
CostToSchool	Cost (in lempiras) to school	-.0232	.0228
TimeToSchool	Time (in minutes) to school	-.119	.0718
CostToCollege	Cost to college	-.424	.361
TimeToCollege	Time to college	-.180	.212
CostToHospital	Cost to hospital	-3.52*	.718
TimeToHospital	Time to hospital	.704*	.263
CostToHealthCtr	Cost to health center	-.194*	.097
TimeToHealthCtr	Time to health center	-.549*	.168
CostToMarket	Cost to market	-.606*	.239
TimeToMarket	Time to market	-.083	.240
CostToPulp	Cost to pulperia	-.126*	.0613
TimeToPulp	Time to pulperia	-.0394	.0704
TimeToTegus	Time to Tegucigalpa	-1.38*	.476
TimeToSPS	Time to San Pedro Sula	.757	.994
TimeToDepCap	Time to departmental capital	-.459	.557
TimeToMunCap	Time to municipal capital	-1.106*	.387
School Attendance			
ChldInSch712	Total number of children aged 7-12 attending school	-.00402	.00470
ChldInSch1318	Total number of children aged13-18 attending school	-.000843	.00375

[1] More detailed definitions of income variables can be found on pages 63-64.

Table ES1. Ordinary-Least-Squares (OLS) Regression Estimate of Average Treatment Effect (ATE) based on Continuous Treatment Variables, for Selected Outcome Measures			
Outcome Variable (with definition)		**Estimate of Impact**	**Standard Error of Estimate**
Use of Health Care Services			
VisHospital	Total number of visits in last 30 days to hospital by all family members	.00675	.00700
VisPrHlthCtr	Total number of visits in last 30 days to private health centers	.00230	.00621
VisPubClinic	Total number of visits in last 30 days to public clinics	-.00513	.01058
VisNonProf	Total number of visits in last 30 days to non-professional health-care providers	-.0255	.00722
VisPharm	Total number of visits in last 30 days to pharmacy	-.00702	.00479
Employment			
WrkdPrevWk	Total number of household members who worked the previous week	.00436	.00649

Note: Income, expense and travel costs measured in Honduran lempiras; travel times measured in minutes.

The preceding results show that the program intervention had a statistically significant effect (of the expected sign) on many of the access times and costs, on some of the indicators concerned with use of health care facilities, and on some of the income indicators. In all cases, the magnitude of estimated impact is small. The principal reason for small magnitude of impact is that it is the expected impact of the project *for a randomly selected household in the country*.

To assess the total national impact of the road-improvement program, the impact estimates are multiplied by the number of households in the nation. The population of Honduras is 8.2 million people (2010 est.), and the average household size is approximately 5, so this corresponds to approximately 1.64 million households. Based on these numbers, the estimated total program impact for the nation is approximately as follows:

Table ES2: Estimated National-Level Impact of the Transportation Program (in Lempiras)

Table ES2. Estimated National-Level Impact of the Transportation Program (table entries are lempiras)				
Outcome Variable	**Estimated Impact per Household**	**Standard error of Estimated Impact per Household**	**Estimated Total Impact for Nation**	**Standard Error of Estimated Total Impact for Nation**
IncEmp	5.00	41.5	8.2M	68M
IncEmpAg	71.9*	17.9	119M*	30M
IncEmpNonAg	-109*	30.1	-180M*	50M
TotHHExp	-25.2	23.0	-42M	38M
NetHHInc	422	609	692M	999M

An approximate 95% confidence interval for the estimated national-level project impact is the estimate plus and minus twice its standard error. For example, an approximate 95% confidence interval for income from agricultural employment (IncEmpAg) is (59M lempiras, 179M lempiras).

Note that in addition to the direct impact estimated here, the Transportation Project may have indirect effects, such as providing the improved transport required for the FTDA project.

The preceding estimates of total national impact are comparable in magnitude to, but small compared to, the cost of the road-improvement project (which was, according to the MCC Compact, 125.7 million dollars).

We conducted a detailed *ex post* statistical power analysis to assess the power associated with the tests of hypothesis associated with the estimates presented in the preceding table. The results of that analysis are presented in Annex 1. The power analysis showed that the evaluation design had high power to detect impacts equal in magnitude to about ten percent of the baseline mean.

Conclusions and Recommendation

Summary of Impact of the Transportation Project

The principal finding of this evaluation is that although the Transportation Project shows some statistically significant effects on a variety of indicators (income and travel times to places of interest), those impacts are very small. A detailed statistical power analysis was conducted, which showed that the small number and size of the statistically significant results is not the result of an underpowered survey, but a result of the small magnitudes of the project effects.

The evaluation design adopted for this impact evaluation was to estimate impact from household-survey data, conditional on project-caused changes in travel time. The travel times are determined by a GIS road network model, using mean vehicle speeds estimated from traffic-survey data. The estimates are conditional on completion and maintenance of the road improvement project. The analysis produced estimates of the mean impact expected for a randomly selected household in Honduras. It was determined that the mean household-level impact of the project, averaged over the nation, is low.

Analysis of the traffic-survey data showed substantial changes in travel speeds and travel times over the project roads, compared to similar non-project roads. While the per-household impact averaged over the nation is low, the effect of the project on the speed of vehicles using the project roads is substantial. The disadvantage of using traffic-survey data alone to assess project impact is that it provides “intermediate” outcomes, not “higher level” impacts such as income, employment, and access to health, medical and other facilities.

The approach used in this report has a number of strengths that argue for the validity of the findings. The results are based on causal modeling, and the assumptions required of the statistical estimation models used to estimate impact are not in doubt. The per-household impact of the Transportation Project is low. While the economic impact of new roads is known to be substantial, the impact of the road improvements implemented in the Transportation Project, on a national level, are not high. On a national level, they represent marginal improvements to the road system and to household access, and they produce marginal impacts. The noticeable effect of the road improvements is on speeds and travel times of users of the improved roads.

In summary, the impact results presented in this report are based on an evaluation design, causal models, and analytical models that are considered to have high validity. An *ex post* statistical power analysis demonstrated that the power associated with the impact estimates is high. Conditional on the soundness of the assumptions described, above, we consider the results of this analysis to be an accurate (valid, reliable and high-power) assessment of the impact of the Transportation Project.

Summary of Assumptions and Limitations

With respect to the "macro-level" causal model and associated statistical model used to estimate impact, the assumptions underlying the estimates of impact are the following:

1. The stable unit treatment value assumption (SUTVA, no-macro-effects assumption, partial equilibrium assumption) is made. Among other things, this assumption implies that the project is not so large that it changes the basic relationship of outcomes of interest to travel time.

2. The estimates of travel time are based on the GIS model of the Honduran road network. This model includes all official Honduran roads, as well as others. The GIS model is highly detailed, and considered to be up-to-date and of high accuracy. The quality of the GIS model is not considered to be a limitation on the quality of the evaluation.

3. The impact estimates are conditional on completion and maintenance of the Transportation Project as finally configured. Under this assumption, the impact estimates refer to this particular project, not to the mean impacts associated with a conceptually infinite population of similar projects in other locations or at other times. The estimate of impact is the average treatment effect of this particular project on a randomly selected household in Honduras, *not* the average treatment effect associated with improving a randomly selected eligible-for-treatment road segment.

4. Although the unit of treatment was the road segment, the unit of analysis was the household. In this case there could be two types of unobserved (hidden) variables, which may introduce biases into the estimates of the model parameters. First, there may be unobserved variables that are time invariant. In a fixed-effects model, these, however, "drop out" for the two-round panel specification. Second, there could be unobserved variables that are not time invariant (though no such variables were identified). It is assumed that such variables, if any exist, are uncorrelated with the explanatory variables.

5. The continuous-treatment-value impact estimates are conditional on the travel speed table derived from the traffic surveys. This table presents estimates of the average speed of a pickup truck over a route by route type (primary, secondary, rural), elevation variation and program intervention status (improved or not improved). The speeds are conditional on the season of the year, day of the week, time of day, and weather conditions under which the traffic surveys were conducted. The speeds are used to calculate travel times to places of interest (using the GIS model). These travel times are used (as described above) to estimate impact. A number of travel times were available (i.e., travel times to

various places of interest from a *caserío* such as to a municipal or department capital, or the nearest town having population 1,000 or more. Attention focused on the one that had the highest relationship to outcomes of interest (i.e., travel time to the nearest town of population 1,000 or more). The continuous-treatment-value estimates are also conditional on manifestation of long-term benefits, as estimated from the partial-treatment-effects model.

Other more specific model assumptions are listed for particular estimation equations in the detailed analysis presented in Annex 1.

The limitations of the evaluation are:

1. The impact estimates constructed in this analysis pertain solely to the Transportation Project as it was finally configured, and when and where it was implemented, not to similar projects in other settings (locations or times). The estimated standard errors reflect sampling variation associated with estimation of characteristics of this project, and do not include variation associated with hypothetical variations in the project location or time, or the higher-level sample units of the household sample survey (*caseríos* and households) (as would be the case for a "random effects" approach). It is expected, however, that similar results would be expected for similar projects in similar settings.

2. More of the resources for primary data collection for this evaluation were invested in the household sample survey, and less in the traffic surveys. The traffic surveys were limited in scope and in sample size. The design of the traffic surveys rested on judgment, not on the methods of statistical sample survey. The estimates of speed were conditioned only on elevation variation, and not on other road variables that may have an important effect on average speed, such as number of lanes, access, road roughness, and curviness. The vehicle speed on which travel times were based for analysis of impact was a pickup truck. The traffic survey data were used as input to the GIS road-network model for estimation of travel times, not for estimation of direct impact (such as via a double-difference estimator applied to the traffic data). The travel-time estimates pertain to a particular vehicle type, season, day of week, time of day, and local weather conditions. The level of correlation among the available travel times was relatively high, so that the limited scope of the traffic surveys is not considered to be a major weakness. In the future, however, if this approach is used, it is recommended that a substantially greater level of resources be allocated to the traffic surveys (at least comparable to that of the household surveys).

3. The study focuses on a variety of indicators to assess impact of the road improvements. Some of them are direct effects, such as travel times to points of interest (education and health facilities), and others are indirect effects such as income and employment. Additional direct effects might have been of interest, such as number of trips or length of trips.

4. As described above, an assumption of the analysis is that are no time-varying unobserved variables that are correlated with any of the explanatory variables, which in this case refers to the treatment variables. While we have investigated and confirmed the soundness of this assumption to the extent feasible, it is not possible to conclusively rule

out the possibility that such variables exist and may be influencing the results. For example, if economic conditions such as labor market characteristics or the level of private investment are changing in systematically different ways that are correlated with travel time,our approach may mistake the impact of the road improvement for the impact of changes in these conditions. This possibility must be considered a limitation of the analysis.

Recommendation for Additional Analysis

This project collected a lot of data from household and traffic surveys. These data were analyzed in accordance with the evaluation design and analysis plan. Under this plan, greater project resources were put into design, collection and analysis of the household survey data than into the design, collection and analysis of the traffic survey data. The traffic survey data were used to estimate travel times and to support an analysis of economic rate of return (ERR) using the Highway Design and Management Version 4 (HDM-4) road planning and analysis computer software package (this work is documented in a separate report). In view of the weak impact results associated with national-level household survey data, it is considered worthwhile to spend some additional effort to conduct a fixed-effects statistical analysis of direct impact using the traffic survey data, even though the traffic surveys were not designed for this purpose. This would include estimation of the impact of the road improvements on traffic speeds, volumes and origin-destination. This option was considered during the course of the project (and recommended by a reviewer and project team member), but resources were not available to conduct a detailed statistical design and analysis effort for both the household survey and the traffic surveys.

One reviewer suggested that a “corridor-level” analysis, focusing on households very near the improved roads, would show stronger impact. While this may be true to some extent, it is considered that basing impact estimation for road-improvement projects on general households will show small impact, even for households close to the project roads (this view is supported by the results for binary-treatment-variable estimates of impact, which, even though they focused attention on households around project roads, were even weaker than the continuous-treatment-variable estimates).

A. INTRODUCTION

This document is the final report for the impact evaluation of the Transportation Project funded by the Millennium Challenge Account – Honduras and the Millennium Challenge Corporation (MCC) in Honduras over the period 2007- 2102. The project was implemented by the Millennium Challenge Account Honduras (MCA-H) under a Compact between the governments of Honduras and the United States of America.

The Goal of the Compact in Honduras, which ended on September 30, 2010, was to stimulate economic growth and poverty reduction. To accomplish this goal, the MCA - Honduras Program aimed to achieve the following objectives:

- Increase the productivity and business skills of farmers who operate small and medium sized farms and their employees (the "Agricultural Objective"); and
- Reduce transportation costs between targeted production centers and national, regional, and global markets (the "Transportation Objective").

Over the course of the Compact, two projects were implemented by MCA - Honduras to achieve these Objectives:

(1) The Rural Development Project, which was comprised of four activities: (I) farmer training and development, (ii) facilitation of access to credit by farmers, (iii) upgrading of farm to market roads (resurfacing of 20 rural roads) and (iv) provision of an agriculture public grants facility.

(2) The Transportation Project, which upgraded two major sections of the CA-5 Logistical Corridor, and paved approximately 65 km of secondary roads.

Under the NORC–MCA - Honduras contract (May 2007 to September 30, 2010) and the follow-on contract between NORC and MCC (October 1, 2010 to December 31, 2013), NORC undertook rigorous impact evaluations of two MCA - Honduras Program activities: the Farmer Training and Development Activity (FTDA), and the Transportation Project. This report discusses and presents the findings of the Transportation Project impact evaluation. A separate report presents the findings of the Farmer Training and Development Activity (FTDA) impact evaluation.

The remainder of this report is organized as follows. Section B presents a brief description of the Transportation Project. Section C discusses in-depth the evaluation design and its implementation. Sections D and F describe the household and traffic surveys conducted to collect the primary data for this impact evaluation, while Section E discusses the geographic information system (GIS) we used for the analysis. Section G presents a summary of results of the impact evaluation. A detailed technical discussion of the impact analysis and results is presented in Annex 1. Annex 2 contains a description of the methodology used to estimate travel times. Annex 3 describes the household sample survey design and Annex 4 describes the statistical power analysis approach to estimation of sample size for the household survey.

B. THE TRANSPORTATION PROJECT

The Transportation Project aimed to reduce transportation costs between targeted production centers and key national markets and ports. The initial scope called for rehabilitating two major sections of Highway CA-5 (totaling 106.9 km in length), upgrading and paving at least 70 km of secondary roads, and developing a vehicle weight control system. Under the Rural Development Project, MCA - Honduras sought to upgrade and pave at 600 km of rural roads (farm-to-market roads). Given that it is part of the national road network, for the purpose of this evaluation we consider the evaluation of the rural roads improvement within the framework of the Transportation Project. Due to increases in costs and a partial re-scoping of the road rehabilitation component of the project, only 65 km of secondary roads and 495 km of rural roads were ultimately upgraded

The timing of road improvements/rehabilitations spans the period May 2008 to December 2012, and is presented in Table 1 below.

Table 1: Timetable of MCA - Honduras Road Improvements

Station ID	Road Segment	Department	Classification	Improvement start date	Improvement end date
70	(CA-5 Sections 3 and 4)Fin Valle de Comayagua – Siguatepeque - Taulabé	Comayagua	Primary	05/01/2008	09/30/2010
66A	(CA-5 Section 1) Tegucigalpa - Río del Hombre	Francisco Morazán	Primary	04/15/2010	07/13/2012 (Contract Completion Date)
66B	(CA-5 Section 2) Río del Hombre - Inicio Valle de Comayagua	Francisco Morazán	Primary	02/13/2009	12/11/2011 (Contract Completion Date)
70A	(CA-5 Section 3) Fin Valle de Comayagua - Siguatepeque	Comayagua	Primary	09/16/2008	05/10/2010
70B	(CA-5 Section 4) Siguatepeque - Taulabé	Comayagua	Primary	09/16/2008	09/15/2010
62	Choluteca - Orocuina	Choluteca	Secondary	02/23/2009	09/15/2010
67	Comayagua - Ajuterique - La Paz	Comayagua	Secondary	09/16/2008	08/25/2010
76	Sonaguera - Km 35	Colón	Secondary	11/27/2008	02/18/2010
27	San Estéban - Toro Muerto	Olancho	Rural	Terminated	Terminated
28	Ilanga - Monte Abajo (S113)	Colón	Rural	09/10/2009	04/30/2010
29	S113 Río Arriba - Los Angeles	Colón	Rural	07/24/2009	05/31/2010
30	El Juncal - Brisas de Olanchito	Yoro	Rural	Terminated	Terminated
31	Guarizama - Tizate	Olancho	Rural	Terminated	Terminated
32	Oculí - Desvío de Cedrales	El Paraíso	Rural	07/15/2009	05/15/2010
34	Yoro (La Aguja) - Guardaraya	Yoro	Rural	Terminated	Terminated
35	La Unión - El Bambú (Ceiba Mocha)	Atlantida	Rural	04/15/2010	06/30/2010
36	Piedra de Diamante - El Aceituno	Choluteca	Rural	09/14/2009	12/18/2009
39	La Corteza - La Catarina	Choluteca	Rural	NA	01/11/2011

Station ID	Road Segment	Department	Classification	Improvement start date	Improvement end date
40	Los Puentes - Pueblo Nuevo	Choluteca	Rural	07/22/2009	10/13/2009
42	Monte Redondo - San Matías	Francisco Morazán	Rural	09/14/2009	12/18/2009
43	Guantillo - Las Lazadas	Francisco Morazán	Rural	07/17/2009	05/17/2009
44	La Esperanza – Monte Sión	Atlantida	Rural	06/01/2010	09/24/2010
45	Soledad - Los Alpes	El Paraíso	Rural	07/15/2009	10/13/2009
46	Las Crucitas - Los Noques	Francisco Morazán	Rural	04/08/2010	05/15/2010
47	Lepaera - Los Coros	Lempira	Rural	01/01/2010	05/15/2010
48	Arada - Las Marías	Santa Barbara	Rural	07/15/2009	06/23/2010
49	Tomalá - Limite Municipio San Esteban	Lempira	Rural	08/03/2009	05/15/2010
50	CA-1 - Cubulero - Alianza	Valle	Rural	07/11/2010	09/23/2010
58	La Germania No 1 - Santa Rosita - Cantillanos (Montañuelos - Cantillanos - La Cueva)	Comayagua	Rural	10/11/2009	06/15/2010
90	Siguatepeque - Caobanal - El Carrizal (CA-5)	Comayagua	Rural	07/17/2009	06/16/2010
91	Lo de Reyna - El Pacon	Comayagua	Rural	07/17/2009	05/10/2010
92	Ajuterique - Quelepa - Playoncito - Misterio - Ajuterique	Comayagua	Rural	10/16/2009	06/24/2009

C. THE TRANSPORTATION PROJECT EVALUATION DESIGN

C.1 EVALUATION GOALS

The Transportation Project sought to improve conditions of the national road network by constructing and upgrading a number of secondary and rural roads, two segments of the CA-5 Highway and implementing a national weight control system. Improved conditions throughout the road network are expected to:

- Lower transport costs and travel time for businesses, including farm households;
- Provide better access to a wider range of job opportunities for individuals (labor market effects);
- Lower price of consumables and inputs by increasing competition and reducing barriers to entry posed by poor transport infrastructure;
- Improve access to health establishments and schools.

The overall expected result of these changes is higher incomes and employment at the household level. We also hypothesize a possible increase in use of health facilities (improved health-seeking behavior) and school attendance.

Based on these hypotheses, we examine changes over time in a number of income and expenditures indicators, as well as travel time to key points of interest including schools and health centers.

C.2 EVALUATION DESIGN

Key Features of the Evaluation Design

To comprehensively evaluate the impact of the MCA - Honduras Transportation project, we used a model-based approach, in which the treatment effect is represented by change in travel time, and the program impact is represented as a function of change in travel time caused by the program intervention. The model relies heavily on geographic information system (GIS) for several purposes, including the estimation of changes in travel times.

Box 1: Key Features of the Evaluation Design

- Estimation of multiple benefit streams – captures widespread benefits of the road projects
- A single integrated network model for all three roads projects – captures network interrelationships
- Measures of project treatment on a continuous scale (instead of a simple binary treatment (treated / untreated) model); estimation of conditional impacts
- Analytical survey design matched to impact estimation goals
- Use of variables from external models to improve accuracy and scope of inference

The evaluation design used for the Transportation Project significantly expanded the scope of the design outlined in the MCA - Honduras 2007 *M&E Plan*. The original design proposed the estimation of a single before-after benefit stream that accrues in the form of decreased vehicle operating costs and decreased travel time for CA-5 and secondary roads, and a separate estimation of changes in income for those households within a specified zone of influence (or "buffer zone") around the rural roads, compared to a comparison group of households outside the zone of influence. The evaluation design that we used for the impact estimation was considerably different and has several important features, which we describe below, along with their associated limitations and benefits.

Evaluation scope. From a technical viewpoint, experimental designs based on randomized selection of units from an eligible population and randomized assignment of those units to treatment levels are generally preferred as the basis for evaluation studies. Randomized assignment to treatment assures that the distributions of variables (other than treatment) related to outcome are the same for the treatment and control groups. Randomization is usually not feasible for large infrastructure projects because infrastructure improvements, such as road rehabilitation, are typically targeted for certain locations and not others for a host of reasons, such as economic potential and/or political considerations. Similarly, the unique nature of the Honduran Highway CA-5, which serves as the main north-south Central American trunk expressway, precluded the possibility of random selection at a project level. In its initial conception, in 2007, the evaluation design did call for using randomization in the selection of rural roads to be upgraded. However, in the implementation of the Transportation Project, MCA - Honduras determined that it was not feasible to employ randomization in the selection of the rural roads to be improved (treated), and added a requirement to the selection criteria for eligible roads that municipalities provide matching funding and/or in-kind contributions towards the road improvements. In summary, the process used to select project roads did not allow for randomized selection from a population of eligible roads. Given the lack of randomization in assignment to treatment, the scope of the evaluation project was to assess the impact of the *particular* road-improvement activities comprising the Transportation Project, not to attempt to estimate the impact of a conceptually infinite population of similar roads projects in other locations at other times.

A single, integrated network model that has greater efficiency and validity. The evaluation model we used recognizes that the Honduran road system functions as a *single, integrated road network*, thereby allowing for network effects to be taken into account. In other words, it takes into account the fact that improvements to a single road section is likely to have impacts that are felt across the entire road network, not just locally, and these impacts may differ depending on where in the road network the improvement section is located, and the degree to which the section serves as a key access point between different sections of the overall network. This new model, which represents the physical road network as an integrated computer/mathematical network (through the GIS), recognizes that in reality, rural households are likely to benefit not only from rural-road improvements, but also from improvements to secondary (or even primary) roads. For example, for a farmer who must travel to a distant location to obtain fertilizer, improvements to his local rural roads may not reduce his travel-time cost nearly as much as secondary-road improvements might. The integrated model captures synergies and interaction effects between improvements made to different parts of the total road network, thereby allowing us to assess the combined impact of different road improvements. Furthermore, this unified approach enables assessment of the nationwide impacts of road improvements.

The integrated network model had the additional benefit of optimizing project resources: all resources that were planned for the three separate roads-project evaluation efforts were allocated to the development of a single model. This not only led to a more valid representation of road-related phenomena (e.g., the interaction of road segments), but also to a substantial increase in the precision of model estimates and the statistical power of tests of hypotheses.

On the cost side, the survey sample size required to develop this integrated model for the three roads projects amounted to approximately the same as that for a sample survey to develop a model for one sub-component project, such as rural roads.

Continuous treatment variables and conditional impact. The design comprehensively evaluates the impact of road improvements by measuring changes in benefit stream variables (such as income and employment) for samples of households, relative to incremental changes in travel time or travel cost (accessibility).

The approach makes use of a mathematical model that describes outcomes of interest (such as income) as continuous functions of travel times to places of interest. This approach embodies the fact that households are affected in varying degrees by road improvements. It enables the analysis to make full use of the power of the GIS to reflect the physical effect of road improvements, and permits the estimation of impact conditional on assumptions about road improvements (through the relationship to travel time).

The continuous (incremental-treatment) approach is a more accurate representation than a binary (dichotomous, zero-one, treatment/no-treatment) approach because it explicitly addresses the fact that the impact of road improvements varies over space as a continuum (as a function of variation in travel-time accessibility to roads and markets).

It should be noted that with respect to treatment units (road segments), the evaluation design is a pretest-posttest-group design – each road segment in the country is either treated (improved) or not treated. With respect to the unit of analysis – the household – however, it is a continuous-treatment-variable design, in which the intensity of treatment is reflected in changes in travel times from the household to points of interest. From this perspective, each experimental unit (household) serves in effect both as a treatment unit and comparison unit.

Analytical model to estimate impact. As a simple conceptual representation, the relationship may be represented as:

> *Vehicle speed = f(project intervention (road improvement), road characteristics (primary, secondary, rural; elevation variation), given vehicle type, season, day of week, time of day, weather)*
>
> *Travel time (for pickup truck) = g(road characteristics, mean vehicle speed given road characteristics)*
>
> *Outcome measure = h(travel-time variables and other variables ("covariates"))*

> *Impact = Expected value (mean) of outcome measure conditional on completion and maintenance of road-improvement project and on covariate means for both survey rounds – Expected value of outcome measure at beginning of project, conditional on covariate means for both survey rounds*

where f(.), g(.), and h(.) denote continuous functional relationships. A more general, unconditional, measure of impact would have been the difference in unconditional means, "*Expected value of outcome measure at end of project, conditional on covariate means for both survey rounds – Expected value of outcome measure at beginning of project, conditional on covariate means for both survey rounds*." It was not useful to estimate the unconditional measure since not all project activities were completed at the end of the evaluation project, and in many instances road improvements did not last because of a lack of maintenance.

The analytical model is referred to as a "continuous-treatment-variable" model or as a "partial-treatment-effects" model. It estimates the change in outcome caused by the project intervention, conditional on setting all covariates (explanatory variables other than treatment) equal to their means over both survey rounds. The model differs from a pretest-posttest-comparison-group design in that every household in the nation is potentially affected by the program intervention, and, in a sense, serves both as a "treatment" and "control."

The travel times are "direct" or "intermediate" results of the project intervention. Impact is measured in terms of indirect, or "ultimate" outcomes of interest, such as income, employment, and access to health and education facilities.

The vehicle speed estimates are obtained from traffic surveys on project roads and comparison roads. The travel-time estimates are estimated using a geographic information system road network model, using the speed estimates and taking into account road characteristics. An impact estimate is obtained by combining the travel-time estimates with data from a nationwide household sample survey.

Note that some project activities were terminated early in the project. The objective of the evaluation is to estimate the impact of the road-improvement activities that were not terminated early.

An advantage of the preceding model is that it expresses outcome as a function of travel time. Because of this, it is possible to estimate conditional impacts, given different values of travel times. This feature can be exploited to estimate the conditional impact of an individual project activity (such as improvement to a particular road segment), given completion and maintenance of the activity. This ability to estimate impact as a function of explanatory variables is not available from the binary-treatment-variable model.

Analytical survey design matched to impact estimation goals. The estimation of impact was based on collection of data from two sources – a household survey and a number of traffic surveys. The household survey data may be used to estimate higher-level outcomes of interest, such as household income and employment, as well as intermediate (direct) effects such as changes in travel times to places of interest. Traffic-survey data may be used only for the latter purpose. More project resources were allocated to the household survey than to the traffic surveys. These surveys were conducted near the beginning and end of the Transportation

Project. The purpose of the household survey was to enable household-level estimation of direct and indirect outcomes of interest, such as changes in travel time and travel behavior and income. The household survey was a national survey stratified by travel times from project roads and other variables believed to be related to outcomes of interest, and the traffic surveys were conducted on project roads, or "treatment" roads, and a matched sample of non-project (non-treatment) roads.

The sample size for the household survey was large, and effective use was made of sample survey design techniques and statistical power analysis in the survey design. The household survey design was an "analytical" survey design that was configured to provide an efficient return of precision and power for estimating impacts of interest and making tests of hypotheses about them. To achieve high precision and power, the design was marginally stratified on variables believed to have a significant effect on outcomes of interest, such as travel times and anticipated changes in travel times associated with the project. The marginal stratification, which was implemented by setting variable probabilities of selection for sample units, assured adequate variation (balance, spread) in explanatory variables believed to have a significant effect on outcomes of interest.

The purpose of the traffic surveys was to provide data to enable estimation of travel times from the GIS model, for use in the statistical analysis of the household-survey data. The primary result of the traffic surveys was a table that showed mean speed of a pickup truck (the most common vehicle in Honduras) over Honduran roads, as a function of road type (primary, secondary, rural), elevation variation, and treatment status (improved / non-improved). Elevation variation was selected as the conditioning road characteristic because this could be determined (in the GIS) for all Honduran roads. To achieve high precision, this table was estimated under controlled conditions (season, day of week, time of day, weather). The survey was conducted for all project roads and a matched sample of comparison roads. The matching was done to reduce the effect of extraneous variables on the conditional estimates of the mean-speed table.

Use of geographic information system. The evaluation effort made use of geographic information system (GIS) data and data processing capabilities, both in the design and analysis phases of the project. The GIS served as the repository of a detailed digital, geo-spatial road network database (which includes detailed and up-to-date data on primary, secondary and rural road networks, and extensive physiographic data (elevation, land cover)). It was used to estimate travel times between any two points in the road network under alternative treatment conditions. The travel-time estimates were used to assist the design of the household sample survey, by enabling stratification on travel-time-related variables (and other variables), including the estimated reduction in travel time caused by the MCA - Honduras road improvements. The travel-time estimates constructed after the traffic survey data were available were used to support the construction of estimates of project impact conditional on project completion and maintenance.

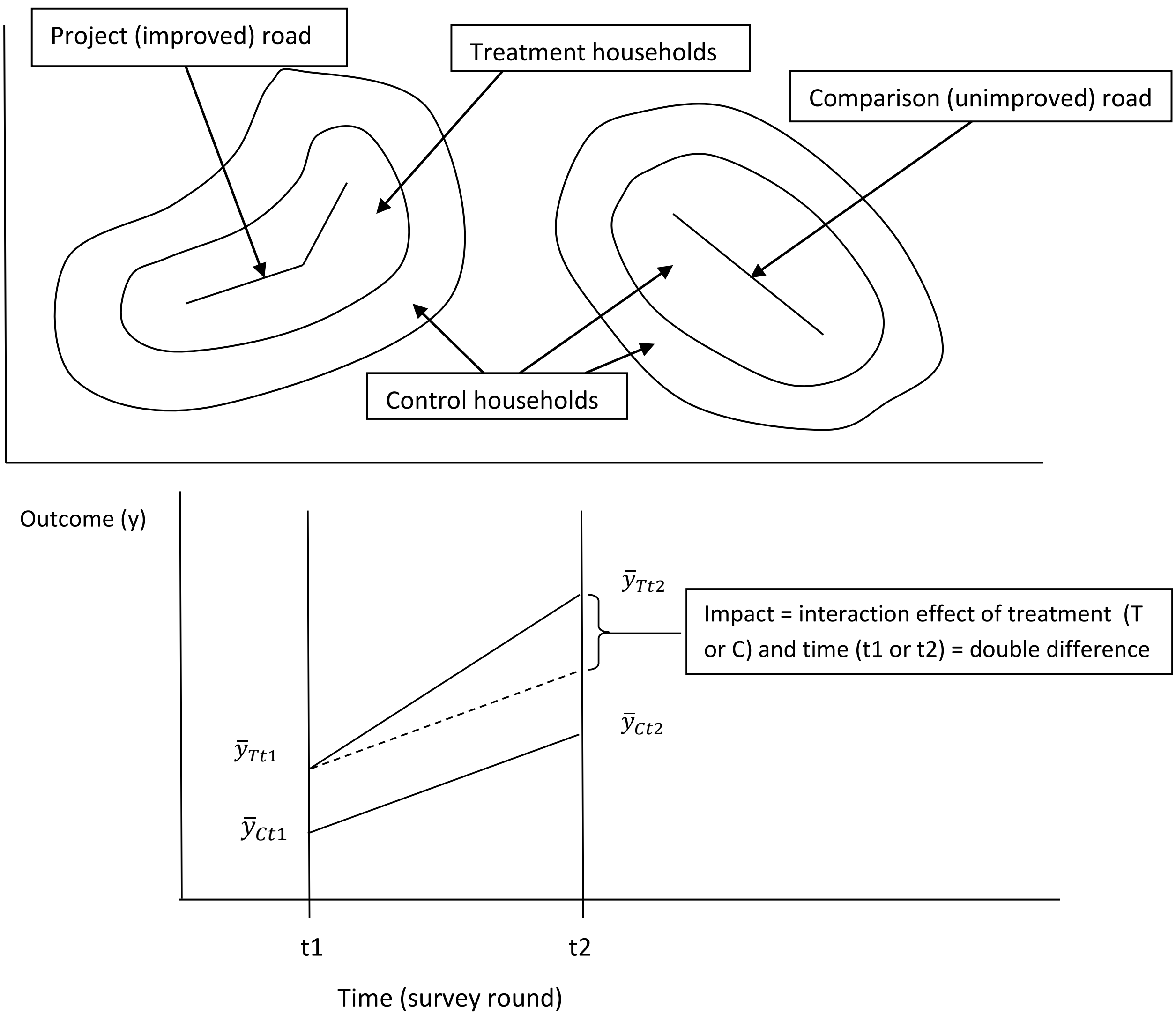
Comparison of the Evaluation Design Used for the Study to the Conventional Approach in Evaluating Road Improvement Projects

Below, we present two graphical representations that illustrate the fundamental characteristics of the approach originally proposed in the MCA - Honduras *Monitoring & Evaluation Plan*, and the approach adopted in this evaluation study. The original approach is the conventional method used to evaluate road improvement projects. It is based on a binary treatment variable (BTV), that is, a treatment variable that has two values – one value (treatment, or "1") represented by households within a certain distance or travel time from project (improved, treated) roads, and the other value (control, or "0"). Figure 1 below depicts this approach.

In the course of the evaluation study, some attention is given to estimation of impact using the conventional method. This is done for two reasons. First, to assist the understanding of the continuous-treatment-variable (CTV) approach that was adopted. Second, because it was represented in the design phase of the evaluation project that the continuous-treatment-variable approach would be superior to the conventional approach, and it is considered important to assess, after the fact, whether this decision was a good one (which it turned out to be).

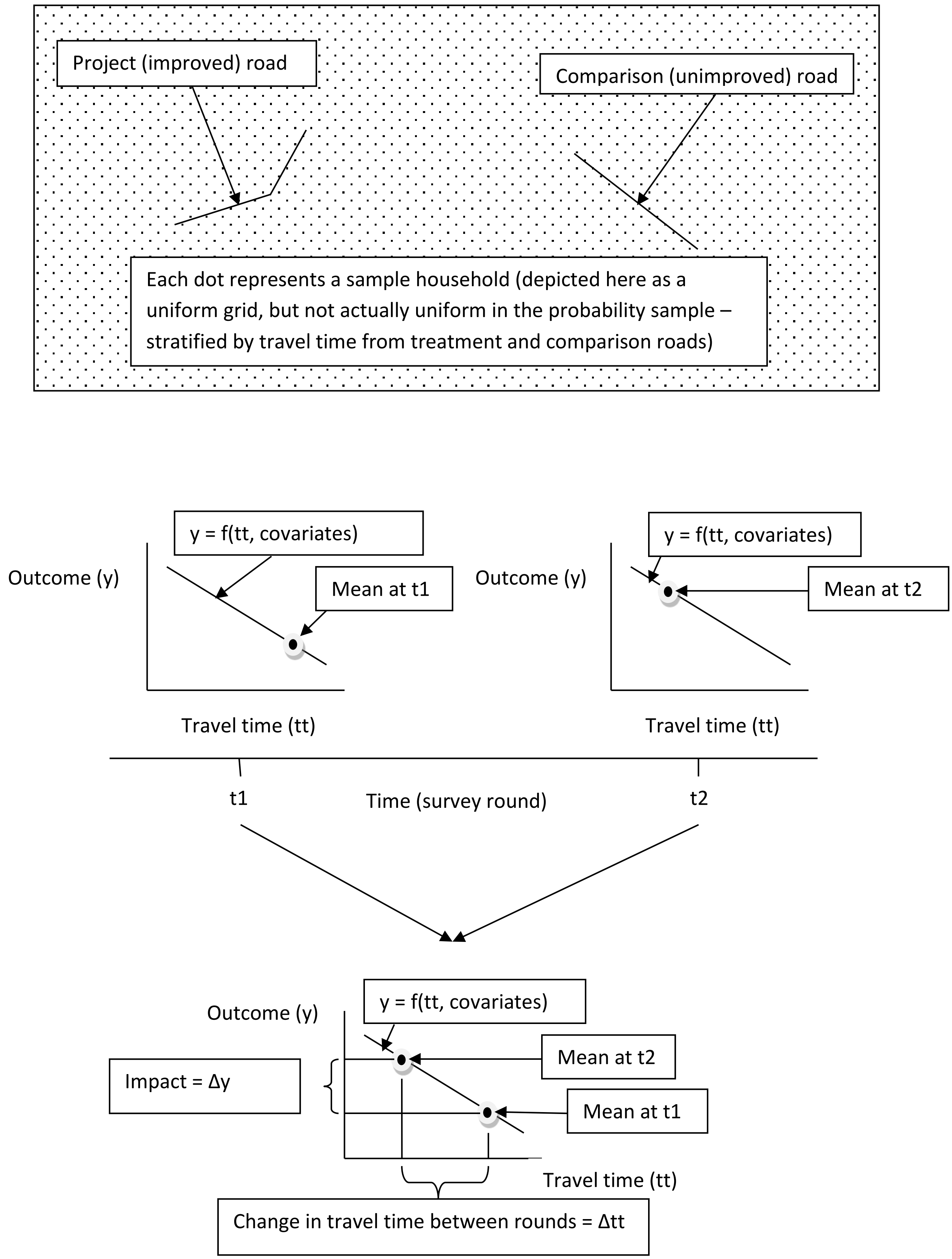
Figure 1: Conventional Approach to Estimating Impact of Road-Improvement Projects

A Pretest-Posttest-Comparison-Group Design Based on Binary Treatment Variable Defined by Zone of Influence, or Buffer Zone



In contrast to the conventional approach, the evaluation design adopted by NORC is based on using continuous treatment variables, such as travel times to places of interest (from a household or *caserío*). Figure 2 illustrates the CTV approach. (The graph shows a very simplified version of the model. The figure depicts a linear relationship, and nonlinear relationships are allowed. Also, the figure depicts the relationship of an outcome variable to a single explanatory variable, and the approach may include multiple explanatory variables. In addition, the figure shows the same relationship at both survey times, and the relationship may be different at the two survey times.) As mentioned earlier, an advantage of the CTV approach is that, not only does it provide a means of estimating the impact of a particular project, but, because it describes the relationship of impact to treatment variables (travel times), it may be used as the basis for analysis of project alternatives (such as completion or non-completion of a particular project activity).

Figure 2: The Approach Used by NORC to Estimate the Impact of the MCC Honduras Transportation Project (Continuous Treatment and Control Variables)



Conceptual Framework for Evaluation

This section describes the conceptual framework for the impact evaluation. It includes the following subsections:

- Causal Model
- Statistical Model Specification
- Analytical Survey Design
- Statistical Power Analysis Used to Determine Sample Size
- Identification of Parameters and Effects Related to Impact
- Estimation Procedures
- Estimation of Standard Errors
- Test-of-Hypothesis Procedures
- *Ex Post* Statistical Power Analysis
- Scope of Inference, External Validity
- Summary of Assumptions and Limitations

The three subsections dealing with estimation and *ex post* statistical power analysis are very brief. Additional details will be provided for these subsections later (in Annex 1).

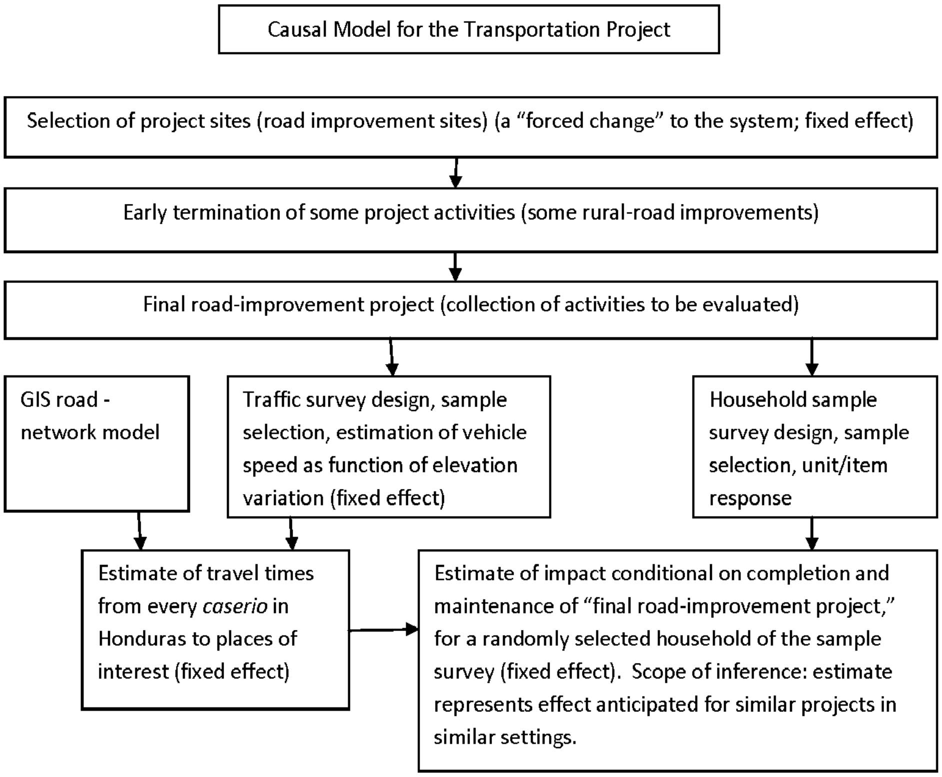
In the discussion of this section, reference will be made to both the binary-treatment-variable model and the continuous-treatment-variable model. This is done to clarify aspects of the continuous-treatment-variable model. The impact estimates that are presented later are based on the CTV model.

Causal Model

A causal model shows the causal relationship among factors or variables relevant to the evaluation. (The term "factors" refers to higher-level constructs, and "variables" to measurable lower-level quantities. For example, "wealth" and "ambition/motivation" are factors, while "value of home," "savings," "hectares owned," "highest grade achieved," and "number of civic offices and awards" are variables. Factors typically refer to constructs that are difficult to measure and cannot be directly observed or measured, and are reflected in (associated with) a number of (measurable) variables. Causal models are specified in terms of both factors and variables.) Causal models are useful because the nature of the relationships among entities relevant to a process under study determines what quantities (e.g., impacts of interest) may be estimated from the available data, and which statistical model specifications and estimation procedures are appropriate. Causal models are represented in various ways, such as equations and graphs. This evaluation will employ both of these methods.

Figure 3 is a high-level causal model (entity-relationship diagram) for the Transportation Project. Each arrow of the figure indicates that a causal relationship exists between the entity at the tail of the arrow and the entity at the head of the arrow. The (high-level) causal model diagram shown in Figure 3 differs from the usual (detailed, low-level) causal model diagrams, which show the relationship of variables in the survey questionnaire to each other and to certain exogenous variables, such as those used to select the project and the survey sample. These latter causal models show causal relationships among survey variables, such as endogeneity of household income and travel times and costs. The causal model diagram shown in Figure 3 addresses entities outside of the household survey questionnaire. Causal relationships within the survey questionnaire will be addressed again, when specific analytical models are considered.

Figure 3: High-Level Causal Model for the Transportation Project



Here follow some comments on the causal model diagram. The diagram is a directed acyclic graph (DAG). The directed arrows imply a time sequence. In this model, no simultaneous (nonrecursive) causal relationships are represented. The diagram corresponds to the conditional relationships discussed earlier, viz., that the estimation of impact is made for the particular road-improvement project that was implemented. Moreover, the impact estimates are based on a "fixed-effects" assumption, in which the times and locations of the higher-level sample units of the traffic and household surveys are considered fixed. This assumption applies to the locations of the *caseríos* and households of the household sample survey, the timing of the survey rounds; and to the season, day of week, time of day, and weather conditions of the traffic surveys.

Note that variables that are considered fixed here may in general certainly be considered to be random variables, and correlated, and mutually causally related. For example, the location and nature of a road-repair project may be determined by taking into account average household incomes in an area. Unconditionally, the project location and household income may have a mutual causal relationship (i.e., be endogenous). In this evaluation, however, impact is estimated *conditional* on the project's being fixed, and the locations of *caseríos* and households of the sample survey also being fixed (the random variation of the model is the within-household variation between survey rounds). This assumption restricts the scope of inference of the analysis, but the lack of a useful sample of road projects whose selection was based on randomization, or could be adequately modelled, necessitates this limitation.

Causal Model Diagrams

Causal models may be specified in different ways, such as by structural equations or by directed graphs (as described in Judea Pearl in *Causality: Models, Reasoning, and Inference*, 2nd ed., (Cambridge University Press, 2009, 1st ed 2000) and by Stephen L. Morgan and Christopher Winship in *Counterfactuals and Causal Inference: Methods and Principles for Social Research* (Cambridge University Press, 2007). A comprehensive theory of causal modeling has been developed by Judea Pearl for directed acyclic graphs (DAGs), which correspond to situations in which variables are not mutually causally related (i.e., to causal models that are recursive). Pearl describes criteria (such as the "back door" criterion) that a causal graph must satisfy for a causal effect to be identifiable (i.e., estimable – may be estimated from data on the variables represented in the graph).

The following figure (adapted from Morgan and Winship op. cit.) presents examples of DAGs. For these graphs, Y denotes an outcome of interest, W denotes treatment, X denotes the set of all variables other than W that affect outcome (i.e., are direct causes of Y), S denotes all variables that affect selection, Z denotes an observed subset of S, and U denotes unobserved variables. In the graphs, a solid directed arrow signifies a causal relationship, and a dashed arrow indicates that the endpoints are affected by common variables. (In the last figure, the term "fix" is used to mean physically (or in a "thought experiment," mentally) setting the value of a variable (as in Pearl's "do" calculus), rather than statistically conditioning on it.)

If selection is based on observed variables (which may or may not affect outcome), the effect of W on Y can be estimated (i.e., an unbiased or consistent estimate is available). If there exist unobserved variables that affect both W and Y, the effect of W on Y cannot be estimated (or, more accurately, a good (unbiased or consistent) estimate of the effect of W on Y is not available), without making certain assumptions about the distribution of the unobserved variables (e.g., that the unobserved variables are time-invariant within households over the term of the study). The purpose of constructing causal model diagrams is to assist determination of whether the effect of W on Y is estimable (i.e., is "identified").

Note that interest focuses on unobserved variables that affect *both* selection and outcome. If a variable affects selection but has no effect on outcome, then it is not relevant (e.g., if selection is based on eye color, and eye color has no effect on outcome, then it may be ignored).

In later discussion, we will make reference to the final panel of Figure 4, Figure 4e. This panel illustrates the fact that the effect of W on Y may be estimated (identified) either by conditioning on S (the set of all variables that affect selection for treatment) or on X (the set of all variables other than W that affect outcome).

Figure 4: Examples of Causal Diagrams

Figure 4a. Causal diagram illustrating nonignorable treatment assignment. W denotes selection for treatment. Y denotes an outcome of interest. Both W and Y are affected by unobserved variables (dashed line). Cannot estimate effect of W on Y (i.e., cannot construct a good (unbiased or consistent) estimate of the effect of W on Y).

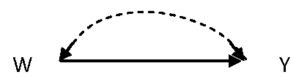


Figure 4b. Causal diagram illustrating ignorable treatment assignment. S denotes all variables that affect selection for treatment, all of which are observed. Can estimate effect of W on Y.

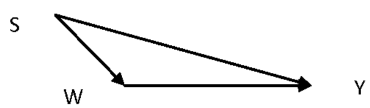


Figure 4c. Causal diagram illustrating selection on observables, S denotes all variables that affect selection for treatment; Z is an observed subset of S; U is an unobserved subset of S. The unobservables, U, do not affect outcome (Y). Can estimate effect of W on Y.

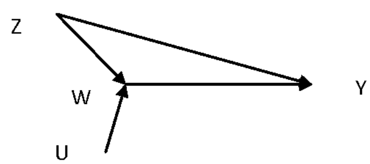


Figure 4d. Causal diagram illustrating selection on unobservables. The unobserved variables may affect both W and Y. Cannot estimate effect of W on Y without additional assumptions about the distribution of U.

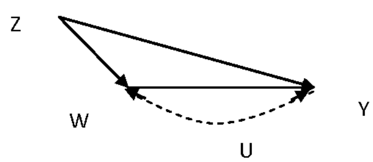
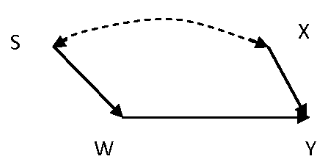


Figure 4e. Causal diagram illustrating fixing on S or X (X is the set of all variables other than W that affect outcome (Y)). If fix S or X, then the causal link between S and X is "broken" – there are no unobserved variables affecting both W and Y (since one of S or X is fixed). Can estimate effect of W on Y (fixing either S or X).



In order to obtain an unbiased estimate of a causal effect, it is necessary to establish the reasonableness of the assumption of conditional independence of the counterfactuals and treatment, given the covariates. Unfortunately, there is no statistical test for this. The standard approach is to specify a causal model in terms of general causal factors, and assess the extent to which those factors are represented by observable variables (i.e., variables from a survey questionnaire or other data sources). If one or more important factors are not represented by, or poorly represented by, observable variables, an assessment must be made about the effect that this situation will have on the impact estimate. In some cases, it is possible to construct a causal estimate even with selection on unobservables, under certain conditions (i.e., if certain assumptions are made, such as time-invariance of unobservables within households over the term of the study). In some cases, it is possible to construct bounds on the estimate by making assumptions about the magnitude of the effect of the unobservables on selection.

The causal diagram that pertains to the present evaluation is Figure 4e. For this evaluation, the goal is to assess the impact of a particular project on all households in Honduras. The project is fixed, and the households are fixed. There are no selection effects. (Note that the sample survey is a (stratified two-stage) probability sample from the population of all households in Honduras. Prior to selecting the sample, the households are random. Once the sample is selected, the households are considered fixed (and the only random variation of interest in the estimation of impact is the within-household variation between survey rounds).) In the diagram, the set S of all variables that affect selection for treatment are fixed. (Note that it is the *road segments* that are selected for treatment, not the households.)

Consideration of Endogeneity

It is instructive to consider the preceding example in somewhat more detail. Suppose that a transportation project consisted of improvements to many roads (meaning road *segments*), and that it was desired to estimate the average treatment effect of improving a road. Since there are many secondary and rural roads, but few primary roads, this example is better suited to the former road types. Suppose that there are well-defined criteria to identify roads that are eligible for improvement (i.e., for treatment). Suppose further that randomized assignment to treatment was feasible, and that a proper randomized experimental design was used as a basis for the evaluation (e.g., a "matched pairs" design might be used, in which eligible roads were matched on characteristics for which data were available, a random (probability) selection of pairs was made, and one member of each pair was assigned to treatment). In this example, an unbiased estimate of program impact may be made by comparing the treated and untreated roads (i.e., by taking the difference in means of the treated and untreated samples). Note that the estimate of impact is the average treatment effect *on a randomly selected eligible road segment*.

Now, consider the case in which the selection of the roads is not done using randomization. Either the procedure for determining eligibility for improvement is not clear, or the decision concerning which of the eligible roads to treat is not based on randomization. This is similar to the situation in the present evaluation, where the decision to improve a road was a joint decision of the central and local governments, based on various technical, political and economic considerations (such as the condition of the road, or the level of congestion, or accident rates, or the potential to increase economic activity, or the willingness of the local government to pay part of the cost of the road improvement). It could be, for example, that the central government examines the mean household incomes in an area, and decides to improve roads in the poorer areas. In this case, there is a mutual causal relationship between mean household income and the location of a road improvement project. All of the characteristics of the road segments (such as travel times) are mutually causally related to household characteristics. In this case, travel times from the household location or *caserío* are endogenous with respect to household income. It does not matter whether the travel times are reported in a household questionnaire or are taken from an independent GIS model – they are endogenous (with respect to household income). Income and travel times are mutually causally related – the model is a nonrecursive model (e.g., a simultaneous-equation model). In this case, it is necessary to use special methods to obtain an unbiased estimate of impact, as measured by the average treatment effect (ATE). An unbiased estimate of impact may be obtained under certain assumptions using the Rosenbaum-Rubin propensity score ("statistical") approach or the Heckman latent-variable ("econometric") approach. A key factor to keep in mind is that the ATE is the average effect of improving a *randomly selected eligible road segment*.

Finally, consider the situation of the present evaluation. The road segments to be improved were selected as in the just-preceding case. Because the selected primary road segments were unique, and the number of selected secondary road segments were small, there was no reasonable hope of developing a selection model, and implementing either the Rosenbaum-Rubin or the Heckman approaches. There is no hope of estimating the average effect of treating a randomly selected eligible road. In this case, a decision was made to estimate the impact of the *particular* transportation project, as it was finally configured (since this quantity is estimable). It is not at all the objective to estimate the average treatment effect associated with a randomly selected eligible road segment. It is the objective to estimate the impact of a particular project – a selection (however made) of road segments of various types, improved at particular times and in particular locations. In technical terms, all aspects of the project (the improved roads) are "fixed effects." It does not matter at all that the selection process used by the central or local government may have selected roads that have tremendous potential for increasing household incomes, or very little potential. There is no "selection bias" in the impact estimates, since the selection process is irrelevant, once the project is fixed.

In summary, it is not the objective to estimate the impact associated with *a randomly selected road* (segment), but to estimate the impact of *the particular project at hand*, however it was selected. In this case, incomes and travel times reported in the survey questionnaire are mutually causally related (because, within the term of the project, a household may switch to a different mode of transport, if its income increases), but there is absolutely no causal relationship between household income and the GIS-model travel times. At the baseline time, the GIS-model travel times (from every *caserío* to any other point in the country) are fixed (they have nothing to do with the project or survey household locations, since these have not yet been determined), and at the endline time, after implementation of the project improvements, they are also fixed (since the project is fixed, and the travel times are conditional on the project). In technical terms, the GIS-model travel times are exogenous (with respect to income), given that the project is fixed. This fact plays a key role in the choice of an estimator, since if the GIS-model travel times are exogenous, they are not correlated with the model error terms (and if no explanatory variables of a regression model are correlated with the model error terms, standard procedures (e.g., ordinary least squares) may be used to obtain unbiased estimates of regression coefficients and other effects of interest).

The points made in the preceding paragraphs are illustrated further by means of an example, in Annex 7.

The term “fixed” is somewhat ambiguous. It may indicate that a variable is nonstochastic, or that it is a conditioning variable, or that it is stochastic and not correlated with a model error term. While the latter is the usual interpretation in econometrics, the former is easier to understand and to justify. The term may also refer to making a forced change in a variable (as in Judea Pearl’s “do” calculus), rather than statistical conditioning. The intended meaning of the term is important, and should be clear from context. Here, we use the term to indicate that the scope of inference is restricted to the Transportation Project as finally configured, to the locations (road segments) and times associated with the traffic surveys (season, time of day, etc.), and to the times and locations of the higher-level sampling units (*caseríos* and households) of the household survey. This means that the effects associated with *caseríos* and households are not stochastic, given the sample. The random variation associated with the household sample survey is the variation *within* households (i.e., within households between survey rounds).

Within a household, mutual (simultaneous, nonrecursive) causal relationships exist. For example, household travel times may affect household income and vice versa. A graphical representation that includes simultaneous causal relationships would be a directed cyclic graph (DCG). When particular estimating equations are considered later, mutual causal relationships will be identified. Some models (e.g., a model containing only treatment and survey round) may contain only fixed effects, and are “fixed-effects” models. Other models, however, may contain household variables that may be considered to be random variables and may be mutually causally related. For example, household income and travel times may be considered to be endogenous (mutually causally related). Such models (containing both fixed effects and random effects) are “mixed-effects” models. Models in which a mutual causal relationship exists between a dependent variable and an explanatory variable are “simultaneous equation models,” or “nonrecursive models.” These circumstances must be addressed in the estimation process, in order to obtain unbiased (or consistent) estimates of effects of interest (such as impact). For specific models involving questionnaire variables, explanation will be provided about model specification and identification, additional to the description provided by the preceding high-level causal model.[2]

The project represents a “forced change,” or intervention, to the road system in Honduras. This intervention was not decided by randomization, but it is a forced change nonetheless, caused by the road planners and local officials of Honduras. (Since the program intervention is a forced change, it is an experiment. It is not an experimental design based on randomized assignment to treatment, but it is an experiment, nonetheless. This is an important consideration. In order to estimate the effect of making a change to a system, it is necessary to base the estimate on data for which a forced change was made to the system. Were it not for the forced changes represented

[2] For more discussion of causal modeling, refer to *Causality: Models, Reasoning and Inference*, 2nd ed by Judea Pearl (Cambridge University Press, 2009, 2000). For a summary, see *Counterfactuals and Causal Inference: Methods and Principles for Social Research* by Stephen L. Morgan and Christopher Winship (Cambridge University Press, 2007). See also “Statistics and Causal Inference” by Paul W. Holland (*Journal of the American Statistical Association*, Vol. 81, No. 396 (Dec., 1986)). Discussion of structural equation modeling is presented in *Linear Causal Modeling with Structural Equations* by Stanley A. Mulaik, (Chapman & Hall / CRC Press, 2009), *Nonrecursive Causal Models* by William D. Berry (Sage Publications, 1984); *Introduction to Structural Equation Models* by Otis Dudley Duncan (Academic Press, 1975); and *Structural Equation Models in the Social Sciences* by Arthur S. Goldberger and Otis Dudley Duncan (Seminar Press, 1973).

by the program intervention, the analysis would be based on "observational data.") Since the project selection was not decided by randomization, and parts of it (CA-5) are unique, an appropriate approach to the evaluation is to estimate the impact of this particular project, rather than to attempt to represent it as a random selection from a conceptually infinite population of similar projects (or similar road segments) at various locations or times, and to estimate impact within a broader context. The scope if inference is that this is the impact that would be anticipated for similar projects in similar settings.

The household sample is a nationwide probability sample, using stratification and multi-stage sampling. Whether the survey estimates are based on a fixed-effects or random-effects assumption (about the higher-level sample units) has little effect on the impact estimates. The impact estimates constructed in this analysis are "fixed effect" estimates, meaning that once the random (probability) sample of *caseríos* and households have been selected, those *caseríos* and households (along with survey round and the project) are considered fixed for the analysis, and the stochastic variation associated with the model is the residual variation within households. While the standard errors of fixed-effects estimators may be substantially greater than the standard errors of random-effects estimators, the power of tests of hypothesis may be much greater when the random variation associated with *caseríos* and households is removed from effects of interest.[3]

The trade-off here is that the power of tests of hypothesis increases, but the scope of inference decreases. The impact estimates are fixed-effects estimates, with respect to project, survey round, *caserío*, and household.

The GIS model is independent of the household and traffic surveys (except for its use of speeds from the traffic surveys, to estimate travel times over specified routes).

Estimation of Travel Times

The traffic surveys were conducted for all project roads and a sample of similar non-project roads. Average speed was estimated for a variety of vehicle types, as a function of elevation variation. This was done for all treatment roads and for a matched sample of non-treated comparison roads. The matching was done to reduce the effect of variables other than elevation variation on the estimate of average speed. Speed estimates were desired for treatment roads and *all other roads*, not just *matching non-treatment roads*. The population of "all other roads" was replaced by the population of matching comparison roads because speed was to be estimated *as a function of elevation variation*, and it was desired to reduce the effect of other variables on the estimate. Had the selection of project roads been different, the speed estimates would have been different because of random variation and because the treatment roads were different. Because the speed estimate is conditional on elevation variation, however, the effect of selecting a different project (on estimated speed, conditional on road type, elevation variation and project

[3] The assumption of whether to use fixed-effects, random-effects, or mixed-effects models makes a substantial difference in the scope of inference and in testing of hypotheses. These notions are discussed at length from an econometric viewpoint in *Econometric Analysis of Cross Section and Panel Data* 2nd ed. by Jeffrey M. Wooldridge (MIT Press, 2010, 2002). From a statistical-analysis viewpoint, references on this topic include *Generalized, Linear and Mixed Models* 2nd ed. By Charles E. McCulloch, Shayle R. Searle and John M. Neuhaus (Wiley, 2008); *Variance Components* by Shayle R. Searle, George Casella and Charles E. McCulloch (Wiley, 1992, 2006); and *Linear Models* by S. R. Searle (Wiley, 1971).

intervention status) on the speed estimates would be expected to be small. In any event, the speed estimates with and without the project intervention pertain to this particular project (since the traffic surveys were conducted on the roads of this particular project, plus the comparison sample).

The product of the traffic surveys was a table showing mean speed of pickup trucks over roads, as a function of treatment condition (improved or not improved), road type (primary, secondary, rural) and elevation variation. There would have been no advantage to constructing a table conditional on additional road characteristics (such as number of lanes, access, roughness, curvature), since these variables were not available in the GIS (for use of the speed table to estimate travel times over *any* route). Of the various road-characteristic variables on which vehicle speed was dependent, data were available for every road in the country only for road type and elevation variation, and so the table was conditional only on treatment status and those two road-characteristic variables.

The data available for matching comparison roads to treatment roads were limited to location, road type (primary or secondary), and the extent to which the road served as a connector between major sections of the road network or between major population centers. The comparison roads were selected to provide extensive geographic coverage over the country, with care to include a good representation of "connector" roads.

Using the speed table produced from the traffic surveys, the GIS model can estimate travel time from every *caserío* to any other point in Honduras. For each household of the sample survey, travel times were estimated for pickup trucks from the *caserío* centroid to ten places of interest from that *caserío*. This was done for each of the two household survey rounds. For the baseline survey, the travel-time estimates represent no project intervention. For the endline survey, the travel-time estimates represent the effect of the project intervention, assuming completion and maintenance of the non-terminated project activities.

Impact was estimated conditional on the preceding baseline and endline travel-time estimates. The estimate of impact is for a randomly selected household of the sample survey (i.e., of the nation), conditional on completion and maintenance of the road-improvement project as finally configured. (The term "randomly selected household" may be misleading. Although the household sample was a probability sample, the estimation procedure assumed that the sample *caseríos* and households, once selected, were fixed for the analysis (the random variation of interest to the analysis is the within-household / between-round variation). Since the sample is a large probability sample for the entire country, the scope of inference is the entire country, even though the fixed-effects estimate is conditional on the selected *caserío* and household sample.)

In this conceptual framework, there are no "selection effects" associated with the project or the households, that need to be taken into account in estimating impact. The impact estimates are conditional on the particular project as finally configured, however the project was determined. It does not matter *at all* how the project roads were selected, or that some of the project roads were terminated early. The evaluation is based on the causal model depicted in Figure 4e, in which impact estimates are conditional on all variables (in set S) affecting selection. The household survey is a nationwide sample of *all* households in the nation, stratified by many variables. There are no selection effects associated with the household sample survey because *all* households in Honduras are subject to probability sampling. All estimates are "fixed-effects" estimates conditional on the particular road project (as finally configured), the times of the baseline and endline surveys, the *caseríos* selected for the household survey, and the households selected within the sample *caseríos*. Conditional on the locations of the project road and household survey sample units (*caseríos* and households), the GIS-model travel times are uncorrelated with all household survey variables, in both survey rounds. (This fact is of importance in assessing the compliance of the sample data with assumptions associated with the estimation procedures to be used, that the explanatory variables of the model (in this case, GIS-model travel times) are uncorrelated with the model error terms. This point may be difficult to grasp. The household variables, such as income, are causally related to the GIS-model travel times. Under the fixed-effects assumption, the GIS-model travel times associated with a household are fixed numbers. The statistical model is $Y(T) = h(T,X) + e$, where Y denotes outcome, T denotes travel time, X denotes covariates, and e denotes a model error term. Under the fixed-effects assumption, T is not correlated with e (it is not even a random variable).)

Unit of Treatment / Unit of Analysis

From a conceptual viewpoint, it is important to maintain the distinction between the unit of treatment and the unit of analysis. The unit of treatment is the road. The unit of analysis is the household. In the conventional approach of constructing a "zone of influence" or "buffer zone" around a project road (or around a comparison road), these notions are conflated. The households within the zone near a treatment road are artificially labelled as "treatment" households, and the households located in other zones are artificially labelled as "control" households. This convention is misleading and conceptually flawed because it is the *roads* that are treatment and controls, not the *households*. Even if there is no selection issue for roads (e.g., because we are estimating impact for a fixed road), this approach introduces the notion of selection of households into treatment. By introducing a binary treatment variable for households, there is a motivation to apply the Neyman-Fisher-Cox-Rubin potential outcomes (counterfactuals) causal model, or the Heckman latent-variable potential outcomes causal model, to the analysis. This approach is conceptually flawed and confused, attenuates the impact, and lessens the power of the sample to detect effects of specified size. Annex 1 presents a detailed description of this approach, and demonstrates the weak results for the BTV approach, compared to the CTV approach. The binary-treatment-variable zone-of-influence approach is conceptually flawed and results in an attenuated estimate of impact. It should not be used to evaluate road improvement projects.

A similar confusion is introduced by comparing the CTV approach with a "dose response" model (or "intensity of treatment" model) from medical applications. The analogy is made that households located far from project roads receive a "lower dose" of treatment, represented in the present application by change in travel time caused by the project intervention. This analogy is misleading and leads to difficulty in interpreting the results. Through this analogy, the binary treatment applied to roads (improved / not improved) is converted into one or more continuous treatment variables applied to households (change in travel time to places of interest). A problem that arises here is that there is a motivation to view the partial treatment effect as the marginal change in outcome produced by a unit change in travel time. The conceptual difficulty that arises is that making changes to travel times cannot be done directly or continuously. A decision may be made to improve or not improve a road, and forced changes may be made in the transportation system via this discrete decision. There is no way to force changes in travel time directly, and it is not possible to make arbitrary incremental changes in the travel-time variables. If more than one travel time is included in a model (i.e., travel times to several places of interest), it is not possible to make independent changes in them – they are highly mutually causally related. All that can be done is to change the counterfactual treatment status of a road (from improved to not improved or vice versa). From a conceptual viewpoint, the travel time may be viewed as an intermediate causal variable that is affected by the program intervention (treating one or more road segments). Travel times to multiple places of interest are highly correlated – if several travel times are included in a model, their effects are hopelessly confounded. The travel time is not a control variable, which use of the term "dose response" implies. The partial-treatment-effect model may be used to assess outcome changes associated with moving from one project counterfactual state to another for roads, not to predict the incremental change in outcome associated with an incremental change in travel times.

Analytical Survey Design

The survey design for this evaluation was an analytical survey design. This type of survey design differs substantially from the survey designs used for descriptive sample surveys. The purpose of descriptive sample surveys is to estimate overall characteristics of a population or subpopulations of interest, such as means, proportions and totals. The purpose of analytical sample surveys is to collect data to enable the construction of analytical models, such as a model that estimates the impact of a program intervention, or of the relationship of impact to explanatory variables.[4]

In a descriptive survey, it is generally attempted to keep the sample selection probabilities as uniform as possible, subject to achieving high precision for estimates for overall population characteristics. For an analytical survey, it is attempted to achieve adequate variation in explanatory variables that are considered to have an important relationship to outcomes of interest, so that those relationships may be estimated with high precision. These two designs are often very different. The selection probabilities for an analytical design usually vary substantially more than for a descriptive survey design.

The sample design took into account *caserío* data that were available from Honduras Census and geographic information system sources[5]. These data included elevation, climatic zone, soil capacity, rainfall, temperature, vegetation cover, protected area status, distance to nearest major river, population, and a variety of travel times to point of interest. We used these data to construct an analytical survey design that had substantial variation on these variables. A two-stage sample design was used, with marginal stratification on the variables just listed. Marginal stratification was implemented using the method of variable probabilities of selection. Some additional information on the sample survey design is presented in the next section. A detailed description of the sample design is presented in Annex 3.

[4] For background information on these two approaches to evaluation and survey design, see the following references: (1) "History and Development of the Theoretical Foundation of Survey Based Estimation and Analysis," by J. N. K. Rao and D. R. Bellhouse, *Survey Methodology*, June 1990, Vol. 16, No. 1, pp. 3-29 Statistics Canada; (2) *Sampling: Design and Analysis* by Sharon L. Lohr (Duxbury Press, 1999); (3) *Sampling*, 2nd edition by Steven K. Thompson (Wiley, 2002); (4) *Practical Methods for Design and Analysis of Complex Surveys*, 2nd edition by Risto Lehtonen and Erkki Pahkinen (Wiley, 2004). (The Lohr book is the most informative.)

[5] *Caseríos* are the smallest local-level governmental administrative unit in Honduras. The country is hierarchically divided into 18 departments, 298 municipalities, 3,721 *aldeas*, and 27,969 *caseríos*.

Statistical Power Analysis Used To Determine Sample Size

For a two-stage sample design, there are two sample sizes – the sample size for the first-stage sample units (primary sample units, PSUs; in this case, *caseríos*), and the sample size for the second-stage units, or households, within each first-stage unit. We determined the sample size for households by taking into account the relative cost of sampling *caseríos* vs. households, and the intra-unit (intra-*caserío*) correlation coefficient. We determined that an efficient household sample size was about 20 households per *caserío*. Based on this analysis, the *caserío* sample size was determined to be 100. A major factor affecting the choice of sample size was the minimum effect to be detected with high power. A detailed description of how this per-*caserío* household sample size was determined is presented in Annex 4.

We conducted a statistical power analysis to determine a *caserío* sample size that would achieve high power for detecting effects of specified size. The *Monitoring and Evaluation Plan* for the project did not specify magnitudes of changes expected for the outcome variables considered to be important for the household survey. Instead, it projected that the economic rate of return (ERR) for the project would be in the range 12-21 percent of the baseline. We determined a sample size, using statistical power analysis, that would detect impacts in this same range with high probability (power), for outcome variables having typical ranges of values for the coefficient of variation and intra-unit (*caserío*) correlation coefficient. Annex 4 presents a summary of the statistical power analysis.

Statistical Model Specification

The causal model underlying the evaluation design is that changes in travel times produce changes in income and other response variables of interest. Based on this model, we used statistical analysis to estimate the relationship of outcomes of interest to travel times to places of interest, and, from that model, we estimated the change in outcomes associated with the changes in travel times caused by the project. Estimates of travel times (and hence, changes in travel times caused by the project) were obtained from a GIS road network model that used vehicle-speed estimates from traffic surveys conducted before and after the road-improvement intervention.

As discussed earlier, a simple conceptual representation of the relationship of outcome to travel time may be represented as:

> *Vehicle speed = f(project intervention (road improvement), road characteristics (primary, secondary, rural; elevation variation), given vehicle type, season, day of week, time of day, weather)*
>
> *Travel time (for pickup truck) = g(road characteristics, mean vehicle speed given road characteristics)*
>
> *Outcome measure = h(travel-time variables and other variables ("covariates"))*
>
> *Impact = Expected value (mean) of outcome measure conditional on completion and maintenance of road-improvement project and on covariate means for both survey rounds – Expected value of outcome measure at beginning of project, conditional on covariate means for both survey rounds*

where f(.), g(.), and h(.) denote continuous functional relationships.

The functional relationship was estimated using the general linear statistical model. Using this approach, an unbiased estimate of the average treatment effect (ATE) may be obtained from a regression model that expresses outcome as a function of explanatory variables. The ATE may be estimated conditional on the values of explanatory variables (**x**), or as an unconditional average (over those variables). The basic regression model on which outcomes of interest are related to explanatory variables, and from which impact estimates are derived, is the following:

$$y_t = \mathbf{x'}_t\boldsymbol{\beta} + e_t,$$

where

t = survey round index (0 for Round 0 or baseline, and 1 for Round 1 or endline)
y_t = outcome variable
$\mathbf{x}_t$ = vector of explanatory variables (design variables, treatment variables, and covariates; one component of which is “1”, representing a constant term)
$\boldsymbol{\beta}$ = vector of parameters (one parameter is a constant term)
e_t = model error term.

To obtain unbiased estimates of the model parameters using the ordinary-least-squares (Gauss-Markov) estimation procedure, it is necessary to make assumptions about the model error terms and the relationship of the explanatory variables to the error terms, such as that the model error terms have zero mean, constant variance and are uncorrelated with each other and the explanatory variables. Such assumptions are reviewed for each model constructed in the course of the analysis (since each model has a different specification (set of explanatory variables and functional form)).[6] The validity of the model parameter estimates and effects of interest is dependent on the correct specification of the outcome model. This means that assumptions about whether explanatory variables are uncorrelated with model error terms, or that there are no hidden variables or errors in variables, must be examined and justified. This will be done on a case-by-case basis, as particular models are examined. As discussed earlier, the causal model assumes that the project is fixed, so selection is not an issue.

For the preceding outcome model, the regression coefficients corresponding to treatment-related variables (i.e., to travel-time-related variables) are partial treatment effects (PTE). This model will be referred to either as an “outcome” model or a “partial treatment effects” model. For this application, the main feature of the model is that all of the explanatory variables outside of the household are considered to be fixed effects, in which case they are uncorrelation with the model error term.

---

[6] The models used for this analysis are described in the book, *Econometric Analysis of Cross Section and Panel Data*, 2nd edition, by Jeffrey M. Wooldridge (Massachusetts Institute of Technology Press, 2010, 2002). For additional information, see *Mostly Harmless Econometrics* by Joshua D. Angrist and Jörn-Steffen Pischke (Princeton University Press, 2009); *Counterfactuals and Causal Inference: Methods and Principles for Social Research* by Stephen L. Morgan and Christopher Winship (Cambridge University Press, 2007); *Micro-Econometrics for Policy, Program, and Treatment Effects* by Myoung-Jae Lee (Oxford University Press, 2005; *Analysis of Panel Data* 2nd edition by Cheng Hsaio (Cambridge University Press, 1986, 2003); *Econometric Analysis of Panel Data* by Baki Baltagi, 4th ed., (Wiley, 2008); and *Econometric Analysis* 7th ed. by William H. Greene (Prentice Hall, 2011).

Differences in the relationship between rounds are addressed by the inclusion of interaction terms (with survey round).

The expected value of y_{et} is given by the same equation as above, omitting the error term:

$$E\ y_t = \mathbf{x}'_t\boldsymbol{\beta}.$$

Identification of Parameters and Effects Related to Impact

An estimate of impact is obtained from the preceding equation by taking the difference of the expectation equation between survey rounds:

$$Impact = \bar{\boldsymbol{x}}_1'\boldsymbol{\beta} - \bar{\boldsymbol{x}}_0'\boldsymbol{\beta}$$

where each component of $\bar{\boldsymbol{x}}_t$, denoted as $\bar{x}_{ti}$, is as follows: for travel-time-related (treatment) variables the component is the GIS-model mean for round t (t = 0 or 1); and for other variables (covariates) the component is the mean over *both* survey rounds (t = 0 and 1).

The preceding general linear statistical model is linear in the parameters ($\boldsymbol{\beta}$). It is also linear in the explanatory variables ($\mathbf{x}$), so that (in this special case) the formula for impact may be represented as:

$$Impact = \Delta\bar{\boldsymbol{x}}_1'\boldsymbol{\beta}$$

where the subscript on $\bar{\boldsymbol{x}}$ denotes survey round, and where Δ denotes the backward difference operator (defined by $\Delta x_t = x_t - x_{t-1}$). In this representation, the components of $\Delta\bar{\boldsymbol{x}}_1$ are zero for all non-travel-time-related variables and equal to the difference in means between survey rounds for the travel-time-related variables. (Although the preceding model is linear in the explanatory variables, nonlinear representations are allowed in terms of the observed variables, e.g., one explanatory variable may be x and another may be x^2.)

The preceding model allows for the treatment variables (travel-time-related variables) to be binary or continuous. In the binary case, for a single treatment variable, the impact estimate is simply the coefficient on that variable. This is easy to see. Let x_1 denote the binary treatment variable, $\mathbf{x}^*$ all of the other variables (design variables, covariates), and β^* the coefficients for the other variables. Then $\bar{x}_{11}$ = 1, $\bar{x}_{10}$ = 0, and $\overline{\boldsymbol{x}^*}_1 = \overline{\boldsymbol{x}^*}_0$ (where the first (or sole) digit in the subscript denotes survey round), so

$$Impact = \bar{\boldsymbol{x}}_1'\boldsymbol{\beta} - \bar{\boldsymbol{x}}_0'\boldsymbol{\beta} = \bar{x}_{11}\beta_1 - \bar{x}_{10}\beta_1 + \overline{\boldsymbol{x}^*}'_1\boldsymbol{\beta}^* - \overline{\boldsymbol{x}^*}'_0\boldsymbol{\beta}^* = \beta_1.$$

For the continuous case with multiple treatment variables, the general formula must be used. That is, the impact is not represented by a single regression coefficient, and the value of the treatment variables are not binary (i.e., they are continuous travel times). The model coefficients are partial treatment effects, and the impact estimate is obtained by multiplying the treatment-variable coefficients by their respective means for the two survey rounds ("before" and "after" the treatment intervention), and summing. In the case in which a single travel-time-related variable is included in the model, the formula is:

$$Impact = \bar{\boldsymbol{x}}_1'\boldsymbol{\beta} - \bar{\boldsymbol{x}}_o'\boldsymbol{\beta} = \bar{x}_{11}\beta_1 - \bar{x}_{10}\beta_1 + \overline{\boldsymbol{x}^*}'_1\boldsymbol{\beta}^* - \overline{\boldsymbol{x}^*}'_0\boldsymbol{\beta}^* = \beta_1\Delta\bar{x}_1.$$

(For the models to be considered later, there will be two treatment (travel-time-related) variables.)

Note that the preceding formulas pertain to a model that is linear in the explanatory variables of the model. If the model is not linear in the explanatory variables, then the impact formula must be calculated as the difference in mean outcomes for the two survey rounds (with the round-specific means for travel-time-related variables and the both-round (overall, grand) means for other variables), and it will not in general be expressible as a function of $\Delta\bar{x}_1$. These special cases are presented here because the models developed in the analysis were in fact linear in the model explanatory variables.

In general, if **d** denotes the vector having the difference in means between rounds for all components associated with treatment (i.e., with GIS-model travel times) and zeros for all other components, then the impact is equal to **d**'**β**.

Estimation Procedures

The preceding formulas have used the true (unknown) value of the model parameters, **β**. In fact, the sample estimate, $\widehat{\boldsymbol{\beta}}$, will be used to estimate impact. The estimate of **β** for the general linear statistical model is given by the following formula. The model (in matrix notation) is

$$y = \boldsymbol{x}'\boldsymbol{\beta} + e$$

for an individual observation, where **x**' = (1, x_2, …, x_k), k denotes the number of parameters, and

$$\boldsymbol{y} = \boldsymbol{X\beta} + \boldsymbol{e}$$

for the sample, where **y**' = (y_1, y_2, …, y_n) and n denotes the sample size. If the e's have mean 0 and variance σ^2 and are uncorrelated with the explanatory variables, an unbiased estimator of $\widehat{\boldsymbol{\beta}}$ is

$$\widehat{\boldsymbol{\beta}} = (\boldsymbol{X}'\boldsymbol{X})^{-\mathbf{1}}\boldsymbol{X}'\boldsymbol{y}\ .$$

Impact is estimated by using the formula given earlier (for the expected impact), replacing the parameter $\boldsymbol{\beta}$ by $\widehat{\boldsymbol{\beta}}$:

$$Estimated\ Impact = \ \Delta\overline{\boldsymbol{x}}_1'\widehat{\boldsymbol{\beta}}\ .$$

In order to obtain unbiased estimates using the preceding estimator, it is necessary that the model assumptions be satisfied. A key assumption is that the explanatory variables not be correlated with the error terms. For the explanatory variables that are design variables (survey round, *caserío* and household), this assumption holds, since they are considered as fixed effects (i.e., not as random variables, so they are not correlated with other variables). The main model of interest will contain only design variables, and so this assumption will hold. (One reason why non-design-variables (covariates) are not included is that they are constant within a household over the term of the study, and so they "drop out" of (i.e., are not estimable in) a fixed-effects model.) Some models that were examined in the course of the analysis also contained other household variables, such as education and size of farm. For these variables to be uncorrelated with the model error term, it is necessary that they not be affected by changes in the dependent variable, over the term of the project. The reasonableness of this assumption will be examined when specific models are constructed, in Annex 1.

Another requirement for the model assumptions to hold is that there be no unobserved (hidden) variables, or that if there are, they are constant (time-invariant) over the term of the study (in which case they drop out of the two-round panel-survey fixed-effects model). Note that many evaluation projects are concerned with "selection effects," in which effects may be biased because of a lack of randomized assignment to treatment (e.g., a farmer training program, in which the decision for a farmer to participate is made by the program implementer and/or the farmer). In that case, it is necessary (for a panel survey, using fixed-effects estimates) to assume that hidden variables that are correlated with selection and outcome be time invariant. Such an assumption is not required here, since the road-improvement project is considered fixed (selection effects are not relevant).

Note that the outcome model is based on both rounds of survey data, and that the project intervention was manifest in the second round. An issue to address is whether this fact affects the outcome model. The project intervention may certainly affect traffic conditions and outcome variables in the second round. From the viewpoint of estimation, an issue to address is whether the presence of the project in the second round might affect the *relationship* of outcome to travel time, and if so, whether and how it is taken into account. The travel times are based on travel speeds estimated from the traffic surveys. The outcome of the traffic surveys was a table that shows the mean speed of a pickup truck over road segments as a function of road type (primary, secondary, rural), elevation variation, and project intervention status (improved or not improved). This table takes into account treatment status. For this reason, the estimate of the relationship of outcome to travel time explicitly takes into account the fact that the program intervention is present in the second survey round.

The outcome model shows the relationship, over the entire country, of outcome to travel times (to places of interest), taking into account road characteristics and project intervention status. Under the assumption that the project is fixed, it does not matter how the project was selected, e.g., that selection of the project location may have taken into account household characteristics (such as income, assets or crop varieties).

Estimation of Standard Errors

For all estimates, standard errors of the estimates were estimated using the standard formulas for the general linear statistical model.

The variance-covariance matrix of $\widehat{\boldsymbol{\beta}}$ is

$$\sigma^2(\boldsymbol{X}'\boldsymbol{X})^{-1}$$

where an unbiased estimate of σ^2 is

$$\hat{\sigma}^2 = \frac{1}{n-p}\left(\boldsymbol{y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}}\right)'(\boldsymbol{y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}})$$

where n denotes the sample size and p denotes the number of estimated parameters. The parameters include all of the design parameters, such as the survey round, *caserío* and household indicator variables, not just those of special interest, such as model coefficients associated with treatment variables and covariates. The preceding formula assumes heteroskedasticity, or uniform variance over the sample units. This assumption may be relaxed (and is, in using the Stata statistical computer program package to perform the estimation calculations).

The variance of a linear function of the parameters, $\mathbf{c}'\boldsymbol{\beta}$, is given by $\mathbf{c}'\mathrm{var}(\widehat{\boldsymbol{\beta}})\mathbf{c}$.

The preceding formulas ignore sampling variation associated with estimation of travel time. They account for sampling variability associated with the household sample survey. The impact estimates are conditional on the estimated travel times.

Test-of-Hypothesis Procedures

The procedures for making tests of hypotheses is substantially simplified by the fixed-effects assumption. The reason for this is that the standard errors of the effects of interest are substantially simplified (the "components of variance" associated with fixed effects includes just the model error variance, not variances associated with random effects).

Ex Post Statistical Power Analysis

It turned out that in this evaluation, the estimates of impact were very small, and in many cases they were not statistically significant. In such cases it is appropriate to ask whether the impact estimates are not statistically significant because they are small in magnitude, or whether the sample size for the evaluation was not sufficiently large to detect effects of the observed size with high power (probability). To address this question, an *ex post* (or *post hoc*) statistical power analysis was conducted.[7] This analysis showed that the study was not underpowered.

Scope of Inference, External Validity

In the preceding discussion, it was mentioned that the issue of considering certain effects to be fixed has implications for the scope of inference, or external validity, of the evaluation. External validity refers to the extent to which the estimates of impact for this study may be generalized to a broader context. These implications are discussed in the following paragraphs.

The issue of external validity is addressed by consideration of the populations of interest, the sample design, and the specification of the stochastic nature of the study variables. These factors will be discussed, in turn.

Populations of Interest and Sample Designs

[7] For a detailed description of *ex post (post hoc)* statistical power analysis, refer to David M. Murray, *Design and Analysis of Group-Randomized Trials*, Oxford University Press, 1998.)

For this study, there are three populations of interest – the population of potential road-improvement projects, the population of roads for the traffic surveys, and the population of households for the household survey. The assumptions made about these populations, and the nature of the sample designs used to make inferences about them, are as follows.

1. Transportation Project. No consideration is given to a population of potential road-improvement projects of which the Transportation Project is a representative. No statistical inferences are to be made about such a population, or to the population of road segments that comprise it. Inferences are to be made about the impact of the particular Transportation Project at hand. It is *not* the goal of the evaluation to estimate the average treatment effect of improving a randomly selected road segment.

2. Traffic Surveys. The (main) purpose of the traffic surveys was to produce a mean-speed table that could be used by the GIS road network model to estimate travel times. The decision was made to estimate mean speeds as a function of road-improvement status, elevation variation, and vehicle type. Other conditions, such as season, day of week, time of day and weather were held constant. With respect to the issue at hand (consideration of population), an issue to address is whether to select a matched sample of comparison roads or a sample of comparison roads from the entire non-treated population (of all non-improved roads in Honduras). The purpose of the comparison sample is to enable estimation of mean speed for unimproved roads. The approach adopted for accomplishing this was to stratify the country (i.e., the country's road segments) by road and environmental characteristics) and estimate the mean speed for each stratum (i.e., cell of the mean speed table). Because project resources were limited, traffic surveys could be conducted for only a small sample of comparison roads. It was decided to conduct a traffic surveys on each project road and on an individually matched sample of comparison roads. This approach afforded the advantage of increasing the precision of the comparison of speeds (and travel times) between counterfactual outcomes (i.e., between the two potential treatment statuses of a road segment), by removing extraneous sources of variability. It is viewed that this sample design was well matched to the evaluation objectives (estimation of impact using a potential-outcomes approach). The impact estimates are conditional on travel times, which in turn are conditional on the variables held constant in the traffic surveys (season, etc.). This conditioning represents a restriction in scope of inference. The alternative, of allowing the conditioning variables (season, etc.) to vary, was not practical to implement, given the level of resources available for the traffic surveys. Since the comparison roads were selected from the entire country, the scope of inference was not restricted by the selection process, other than by the matching process.

3. Household Sample Survey. The primary goal of the evaluation was to assess the impact of the Transportation Project on a randomly selected household in Honduras. A secondary objective was to estimate the total national-level impact of the project. Given this goal, the household sample survey was designed as a nationwide probability sample, marginally stratified to achieve high precision for estimating the average treatment effect on a randomly selected household in Honduras (and high power for making tests of hypotheses about impact). The surveyed population was all households in Honduras.

The survey was a nationwide probability sample using two-stage sampling in which the first-stage sample unit was *caserío* and the second-stage sample unit was household within the selected sample *caseríos*. The *caserío* was selected as the primary sampling unit for two reasons – first, because of the efficiency advantage afforded by two-stage sampling; and second, because much data useful for survey design were available for this sampling unit (especially from the GIS database). The sample design was marginally stratified to assure variation in explanatory variables (available for the survey design) believed to have a substantial effect on outcomes of interest. Because of this stratification, the selection probabilities for *caseríos* vary substantially. While the sample design was very good for estimation of impact at the household level (high precision for estimates of impact and high power for tests of hypotheses about impact), it was not so good for estimation of impact at the national level. The primary estimate of interest is the average treatment effect (i.e., effect of the Transportation Project) on a randomly selected household in the country. This estimate is simply the difference in means for the two counterfactual states, project or no-project. To estimate the mean impact for the country, it is necessary to estimate the mean change in travel times over all households in the entire country. Since the household survey is a nationwide probability sample, this effect is estimable, but, because of the high level of variation in the sample selection probabilities, it cannot be estimated with high precision. Hence, with respect to scope of inference, the household sample survey is better suited to estimation of project impact for a randomly selected household in Honduras, than for estimating the total impact of the project over all households in Honduras.

Stochastic Nature of Study Variables

In this evaluation, the following effects are considered fixed:

1. Transportation Project: All of the locations, times and characteristics of the road improvements.
2. Traffic Surveys: All of the locations, times and procedures of the traffic surveys. Note, in particular, that the road improvement (treatment) status, elevation variation, season, day of week, time of day, and weather were fixed.
3. Household Survey: The timing of the household survey rounds, and the locations of the higher-level sampling units (*caseríos* and households).

The random variables associated with the project, traffic surveys, and household survey are the following:

1. Transportation Project: There are no random variables associated with the Transportation Project. Every aspect is considered fixed.
2. Traffic Surveys: The vehicles observed, and the measurements made on them (i.e., vehicle type, speed, origin-destination).
3. Household Surveys: The within-household characteristics (i.e., the observations made at the two different survey times, on the same household). (For a fixed-effects analysis, a household must be observed in both survey rounds.)

Note that it is not necessary to consider all of the preceding effects as fixed. For example, the project could be considered fixed and the locations of the *caseríos* and households of the household survey could be considered random. The decision about which effects are considered fixed and which random has implications on the scope of inference of the evaluation study, and these implications are different for the different categories of fixed effects.

Unconditionally (i.e., prior to the selection and implementation of the Transportation Project and the (traffic and household) surveys of this evaluation), all of the factors listed above are random variables. With respect to considering an effect to be fixed, an issue to be considered is whether the goal of the evaluation is to make inferences about the populations (hypothetical or real) on which those random variables were observed, or to make inferences about the particular sample units that were selected, once they have been selected. This choice is affected both by which is the more useful evaluation product and by the precision of the estimates in the two situations (i.e., a random-effects estimator may be more desirable on technical grounds, but of such low precision that it is unlikely to detect impacts of anticipated magnitude).

With respect to considering the Transportation Project fixed, the implication for the evaluation is that the estimates of impact refer to this particular project. Although the project involves improvements to a selection of road segments, it is not the goal to estimate the average treatment effect of making improvements in the future to a randomly selected road segment. (This is not the goal since it is not possible to do it properly in this evaluation study – randomization was not used to assign road segments to treatment and it is not feasible to develop a selection model (the selection process is ill-defined, and even if it were not, the road sample sizes are too small). It is, of course, desired to assess the degree to which the impact associated with this particular Transportation Project might be generalized. In general, the impact estimates constructed in this evaluation would represent the impact expected for a similar project in similar circumstances. This is "logical" (deductive) inference, not "statistical" (inductive) inference. The term "circumstances" is broad, referring not only to the physical and economic characteristics of the project location, but also the manner in which the central and local governments determined which road segments would be improved. For example, if this project was implemented in areas in which it was expected to have the greatest impact, it would be unreasonable to conclude that the impact associated with this project would be similar to that of a project located in other places of the country.

With respect to the estimation of vehicle speeds from the traffic surveys, the fixed-effects assumption has little effect. The main product of the traffic surveys was a table of mean speeds for pickup trucks conditional on a number of road and environmental variables (road improvement status, elevation variation, season, day of week, time of day, weather). This estimation was done for all treatment roads and a matched sample of comparison roads. The impact estimates are conditional on the mean speeds, which in turn are conditional on the fixed environmental conditions and the fixed sample of (treatment and matching) roads. Were some of the conditioning variables considered random rather than fixed, the standard errors of impact estimates would be increased. The project scope would increase somewhat (to reflect variation in the conditioning variables), but the power of tests of significance would decrease. There is no evident rationale for doing this. The major feature of the scope of inference for the evaluation would still be the fact that the impact estimates refer to a particular project. Expanding the scope to that same project plus variation in some environmental variable would serve no evident purpose. In summary, restriction of the mean-speed estimates to fixed-effects estimators has little effect on the evaluation scope of inference.

With respect to the household sample survey, the assumption of fixed effects has little effect on the scope of inference. The reason for this is that the survey is a large, nationwide probability sample, designed to produce unbiased national estimates. (Both the *caserío* sample size and the household sample size are large.) Whether the analysis is restricted to the households of the sample, or intended to reflect the total national household population, the estimates are similar. The magnitudes of the effects differ for some outcomes of interest, depending on whether a fixed-effects or random-effects model is used. A significant conceptual difference between the fixed-effects and random-effects estimators is that the household-to-household variation and the *caserío*-to-*caserío* variation is included in the standard errors of estimates for the random-effects estimates, but not for the fixed-effects ("within household") estimates. The analysis of impact is based on the assumption of fixed effects for households and *caseríos*, but this assumption has little effect on the scope of inference for the evaluation. (In the course of the analysis, both fixed-effects and random-effects models were investigated. Differences between the fixed-effects and random-effects models may be tested with a Hausman test, which is a test of whether the unobserved time-invariant variables are correlated with the covariates (if these variables were to be considered as random variables). Even if impact is estimated under a fixed-effects assumption, the random-effects models are useful for investigating the effects of variables that are time-invariant within households but which have variation among households.)

(Under the fixed-effects assumption for the household survey data, the analysis uses data only from households surveyed in both survey rounds, and the estimates are based solely on the within-household variation. If there is no variation within households for a particular variable, then the effect associated with that variable cannot be estimated from the fixed-effects analysis. If a household is observed in only one survey round, it is omitted from the fixed-effects estimator.)

In summary, the various assumptions made about the surveyed populations, the sample designs, and the stochastic nature of study variables have an influence on the scope of inference of the study. The major factor limiting the scope of inference of the evaluation is the fact that the estimates of impact are for the particular Transportation Project, and not for some more general class of road-improvement projects. The primary impact estimate produced by the study is the average treatment effect of this particular project intervention on a randomly selected household in Honduras. It is *not* the average treatment effect (on a household) associated with applying the improvement to a similarly sized random sample of roads. Generalization of the results (assessment of external validity) must rest on consideration of the procedures used to select this particular project, and of its idiosyncratic physical and economic features. The impact observed in this study would be expected to be similar for similar road improvement projects implemented in similar contexts (i.e., selected in similar ways, and implemented in locations that are similar economically and physically). In the absence of randomized assignment to treatment, and in the absence of sufficient information and sample data to develop a credible selection model, this restriction in scope is considered necessary. Within this restricted scope of inference, the impact estimates are considered to have high internal validity, i.e., the impact estimates are considered to be accurate estimates of the effect of the Transportation Project on a randomly selected Honduran household. The average impact may differ depending on the characteristics of a household, such as distance from the project, but the analysis did not show this.

Summary of Assumptions and Limitations

With respect to the "macro-level" causal model and associated statistical model used to estimate impact, the assumptions underlying the estimates of impact are the following:

1. The stable unit treatment value assumption (SUTVA, no-macro-effects assumption, partial equilibrium assumption) is made. Among other things, this assumption implies that the project is not so large that it changes the basic relationship of outcomes of interest to travel time.

2. The estimates of travel time are based on the GIS model of the Honduran road network. This model includes all official Honduran roads, as well as others. The GIS model is highly detailed, and considered to be up-to-date and of high accuracy. The quality of the GIS model is not considered to be a limitation on the quality of the evaluation.

3. The impact estimates are conditional on completion and maintenance of the Transportation Project as finally configured. Under this assumption, the impact estimates refer to this particular project, not to the mean impacts associated with a conceptually infinite population of similar projects in other locations or at other times. The estimate of impact is the average treatment effect of this particular project on a randomly selected household in Honduras, not the average treatment effect associated with improving a randomly selected eligible-for-treatment road segment.

4. Although the unit of treatment was the road segment, the unit of analysis was the household. In this case there could be two types of unobserved (hidden) variables, which may introduce biases into the estimates of the model parameters. First, there may be unobserved variables that are time invariant. In a fixed-effects model, these, however, "drop out" for the two-round panel specification. Second, there could be unobserved variables that are not time invariant (though no such variables were identified). It is assumed that such variables, if any exist, are uncorrelated with the explanatory variables.

5. The continuous-treatment-value impact estimates are conditional on the travel speed table derived from the traffic surveys. This table presents estimates of the average speed of a pickup truck over a route by route type (primary, secondary, rural), elevation variation and program intervention status (improved or not improved). The speeds are conditional on the season of the year, day of the week, time of day, and weather conditions under which the traffic surveys were conducted. The speeds are used to calculate travel times to places of interest (using the GIS model). These travel times are used (as described above) to estimate impact. A number of travel times were available (i.e., travel times to various places of interest from a caserío such as to a municipal or department capital, or the nearest town having population 1,000 or more. Attention focused on the one that had the highest relationship to outcomes of interest (i.e., travel time to the nearest town of population 1,000 or more). The continuous-treatment-value estimates are also

conditional on manifestation of long-term benefits, as estimated from the partial-treatment-effects model.

Other more specific model assumptions are listed for particular estimation equations in the detailed analysis presented in Annex 1.

The limitations of the evaluation are:

1. The impact estimates constructed in this analysis pertain solely to the Transportation Project as it was finally configured, and when and where it was implemented, not to similar projects in other settings (locations or times). The estimated standard errors reflect sampling variation associated with estimation of characteristics of this project, and do not include variation associated with hypothetical variations in the project location or time, or the higher-level sample units of the household sample survey (*caseríos* and households) (as would be the case for a "random effects" approach). It is expected, however, that similar results would be expected for similar projects in similar settings.

2. More of the resources for primary data collection for this evaluation were invested in the household sample survey, and less in the traffic surveys. The traffic surveys were limited in scope and in sample size. The design of the traffic surveys rested on judgment, not on the methods of statistical sample survey. The estimates of speed were conditioned only on elevation variation, and not on other road variables that may have an important effect on average speed, such as number of lanes, access, road roughness, and curviness. The vehicle speed on which travel times were based for analysis of impact was a pickup truck. The traffic survey data were used as input to the GIS road-network model for estimation of travel times, not for estimation of direct impact (such as via a double-difference estimator applied to the traffic data). The travel-time estimates pertain to a particular vehicle type, season, day of week, time of day, and local weather conditions. The level of correlation among the available travel times was relatively high, so that the limited scope of the traffic surveys is not considered to be a major weakness. In the future, however, if this approach is used, it is recommended that a substantially greater level of resources be allocated to the traffic surveys (at least comparable to that of the household surveys).

3. The study focuses on a variety of indicators to assess impact of the road improvements. Some of them are direct effects, such as travel times to points of interest (education and health facilities), and others are indirect effects such as income and employment. Additional direct effects might have been of interest, such as number of trips or length of trips.

4. As described above, an assumption of the analysis is that are no time-varying unobserved variables that are correlated with any of the explanatory variables, which in this case refers to the treatment variables. While we have investigated and confirmed the soundness of this assumption to the extent feasible, it is not possible to conclusively rule out the possibility that such variables exist and may be influencing the results. For example, if economic conditions such as labor market characteristics or the level of private investment are changing in systematically different ways that are correlated with travel time, our approach may mistake the impact of the road improvement for the impact

of changes in these conditions. This possibility must be considered a limitation of the analysis.

C.3 CHALLENGES ENCOUNTERED IN IMPLEMENTING THE EVALUATION DESIGN

The main challenge encountered in implementing the Transportation Project evaluation pertained to repeated delays in scheduled road construction activities. As indicated in Table 1, the final road improvement activities were not completed until the end of 2012. As a result, follow-on data collection for the household survey, which took place in February/March 2011, occurred either before or very soon after road construction terminated on some project roads, leaving barely any time for impacts to manifest themselves. In fact, construction work could have caused delays that negatively affected travel times across the road network.

The issue of delays in completion of project activities was addressed by constructing impact estimates that were conditional on project completion and maintenance and on a partial-treatment-effect model. This was feasible to do by identifying a causal model and an associated statistical model that permits estimation of impact conditional on travel times estimated from the GIS model, and these travel times could be specified to correspond to project completion and maintenance. Without the GIS travel-time model, construction of these conditional estimates would not have been possible. (In the course of the analysis, impact estimates were constructed based on binary treatment indicators / zones of influence. These estimates were based solely on questionnaire variables, not on the GIS-model travel times. These estimates of impact may be affected by the delay in project implementation and the time lag for manifestation of higher-order benefits The CTV estimates, based on the partial-treatment-effect model, are not affected by delays in manifestation of higher-level benefits.

D. HOUSEHOLD SURVEY

We used a national household survey as the primary data source for the Transportation Project impact evaluation. As mentioned, the design for the household survey was an “analytical” survey design, where the goal was to obtain a sample that is a good basis for estimating parameters of the analytical model representing the process under study. Briefly, this means that the sample contained (to the extent feasible) substantial variation in the explanatory variables of the analytical model, and low correlation among them.

Target population. The target population for the household survey was the population of all households in Honduras at the beginning and end of the project. We used a sample frame constructed for the most recent National Census. Because that Census was conducted a number of years ago (2001), the survey field procedures included procedures to ensure that all current households are subject to sampling (e.g., use of systematic sampling over all of the current households of the entire Census segment). The unit of analysis was the household.

Variables of interest. The variables of primary interest in the survey were household income and employment, but data were collected for other variables, including: access to and use of educational and health facilities; travel time and travel costs to these facilities and other points of interest; key socioeconomic and demographic indicators; and household consumption and expenditures.

Sample design and stratification. We used a panel survey with a stratified two-stage sample design in which the first-stage sample units (primary sample units) are *caseríos*. *Caseríos* were selected for use as the first-stage sample unit not only because they (like Census segments) are an efficient size for sampling, but also because a substantial amount of GIS data are available for them. Two-stage sampling is widely used in surveys of households, because it affords a high return of precision for sampling effort expended, and because it is often relatively easy to implement because much data are available for potential primary sample units.

The design of an analytical survey seeks variation in variables that have a substantial effect on impact (in this case, mainly travel-time related variables, but a number of other explanatory variables as well). We achieved this variation through stratification of the sample by changes in travel time to various points of interest (as estimated using a GIS model), urban/rural status and a number of geophysical variables from GIS data sources.

The panel survey design, where every effort is made in the second survey round to re-interview the same households interviewed in the initial survey, promotes local control – the precision of estimates of change are substantially higher for a panel or longitudinal approach than for a “repeated cross-sectional” approach of interviewing an independent sample of households the second time. Based on previous experience in Honduras, we anticipated that about 10-20 percent of the households would be different from those interviewed at the beginning of this project. In these few cases, the household currently occupying the dwelling were interviewed on the rationale that its road-related behavior is probably similar to those of the previous occupants, and that the accuracy obtained by including it is probably greater than if it were dropped from the survey. Additionally, we included questions that attempt to identify why the previous occupants moved, and obtained location information for their new residence. If they had moved within the same Census segment or *caserío*, interviewers attempted to locate and reinterview them. To keep survey costs down and stay within time constraints, no attempt was made to locate and interview households that no longer live in the primary sampling unit of the sample.

Annex 3 contains a detailed description of the sample design and sample selection methodology for the *caserío* sample. Here follows some additional description of the sample design and selection process.

In order to construct national-level estimates of the relationship of outcome variables of interest to GIS model travel times, both the GIS road network model and the household survey had to be national in scope. For the household survey, all *caseríos* of the nation were subject to sampling (i.e., had a nonzero probability of selection). The sample was stratified by many variables, the most important of which were anticipated change in travel times (to places of interest from each *caserío*) associated with the road improvements of the Transportation Project. The stratification procedure used was marginal stratification, not ordinary stratification. In marginal stratification, the stratification is implemented for the *marginal* distribution of each design variable, rather than on the *joint* distribution. (Ordinary stratification is usually implemented on the joint distribution, and is limited to a few variables of stratification because of the "curse of dimensionality" (in this case, the proliferation of "cells" associated with cross-stratification). It is useful for precision improvement for a small number of estimates (such as population and subpopulation means), but not for promoting precision for estimation of relationships to a large number of explanatory variables related to outcomes of interest.) With ordinary stratification, the sample size is controlled exactly in each stratum. With marginal stratification, the probabilities of selection of each primary sample unit (*caserío*) are adjusted so that the *expected* sample size is close to desired, for each (marginal) stratum.

Sample size determination. The *caserío* sample size and the number of households to be selected from each sample *caserío* were determined by a detailed statistical power analysis. Based on this analysis, we decided to select a sample of 20 households from each of 100 sample *caseríos*. That is, the sample design called for a sample of 2,000 households located in 100 *caseríos*, in each of two survey rounds (before and after the program intervention). The same households were to be interviewed in both survey rounds. Because many *caseríos* were small, the sample size of 2,000 was not realized. The final sample size for the baseline sample was 1,600 households in 116 *caseríos* (some additional *caseríos* were added to a stratum near project roads). Of these households, 1,408 households, or about 88 percent, yielded completed interviews in the second round of data collection. Annex 4 presents a summary of the statistical power analysis used to determine household and *caserío* sample sizes.

Survey instrument. The questionnaire was approximately two hours in length. The head of the household, if available, was the primary survey respondent, though often another individual, usually the spouse or mother, would provide information on household consumption and expenditures if that person was better informed. After two attempts were made to locate and interview the head of the household, a proxy adult within the household was selected to complete the survey. If a household moved since the baseline, interviewers were instructed to interview the new household resident in the structure, if possible. If a household refused or if a key respondent could not be identified, it was not replaced in the sample.

Categories of data items included in the survey instrument are presented in Table 2. The major data elements of the analysis are listed in Section G.

Table 2: Household Survey Elements

Key Elements	Item Description
1. Labor and Income	Detailed information on employment activity and household income and their sources
2. Consumption and expenditures	Retrospective family consumption and expenditure measures (week, month, quarter and year). Health and education measures.
3. Travel information	Travel times, cost, access to major employment, highways, markets, school, clinics, etc. Collect names and locations of high schools, health centers, hospitals and markets visited.
4. Micro enterprises and agriculture	Involvement in micro enterprises including the informal sector. Agricultural practices and products, changes, and additional items for program farmers.
5. Housing costs and prices	Land value items including "How much did you pay for your home/land?" "If you were to sell this land today, how much do you think a buyer would be willing to pay you for it?"
6. Loans and credit	Sources and uses of credit, value of loans, etc.
7. SES/demographics	Basic household demographic information. Relying upon many standard Census and national household survey items.
8. Perceptions of MCA-Program[8] elements	Qualitative questions on impact of program activities, negative consequences, etc.

In the first phase of questionnaire development, NORC conducted a systematic review of existing questionnaires that collected data on subject areas similar to those proposed for the Transportation Project survey. Preference was given to surveys that had been applied and field-tested in the region. The team determined that the ENCOVI (*Encuesta de Condiciones de Vida*), a national survey of the conditions of life of Honduran households, was the best source of existing items for the survey because it focused on many of the same content areas that we proposed to include. Furthermore, our local data collector, the *Instituto Nacional de Estadística* (INE), had experience with ENCOVI, having just fielded the survey nationally in Honduras in 2006.

While a significant percentage of the items incorporated into the final household questionnaire were taken directly from previous surveys, many items, particularly those on transportation and travel times, household consumption, and agricultural production, were modified or expanded to collect the more detailed data deemed necessary to construct particular impact indicators. Response categories were modified and adjustments were made to ensure adherence to local norms. INE assisted with many adjustments to the "language" and "terminology" used in instructions, items, and response categories to ensure that we were using appropriate terms and a level of language that was accessible to respondents with lower levels of education.

For the follow-on (endline) survey instrument, we included additional questions on the names and locations of key points of interest (schools, clinics, and markets) since we determined that this would facilitate the analysis of travel times using GIS.

[8] These questions on the Transportation Project were included only in the second round of the survey.

The questionnaire was pilot-tested and modified based on the pilot test. INE collected baseline in August 2008 and endline data in March 2011. The data collection for the baseline, as well as for the endline, was completed during a two-week period.

E. GEOGRAPHIC INFORMATION SYSTEM DATA

Honduras possesses exceptionally comprehensive and high-quality GIS data compared to most developing countries. GIS data generated in multiple Honduran government agencies – and also collected by private firms and in non-governmental research studies – have already been pooled centrally into a separate, stand-alone entity backed by a 12-year World Bank financial support program. This entity –*Programa de Administración de Tierras de Honduras ("PATH")* Digital Land Information System – has pooled into GIS databases, stored on powerful servers extensive and comprehensive Honduran GIS datasets including:

- National and local political boundaries (*departamentos*, *municípios*, *aldeas* and *caseríos*);
- National Census and survey data already linked to *caserío, aldea* and *munícipio* locations (including aggregated data by political district);
- Extensive primary, secondary and rural road network data. Table 4 below presents details about the availability and sources, both PATH and other, of GIS data on the Honduran road network.
- Elevation;
- Hydrology (rivers, streams, lakes, watersheds, etc.);
- Extensive collections of high-resolution digital orthophotos and geo-rectified satellite imagery;
- Derived land cover;
- Extensive GIS datasets on climate, temperature, rainfall variation, humidity, and other factors. (PATH climate data includes sub-annual (e.g. monthly and daily) data.) ;
- For a subset of major cities, high-resolution GIS data providing the digital polygons of buildings, detailed street networks, electric and sewer networks, and household locations (linked to Census data);

GIS specialists from the NORC team toured and evaluated the PATH GIS archive, and secured an agreement for full GIS data sharing and usage. Crucial to our evaluation of the extent and quality of these data was the evaluation of the GIS national road network, including primary, secondary and (most importantly and typically difficult to obtain) rural-road networks, since our Transportation Program evaluation framework is premised on a reasonably complete GIS Honduran road network. We concluded that the GIS road network for Honduras was almost complete, including more than 70 percent of the rural-road network, and therefore, sufficient for our impact analysis.

Table 3: Honduran GIS Road Network Data: Sources and Extent

Road Data Description	Data Format and Extent	Data Sources
Primary Roads	Complete Honduran primary road network coverage available in GIS format	PATH (Programa de Administración de Tierras de Honduras); SOPTRAVI (Ministry of Transportation)
Secondary Roads	Complete Honduran secondary road network coverage available in GIS format	PATH (Programa de Administración de Tierras de Honduras); SOPTRAVI (Ministry of Transportation)
Rural/Tertiary Roads	Approximately 70-80 percent of rural road GIS coverage available, obtained from GIS data, maps, private sector projects, etc.	CIES (Centro de Investigaciones Económicas y Sociales de la Universidad Jose Cecilio del Valle) and COHEP (Consejo Hondureño de la Empresa Privada) CIAT (Centro Internacional de Agricultura Tropical)
Location of Improved Primary, Secondary and Rural road segments	Improved road segments can be labeled in the GIS database, for display or analysis	MCA, WORLD BANK, Road Coordination Committee, SOPTRAVI
Periodic updates over time for changes to Primary, Secondary and Rural road locations or improvements, and to locations of improved road segment	Data can be entered periodically into the GIS database, and then display or analysis modified or re-run	PATH, SOPTRAVI, CIES, COHEP, Road Coordination Committee, SOPTRAVI, WORLD BANK, MCA, etc.

NOTE: NORC personnel in coordination with ESA Consultores conducted a field visit to Honduras in August, 2007, and identified the current extent and availability of Honduran GIS road network data.

F. TRAFFIC SURVEY DATA

F.1 THE ROLE OF THE TRAFFIC SURVEY DATA IN THE EVALUATION

The conceptual framework for the evaluation included two sources of data – the household survey questionnaire and the traffic surveys. It was planned to invest more resources in the household survey than in the traffic surveys. The household survey was a large, national-level probability sample. Every household in the country was subject to sampling, with a known, nonzero probability of selection. Statistical power analysis was used to construct the design of the survey and to determine sample size. The traffic surveys were conducted on project roads and on matched sample of comparison roads under controlled conditions. The purpose of the traffic surveys was to provide estimates of the average travel time of vehicles over roads with and without improvement (treatment), by road type (primary, secondary, rural) and road characteristic (elevation variation and possibly other characteristics). Relative to this evaluation, the product of the traffic survey was a table that showed mean speed of a pickup truck over each type of road and three levels of elevation variation, with and without treatment. This mean speed table was input to the GIS model to estimate travel times (for pickup trucks) from sample *caseríos* to points of interest (e.g., the nearest town of population 1,000 or more). (The traffic surveys were also used to support an analysis of the economic rate of return (ERR) of road improvements, using the Highway Development and Management Model HDM-4 road planning and analysis computer software package. That analysis is documented in a separate report.)

The traffic survey sample consisted of the project roads and a judgmentally matched sample of comparison roads. The purpose of the matched comparison sample was to facilitate the construction of a "potential outcomes" estimate of the effect of the road improvements on mean vehicle speed. (Matching was used to increase the precision of the mean speed estimates, by removing sources of variation other than the conditioning variables of interest (vehicle type, road type, elevation variation and treatment status). The surveys were conducted under controlled conditions (season, day of week, time of day, weather), also to increase precision.) Although the speed table is a "parametric" table showing mean speed as a function of vehicle type, road type, and elevation variation, the treatment roads were selected from this particular Transportation Project, and so the speed estimates are specific to this project. The speed estimates are "fixed effect" estimates, i.e., they are means associated with this particular project, not with a hypothetical infinite population of similar projects in various locations. The traffic surveys were conducted under special conditions (season of the year, day of week, time of day, weather), and the speed estimates are conditional on those special conditions. They are not estimates of mean speed averaged over seasons, times of day or weather. They are intended to provide summary indicators of the effect of the road improvement project on project roads, and enable (using the GIS road network model) the construction of *conditional* estimates of mean travel times from *caserío* centroids to places of interest, under very specific conditions. The speed estimates are "fixed effect" estimates, that pertain to the particular set of treatment and comparison roads (and other conditions, such as season and time of day) selected for the traffic surveys.

Note that, in general, a traffic survey *can be used* to provide estimates of project impact, e.g., using a double-difference or regression-adjusted estimator, if it is properly designed for that purpose and has a sufficiently large sample size. If this had been done for the present project, it would have been done under a fixed-effects model assumption (since the project roads were selected by judgment, not randomization). This use of the traffic surveys was not planned at the start of the project – no statistical power analysis or other survey design effort was allocated to the traffic surveys to assure useful results for use in construction of direct estimates of impact, as was done for the household survey; and time and labor sufficient to conduct a competent statistical analysis of impact from the traffic-survey data, in addition to that done for the household-survey data, was not allocated. Construction of direct estimates of impact was not the goal of the traffic surveys conducted in this evaluation – they were used to provide estimates of mean vehicle speed under controlled conditions, for use in construction of estimates of impact conditional on travel times, from the household survey data. (In future roads-project evaluations, this decision should be reconsidered. In view of the fact that mean vehicle speed is a sensitive and strong indicator of the effect of a road improvement, and the various indicators derived from the household survey data are not, consideration should be given, in the planning of future roads-project evaluations, of basing the evaluation on estimates of outcome indicators from statistically designed traffic surveys, rather than on impacts estimated from household sample surveys. The cost of the traffic surveys was somewhat less than, but comparable in magnitude to, the cost of the household survey.)

A detailed description of the process used to estimate travel times from estimated speeds is presented in Annex 2.[9]

We used the GIS-model travel time estimates in two ways for this evaluation: first, to assist sample design, by providing variables that were used as variables of stratification; and second, in the analysis, by enabling the construction of impact estimates that take into account the changes in travel times (for every official road segment in the country) near the beginning and end of the program intervention. (For the survey design, subjective estimates were used for vehicle speeds before and after the project intervention. For the data analysis, the mean vehicle speeds obtained from the traffic surveys were used.)

F.2 TRAFFIC SURVEY DATA COLLECTION METHODOLOGY

Data collection took place on 65 road segments, 29 of which were MCA - Honduras project roads, and 37 of which were 37 comparison roads that were matched to the treatment segments. The comparison segments were selected in order to satisfy a set of criteria specified and designed to ensure that the comparison roads would capture network-wide traffic variation across Honduras. The selected comparison roads represent widespread spatial variation throughout Honduras. These roads were chosen because they also serve as “connectors” in the overall road network, connecting major sections of the road network to other sections, or providing connections between major populated areas of Honduras.

The sample of road segments for the baselines and endlines of the traffic survey are shown in Table 4. The traffic survey was conducted on all MCA - Honduras road improvement sections (two CA-5 primary sections, three secondary road-improvement sections, and 24 rural unpaved-road sections), as well as on 37 comparison roads (six primary comparison segments, six secondary comparison segments, and 25 rural comparison segments).

[9] The Geographic Information System (GIS) Model of the Honduras road network is based on data provided by the Sistema Nacional de Información Territorial de la República de Honduras (SINIT) - http://www.sinit.hn/sinit/

The traffic survey had three separate data collection components: a survey of traffic volumes, broken down by vehicle type; a survey of travel speeds, broken down by vehicle type; and an origin-destination survey. The traffic survey was conducted in 3 rounds: a pre-improvement round in 2009, a mid-line round in 2010, and a post-improvement round in 2011.

Table 4: Sample of Roads for Traffic Survey

Road Segment Type	MCA - Honduras Program Roads	Non-program roads
Primary Roads	2 (total of 4 data collection sites)	6 (one data collection site per road)
Secondary Roads	3 (one data collection site per road)	6 (one data collection site per road)
Rural Roads	24 (one data collection site per road)	25 (one data collection site per road)

Volumetric traffic counts and origin-destination data were collected for all road segments, while speed data were recorded only for a subset of these roads: all primary program and non-program roads; three secondary program roads and three secondary comparison roads; and seven rural program roads and seven rural comparison roads. (Note that not all of the data collected in the traffic surveys were used in the evaluation documented in this report. Some were used in support of the ERR study. The data used for this analysis was the mean vehicle speed for pickup trucks, under various conditions.)

Data collection instruments included a form to gather volumetric data, a one-page questionnaire for the origin-destination survey, and a form to record vehicle speed data. The data collection form for the traffic counts (volumetric data) was designed to obtain the traffic counts (total number of vehicles) classified by vehicles type – bicycles, motorcycle, pick-up trucks, standard cars (e.g., sedans), buses, trucks of two and three axles and large containers trucks – every hour, over a 12-hour period.

The methodology used for each type of data collection is described below:

- **Traffic Counts.** On all primary road segments, both program and non-program, diurnal counts were registered at each station during seven consecutive days during a 12 hour period (6 a.m. to 6 p.m.). Nocturnal counts were recorded at each station between 6 p.m. and 6 a.m. on Wednesday, Friday and Sunday, with traffic counts spanning a 24-hour period on these three days. Similarly, on secondary and rural roads, volumetric data were collected during a 12-hour period (6 a.m. to 6 p.m.) during the day for a seven-day period; however, no nocturnal counts were collected on these roads on account of low traffic volume. Traffic counts were recorded for both directions on all roads. Counts were recorded for bicycles, motorcycle, pick-up, passenger vehicles, buses, trucks of two and three axles, and container trucks, as appropriate for each road type.
- **Origin-Destination Surveys.** There were two stations for collecting origin and destination (O-D) data on the CA-5 (primary) project roads, and one intercept point on each of the six non-program roads.

  For primary roads, enumerators stopped and interviewed at least six vehicles per hour, attempting to interview at least one of each of the seven aforementioned vehicle types each hour, if possible. The methodology to be used for conducting the O-D surveys on the rural and secondary roads was similar to that used for primary roads. However, given the lower traffic volumes, the goal was to interview 3-6 vehicles per hour.

- **Vehicle Speed Data**. For primary road segments, the enumerators tagged and closely followed different vehicle types over the length of the road segments, recording the time taken to traverse the segment during both peak and non-peak hours. The vehicle followed cars, buses and trucks during a 12-hour period for 7 days. The tagging process was continuous during the 12-hour period. Vehicle speed was calculated by dividing the travel time by length of road segment. A different methodology was used to collect speed data on secondary and rural roads. An enumeration vehicle and driver drove the length of selected secondary and tertiary/rural road segments at a “safe maximum” speed: that is, a speed that was considered to be the maximum possible on that segment while still maintaining complete vehicle safety, both to the vehicle being driven and to any other vehicles, pedestrians, travelers or bystanders on the road. The total driving time and distance were recorded for each segment driven. Vehicle speed measures were taken twice during peak hours and twice during off-peak hours on these road segments, once in each direction. These measurements were obtained only during daytime (6:00 a.m. - 6:00 p.m.) hours.

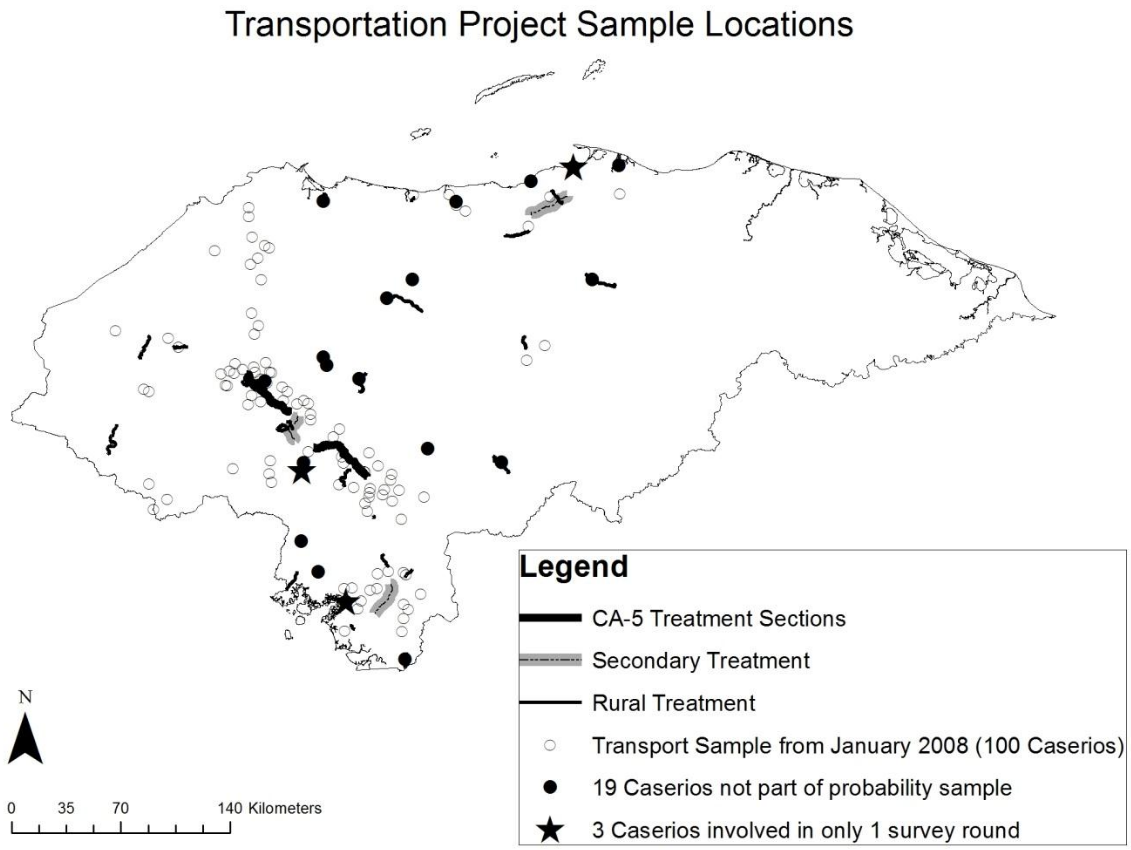
Here follows the table (Table 8) of mean travel speeds used to estimate travel times for the analysis. This table is the same as Table A.18 of Annex 2. The table presents the mean speed for pickup trucks, under various conditions (road type, elevation variation, treatment status). The reasons why pickup truck speeds were used were:

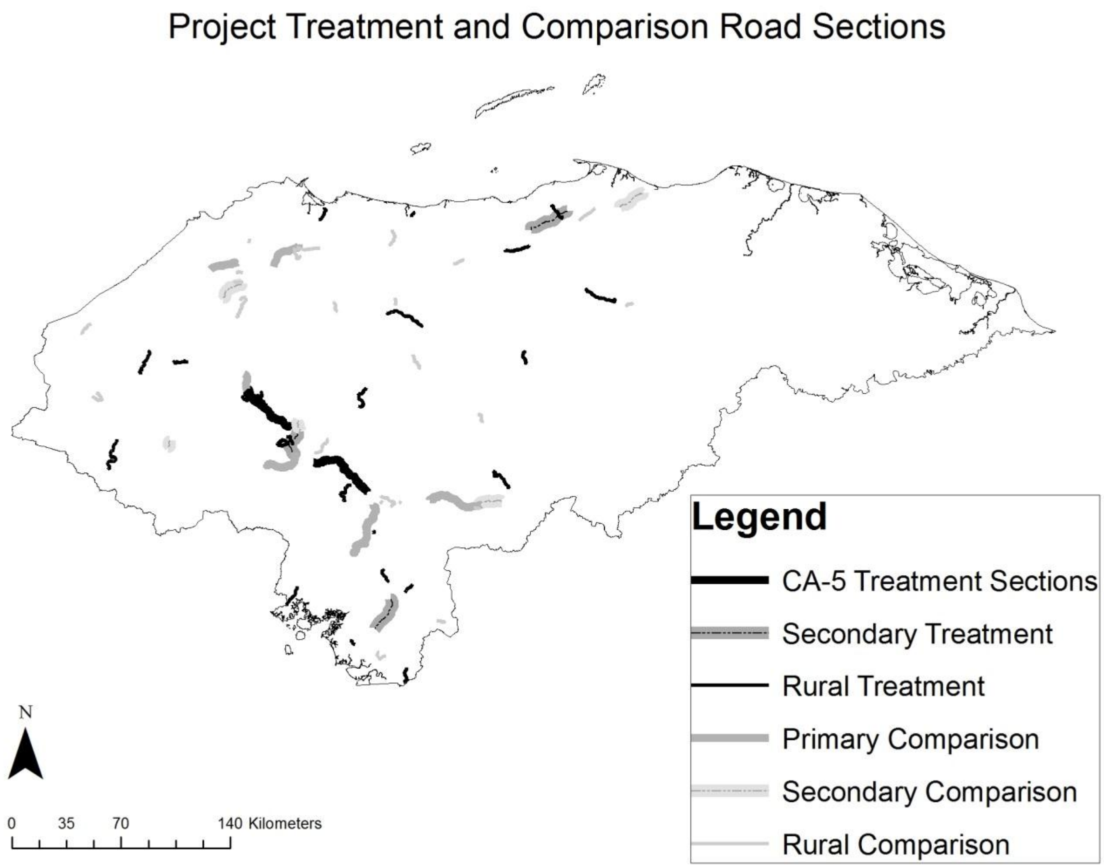
1. They were the most common vehicle type on all road types.
2. Variation in vehicle speeds by vehicle type was measured on the primary roads, but not on the secondary or rural roads. For the secondary and rural roads, a pickup truck was used by the traffic survey company to drive the length of the segment and record the average speed. So for the secondary and rural roads we had only pickup truck speeds measured.

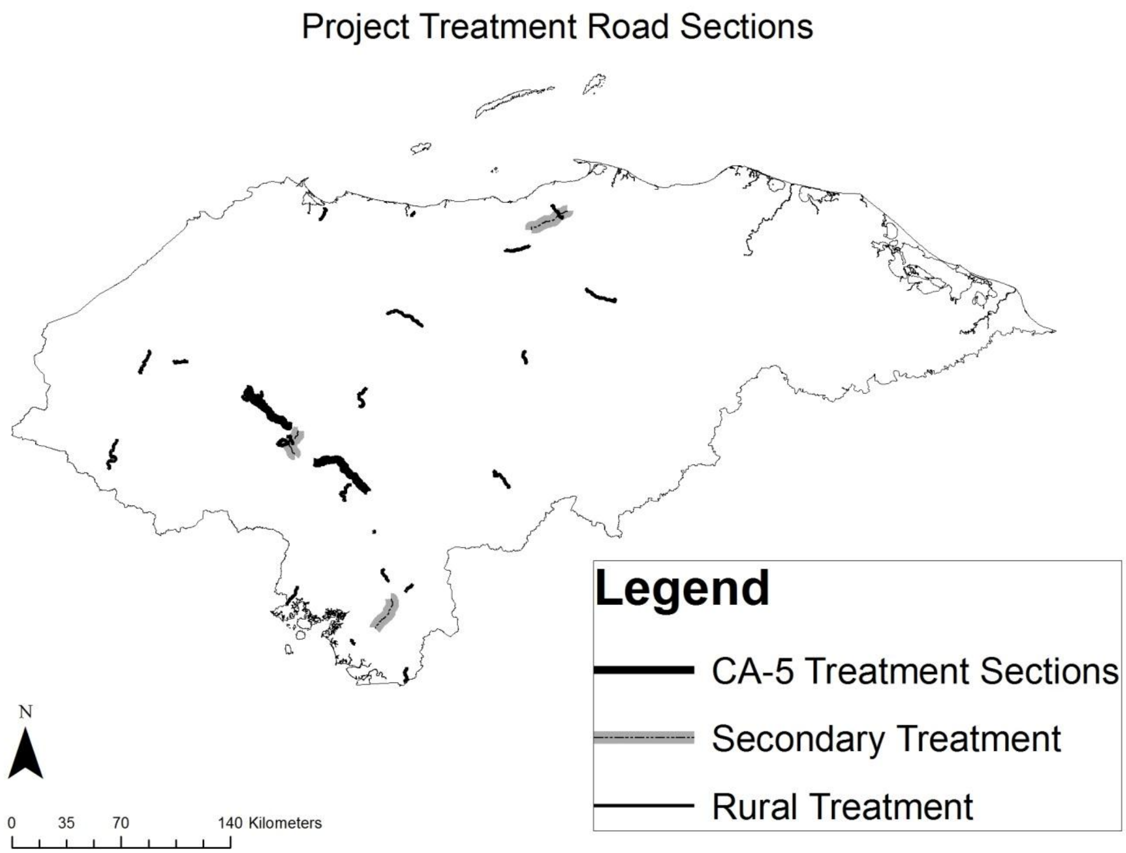
Table 5: Measured Honduras Pickup-Truck Mean Travel Speeds

Measured Pick-Up Truck Travel Speeds Honduras Road Network	Pre-Improvement (Round 1: 2009)	Post-Improvement (Round 3: 2011)
Primary Treatment Roads, Greatest Elevation Variation	**53.2**	**54.7**
Primary Comparison Roads, Greatest Elevation Variation	**50.4**	**50.9**
Primary Treatment Roads, Some Elevation Variation	**61.2**	**62.4**
Primary Comparison Roads, Some Elevation Variation	**57.6**	**58.1**
Primary Treatment Roads, Low Elevation Variation	**67.9**	**69.7**
Primary Comparison Roads, Low Elevation Variation	**64.2**	**64.6**
Secondary Treatment Roads, Greatest Elevation Variation	**41.2**	**70.3**
Secondary Comparison Roads, Greatest Elevation Variation	**54.3**	**55.2**
Secondary Treatment Roads, Some Elevation Variation	**44.2**	**74.1**
Secondary Comparison Roads, Some Elevation Variation	**58.9**	**59.6**
Secondary Treatment Roads, Low Elevation Variation	**48.1**	**77.5**
Secondary Comparison Roads, Low Elevation Variation	**62.7**	**63.1**
Rural Treatment Roads, Greatest Elevation Variation	**23.7**	**34.2**
Rural Comparison Roads, Greatest Elevation Variation	**18.3**	**18.7**
Rural Treatment Roads, Some Elevation Variation	**30.2**	**39.9**
Rural Comparison Roads, Some Elevation Variation	**20.4**	**20.6**
Rural Treatment Roads, Low Elevation Variation	**33.8**	**43.5**
Rural Comparison Roads, Low Elevation Variation	**23.1**	**23.4**

Here follow several maps showing the locations of the project roads, the comparison roads, and the *caseríos* of the household sample survey.


Transportation Project Sample Locations
Legend
CA-5 Treatment Sections
Secondary Treatment
Rural Treatment
Transport Sample from January 2008 (100 Caserios)
19 Caserios not part of probability sample
3 Caserios involved in only 1 survey round
N
0
35
70
140 Kilometers


Project Treatment and Comparison Road Sections
Legend
CA-5 Treatment Sections
Secondary Treatment
Rural Treatment
Primary Comparison
Secondary Comparison
Rural Comparison
N
0
35
70
140 Kilometers


Project Treatment Road Sections
Legend
CA-5 Treatment Sections
Secondary Treatment
Rural Treatment
N
0 35 70 140 Kilometers

G. PROGRAM IMPACT: SUMMARY OF RESULTS

This section summarizes the results of the impact analysis of the Transportation Project. A detailed description of the analysis is presented in Annex 1.

The primary data source for the impact analysis was 3,008 completed interview questionnaires, of which 1,600 were from the first survey round or baseline ("Round 0") and 1,408 were from the second survey round ("Round 1"). Analysis of the data was facilitated through use of the Stata statistical program package. We conducted the data analysis at the household level. Data below the household level, such as for individual household members, items of consumption or specific crops, were aggregated to the household level, and the unit of analysis was the household.

G.1 OUTCOME VARIABLES

The primary objective of this evaluation was to assess the impact of the program in affecting income and employment, access to places of interest, impact on school attendance, and use of health care services. To this end, the evaluation focused on the following categories of indicators.

Indicators Related to Income and Employment

Income Generated through Employment *(monthly amounts)*

— Income from labor-market employment in the agricultural sector (IncEmpAg)
— Income from labor-market employment in non-agricultural sectors (IncEmpNonAg)
— Income from labor-market employment = IncEmpAg + IncEmpNonAg (IncEmp)

Income Generated through Production and Sale of Agricultural Crops

For basic grains (BG) (annual amounts):
— Income from basic grains (including used for own consumption) (IncBG)
— Expenses for inputs for basic grains (FactorBG)
— Transportation expenses for basic grains (TranspBG)
— Other costs for basic grains (OthCostBG)
— Labor expense for basic grains (measure of employment associated with BG) (LabExpBG)
— Total expenses, basic grains (ExpBG) = FactorBG + TranspBG + OthCostBG + LabExpBG
— Net income from basic grains (NetBG) = IncBG – ExpBG

For other crops (OC) – horticultural crops (annual amounts):
— Income from other crops (including used for own consumption) (IncOC)
— Expenses for inputs for other crops (FactorOC)
— Transportation expense for other crops (TranspOC)
— Other costs for other crops (OthCostOC)
— Labor expense for other crops (measure of employment associated with OC) (LabExpOC)

- Total expenses, other crops (ExpOC) = FactorOC + TranspOC + OthCostOC + LabExpOC
- Net income from other crops (NetOC) = IncOC – ExpOC

Overall Household Expenditure and Income (from all sources)

- Total household expenditures (TotHHExp) (monthly amount)
- Net household income (NetHHInc) = NetBG + NetOC + IncTotalHH*12 (annualized amount), where IncTotalHH = monthly non-agricultural household income from all sources, IncEmp + Inc Remit + IncOther

Indicators Related to Travel Costs and Travel Times

- Cost (in lempiras) to school (CostToSchool)
- Time (in minutes) to school (TimeToSchool)
- Cost to college (CostToCollege)
- Time to college (TimeToCollege)
- Cost to hospital (CostToHospital)
- Time to hospital (TimeToHospital)
- Cost to health center (CostToHealthCtr)
- Time to health center (TimeToHealthCtr)
- Cost to market (CostToMarket)
- Time to market (TimeToMarket)
- Cost to pulperia (CostToPulp)
- Time to pulperia (TimeToPulp)
- Time to Tegucigalpa (TimeToTegus)
- Time to San Pedro Sula (TimeToSPS)
- Time to departmental capital (TimeToDepCap)
- Time to municipal capital (TimeToMunCap)

Indicators Related to School Attendance

- Total number of children aged 7-12 attending school (ChldInSch712)
- Total number of children aged13-18 attending school (ChldInSch1318)

Indicators Related to Use of Health Services

- Total number of visits in last 30 days to hospital by all family members (VisHospital)
- Total number of visits in last 30 days to private health centers (VisPrHlthCtr)
- Total number of visits in last 30 days to public clinics (VisPubClinic)
- Total number of visits in last 30 days to non-professional health-care providers (VisNonProf)
- Total number of visits in last 30 days to pharmacy (VisPharm)

Indicators Related to Employment

- Total number of household members who worked the previous week (WrkdPrevWk)

G.2 IMPACT ESTIMATORS (FORMULAS)

This section summarizes the procedures used to estimate impact. The description here is more specific than the general description presented earlier. The four steps of the process are as follows.

Step 1. Estimation of Travel Times.

Travel times to places of interest from any caserío in the nation are estimated from a table that shows mean pickup speed as a function of road type (primary, secondary, rural), elevation variation, and project intervention (treatment) status (improved or not improved). The mean-speed table was estimated from the traffic surveys. It is presented as Table 5, above.

Step 2: Estimation of the Outcome (PTE) Model.

Estimate the relationship of outcome to travel-time-related variables. This is the partial-treatment-effect model

$$E\ y_t = \mathbf{x}'_t\boldsymbol{\beta}.$$

A general description of the general linear statistical model used in the analysis was presented earlier. Below we present a description of the final model we used as a basis for impact estimation. We used the GIS model to construct estimates of travel times from each *caserío* to ten points of interest. These travel times were highly correlated, and it was not productive to include all of them as treatment variables in the response models. Instead, what worked best was to include a single travel-time variable, namely, the travel time to the nearest town of population 1,000 or more, town1000tt.

The model used as a basis for estimating the relationship of outcome to explanatory variables (treatment variable, design variables, and covariates, if any) is:

$$y_t = \beta_1\ \mathrm{trt} + \beta_2\ \mathrm{Rtrt} + \beta_3\ \mathrm{Round} + \beta_0 + e_t,$$

where

y_t denotes an output variable;

trt is the treatment variable, equal to town1000tt = mean travel time in minutes to the nearest town of population 1,000 or more (as estimated by the GIS travel-time model);

Round = survey round (0 or 1);

Rtrt = RoundStd*trtStd, where RoundStd and trtStd denote the demeaned Round and trt variables;

e_t denotes the model error term;

and the β's are the model parameters (regression coefficients).

The β's are estimated as described earlier, with the estimates denoted as $\hat{\boldsymbol{\beta}}$, or, for simplicity, as **b**. Note that in this model the treatment variables (travel-time-related variables) are trt and Rtrt.

Also note that no covariates are included in this model (there might have been, but including them in fact afforded no advantage).

Step 3: Estimation of $\Delta\bar{\boldsymbol{x}}_1'$.

Estimate the means, by survey round, of all travel-time-related variables. These are national estimates, using the survey "weights" of the nationwide household sample survey (the survey "base weights" are the reciprocals of the probabilities of selection of the households). The formulas for these estimated mean of the treatment variable, trt, is:

$$\overline{trt}_t = \sum_{i=1}^{n} w_i trt_{ti}$$

where

t = survey round (0 or 1)

n = number of households interviewed in both survey rounds

I = household index (I = 1, 2, …, n)

w_i = sample weight for i-th household (normalized to sum to one, over the sample)

trt_{ti} = treatment value (travel time) for i-th household in t-th survey round

$\overline{trt}_t$ = mean treatment value (travel time) for t-th survey round.

The formula for the mean of Rtrt is similar.

The components of $\Delta\bar{\boldsymbol{x}}_1'$ are the differences between survey rounds of the means for the treatment variables and zeros for the non-treatment variables. That is, if trt and Rtrt are the first two components of $\bar{\boldsymbol{x}}_1'$, then

$$\Delta\bar{\boldsymbol{x}}_1' = (\overline{trt}_1 - \overline{trt}_0, \overline{Rtrt}_{\mathbf{1}} - \overline{Rtrt}_0, 0,0).$$

Step 4: Estimation of Impact.

Estimate impact from the formula

$$Estimated\ Impact = \Delta\bar{\boldsymbol{x}}_1'\widehat{\boldsymbol{\beta}} = \Delta\bar{\boldsymbol{x}}_1'\boldsymbol{b}.$$

G.3 IMPACT ESTIMATES

We shall now apply the four preceding steps to every outcome of interest, to estimate impacts for each outcome variable.

Step 1: Estimation of Travel Times.

This was described above.

Step 2: Estimation of the Outcome (PTE) Model.

There is a separate outcome (PTE) model for each outcome variable, each with its own set of β's. The full regression output for those models is not presented here, but are included in the Stata .log file that accompanies the project documentation. Table 6 presents key model parameters (coefficients of treatment-related parameters). The table presents the values of β_1 and β_2 for each outcome variable, along with their standard errors. Two partial-effects parameters are involved, both involving trt (i.e., trt and Rtrt). The "main effect" of treatment is β_1, with β_2 representing an adjustment (associated with round). The coefficient β_1 (trt) varies in sign over the various outcomes. The trt parameter represents the partial effect of treatment, as measured by town1000tt. The coefficient is the estimated change in outcome per unit change in town1000tt. That coefficient along does not represent impact – impact is estimated by estimate $\Delta\boldsymbol{x}'\widehat{\boldsymbol{\beta}}$ (or $\Delta\boldsymbol{x}'\boldsymbol{b}$).

Table 6: Key Model Parameters for Outcome (Partial Treatment Effect) Regression Model with No Covariates

Table 6. Key Model Parameters for Outcome (Partial Treatment Effect) Regression Model with No Covariates				
Outcome Variable	**β_1 (trt)**		**β_2 (Rtrt)**	
	Estimate	**Standard Error**	**Estimate**	**Standard Error**
IncEmp	135	243	15.6	17.1
IncEmpAg	-293*	105	16.6*	7.39
IncEmpNonAg	610*	176	-9.85	12.4
TotHHExp	218	135	4.89	9.47
NetHHInc	1747	3561	421	251
CostToSchool	.087	.134	-.0061	.0094
TimeToSchool	.628	.420	-.0143	.0245
CostToCollege	5.59*	2.11	.262	.149
TimeToCollege	2.38*	1.34	.112	.088
CostToHospital	4.39	4.20	-1.74*	.296
TimeToHospital	.352	1.54	-.397*	.108
CostToHealthCtr	.614	.568	-.062	.040
TimeToHealthCtr	3.65*	.979	-.0044	.0692
CostToMarket	2.88*	1.40	-.102	.099
TimeToMarket	-1.35	1.40	-.176	.099
CostToPulp	.905*	.359	-.0073	.0254
TimeToPulp	.040	.412	-.020	.029
TimeToTegus	-1.59	2.88	-.991*	.187
TimeToSPS	-.032	5.91	.459	.381
TimeToDepCap	7.18*	3.21	-.950*	.214
TimeToMunCap	3.42	1.80	-.357*	.158

Table 6. Key Model Parameters for Outcome (Partial Treatment Effect) Regression Model with No Covariates				
Outcome Variable	β_1 (trt)		β_2 (Rtrt)	
	Estimate	Standard Error	Estimate	Standard Error
ChldInSch712	.022	.0275	-.0045	.0019
ChldInSch1318	-.030	.0220	-.0033*	.0015
VisHospital	-.004	.041	.0045	.0029
VisPrHlthCtr	-.011	.036	-.0024	.0025
VisPubClinic	-.078	.052	-.0104*	.0044
VisNonProf	.086*	.042	-.0076*	.0030
VisPharm	.010	.028	-.0033	.0020
WrkdPrevWk	.066	.038	.0088*	.0027

Step 3: Estimation of $\Delta\bar{\boldsymbol{x}}_1'$.

The parameters in Table 6 are the partial treatment effect of treatment on each of the outcomes of interest. The partial treatment effects are the coefficients (β_1 and β_2) of the regression models. These models do not show impact directly. The impact is obtained, as described earlier, by multiplying each partial treatment effect (the regression coefficient) by the average change in the value of the corresponding variable between survey rounds, and summing.

The means of trt and Rtrt are the same for all outcome variables. For Round 0, the means of trt and Rtrt are 14.9465 and -.8042, and for Round 1 the means are 14.7840 and .8328, so that the difference in means between the two rounds for trt and Rtrt are -.1525 and 1.6370. These values are the values for the first two components of $\Delta\bar{\boldsymbol{x}}_1'$, and the remaining components are zero. Hence

$$\Delta\bar{\boldsymbol{x}}_1' = (-.1525, 1.6370, 0, 0).$$

The value -.1525 is the average change in trt = town1000tt over the entire country, between survey rounds. The value -.1525 indicates that town1000tt decreased by an average of .1525 minutes over the entire country, between survey rounds. The value of β_1 is multiplied by this to obtain an unadjusted estimate of impact. This is added to the value of β_2 multiplied by 1.6370 (the mean change in Rtrt between survey rounds) to obtain the full impact.

Step 4: Estimation of Impact.

The estimated impact, presented in Table 7, is given by

$$\boldsymbol{\Delta x}'\,\mathbf{b} = \text{trtd } b_1 + \text{Rtrtd } b_2 = -.1525\, b_1 + 1.6370\, b_2,$$

where the “d” suffixed variables are the differences in means for the explanatory variables of the regression model (e.g., trtd = trt1 – trt0 where trt0 denotes the mean of trt in Round 0 and trt1 denotes the value of trt in Round 1.

For example, in the case of NetHHInc, the impact is -.1525 (1747) + 1.6370 (421) = 422. The standard error of the estimated impact cannot be obtained from just the standard deviations shown in this table – it also depends on the covariance between the coefficient estimates, which is not shown in this table. (The covariances are shown in the .log file.)

Table 7 presents estimates of the total treatment effect for the complete set of response indicators analyzed in detail. Impact estimates that are statistically significant (two-sided or one-sided, as appropriate, .05 significance level) are marked with an asterisk (*). Note that the components of IncEmp (IncEmpAg and IncEmpNonAg) do not sum to IncEmp, because they are estimated independently.

Table 7: Ordinary-Least-Squares (OLS) Regression Estimate of Average Treatment Effect (ATE) based on Continuous Treatment Variables, for Selected Outcome Measures

Table 7. Ordinary-Least-Squares (OLS) Regression Estimate of Average Treatment Effect (ATE) based on Continuous Treatment Variables, for Selected Outcome Measures		
Outcome Variable	**Estimate of Impact**	**Standard Error of Estimate**
Household Income and Expenditure		
IncEmp	5.00	41.5
IncEmpAg	71.9*	17.9
IncEmpNonAg	-109*	30.1
TotHHExp	-25.2	23.0
NetHHInc	422	609
Access		
CostToSchool	-.0232	.0228
TimeToSchool	-.119	.0718
CostToCollege	-.424	.361
TimeToCollege	-.180	.212
CostToHospital	-3.52*	.718
TimeToHospital	.704*	.263
CostToHealthCtr	-.194*	.097
TimeToHealthCtr	-.549*	.168
CostToMarket	-.606*	.239
TimeToMarket	-.083	.240
CostToPulp	-.126*	.0613
TimeToPulp	-.0394	.0704
TimeToTegus	-1.38*	.476
TimeToSPS	.757	.994
TimeToDepCap	-.459	.557
TimeToMunCap	-1.106*	.387
School Attendance		
ChldInSch712	-.00402	.00470
ChldInSch1318	-.000843	.00375

Table 7. Ordinary-Least-Squares (OLS) Regression Estimate of Average Treatment Effect (ATE) based on Continuous Treatment Variables, for Selected Outcome Measures		
Outcome Variable	**Estimate of Impact**	**Standard Error of Estimate**
Use of Health Care Services		
VisHospital	.00675	.00700
VisPrHlthCtr	.00230	.00621
VisPubClinic	-.00513	.01058
VisNonProf	-.0255	.00722
VisPharm	-.00702	.00479
Employment		
WrkdPrevWk	.00436	.00649

Note: Income, expense and travel costs measured in Honduran lempiras; travel times measured in minutes.

The preceding results show that the program intervention had a statistically significant effect (of the expected sign) on many of the access times and costs, on some of the indicators concerned with use of health care facilities, and on some of the income indicators. In all cases, the magnitude of estimated impact is small. The principal reason for small magnitude of impact is that it is the expected impact of the project *for a randomly selected household in the country*.

To assess the total national impact of the road-improvement program, the impact estimates are multiplied by the number of households in the nation. The population of Honduras is 8.2 million people (2010 est.), and the average household size is approximately 5, so this corresponds to approximately 1.64 million households. Based on these numbers, the estimated total program impact for the nation is approximately as follows:

Table 8: Estimated National-Level Impact of the Transportation Program (in Lempiras)

Table 8. Estimated National-Level Impact of the Transportation Program (table entries are lempiras)				
Outcome Variable	**Estimated Impact per Household**	**Standard error of Estimated Impact per Household**	**Estimated Total Impact for Nation**	**Standard Error of Estimated Total Impact for Nation**
IncEmp	5.00	41.5	8.2M	68M
IncEmpAg	71.9*	17.9	119M*	30M
IncEmpNonAg	-109*	30.1	-180M*	50M
TotHHExp	-25.2	23.0	-42M	38M
NetHHInc	422	609	692M	999M

An approximate 95% confidence interval for the estimated national-level project impact is the estimate plus and minus twice its standard error. For example, an approximate 95% confidence interval for income from agricultural employment (IncEmpAg) is (59M lempiras, 179M lempiras).

Note that in addition to the direct impact estimated here, the Transportation Project may have indirect effects, such as providing the improved transport required for the FTDA project.

The preceding estimates of total national impact are comparable in magnitude to, but small compared to, the cost of the road-improvement project (which was, according to the MCC Compact, 125.7 million dollars).

We conducted a detailed *ex post* statistical power analysis to assess the power associated with the tests of hypothesis associated with the estimates presented in the preceding table. The results of that analysis are presented in Annex 1. The power analysis showed that the evaluation design had high power to detect impacts equal in magnitude to about ten percent of the baseline mean.

G.4 CONCLUSIONS AND RECOMMENDATION

Summary of Impact of the Transportation Project

The principal finding of this evaluation is that although the Transportation Project shows some statistically significant effects on a variety of indicators (income and travel times to places of interest), those impacts are very small. A detailed statistical power analysis was conducted, which showed that the small number and size of the statistically significant results is not the result of an underpowered survey, but a result of the small magnitudes of the project effects.

The evaluation design adopted for this impact evaluation was to estimate impact from household-survey data, conditional on project-caused changes in travel time. The travel times are determined by a GIS road network model, using mean vehicle speeds estimated from traffic-survey data. The estimates are conditional on completion and maintenance of the road improvement project. The analysis produced estimates of the mean impact expected for a randomly selected household in Honduras. It was determined that the mean household-level impact of the project, averaged over the nation, is low.

Analysis of the traffic-survey data showed substantial changes in travel speeds and travel times over the project roads, compared to similar non-project roads. While the per-household impact averaged over the nation is low, the effect of the project on the speed of vehicles using the project roads is substantial. The disadvantage of using traffic-survey data alone to assess project impact is that it provides “intermediate” outcomes, not “higher level” impacts such as income, employment, and access to health, medical and other facilities.

The approach used in this report has a number of strengths that argue for the validity of the findings. The results are based on causal modeling, and the assumptions required of the statistical estimation models used to estimate impact are not in doubt. The per-household impact of the Transportation Project is low. While the economic impact of new roads is known to be substantial, the impact of the road improvements implemented in the Transportation Project, on a national level, are not high. On a national level, they represent marginal improvements to the road system and to household access, and they produce marginal impacts. The noticeable effect of the road improvements is on speeds and travel times of users of the improved roads.

In summary, the impact results presented in this report are based on an evaluation design, causal models, and analytical models that are considered to have high validity. An *ex post* statistical power analysis demonstrated that the power associated with the impact estimates is high.

Conditional on the soundness of the assumptions described, above, we consider the results of this analysis to be an accurate (valid, reliable and high-power) assessment of the impact of the Transportation Project.

Summary of Assumptions and Limitations

With respect to the “macro-level” causal model and associated statistical model used to estimate impact, the assumptions underlying the estimates of impact are the following:

1. The stable unit treatment value assumption (SUTVA, no-macro-effects assumption, partial equilibrium assumption) is made. Among other things, this assumption implies that the project is not so large that it changes the basic relationship of outcomes of interest to travel time.

2. The estimates of travel time are based on the GIS model of the Honduran road network. This model includes all official Honduran roads, as well as others. The GIS model is highly detailed, and considered to be up-to-date and of high accuracy. The quality of the GIS model is not considered to be a limitation on the quality of the evaluation.

3. The impact estimates are conditional on completion and maintenance of the Transportation Project as finally configured. Under this assumption, the impact estimates refer to this particular project, not to the mean impacts associated with a conceptually infinite population of similar projects in other locations or at other times. The estimate of impact is the average treatment effect of this particular project on a randomly selected household in Honduras, not the average treatment effect associated with improving a randomly selected eligible-for-treatment road segment.

4. Although the unit of treatment was the road segment, the unit of analysis was the household. In this case there could be two types of unobserved (hidden) variables, which may introduce biases into the estimates of the model parameters. First, there may be unobserved variables that are time invariant. In a fixed-effects model, these, however, “drop out” for the two-round panel specification. Second, there could be unobserved variables that are not time invariant (though no such variables were identified). It is assumed that such variables, if any exist, are uncorrelated with the explanatory variables.

5. The continuous-treatment-value impact estimates are conditional on the travel speed table derived from the traffic surveys. This table presents estimates of the average speed of a pickup truck over a route by route type (primary, secondary, rural), elevation variation and program intervention status (improved or not improved). The speeds are conditional on the season of the year, day of the week, time of day, and weather conditions under which the traffic surveys were conducted. The speeds are used to calculate travel times to places of interest (using the GIS model). These travel times are used (as described above) to estimate impact. A number of travel times were available (i.e., travel times to various places of interest from a caserío such as to a municipal or department capital, or the nearest town having population 1,000 or more. Attention focused on the one that had the highest relationship to outcomes of interest (i.e., travel time to the nearest town of population 1,000 or more). The continuous-treatment-value estimates are also

conditional on manifestation of long-term benefits, as estimated from the partial-treatment-effects model.

Other more specific model assumptions are listed for particular estimation equations in the detailed analysis presented in Annex 1.

The limitations of the evaluation are:

1. The impact estimates constructed in this analysis pertain solely to the Transportation Project as it was finally configured, and when and where it was implemented, not to similar projects in other settings (locations or times). The estimated standard errors reflect sampling variation associated with estimation of characteristics of this project, and do not include variation associated with hypothetical variations in the project location or time, or the higher-level sample units of the household sample survey (*caseríos* and households) (as would be the case for a "random effects" approach). It is expected, however, that similar results would be expected for similar projects in similar settings.

2. More of the resources for primary data collection for this evaluation were invested in the household sample survey, and less in the traffic surveys. The traffic surveys were limited in scope and in sample size. The design of the traffic surveys rested on judgment, not on the methods of statistical sample survey. The estimates of speed were conditioned only on elevation variation, and not on other road variables that may have an important effect on average speed, such as number of lanes, access, road roughness, and curviness. The vehicle speed on which travel times were based for analysis of impact was a pickup truck. The traffic survey data were used as input to the GIS road-network model for estimation of travel times, not for estimation of direct impact (such as via a double-difference estimator applied to the traffic data). The travel-time estimates pertain to a particular vehicle type, season, day of week, time of day, and local weather conditions. The level of correlation among the available travel times was relatively high, so that the limited scope of the traffic surveys is not considered to be a major weakness. In the future, however, if this approach is used, it is recommended that a substantially greater level of resources be allocated to the traffic surveys (at least comparable to that of the household surveys).

3. The study focuses on a variety of indicators to assess impact of the road improvements. Some of them are direct effects, such as travel times to points of interest (education and health facilities), and others are indirect effects such as income and employment. Additional direct effects might have been of interest, such as number of trips or length of trips.

4. As described above, an assumption of the analysis is that are no time-varying unobserved variables that are correlated with any of the explanatory variables, which in this case refers to the treatment variables. While we have investigated and confirmed the soundness of this assumption to the extent feasible, it is not possible to conclusively rule out the possibility that such variables exist and may be influencing the results. For example, if economic conditions such as labor market characteristics or the level of private investment are changing in systematically different ways that are correlated with travel time, our approach may mistake the impact of the road improvement for the impact

of changes in these conditions. This possibility must be considered a limitation of the analysis.

Recommendation for Additional Analysis

This project collected a lot of data from household and traffic surveys. These data were analyzed in accordance with the evaluation design and analysis plan. Under this plan, greater project resources were put into design, collection and analysis of the household survey data than into the design, collection and analysis of the traffic survey data. The traffic survey data were used to estimate travel times and to support an analysis of economic rate of return (ERR) using the Highway Design and Management Version 4 (HDM-4) road planning and analysis computer software package (this work is documented in a separate report). In view of the weak impact results associated with national-level household survey data, it is considered worthwhile to spend some additional effort to conduct a fixed-effects statistical analysis of direct impact using the traffic survey data, even though the traffic surveys were not designed for this purpose. This would include estimation of the impact of the road improvements on traffic speeds, volumes and origin-destination. This option was considered during the course of the project (and recommended by a reviewer and project team member), but resources were not available to conduct a detailed statistical design and analysis effort for both the household survey and the traffic surveys.

One reviewer suggested that a “corridor-level” analysis, focusing on households very near the improved roads, would show stronger impact. While this may be true to some extent, it is considered that basing impact estimation for road-improvement projects on general households will show small impact, even for households close to the project roads (this view is supported by the results for binary-treatment-variable estimates of impact, which, even though they focused attention on households around project roads, were even weaker than the continuous-treatment-variable estimates).

ANNEX 1: ESTIMATION OF IMPACT: A DETAILED DESCRIPTION OF THE ANALYSIS AND RESULTS

I. INTRODUCTION

This annex presents details on the analysis used to construct the impact estimates presented in the main text. It includes a description of the data processing that was done prior to the statistical analysis. It includes discussion of the binary-treatment-variable estimates of impact. That discussion is included as background for the analysis of the continuous-treatment-variable estimator. Also, the binary-treatment-variable approach was suggested in the MCA – Honduras *M&E Plan*, and it is of interest to show how the two approaches compare. In the main text, only the results for the continuous-treatment-variable approach are presented.

A significant aspect of the binary-treatment-variable estimators is that they do not involve the GIS travel-time estimates, other than for definition of the boundaries of the zones of influence around the project roads. Also, they assess impact at the time of the second survey round. The continuous-treatment-variable estimates are conditional on completion and maintenance of the project roads, and manifestation of indirect benefits per the partial-treatment-effect model.

This annex includes description of the binary-treatment-variable and continuous-treatment-variable estimates of impact, given the GIS travel time estimates. A description of the traffic surveys and procedures used to estimate travel times is presented in Annex 2.

II. PRELIMINARY DATA PROCESSING

Prior to conducting the data analysis (using the Stata statistical program package, version 10.0), NORC's data analysts conducted a rigorous quality review, cleaning and aggregation of the "raw" survey data. Since the primary unit of analysis for the survey data was the household, a major aspect of the initial data processing was aggregation of the detailed information included on the survey questionnaire into household-level data for analysis. This included aggregation of data for individual family members and items of income and expense, for various crops and for the household in general. The result of this initial data processing was a "flat file" (table) that included aggregated household-level data (one file record (row) per household)[10],[11]

[10] All of the initial data processing and analysis steps are documented in detail in Stata command files ("do" or ".do" files). The output from each .do file is a "log" file (or ".log" file). The processing and analysis may be replicated by executing the .do file, in which case the results will be presented in an associated .log file. For the Transportation evaluation project, .do files named Do1* - Do11* (where * denotes additional text) were used to clean and aggregate the questionnaire data to household level and Do12TranImpactEstimation.do was used to construct the impact estimates.

[11] One may ask why the questionnaire collected disaggregated data, when the data were aggregated for the impact analysis. The primary reason for collecting disaggregated data (e.g., data for individual family members or separate crops) is that collecting the detailed data and aggregating it is generally considered to produce more accurate aggregate measures than simply asking for aggregate amounts in the questionnaire. A secondary reason is that analysis of the disaggregated data may provide additional insight into the mechanisms of impact, such as relationships to family-member characteristics or effects for individual crops. The scope of the evaluation contract

II.A. Indicators of Interest: Distributions and Summary Statistics

The household survey for the Transportation Project evaluation consisted of a total of 3,200 sample units (households) in both survey rounds, of which 1,600 are in Round 0 (baseline) and 1,600 in Round 1 (endline or follow-up). The data were collected using a two-stage sample design consisting of a first-stage sample of 100 *caseríos* and a second-stage sample of approximately 16 households per sample *caserío*. Table A.1 shows the number of sample households by response category. The number of nonrespondents (all table lines after the first) is 0 for Round 0 and 192 for Round 1. Only completed questionnaires (line 1 of the table) were retained for the analysis.

Table A.1. Survey Responses

Response Category	Round		Total
	0	1	
Interviewed	1,600	1,408	3,008
Absent	0	41	41
Incomplete	0	9	9
Home Unoccupied	0	69	69
Home Destroyed	0	11	11
Refused	0	33	33
Moved	0	1	1
Unknown/Not Located	0	24	24
Duplicate Household	0	2	2
Incapacitated	0	1	1
Unknown	0	1	1
Total	1,600	1,600	3,200

As described in Section B of the main report, the objective of the MCA - Honduras Transportation Project was to foster economic growth and thereby reduce poverty by means of transportation infrastructure improvements. The intervention also sought to improve access and use of education and health facilities by reducing travel time to schools and health facilities. The evaluation questionnaire collected data on household income and expenses in a number of categories, including income derived through different types of employment, as well as data on travel times and use of educational and health services.

In this section, we present sample characteristics for various impact indicators of interest including income from employment (agricultural and non-agricultural), income from agricultural production, travel costs and time, and use of social services. We also present frequency distributions for type on jobs that people are engaged in rounds 0 and 1.

was to estimate overall program impact, and it did not include time or resources to conduct extensive analysis of disaggregated data.

Tables A.2A and A.2B show sample characteristics of indicators related to employment income, while Tables A.3A and A.3B show some basic distributions for respondents' employment status and occupation types by survey round.

Note that household income reported in Table A.2A does not include income from production and sale of agricultural crops, but just household income from labor-market employment. Agricultural income is presented separately in Tables A.4A and A.4B.

In all of the summary tables presented in this section, the means and standard deviations are simple unweighted values that do not take into account the design characteristics. The values presented in the tables are simply sample statistics, calculated using standard data-summary procedures such as Stata's *summarize* or *tabulate*. They should not be interpreted as estimates of population means or standard deviations – they are simply estimates of characteristics of the sample. In the impact estimation presented later, the design *is* taken into account, and the estimates and standard errors have desirable properties, such as unbiasedness or consistency.

1. Indicators Related to Income and Employment

1.A. Income Generated through Employment

The household income categories included in Table A.2A are the following:

Income from labor-market employment in the agricultural sector (IncEmpAg)
Income from labor-market employment in non-agricultural sectors (IncEmpNonAg)
Income from labor-market employment = IncEmpAg + IncEmpNonAg (IncEmp)
Income from foreign remittances (IncRemit)
Other income (pension, rent, Bono 80) (IncOther)
Total household income (monthly) from all sources (exclusive of income from agricultural production) = IncEmp + IncOther + IncRemit (IncTotalHH)

The focus of this evaluation is impact of the road improvements on income from employment and production of agricultural crops. There is no reason to hypothesize that the project intervention would lead to higher incomes from pensions, rent, investment funds, or foreign remittances. For this reason, some basic statistics are presented for all income categories, but most of the analysis disregards these latter categories.

Note that the sample of households was stratified by estimated time from project roads and by estimated change in travel time to be caused by road improvements, and for this reason it is not *uniformly* representative of the general Honduran population (i.e., sample units were selected with varying probabilities, and hence have different weights). It *is* a probability sample, however, and the sample data could be used to construct unbiased estimates for the nation (all *caseríos* were subject to sampling, except for some remote areas). The survey design was not oriented, however, to producing estimates of overall characteristics (means or totals) of the population. The design was not a "descriptive" survey design, but an "analytical" survey design, intended to produce estimates of the relationship of outcomes of interest to explanatory variables. To this end, the survey design was configured to promote variation in explanatory variables that were considered to have an effect on outcomes of interest. In achieving this result, the selection probabilities vary widely, and the precision of estimates of overall population characteristics would be low. For this reason, the summary sample statistics presented in the tables that follow are not "weighted" estimates of the population, but simply characteristics of the sample. (The weights are used for some of the estimates constructed in the data analysis.)

Table A.2A. Basic Characteristics of the Distribution for Household Income (excluding income from agricultural production), (Honduran Lempiras, monthly) Baseline (Round =0), N=1,600

Indicator	Mean	Std. Dev	Min	Max
Total Household Income from all sources (exclusive of income from agricultural production) (IncTotalHH)	6631	18359	0	343478
Income from labor-market employment (IncEmp)	4781	7600	0	14075
Income from labor-market employment in the agricultural sector (IncEmpAg)	788	2618	0	41183
Income from labor-market employment in non-agricultural sectors (IncEmpNonAg)	3993	7186	0	140170
Income from foreign remittances (IncRemit)	459	1793	0	25890
Other Income (IncOther)	1392	16663	0	333933

Table A.2B. Basic Characteristics of the Distribution for Household Income (excluding income from agricultural production), (Honduran Lempiras, monthly) Endline (Round =1), N=1,408

Indicator	Mean	Std. Dev	Min	Max
Total Household Income (exclusive of income from agricultural production) (IncTotalHH)	8270	19010	0	372083
Income from labor-market employment (IncEmp)	7172	13379	0	303150
Income from labor-market employment in the agricultural sector (IncEmpAg)	2413	11899	0	300150
Income from labor-market employment in non-agricultural sectors (IncEmpNonAg)	4759	7419	0	87000
Income from foreign remittances (IncRemit)	279	1411	0	22824
Other income (IncOther)	819	12591	0	334633

Note that the values for components of IncEmp (IncEmpAg and IncEmpNonAg) sum to the value for IncEmp. This property of the raw data does not hold for estimates to be considered later.

Observe the substantial difference in many of the indicators, by survey round. Part of this difference is attributable to the fact that 192 of the 1,600 Round 0 sample households were not interviewed in Round 1. Table A.2C presents Basic Characteristics for Round 0, for those households that were interviewed in both survey rounds. For this particular selection of variables, it is seen that there is not much difference between the entire Round0 sample and the sample of those Round 0 units that were interviewed in both rounds. The situation is similar for other variables, and so no additional tables similar to Table A.2C (of Round 0 means for households interviewed in both survey rounds) are presented.

Table A.2C. Basic Characteristics of the Distribution for Household Income (excluding income from agricultural production), (Honduran Lempiras, monthly) Baseline (Round =0), for households interviewed in both survey rounds, N=1,408				
Indicator	**Mean**	**Std. Dev**	**Min**	**Max**
Total Household Income from all sources (exclusive of income from agricultural production) (IncTotalHH)	6588	17248	0	343478
Income from labor-market employment (IncEmp)	4842	7518	0	14075
Income from labor-market employment in the agricultural sector (IncEmpAg)	814	2693	0	41183
Income from labor-market employment in non-agricultural sectors (IncEmpNonAg)	4028	7049	0	140170
Income from foreign remittances (IncRemit)	459	1846	0	25890
Other Income (IncOther)	1286	15396	0	333933

1.B. Employment Status and Occupation Types

Tables A3.A and A3.B respectively present frequency distributions for respondents' employment status (labor force participation) and principal occupation by survey round.

Table A.3A. Frequency Distribution of Employment Status				
Labor Force Participation Status	**Baseline (Round 0)**		**Endline (Round 1)**	
	Frequency	**Percent**	**Frequency**	**Percent**
Employed	1,357	84.81	1,190	84.64
Unemployed	9	0.56	10	0.71
Inactive	234	14.63	206	14.65
Totals	**1,600**	**100**	**1,406**	**100**

Table A.3B. Frequency Distribution of Principal Occupation				
Principal Occupation	**Baseline (Round 0)**		**Endline (Round 1)**	
	Frequency	**Percent**	**Frequency**	**Percent**
Managers, Public & Private Sector	71	5.23	64	5.38
Professionals (Scientists/Academics)	15	1.11	13	1.09
Technicians and mid-level professionals	33	2.43	27	2.27
Clerical workers, Public & Private Sector	5	0.37	7	0.59
Vendors/Service providers	28	2.06	17	1.43
Skilled agricultural/forestry/fishery workers	649	47.83	581	48.82
Artisans & workers in related trades	199	14.66	170	14.29
Operators of heavy equipment/machinery	55	4.05	50	4.20
Unskilled laborer (including in agriculture)	302	22.25	261	21.93
Totals	**1,357**	**100**	**1,190**	**100**

Over 80 percent of survey respondents are employed, and almost 50 percent are engaged in some type of agricultural work. Given this preponderance of agricultural labor, the impact analysis that follows presents breakdowns of income by household labor-market income (IncEmp) and production of basic grains and horticultural crops.

1.C. Income Generated through Production and Sale of Agricultural Crops

This sub-section presents basic characteristics (summary statistics) for income and expenses related to the production of crops. The key outcome indicators associated with agricultural income and expenses are the following:

For basic grains (BG) (annual amounts):

- Income from basic grains (including used for own consumption) (IncBG)
- Expenses for inputs for basic grains (FactorBG)
- Transportation expenses for basic grains (TranspBG)
- Other costs for basic grains (OthCostBG)
- Labor expense for basic grains (measure of employment associated with BG) (LabExpBG)
- Total expenses, basic grains (ExpBG) = FactorBG + TranspBG + OthCostBG + LabExpBG
- Net income from basic grains (NetBG) = IncBG – ExpBG

For other crops (OC) – horticultural crops (annual amounts):

- Income from other crops (including used for own consumption) (IncOC)
- Expenses for inputs for other crops (FactorOC)
- Transportation expense for other crops (TranspOC)
- Other costs for other crops (OthCostOC)
- Labor expense for other crops (measure of employment associated with OC) (LabExpOC)
- Total expenses, other crops (ExpOC) = FactorOC + TranspOC + OthCostOC + LabExpOC
- Net income from other crops (NetOC) = IncOC – ExpOC

The indicators LabExpBG and LabExpOC are measures of employment. Since reported income is often not considered accurate, the expense measures (ExpBG and ExpOC) may constitute better measures of program impact than the reported income measures (IncBG and IncOC).

Tables A.4A and A.4B present basic characteristics of the distribution of income, expense and net income from basic grains (BG) and other crops (OC). The units in the table (and most other tables that follow) are Honduran lempiras. The current exchange rate for the lempira is approximately 19 lempiras to the US Dollar. Note that income and expense amounts for crops are annual.

Table A.4A. Basic Characteristics of the Distribution for Household Income from Agricultural Production (Honduran Lempiras, annual)
Baseline (Round =0), N=1,600

Indicator	Mean	Std. Dev	Min	Max
Income, basic grains (IncBG)	4125	8959	0	130000
Expenses for inputs for basic grains (FactorBG)	987	2144	0	30500
Transportation expenses for basic grains (TranspBG)	68	281	0	5000
Other costs for basic grains (OthCostBG)	56	405	0	9300
Labor expense for basic grains (LabExpBG)	801	7954	0	301000
Total expenses, basic grains (ExpBG)	1912	8756	0	304270
Net income, basic grains (NetBG)	2213	9247	-264270	74540
Income, other crops (IncOC)	9538	58764	0	1271838
Expenses for inputs for other crops (FactorOC)	1216	6238	0	136500
Transportation expenses for other crops (TranspOC)	138	1058	0	30000
Other costs for other crops (OthCostOC)	133	2638	0	96000
Labor expense for other crops (LabExpOC)	4178	47054	0	1267200
Total expenses, other crops (ExpOC)	5665	52330	0	1275875
Net income, other crops (NetOC)	3873	52066	-1211595	1157950

Note: All units of measure for the indicators listed above are in Lempiras per year, with the exception of Labor Market Employment (IncEmp).

Table A.4B. Basic Characteristics of the Distribution for Household Income from Agricultural Production (Honduran Lempiras, annual)
Endline (Round =1), N=1,408

Indicator	Mean	Std. Dev	Min	Max
Income, basic grains (IncBG)	4866	13628	0	217750
Expenses for inputs for basic grains (FactorBG)	1193	2683	0	30600
Transportation expenses for basic grains (TranspBG)	50	200	0	2200
Other costs for basic grains (OthCostBG)	27	346	0	8000
Labor expense for basic grains (LabExpBG)	616	2742	0	40000
Total expenses, basic grains (ExpBG)	1886	4899	0	60000
Net income, basic grains (NetBG)	2980	11437	-35000	213250
Income, other crops (IncOC)	16236	97501	0	2448155
Expenses for inputs for other crops (FactorOC)	1497	9011	0	227740
Transportation expenses for other crops (TranspOC)	106	1041	0	30000
Other costs for other crops (OthCostOC)	36	490	0	11250
Labor expense for other crops (LabExpOC)	3839	38162	0	1080000
Total expenses, other crops (ExpOC)	5478	45323	0	1250000
Net income, other crops (NetOC)	10758	88061	-1069400	2312355

1.D. Overall Household Expenditure and Income (from all sources)

Indicators for overall income and expenditures at the household level include:

- Total household expenditures (TotHHExp) (monthly amount)
- Net household income (NetHHInc) = NetBG + NetOC + IncTotalHH*12 (annualized amount), where IncTotalHH = monthly non-agricultural household income from all sources, IncEmp + Inc Remit + IncOther

Household expenditures (TotHHExp) is monthly, and NetHHInc is annualized, in the table below.

Table A.5. Basic Characteristics of the Distribution for Overall Household Expenditures and Income (from all sources) (Honduran Lempiras)

Indicator	Mean	Std. Dev	Min	Max
Baseline (Round =0), N=1,600				
Total hhold expenditures (TotHHExp, monthly)	5268	4360	0	77175.08
Net household income (NetHHInc, annual)	85929	228685	-1166525	4152988
Endline (Round =1), N=1,408				
Total hhold expenditures (TotHHExp, monthly)	6384	4915	0	39705.43
Net household income (NetHHInc, annual)	114014	265739	-990800	4716996

It is seen from Table A.5 (from the mins and maxes) that the sample contains some extreme values. A later section describes the procedures used to handle extreme and missing values.

2. Indicators Related to Travel Costs and Travel Times

In addition to detailed information about income from agricultural and non-agricultural sources, the household survey also collected data on access to places of interest as measured by travel time and travel cost. The hypothesis is that improvements in the road network – primary, secondary and tertiary roads – would improve access to markets, city centers, schools and health centers by reducing travel times and associated costs. Improved access to these centers of business and social activity would in turn provide opportunities to increase income (through easier access to jobs and markets) and encourage the use of health centers and schools.

Key outcome indicators associated with improved access are the following:

Cost (in lempiras) to primary school (CostToSchool)
Time (in minutes) to primary school (TimeToSchool)
Cost to secondary school[12] (CostToCollege)
Time to secondary school (TimeToCollege)
Cost to hospital (CostToHospital)
Time to hospital (TimeToHospital)
Cost to health center (CostToHealthCtr)
Time to health center (TimeToHealthCtr)

[12] Note that the term "College" or "colegio" in Spanish refers to secondary education, not tertiary education.

Cost to market (CostToMarket)
Time to market (TimeToMarket)
Cost to pulperia (CostToPulp)
Time to pulperia (TimeToPulp)
Time to Tegucigalpa (TimeToTegus)
Time to San Pedro Sula (TimeToSPS)
Time to departmental capital (TimeToDepCap)
Time to municipal capital (TimeToMunCap)

Tables A.6A and A.6B presents basic characteristics of the distribution of these access indicators for Rounds 0 and 1. The units for travel costs are in Honduran lempiras, and travel-time units are minutes.

Table A.6A. Basic Characteristics of the Distribution for Travel Cost and Travel Time Indicators (Honduran Lempiras and Minutes)
Baseline (Round =0)

Indicator	Obs	Mean	Std. Dev	Min	Max
Cost to school (CostToSchool)	1600	.12	1.70	0	50
Time to school (TimeToSchool)	1596	14.28	13.56	1	120
Cost to college (CostToCollege)	1600	10.04	63.81	0	2500
Time to college (TimeToCollege)	1579	53.32	44.56	1	300
Cost to hospital (CostToHospital)	1600	51.99	96.78	0	1000
Time to hospital (TimeToHospital)	1585	111.96	65.08	5	600
Cost to health center (CostToHealthCtr)	1600	5.39	16.49	0	450
Time to health center (TimeToHealthCtr)	1578	43.5	36.52	1	300
Cost to market (CostToMarket)	1600	25.22	36.77	0	500
Time to market (TimeToMarket)	1580	82.17	58.53	1	540
Cost to pulperia (CostToPulp)	1600	.28	5.85	0	200
Time to pulperia (TimeToPulp)	1599	7.91	12.01	0	180
Time to Tegucigalpa (TimeToTegus)	1109	241.36	137.55	1	720
Time to San Pedro Sula (TimeToSPS)	956	268.67	159.25	4	845
Time to departmental capital (TimeToDepCap)	1501	112.36	67.89	0	540
Time to municipal capital (TimeToMunCap)	1578	68.82	53.96	0	480

Table A.6B. Basic Characteristics of the Distribution for Travel Cost and Travel Time Indicators (Honduran Lempiras and Minutes)
Endline (Round =1)

Indicator	Obs	Mean	Std. Dev	Min	Max
Cost to school (CostToSchool)	1407	.32	4.12	0	100
Time to school (TimeToSchool)	1407	13.77	13.57	0	127
Cost to college (CostToCollege)	1408	8.99	22.22	0	500
Time to college (TimeToCollege)	1405	48.57	41.77	1	270
Cost to hospital (CostToHospital)	1408	54.91	112.65	0	1512
Time to hospital (TimeToHospital)	1405	105.9	61.85	0	481
Cost to health center (CostToHealthCtr)	1408	8.4	17.75	0	200
Time to health center (TimeToHealthCtr)	1407	39.62	32.25	0	210
Cost to market (CostToMarket)	1404	29.22	41.81	0	525
Time to market (TimeToMarket)	1398	80.55	56.21	0	390
Cost to pulperia (CostToPulp)	1403	.66	10.286	0	321
Time to pulperia (TimeToPulp)	1403	7.47	11.10	0	120
Time to Tegucigalpa (TimeToTegus)	1126	253.82	142.65	0	990
Time to San Pedro Sula (TimeToSPS)	937	264.83	168.19	0	1230
Time to departmental capital (TimeToDepCap)	1320	120.38	105.66	0	1230
Time to municipal capital (TimeToMunCap)	1332	72.04	76.69	0	1410

3. Indicators Related to School Attendance, Use of Health Services, and Employment

The preceding two subsections have described a variety of indicators related to income and travel time. This subsection describes several of indicators related to school attendance, use of health services and employment.

Outcome indicators related to these areas of interest are the following:

School attendance:

Total number of children aged 7-12 attending school (ChldInSch712)
Total number of children aged13-18 attending school (ChldInSch1318)

Use of health-care services:

Total number of visits (by all household members) in last 30 days to hospital by all family members (VisHospital)
Total number of visits in last 30 days to private health centers (VisPrHlthCtr)
Total number of visits in last 30 days to public clinics (VisPubClinic)
Total number of visits in last 30 days to non-professional health-care providers (VisNonProf)
Total number of visits in last 30 days to pharmacy (VisPharm)

Employment:

Total number of household members who worked the previous week (WrkdPrevWk)

Tables A.7A and A.7B presents basic characteristics of the distribution of these indicators for Rounds 0 and 1.

Table A.7A. Basic Characteristics of the Distribution for Indicators Related to School Attendance, Use of Health-Care Services and Employment Baseline (Round =0), N=1,600

Indicator	Mean	Std. Dev	Min	Max
Children age 7-12 in school (ChldInSch712))	.874	.997	0	5
Children age 13-18 in school (ChldInSch1318))	.371	.666	0	4
Visits to hospital in last 30 days (VisHospital)	.306	1.11	0	16
Visits to private health center in last 30 days (VisPrHlthCtr)	.303	.977	0	12
Visits to public clinics in last 30 days (VisPubClinic)	.64	1.55	0	13
Visits to non-professional health-care providers in last 30 days (VisNonProf)	.222	1.29	0	26
Visits to pharmacy in last 30 days (VisPharm)	.13	.631	0	8
Total number of HH members who worked the previous week (WrkdPrevWk)	1.80	1.23	0	9

Note: Indicators related to health facility visits refer to all household members

Table A.7B. Basic Characteristics of the Distribution for Indicators Related to School Attendance, Use of Health-Care Services, and Employment Endline (Round =1), N=1,408

Indicator	Mean	Std. Dev	Min	Max
Children age 7-12 in school (ChldInSch712))	.815	.960	0	7
Children age 13-18 in school (ChldInSch1318))	.394	.689	0	4
Visits to hospital in last 30 days (VisHospital)	.352	1.09	0	12
Visits to private health center in last 30 days (VisPrHlthCtr)	.261	.865	0	14
Visits to public clinics in last 30 days (VisPubClinic)	.568	1.69	0	24
Visits to non-professional health-care providers in last 30 days (VisNonProf)	.043	.445	0	12
Visits to pharmacy in last 30 days (VisPharm)	.153	.710	0	12
Total number of household members who worked the previous week (WrkdPrevWk)	1.88	1.302	0	8

Note: Indicators related to health facility visits refer to all household members

II.B. Treatment of Extreme or Missing Values

Virtually any large sample survey contains some extreme or missing responses. Some of those extreme or missing values may unduly influence the results, and decisions must be made on how to handle them to promote high accuracy of the sample estimates and high power for tests of hypotheses. Standard alternatives for addressing this issue are imputation of missing values, censoring of extreme values and deletion (dropping) of observations containing missing or extreme values.

Casewise deletion of observations is routinely done by statistical software (such as Stata) during the course of model development (such as regression analysis), unless the missing values are imputed. Therefore, deletion of observations containing missing values or imputation of missing values is usually unavoidable at some point in the development of analytical models. The approach adopted here is to retain all observations in the model, and allow deletion of them only by the model-development software in cases in which missing values are not imputed. Missing values were imputed for income and expense for agricultural production (of basic grains and other crops). After this, very few missing values occurred in other variables used in regression models. If imputation of missing values was considered necessary in a regression model, the missing values were imputed by means.

Censoring of extreme values is problematic in the present application because some variables are interrelated (i.e., if a value is imputed for one variable, it must be consistent with the values of all related variables). In this analysis we examined the distribution of all components of income and expense for each of the two crop sources of income (BG and OC) and identified observations (households) for which any of the income or expense components exceeded the 99^{th} percentile. For identified observations, we replaced the income value by the 99^{th} percentile and the expense values by a value determined from a regression of the expense value on the income value. This procedure assures the consistency of all imputed income and expense components[13]. Censoring was done for all income and expenditure variables, i.e., for all "BG" and "OC" variables, IncTotalHH, IncEmp, IncEmpAg, IncEmpNonAg, IncRemit, IncOther and TotHHExp.

The process of censoring is not without drawbacks. Some extreme observations are valid, and they are censored along with erroneous ones. Although censoring reduces bias by moderating the values of erroneous extreme values, it may introduce bias by altering values of legitimate extreme values. There is hence a trade-off between censoring at too high or too low a value. In the present study, all of the impact estimates involve the use of regression models, and we

[13] The procedure used in the censoring is follows. If any of the components of income or expense exceeds the 99^{th} percentile, then the values of all components were censored according to the following rules:

FactorBG = .19 IncBG

TranspBG = .017 IncBG

OthCostBG = .019 IncBG

LabExpBG = .22 IncBG

FactorOC = .080 IncOC

TranspOC = .012 IncOC

OthCostOC = .018 IncOC

LabExpOC = .40 IncOC

LabExpOC = .40 IncOC.

consider that a somewhat stringent censoring is appropriate. Some legitimate large values of incomes and expense may be wrongly censored, but the nature of the relationships represented in the regression models is not unduly affected. The observations that are censored in error tend to be “well-off” households, and the focus of the program intervention is to reduce poverty, i.e., poorer households.

In addition to its role in reducing bias, censoring also has an effect on reducing variation, i.e., it is expected to reduce standard errors of estimates somewhat. Bias and precision (reliability) are two components of accuracy. Both are of concern, and it is viewed that the censoring contributed to improvements in both aspects in the present evaluation. Note that in the analysis, a particular variable may appear in one instance as an explained variable (“dependent” variable) in a model and in another instance as an explanatory variable (“independent” variable) (and even sometimes as both, e.g., an endogenous variable). Once the decision was made to censor a variable, the censored values were used throughout the analysis, regardless of the role of the variable in a model (dependent or independent).

Tables A.8A and A.8B show the same information as Table A.4A and A4.B, but for the censored data, and only for the items of income and expense other than travel costs and times (since those were not affected by the imputation). This table shows that the censoring caused a modest reduction in the means of the outcome variables, and a substantial reduction in the standard deviations. The extreme maxima and minima observed in Table A.5 have been substantially reduced.

Note that, after censoring, the maxima or minima may be exactly the same for both survey rounds.

Censoring was applied just to the income and expense variables specified above. There may be other extreme values in the data set. As models were developed, care was taken that the variables included in the model did not contain unreasonably extreme values, that might cause undue influence on the results.

Table A.8A. Basic Characteristics of the Distribution for Key Outcome Variables for Censored Data (Honduran Lempiras, annual unless otherwise indicated)
Baseline (Round =0), N=1,600

Indicator	Mean	Std. Dev	Min	Max
Income, basic grains (IncBG)	4020	8050	0	57500
Expenses for inputs for basic grains (FactorBG)	883	1653	0	11030
Transportation expenses for basic grains (TranspBG)	47	154	0	1300
Other costs for basic grains (OthCostBG)	24	115	0	1300
Labor expense for basic grains (LabExpBG)	454	1433	0	12222
Total expenses, basic grains (ExpBG)	1408	2893	0	25301.73
Net income, basic grains (NetBG)	2612	5933	-5560	55520
Income, other crops (IncOC)	7333	28394	0	244090
Expenses for inputs for other crops (FactorOC)	734	2634	0	26200
Transportation expenses for other crops (TranspOC)	66	337	0	3000
Other costs for other crops (OthCostOC)	19	118	0	1000

Table A.8A. Basic Characteristics of the Distribution for Key Outcome Variables for Censored Data (Honduran Lempiras, annual unless otherwise indicated)
Baseline (Round =0), N=1,600

Indicator	Mean	Std. Dev	Min	Max
Labor expense for other crops (LabExpOC)	1434	7171	0	75000
Total expenses, other crops (ExpOC)	2253	9309	0	98640.82
Net income, other crops (NetOC)	5080	20855	-60100	214290
Tot. hh inc. from all sources (IncTotalHH)	5779	6829	0	51019
Labor market income (IncEmp, monthly)	4632	6033	0	45142
Agricultural labor market income (IncEmpAg, monthly)	753	2218	0	21166
Nonag. labor market income (IncEmpNonAg, monthly)	3779	5058	0	32000
Total hhold expenditures (TotHHExp, monthly)	5245	4066	0	40000
Net household income (NetHHInc)	77052	89246	0	819428

Table A.8B. Basic Characteristics of the Distribution for Key Outcome Variables for Censored Data (Honduran Lempiras, annual unless otherwise indicated)
Endline (Round =1), N=1,408

Indicator	Mean	Std. Dev	Min	Max
Income, basic grains (IncBG)	4391	9896	0	57500
Expenses for inputs for basic grains (FactorBG)	1035	1954	0	11030
Transportation expenses for basic grains (TranspBG)	42	151	0	1300
Other costs for basic grains (OthCostBG)	7.53	66.66	0	1065.98
Labor expense for basic grains (LabExpBG)	451	1556	0	12222
Total expenses, basic grains (ExpBG)	1536	3178	0	24317.97
Net income, basic grains (NetBG)	2855	7762	-12545	56232
Income, other crops (IncOC)	10769	38124	0	244090
Expenses for inputs for other crops (FactorOC)	906	3051	0	25000
Transportation expenses for other crops (TranspOC)	51	297	0	3000
Other costs for other crops (OthCostOC)	9.07	86.02	0	1000
Labor expense for other crops (LabExpOC)	1705	9123	0	75000
Total expenses, other crops (ExpOC)	2671	11544	0	97640.82
Net income, other crops (NetOC)	8098	29161	-28540	241150
Tot. hh inc. from all sources (IncTotalHH)	7290	8706	0	51019
Labor market income (IncEmp)	6553	8167	0	45142
Agricultural labor market income (IncEmpAg, monthly)	1664	3854	0	21166
Nonag. labor market income (IncEmpNonAg, monthly)	4546	6115	0	32000
Total hhold expenditures (TotHHExp, monthly)	6384	4915	0	39705.43
Net household income (NetHHInc)	98431	121119	-4700	884958

III. ESTIMATION OF IMPACT

III.A Outcome Variables of Interest

In the impact analysis presented in the following sections of this annex, we examine the following impact indicators discussed in Section II.A of this annex: IncEmp, IncEmpAg, IncEmpNonAg, TotHHExp, NetHHInc and all of the travel-time and travel-cost indicators listed in Section II.A, subsection 2.

Standard errors are presented for all statistical estimates. To approximately assess the statistical significance of an estimate, divide the estimate by its standard error. Results exceeding two in magnitude are of moderate statistical significance in the context of the assumed model (the likelihood that the estimated effect size exceeds its standard error in magnitude by a factor of two is about one in twenty, if the effect is in fact zero). An approximate 95 percent confidence interval for the estimate is defined by the estimate plus and minus two standard errors. (More precisely, an effect is considered statistically significantly different from zero if it differs from zero by more than 1.96 times its standard error, for two-sided tests of hypothesis (i.e., the effect may be either positive or negative), or by more than 1.645 times its standard error, for one-sided tests of hypothesis (i.e., the sign of the effect is specified). On the indication of statistical significance, the interpretation is that over many independent investigations, the probability that the confidence interval includes the true value of the parameter is approximately .95. Confidence intervals within the same investigation are correlated.)

III.B Treatment Variables and Covariates

Travel-time (or travel-cost) variables were obtained from two sources: from the questionnaire and from the geographic-information-system (GIS) network travel-time model. The GIS-model travel times are used as explanatory variables in models. The questionnaire travel times are used only as response variables (because of endogeneity with other questionnaire variables, such as income).

The travel-time and travel-cost variables obtained from the questionnaire are listed above in Section II.A, subsection 2.

The travel-time variables produced by the geographic information system (GIS) network travel-time model were the following:

Travel time (in minutes) to the nearest point on the nearest MCA project primary road (all of which are segments of Highway CA-5) (mcapritt)
Travel time to the nearest point on the nearest MCA project secondary road (mcasectt)
Travel time to the nearest point on the nearest MCA project rural road (mcarurtt)
Travel time to nearest point on nearest primary road (major highway) (pritt)
Travel time to nearest point on nearest secondary road (sectt)
Travel time to nearest point on nearest rural (tertiary) road (rurtt)
Travel time to Tegucigalpa (tegustt)
Travel time to San Pedro Sula (sanpedtt)
Travel time to nearest *caserío* of population 1,000 or more (town1000tt)

Travel time to the nearest of the top 10 Honduran cities, ranked by population - Tegucigalpa, La Ceiba, Comayagua, San Pedro Sula, Choloma, Puerto Cortes, La Lima, Choluteca, Danli and El Progresso (top10tt)

The preceding travel times were calculated before and after the road improvements. The GIS-model travel times for Round 0 are labeled the same as above, but with suffix "0" appended (e.g., mcapritt0, mcasectt0, etc.). The changes in travel time between Round 0 and Round 1 estimated by the GIS travel-time model are labeled in the same way, by adding the suffix "d" (e.g., mcaprittd, etc.). The GIS model also produced the Euclidean (straight-line) distances corresponding to the ten travel-time measures listed above (labeled with suffix "dis" instead of "tt", e.g., mcapridis instead of mcapritt).

The preceding GIS-model travel times are shown for the two survey rounds in Tables A.9A and A.9B. Note that relatively little change occurs in these travel times between the two survey rounds (i.e., before and after the road-improvement intervention)[14].

Table A.9A. Basic Characteristics of the Distribution of Travel Times from the GIS Travel-Time Model, Baseline (Round=0), N=1557

Indicator	Mean	Std. Dev.	Min	Max
mcapritt	96.32	88.64	.71	374.89
mcasectt	84.42	49.52	12.25	210.94
mcarurtt	43.40	30.58	0	137.38
pritt	18.05	19.32	0	71.87
sect	22.05	21.90	.00034	102.32
rurtt	4.03	5.94	0	48.13
tegustt	152.25	91.35	8.98	401.02
sanpedtt	204.17	105.25	17.09	436.81
town1000tt	13.26	13.36	0	54.10
top10tt	60.69	46.86	0	249.11

Table A.9B. Basic Characteristics of the Distribution of Travel Times from the GIS Travel-Time Model
Endline (Round=1), N=1367

Indicator	Mean	Std. Dev.	Min	Max
mcapritt	96.56	87.51	.71	359.73
mcasectt	82.94	48.91	12.25	207.18
mcarurtt	42.48	30.81	0	137.38
pritt	17.58	18.70	0	71.87
sect	22.12	21.96	.00033	102.32
rurtt	3.88	5.70	0	48.13

[14] Note that in one instance (sectt), the mean travel time *increased* between survey rounds (baseline and endline). This occurred because of nonresponse in the second round. If the estimates are based only on households that respond in both survey rounds (as is the case for the fixed-effects estimates), the GIS-model travel times in Round 1 are uniformly less than those in Round 0.

Table A.9B. Basic Characteristics of the Distribution of Travel Times from the GIS Travel-Time Model
Endline (Round=1), N=1367

Indicator	Mean	Std. Dev.	Min	Max
tegustt	149.90	89.48	8.98	401.02
sanpedtt	200.78	103.80	17.09	431.36
town1000tt	13.17	13.25	0	54.10
top10tt	59.69	45.79	0	243.65

Definition of Binary Travel-Time Variables

The travel-time variables listed above are all continuous variables. In some cases, we used these variables directly in models. In other cases, we obtained binary variables from these variables, and used them in place of the continuous variables. This was done to simplify the model, and thereby promote understanding. Construction of binary travel-time variables was done only for the GIS travel-time variables (not for the questionnaire travel-time variables).

For three of the continuous-treatment variables, we estimated a corresponding binary-treatment variable as follows:

> Trtprit = 1 if mcapritt < 60, 0 otherwise, where mcapritt = travel time (in minutes) to nearest MCA-project primary road (i.e., an improved segment of Highway CA-5)
>
> Trtsect = 1 if mcasectt < 60, 0 otherwise, where mcasectt = travel time to nearest MCA-project secondary road
>
> Trtrurt = 1 if mcarurtt < 30, 0 otherwise, where mcarurtt = travel time to nearest MCA-project rural road
>
> Trtprid = 1 if mcapridis < 24,000, 0 otherwise, where mcapridis = distance (in meters) to nearest MCA-project primary road
>
> Trtsecd = 1 if mcasecdis < 24,000, 0 otherwise, where mcasecdis = distance to nearest MCA-project secondary road
>
> Trtrurd = 1 if mcarurdis < 12,000, 0 otherwise, where mcarurdis = distance to nearest MCA-project rural road.

The time and distance limits used to define the preceding indicator variable were determined by examining a range of values (e.g., for times, 15 minutes, 30 minutes, 45 minutes and 60 minutes) and selecting the value that best discriminated between households within and outside of the boundary, for a variety of outcome variables).

In addition to the preceding indicators, we defined and examined two “combined” treatment indicators: Treatedt = 1 if any of the three road-type-specific travel-time indicators is 1 (and 0 otherwise); and Treatedd = 1 if any of the three road-type-specific distance indicators is 1 (and 0 otherwise). These combined estimators did not prove useful.

Other Explanatory Variables (Covariates)

The most important explanatory variables in the various models from which impact estimators are derived are the travel-time or travel-cost variables described above. These are the treatment variables associated with the program intervention. The survey questionnaire contains a large number of variables that may be considered as explanatory variables in causal models. The following table presents basic statistics for a number of potential explanatory variables.

Table A.10A. Summary Characteristics of Questionnaire Variables Considered for Use as Covariates in Impact Models (Baseline, Round=0)

Indicator	Obs	Mean	Std. Dev	Min	Max
Harvested horticultural crops in last 12 months (Horticulture)	861	1.91	.27	1	2
No. persons in hhold (HouseholdSize)	1600	5.00	2.39	1	23
Sex of head of household, male=1, female=2 (SexHead)	1600	1.23	.42	1	2
Male hhold head (MaleHead)	1600	.77	.42	0	1
Age of head of household (AgeHead)	1599	47.54	16.49	15	101
Mean age of family members (MeanAge)	1600	27.55	14.82	7.2	98.5
Head of hhold is married (Married)	1600	.77	.42	0	1
Head of hhold is literate, yes=1, no=2 (Literate)	1600	1.32	.47	1	2
Years of formal ed of head of hhold (FormalEdHead)	1594	3.40	3.25	0	20
Avg No. years of education of hhold members (MeanEduc)	1600	3.34	2.22	0	19
No. agricultural employees in hhold (AgEmployees)	1600	.07	.31	0	4
No. of employed persons in hhold (NumEmployed)	1600	1.86	1.24	0	9
No. of unemployed persons in hhold (NumUnempl)	1600	.02	.16	0	3
Basic Necessities Index (NBIviv)	1600	0	0	0	0
Average no. of persons per room (PersPerRoom)	1600	1.94	1.55	.17	23
PersPerEmp~d	1501	3.07	1.64	1	11
No. of Basic Necessities (NumOfNBI)	1600	1.018	1.11	0	5
Time (min) to where purchase majority of food (TimeToFoodStore)	1438	31.73	44.48	1	600
Time from farm to where sell production (TimeToWhereSellProduce)	409	63.25	46.82	0	240
Total number of hectares of own farm (TotHaOwnFarm)	1600	1.66	6.58	0	142
Value (lps) of own farm (ValueOwnFarm)	1600	82726.21	461522.3	0	1.19e+07
Rental value (lps) of own farm (RentValueOwnFarm)	1600	5401.42	59860.86	0	2160000
Total number of hectares rented (TotHaRented)	1600	.31	.72	0	9.23
Value (lps) of land (LandValue)	1600	208.73	875.27	0	17450
Value (lps) of agricultural equipment (EqptValue)	1600	7088.06	35924	0	930900

Table A.10A. Summary Characteristics of Questionnaire Variables Considered for Use as Covariates in Impact Models (Baseline, Round=0)					
Indicator	**Obs**	**Mean**	**Std. Dev**	**Min**	**Max**
Rent received on agricultural equipment in last 12 months (EqptRentalValue)	1600	51.67	1099.26	0	40000
Rental value (lps) of farm installation (InstRentalValue)	1600	52.32	588.24	0	16365
Sale value (lps) of all animals (AnimalValue)	1600	13129.6	126472.4	0	4536500

Table A.10B. Summary Characteristics of Questionnaire Variables Considered for Use as Covariates in Impact Models (Endline, Round=1)					
Indicator	**Obs**	**Mean**	**Std. Dev**	**Min**	**Max**
Harvested horticultural crops in last 12 months (Horticulture)	772	1.926166	.2616704	1	2
No. persons in hhold (HouseholdSize)	1408	5.229403	2.456863	1	17
Sex of head of household, male=1, female=2 (SexHead)	1408	1.204545	.4035124	1	2
Male hhold head (MaleHead)	1408	.7954545	.4035124	0	1
Age of head of household (AgeHead)	1408	50.09801	16.12126	16	103
Mean age of family members (MeanAge)	1408	29.20735	14.95649	9.2	103
Head of hhold is married (Married)	1408	.7542614	.4306773	0	1
Head of hhold is literate, yes=1, no=2 (Literate)	1408	1.307528	.4616341	1	2
Years of formal ed of head of hhold (FormalEdHead)	1407	3.511016	3.439539	0	22
Avg No. years of education of hhold members (MeanEduc)	1408	3.867458	2.389666	0	19.5
No. agricultural employees in hhold (AgEmployees)	1408	.0582386	.2682228	0	4
No. of employed persons in hhold (NumEmployed)	1408	1.963068	1.29788	0	8
No. of unemployed persons in hhold (NumUnempl)	1408	.053267	.259852	0	3
Basic Necessities Index (NBIviv)	1408	.0007102	.0266501	0	1
Average no. of persons per room (PersPerRoom)	1406	2.002338	1.635139	.1428571	14
PersPerEmp~d	1334	3.095384	1.671702	1	14
No. of Basic Necessities (NumOfNBI)	1408	.9190341	1.035534	0	5
Time (min) to where purchase majority of food (TimeToFoodStore)	1280	48.88801	259.4395	0	7800
Time from farm to where sell production (TimeToWhereSellProduce)	282	63.66667	55.37561	2	480
Total number of hectares of own farm (TotHaOwnFarm)	1408	1.649318	6.654195	0	120.7

Table A.10B. Summary Characteristics of Questionnaire Variables Considered for Use as Covariates in Impact Models (Endline, Round=1)					
Indicator	**Obs**	**Mean**	**Std. Dev**	**Min**	**Max**
Value (lps) of own farm (ValueOwnFarm)	487	294307.3	1240588	0	2.45e+07
Rental value (lps) of own farm (RentValueOwnFarm)	487	18110.85	78525.68	0	1200000
Total number of hectares rented (TotHaRented)	1408	.3531988	1.139047	0	28.4
Value (lps) of land (LandValue)	357	2210.466	7023.296	0	96000
Value (lps) of agricultural equipment (EqptValue)	771	16796.19	55422.22	0	710000
Rent received on agricultural equipment in last 12 months (EqptRentalValue)	771	13.06615	125.1601	0	2500
Rental value (lps) of farm installation (InstRentalValue)	771	103.7988	600.8305	0	12000
Sale value (lps) of all animals (AnimalValue)	771	56137.26	1013190	0	2.80e+07

During the course of the analysis, we considered a number of alternative model specifications, based on different selections of explanatory variables. The particular choice of explanatory variables is rather flexible, because most of them are correlated and their effects confounded. After a few of the more (causally) important explanatory variables are included in a model, there is no advantage to adding additional variables (and it may in fact be disadvantageous). Because no experimental (forced-change) control was applied to the covariates (non-treatment variables) of the questionnaire, it is not possible to attribute causal influence to individual covariates – all that the analysis shows relative to these variables is statistical associations, not causal relationships/effects.

The particular selection of explanatory variables used in a model derives from a causal model that relates an outcome (response) variable to explanatory variables. Although we examined a wide variety of models, only a few are presented in this annex. Many models were examined using the following selection of covariates:

HouseholdSize (number of persons in household)
MaleHead (head of household is male)
AgeHead (age of head of household)
MeanAge (mean age of family members)
Married (head of household is married)
MeanEduc (average number of years of education of persons in household)
AgEmployees (number of agricultural-sector employees in household)
TotHaOwnFarm (total number of hectares of own farm)
TotHaRented (total number of hectares rented).

Of the preceding questionnaire variables, only HouseholdSize, MeanEduc and TotHaOwnFarm were highly statistically significant for models of income and expenditure response variables, and the income and expense models of this annex include just those covariates (if they include any covariates at all). For models relating to access and the other response variables, MeanEduc was the only covariate included (if any covariates were included at all).

The ten GIS-model travel times listed earlier may be used either as treatment variables or as covariates. If used as covariates, only the Round 0 values are used – it is the "levels" of the variables that is of interest. If used as treatment variables, the values from both rounds are used – it is the change in the variables between rounds that is of interest. As discussed in the main text, in the fixed-effects framework adopted for this evaluation, the GIS-model travel times from both survey rounds are exogenous with respect to household response variables.

A number of "exploratory" analyses were done using selections of all of the Round 0 GIS-model travel times as covariates. In the end, it was seen that including a single one of them, town1000tt0, produced the best results. This was the only GIS-model travel time covariate included in the models presented later (income, access and other), if covariates were included at all. (As a group, the Round 0 GIS travel times are highly intercorrelated, and including many of them leads to unstable models. As part of the analysis of them, a principal components analysis was done, and a canonical correlation analysis was conducted relating the GIS-model and questionnaire travel times. These analyses will be described later.)

While the preceding variables may have an important effect on a response variable, the primary relationship of interest is the relationship of (changes in) the response variable to ((forced) changes in) the treatment variables. In most instances, the effect of covariates that may have a strong relationship to an outcome variable of interest is "washed out" by differencing over time (or using a fixed-effect estimate) (since many household variables are time-invariant, within a household), and whether these covariates are included in the model or not has little effect on estimation of impact. (Even if an explanatory variable has a strong relationship to outcome, if there is no variation with a household, the fixed-effects-model coefficient for that variable is not estimable.) This situation is particularly relevant to the present application, which involves a two-round panel survey in which most households are interviewed in both survey rounds: many variables (such as *caserío* or household characteristics) remain the same or change very little between survey rounds. Therefore, while covariates may have an important effect on the *level* of a response variable at a point in time, they may have little effect *on impact* over time, and it is not surprising when they do not "show up" in an impact model (such as an estimate of a difference measure of impact in a before-and-after study). What matters is not whether a response variable is affected by explanatory variables other than treatment variables, but whether the *nature of the relationship* of an effect of interest (such as a measure of impact) to treatment variables is dependent on other explanatory variables (covariates), and often it is not.

In the analysis that follows, primary attention focuses on regression-model coefficients that are related to treatment. In most instances, for the panel-survey-based estimates of interest in this evaluation, the values of those coefficients depended very little on whether or not other explanatory variables (covariates) were included in the model. (As discussed, this is a reflection of the fact that the data are two-round panel data and fixed-effects estimators are used. Many covariates change little or not at all between survey rounds, so they drop out of the models.)

To facilitate understanding of the analysis methodology, detailed descriptions of the procedures (formula or regression-analysis procedure or computer-program output) for obtaining the impact estimates will be presented in the case of an exemplar outcome measure (NetHHInc). Readers familiar with Stata may execute the Do12TranImpactEstimation.do command file (or view the .log file) to obtain detailed information for the other outcome measures.

The following sections present estimates of impact for the binary-treatment-variable (BTV) model and the continuous-treatment-variable (CTV) model. The analysis of the BTV model is provided as background for the CTV model. The impact results presented in the main text are for the CTV model only.

III.C. Impact Estimators in the Case of Binary Treatment Variables

The main text provides detailed discussion about impact estimators for the case of continuous treatment variables. This section provides some discussion about impact estimators in the case of binary treatment variables, for the case in which the evaluation design is a pretest-posttest-comparison-group design (which corresponds to the "zone of influence" design described in the main text).

The approach suggested in the *M&E Plan* was to use a pretest-posttest-comparison group design, and to use a double-difference measure of program impact. The approach that was implemented in this evaluation differed from that approach in that it included multiple treatment variables – changes in travel time to different places of interest caused by the road improvements – and that those treatment variables were continuous. With the collected data, both approaches could be implemented. The binary-treatment-variable approach was implemented by deriving binary treatment variables from the continuous treatment variables. In either case, the impact estimators are similar in concept – they are represented by terms (usually interaction terms) in a regression model. The simplification that occurs in the case of a single binary treatment variable is that the impact estimate corresponds to a single regression coefficient.

In an evaluation design that involves a binary treatment variable (treated vs. untreated), a double-difference estimate is the difference, between the treatment sample and the control sample, of the difference in means of an outcome measure between the baseline and endline (or follow-up) surveys. The standard evaluation design used to obtain data for constructing double-difference estimates of program impact is the pretest-posttest-randomized-control-group design.

In the present evaluation, the design was constructed to provide high precision for estimating impact represented as a partial treatment effect in a multiple regression model. This was accomplished by stratifying the sample in such a way as to provide adequate variation in variables believed to affect outcomes of interest; the sample was stratified by estimated changes in travel time to be caused by the road improvements and a number of other variables related to outcomes of interest. The design involved data collection at two points in time – the baseline time ("time 0", "before", "pretest") and the endline ("time 1", "after", "posttest", "follow-up"). While this design is oriented to construction of a CTV estimate of impact, it works well for BTV estimator, since an important variable of stratification is distance from project road.

It is important to distinguish between the double-difference *measure* of impact and the double-difference *estimator*. The double-difference *measure* is the double difference of the four group true (population) means, $(\mu_{11} - \mu_{10}) - (\mu_{01} - \mu_{00})$, where μ_{11} = mean of treatment group at endline, μ_{10} = mean of treatment group at baseline, μ_{01} = mean of control group at endline, and μ_{00} = mean of control group at baseline. The double-difference *estimator* is a sample-based function such as the double difference of the four sample means or a regression coefficient. That is, the double-difference measure is a *population* characteristic (parameter), and the double-difference estimator is a statistic based on a *sample*. The double-difference measure may be estimated by a statistic that is not equal to the double-difference in sample means (e.g., the double-difference measure may be estimated by the regression coefficient of a treatment variable in a model that contains a number of covariates). We further describe these concepts below.

For a pretest-posttest-comparison-group design, the (unadjusted, "raw") double-difference estimate is given by the following formula:

$$DD_{raw} = (\bar{y}_{t1} - \bar{y}_{t0}) - (\bar{y}_{c1} - \bar{y}_{c0})$$

where

DD_{raw} = double-difference estimate (raw, unadjusted)
$\bar{y}_{t1}$ = mean outcome for treatment sample at time 1
$\bar{y}_{t0}$ = mean outcome for treatment sample at time 0
$\bar{y}_{c1}$ = mean outcome for control sample at time 1
$\bar{y}_{c0}$ = mean outcome for control sample at time 0.

The preceding statistic is also called the observed treatment effect (OTE). In the preceding, the variable $\bar{y}_{ij}$ refers to any outcome variable of interest, such as income. The means (averages) referred to are "design-based" ("weighted") sample estimates that take into account the nature of the probability sampling used in the sample survey used to collect the data (e.g., stratification, multi-stage sampling, and selection with varying probabilities). For example, the means and their estimated variances may be estimated using Horvitz-Thompson estimation procedures.

For a pretest-posttest-randomized-control-group design, the double-difference estimator is an unbiased estimate of the double-difference measure. For pretest-posttest designs that are not based on randomized assignment to treatment, the double-difference estimator is not necessarily an unbiased or consistent estimate of the double-difference measure, and more complicated estimators, such as regression estimators, must be used to obtain an unbiased estimate of the double-difference measure.

Estimation of Double-Difference Measure When Randomized Assignment to Treatment Is Used

For design-based estimates, the mathematical (statistical) model used as a basis for constructing estimates of interest and conducting tests of hypotheses of interest describes the sample design and sample selection procedures used to collect data for the evaluation design. If a randomized experimental design is used (and if it is assumed that no variables that affect outcome change over time differently for the treatment and control samples), there is no need to specify an explicit causal model to obtain causal estimates – causal estimates of impact may be estimated directly from the sample, using design-based estimates (since randomized assignment to treatment assures that the distributions of explanatory variables other than treatment are the same for the treatment and control groups).

Estimation of Double Difference Measure When Assignment to Treatment Is Not Based on Randomization

For a design based on a binary treatment variable, if the experimental units (road segments) are not selected using randomization it is necessary to describe the selection process in the causal model. The standard approach to causal modelling in this case is the potential outcomes model (or counterfactuals model, or Neyman-Fisher-Cox-Rubin approach). This approach may be implemented either using the Rosenbaum-Rubin approach (or "statistical" approach) or the Heckman latent-variable approach (or "econometric" approach). This model applies to the road segments of the traffic surveys (since the road segments are the units of treatment). It may also be applied to the binary-treatment-variable ("zone of influence") representation of the household survey, but, as discussed in the main text, this approach has conceptual difficulties associated with it (because treatment is applied to roads, not to households, resulting in an artificial, "fuzzy" classification of "treatment" and "control" households, and attenuated impact / low power). Under this conceptual framework, each sample unit (household in a particular survey round, or road in a traffic survey) is considered to possess two alternative possible outcomes, conditional on the project intervention. The difficulty associated with this approach is that for a particular sample unit, only one of the two potential outcomes can be directly observed – whichever is observed is a "counterfactual" for the other. The estimate of impact is obtained by comparing groups of similar individuals under the alternative treatment specification (project intervention or no intervention). With randomized assignment of treatment, it is straightforward to construct unbiased estimates of impact. With a lack of randomization, the properties of impact estimates depend on assumptions made about the model (such as conditional independence of the counterfactual responses and treatment, given the values of covariates), and the estimators are more complicated (e.g., matching estimators and regression-adjusted estimators).[15]

III.D Impact Estimators of Interest[16]

We use the following impact estimators for this analysis:

1. Ordinary-Least-Squares (OLS) regression estimator of the average treatment effect (ATE) based on *binary treatment variables* (BTVs, derived from the GIS travel-time model (which is used to define the boundaries of the zone of influence of the project roads))
2. OLS regression estimator of partial treatment effect (PTE) based on *continuous treatment variables* (CTVs, derived from the GIS travel-time model and the traffic survey data), and an estimator of the average treatment effect derived from this model

Note that since the CTV estimates are based on the traffic survey data and the GIS travel-time model, they are *conditional* estimates. Vehicle speeds were estimated from a sample of project

[15] The counterfactual approach to impact estimation is described in *Causality: Models, Reasoning, and Inference, 2nd edition* by Judea Pearl (Cambridge University Press, 2009, 2000). For a summary, see *Counterfactuals and Causal Inference: Methods and Principles for Social Research* by Stephen L. Morgan and Christopher Winship (Cambridge University Press, 2007). See also "Statistics and Causal Inference" by Paul W. Holland (*Journal of the American Statistical Association*, Vol. 81, No. 396 (Dec., 1986)).

[16] The estimators used to assess program impact were discussed in detail in the *Analysis Plan*, and were summarized in the introduction to this chapter. The mathematical notation used for the estimation formulas follows *Econometric Analysis of Cross Section and Panel Data*, 2nd edition, by Jeffrey M. Wooldridge (MIT Press, 2010, first edition 2002)). Many of the formulas presented in this reference pertain to the case of a single cross-section of data, and must be modified as appropriate for panel data.

roads for which the improvements had been completed, and a similar sample of untreated roads. These estimates are conditional on completion of the road improvement, and maintenance of the roads to maintain the improvement. The speed estimates from the traffic surveys are also conditional on a number of variables, including season, day of week, time of day, and weather. The CTV estimates are also conditional on manifestation of indirect benefits as estimated from the PTE model.

Both of these estimators can be obtained by linear regression – either ordinary least squares or two-stage least squares (if instrumental variables are involved). The regression models include the design parameters associated with the sample-survey design, such as time (baseline or endline), *caserío* and household, and other explanatory variable (covariates) such as family size, education, assets and travel distances, times or costs to places of interest.

The BTV and CTV estimates are of interest for different reasons. The CTV estimator was selected as the estimator of choice in project design meetings held at the beginning of the project. The primary rationale for using this estimator is that it is considered to be a more accurate representation of the system under study. It recognizes that households are affected by road improvements to a continuously varying degree, depending on their location in the road network. This representation has greater face validity than the binary-treatment-variable representation, which classifies a household simply as “treated” or “control” based on its time or space distance from a project road, ignoring the fact that in a road network the travel times of all households to points of interest are affected to some extent by road improvements (anywhere in the network), and there are no unaffected controls.

With the BTV approach, it is not possible to have a clear demarcation between treatment units and control units. (As discussed in the main text, this difficulty arises since the unit of treatment is the road, not the household. The device of classifying households as “treatment units” or “control units” by means of “buffer zones” around treatment and comparison roads is artificial and imprecise.) All households in the road network experience some effect of the program intervention. With the CTV approach, in a sense, *every household serves both as a treatment and as a control compared to every other household*. The CTV approach recognizes that the effect of the road improvement varies over the population, and it explicitly represents this in a continuous-treatment-variable model. The face validity of the CTV model is substantially higher than that of the BTV model.

Despite the preference for the CTV model on conceptual grounds, we nevertheless also consider estimators based on binary treatment variables, for several reasons. First, for the binary treatment variables, the impact estimates are based solely on the household survey data (given the boundaries of the zones of influence), whereas for the continuous treatment variables the estimates are based also on the traffic surveys and the GIS travel-time model (which used the traffic survey data to estimate speeds as a function of road characteristics, and then to estimate travel times from the speeds). Second, the BTV estimators are easy to explain and understand, although they are not as efficient as the CTV estimators (since each binary treatment variable contains less information than the continuous treatment variable from which it is derived). Third, they are the traditional (conventional) basis for estimating the impact of road-improvement projects (in which households within a certain distance or time (zone of influence, “buffer zone”) are considered to be treatment units and those beyond that limit are control units. Finally, the BTV approach was suggested in the *M&E Plan*. Since an alternative approach was suggested (i.e., the CTV approach), it is of interest to compare the two approaches, to assess whether this was a good decision.

Both the BTV and CTV estimators may be used to estimate impact by road type. The BTV estimators are presented in this report by road type, but the CTV estimators are presented only for all road-improvement projects combined. The CTV estimators could have been presented by road type, but a decision was made early in the project to construct a single set of travel time estimates, corresponding to the changes induced by all road improvement projects combined. (In view of the weak impacts observed this was, in retrospect, a good decision.)

The analysis that follows presents results for both of the estimators listed above. For a number of reasons, more confidence is placed in the CTV estimates. These reasons include: (1) the reasonableness (face validity) of the causal model; (2) statistical tests of the validity and reliability of the statistical models (e.g., goodness-of-fit tests; specification tests); and (3) an *ex post* statistical power analysis, which showed that the tests of hypothesis based on the CTV models were substantially more powerful than those based on the BTV models. Only the CTV estimates are presented in the main text of this report (although both the BTV and CTV estimators are discussed).

Under certain conditions (such as conditional independence), both of the impact estimators defined above are consistent estimators of impact (i.e., of the average treatment effect or the partial treatment effect). Under reasonable assumptions that apply to this project, the preceding estimators are consistent estimates of impact (the expected value of the sample estimate converges to the desired population value as the sample size becomes large).

For descriptive surveys (intended to provide estimates of population means and totals), the formulas (or numerical algorithms) for the estimates are fixed (determined) by the survey design. For analytical surveys (as in the present application), the objective is to estimate parameters of an associated causal model, and there is no single mathematical formula or algorithm for accomplishing the model development. A number of estimators are available, and their performance depends on the application. While it may be confusing to examine a number of different estimators, doing so is desirable (since some work better than others in different situations, and which ones work better in the present application is not known in advance of conducting the analysis). It is because no fixed formula is available for constructing an analytical model that detailed description is presented in this report about the model development and model-based estimation process.

Regardless of the estimator used, it is necessary that it take into account the design features. In many cases the estimator has the same value whether the design is correctly accounted for or not, but to obtain correct estimates of the standard errors of the estimates, the design characteristics must be correctly represented in the model. Apart from the selection of the project roads, the principal design feature for this evaluation is the fact that (in most instances) the same households are interviewed in both survey rounds (i.e., the design is a “strongly balanced” panel design). Once this longitudinally-matched-pairs feature has been taken into account, most other design features (e.g., *caserío*) and covariates (e.g., whether a farmer owns his own land) are of less importance (mainly, because the covariates do not change much over time (within a household)).

In the sample design used for this project, the probabilities of selection of the primary sample units (*caseríos*) are variable, and are a prominent feature of the evaluation design. They are determined by the stratification of households according to estimated travel-time change (to be caused by the program intervention) and other explanatory variables. The selection probabilities are used in two ways. First, models are constructed with and without consideration of the selection probabilities (i.e., with and without "weights," where the weight for a household is the reciprocal of its probability of selection, normalized), and compared. If the two models are similar, this may serve as evidence that the model specification is correct (or, more properly, as a lack of evidence that the model specification is not correct)[17]. If the two models differ substantially, this is taken as evidence that the model specification is not correct, and a better model specification is sought. If a better specification cannot be found (either in terms of the same variables, or by adding other variables), then consideration is given to use of weighted estimates. The Stata *xtreg* program that was used extensively in the analysis does not accommodate weights, and so using weights is done only in particular circumstances (e.g., in analysis of a single survey panel, or by transforming the data using procedure *xtdata*, and using non-panel procedures (such as *regress*) that allow weights).

The second way in which the survey weights are used is to estimate changes in the levels of explanatory variables over time, for use in making conditional estimates of impact for the CTV model (this will be explained in detail later). The estimates of the mean levels of explanatory variables in the two survey rounds are constructed using the sample weights (which are the reciprocals of the probabilities of selection, normalized).

All of the impact estimators considered here take into account the fact that the evaluation design is a pretest-posttest (longitudinal, panel, two-round) design in the same way; however, the way that each estimator accommodates the treatment variable differs somewhat depending on whether treatment is represented by a binary or continuous variable, or by a single variable versus multiple variables. The basic statistical model is the same – a general linear statistical model – but the model configuration differs substantially in the various cases. In all cases, however, the estimators are similar to double-difference estimators of impact (or the interaction effect of treatment variables and time), particularly since inclusion of covariates in models made little difference.

For the binary indicators, the process of double differencing removes the mean levels of variables in the four design groups (treatment before, treatment after, control before, control after). It is important to realize that the impact estimators measure an effect that is similar to a double difference, or interaction of treatment and time. This effect is not a level, and it is not an increase. If it is reported that the income effect of treatment is 10,000 lempiras, this does not mean that income increases on average by this amount for households near project roads if a road improvement is made. Rather, it means, for the BTV-based models, that the incomes of the near-project-road households after four years are about 10,000 lempiras more than the incomes of far-from-project-road farmers. By that same token, if an impact estimator is negative, it does not imply that the program caused, or is even associated with, a decrease in the associated outcome variable. It is important to keep this distinction in mind when reviewing the impact tables presented in this report. (For the CTV-based models, the impact represents the expected change

[17] Note that there are other tests for specification, such as the "principle of conditional error" or the so-called "Hausman" test, which compares model parameters for random-effects and fixed-effects specifications.

(partial treatment effect) associated with a unit (forced) change in an explanatory variable for a randomly selected household in the country.)

An impact indicator variable that is strongly correlated with another may be used as a surrogate, or alternative, estimate for it. For example, since income from basic grains is about two to three times total expense for basic grains, the income effect for basic grains is about two to three times as large as the expense effect. Since reported income is not considered to be as accurate as reported expense, the expense impact times three may be a better estimate of the income effect than the income effect estimated from reported incomes.

III.E Assumptions about the Stochastic Nature of Explanatory Variables

As discussed in the main text, estimates of impact are based on fixed-effects estimators. For all panel regression models considered in the analysis (i.e., all models estimated using Stata procedure *xtreg*), a Hausman test was applied to test the equivalence of fixed-effects and random-effects models. The Hausman test is a test of whether the unobserved variables of the model are correlated with the explanatory variables. The results of these test showed that the two model specifications were similar in many instances, but also differed in many instances. To estimate the effect of a variable that does not change within households it is necessary to use a random-effects specification (the level of such variables may be represented through interaction terms of the variable with survey round).

III.F. Impact Estimates

1. OLS Regression Estimator of the Average Treatment Effect (ATE) based on Binary Treatment Variables (OLS/ATE/Binary-Treatment-Variable model)

As discussed earlier, an unbiased estimate of the ATE may be obtained from a regression model that expresses outcome as a function of explanatory variables. The basic regression model on which impact estimates are based is the following:

$$y_t = \mathbf{x'}_t\boldsymbol{\beta} + \theta d_t + \phi w_t + \delta d_t w_t + e_t, \qquad (1)$$

where

t = survey round index (0 for Round 0 and 1 for Round 1)
y_t = outcome variable (explained variable, response variable, dependent variable)
$\mathbf{x}_t$ = vector of explanatory variables (the first component is one)
$\boldsymbol{\beta}$ = vector of parameters (the first parameter is a constant term)
d_t = indicator variable for survey round, = 0 for Round 0 and 1 for Round 1
θ = round effect
w_t = treatment variable
ϕ = treatment effect (not the impact, but the average difference in means between the treatment and control groups at baseline)
δ = impact (interaction effect of treatment and round)
e_t = model error term.

(The usual convention of representing row vectors in boldface and denoting a column vector or matrix transpose with a prime will be adhered to.) The model error term is assumed to have mean zero, constant variance, and be uncorrelated with the explanatory variables. As discussed in the main text, the condition of zero correlation of the explanatory variables and the model residuals follows from the assumption of a fixed-effects framework (for explanatory variables outside the household), and the fact that the estimates are made conditional on the baseline conditions (e.g., a fixed project). In the model, explanatory variables outside of the household are considered as fixed variables, not as stochastic variables. The stochastic properties of within-household variables will be discussed later, for each specific model considered. Unobserved variables are assumed to be time-invariant. (In general, the presence of unobserved variables biases the estimates of the regression coefficients. In a fixed-effects, two-round panel-data model, unobserved variables that are time-invariant drop out of the estimation model, and do not contribute to bias.)

Measurement error (errors in variables) is not considered to be a serious issue in this application. (Measurement error, if present, introduces an attenuation bias in the regression coefficient estimates.)

In this application (binary treatment variables), the treatment variable, w_t, is one of the binary travel-time variables Trtprit, Trtsect or Trtrurt (as defined in Section III.B). In this model formulation, the value of the treatment indicator variable, w_0, varies over the Round 0 sample units depending on whether a household is located within a specified travel time (or distance) of a project road, and the Round 1 value (w_1) is identical to the Round 0 value for a particular household. In this model formulation, the estimate of impact is the coefficient of the interaction effect of treatment and round. In the preceding specification, with rounds 0 and 1, $d_t = t$.

The preceding model formulation is appropriate if there is no interaction between the treatment variable and the other explanatory variables, or "covariates." If this assumption is not valid, then it is necessary to include interaction terms between treatment and the covariates. If this is done, the covariate factor of the interaction term must be the deviation from the mean. *This is very important.* If the covariate factor is not demeaned, then the coefficient of the interaction of treatment and round will not be an unbiased estimate of impact. If $\boldsymbol{\mu}$ denotes the mean of the covariate, $\mathbf{x}$ (i.e., $E(\mathbf{x}) = \boldsymbol{\mu}$), then the additional term is $d_t w_t(\mathbf{x}_t - \boldsymbol{\mu})$.

In the present application, we examined models with and without the interaction terms between treatment and demeaned covariates, and determined that in many cases the covariate interaction terms were not necessary (or even desirable). (Note that we are discussing *covariate* interaction terms here – many models include an interaction effect of treatment and Round.)

Some additional comments about the preceding models are the following. The (vector) parameter $\boldsymbol{\beta}$ contains not only substantively (economically) meaningful explanatory variables, such as farm size or educational level of the head of household, but also design parameters, such as survey round, *caserío* and household. The sample consists of 1,600 households in round 0 and 1,408 in round 1, so (in a fixed-effects model) there are hence 1,408 household parameters (coefficients). The particular values of these parameters are of no interest, but they are essential to include in the model in order to obtain a correct estimate of the standard error of the parameter of interest, viz., δ. There is one household indicator variable for each household. These parameters are "nuisance" parameters. They are explicitly represented in the undifferenced model described above, but not in a first-difference model (any variable that has the same value in both rounds falls out of the differenced model).

The situation is similar for the *caserío* sample units (100 nuisance parameters). Once the household indicator variables are included in a model, the importance of the *caserío* indicator variables is diminished. A number of regressions were done with and without the *caserío* indicator variables, and very little difference was observed in the estimates and their standard errors. In the interest of model simplicity and reasonableness, the *caserío* indicator variables were dropped (there are about 100 *caseríos*, so there are about 100 *caserío* indicator variables, and a loss of 100 degrees of freedom associated with estimation of their effects). (The *caserío* indicator variables might be of greater significance in a random-effects model, where the component of variance associated with *caseríos* might be substantial, even after inclusion of the household variables.)

In this (BTV) model, the GIS-model travel times are used in two ways:

(1) To define the binary treatment variables for the three road types (by defining the boundaries of the zones of influence around project roads); and
(2) To construct exogenous travel-time variables (either treatment variables such as town1000tt (in both Rounds) or covariates such as town1000tt0 (from Round 0 only – the covariates may or may not be included in models).

In the course of the analysis, regression models were considered that included a number of covariates, i.e., explanatory variables other than the treatment variables (models were also developed that included no covariates). For income and expenditure response variables, the final set of covariates included (if covariates were include in the model at all) household size (HouseholdSize), education (MeanEduc), farm size (TotHaOwnFarm), and travel time to nearest town of population 1,000 or more (town1000tt0). For access response variables and the other response variables (school attendance, use of health care facilities, and employment), these included just MeanEduc and town1000tt0.

Because of intercorrelations among explanatory variables, there are a variety of model specifications available, and it is often not possible to place a lot of confidence in a particular regression coefficient estimate for a covariate (the effects are confounded). The main purpose in including a selection of covariates in a model is to increase the precision of the coefficients on treatment-related variables, not to estimate the regression coefficient for a particular covariate.

The decision to include town1000tt0 as a single travel-time covariate was based on a number of considerations. First, the collection of GIS-model travel-time covariates were highly intercorrelated. Including all or many of them resulted in impact models that were unstable (correlated coefficients with large standard errors). A principal components analysis was conducted to assess the nature of the correlation among them. Some output of the principal components analysis of seven of the travel times is shown in the table that follows (the mca* travel-time variables were not included in this analysis, since they did not represent places of high interest to most survey respondents). (Note: Figure titles and numbers will not be assigned to small tables, such as the one presented below, that are not referenced later in the text.) This analysis showed that the first principal component (denoted as “Comp1”) accounted for 42.17% of the variation of the set, as measured by the size of the eigenvalues of the correlation matrix.

```
. pca pritt sectt rurtt tegustt sanpedtt town1000tt top10tt if Round==0

Principal components/correlation                  Number of obs    =      1557
```

```
                                          Number of comp.  =          7
                                          Trace            =          7
Rotation: (unrotated = principal)         Rho              =     1.0000

    --------------------------------------------------------------------------
       Component |   Eigenvalue   Difference         Proportion   Cumulative
    -------------+------------------------------------------------------------
           Comp1 |      2.95204      1.64339             0.4217       0.4217
           Comp2 |      1.30865      .433791             0.1869       0.6087
           Comp3 |      .874858      .051016             0.1250       0.7336
           Comp4 |      .823842      .400211             0.1177       0.8513
           Comp5 |      .423631     .0591129             0.0605       0.9119
           Comp6 |      .364518      .112055             0.0521       0.9639
           Comp7 |      .252464            .             0.0361       1.0000
    --------------------------------------------------------------------------
```

The multiple correlation coefficient of each of the seven GIS-model travel-time variables was determined by regressing each of them on the others. All seven had high multiple correlations with the others, and the one that had the highest multiple correlation with the others was town1000tt0. The regression model for town1000tt0 on the other travel times is shown in following table.

```
. regress town1000tt pritt sectt rurtt tegustt sanpedtt top10tt if Round==0

      Source |       SS       df       MS              Number of obs =    1557
-------------+------------------------------           F(  6,  1550) =  391.33
       Model |  167310.461     6  27885.0768           Prob > F      =  0.0000
    Residual |  110448.383  1550  71.2570215           R-squared     =  0.6024
-------------+------------------------------           Adj R-squared =  0.6008
       Total |  277758.844  1556  178.508255           Root MSE      =  8.4414

------------------------------------------------------------------------------
  town1000tt |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       pritt |    .347923   .0147239    23.63   0.000     .3190422    .3768038
       sectt |   .2031283   .0128186    15.85   0.000     .1779846     .228272
       rurtt |  -.0149509   .0379406    -0.39   0.694    -.0893712    .0594695
     tegustt |  -.0202452   .0028112    -7.20   0.000    -.0257595    -.014731
    sanpedtt |   .0233301   .0023323    10.00   0.000     .0187553     .027905
     top10tt |  -.0323884   .0057329    -5.65   0.000    -.0436335   -.0211432
       _cons |   2.850112   .5986261     4.76   0.000      1.67591    4.024315
------------------------------------------------------------------------------
```

The R^2 value (coefficient of determination) presented in the table is the square of the multiple correlation coefficient. Of all of the seven GIS-model travel times, town1000tt had the largest value of R^2 (.6024) when regressed on the others. (The Round 0 travel times are indicated with suffix “0”. With the restriction to Round 0, the regression model is identical whether the suffixed or unsuffixed travel times are used. The regression model presented above uses the unsuffixed variables, but the results would have been identical had the suffixed variables been used (since it is restricted to Round = 0).) Based on this fact, and also the fact that town1000tt0 appeared to bear the strongest relationship to a wide variety of outcome indicators of interest, town1000tt0 was selected as a single GIS-model travel-time covariate to include in models. It is recognized that the travel-time variable having the strongest relationship to an outcome variable may differ from outcome variable to outcome variable. Because of the high intercorrelation among travel-time variables, however, this difference is not great, and in the interest of simplicity, since so many models were examined, the same travel-time covariate was used for all models.

We developed regression models using the OLS/ATE/binary-treatment-variable estimator for all of the outcome variables, by road type. Before presenting the results of these models for all outcome variables, we shall present a detailed analysis for one outcome variable, NetHHInc, in the case of primary roads. The binary treatment variable for primary roads is Trtprit.

Here follows a table of means of NetHHInc for households in the treatment and control groups defined by the binary treatment variable Trtprit (the BTV for primary roads), by survey round.

```
. tabulate Round Trtprit, summarize (NetHHInc)

          Means, Standard Deviations and Frequencies of NetHHInc

           |       Trtprit
     Round |         0          1 |     Total
-----------+----------------------+----------
         0 |  82179.38  70391.857 | 77051.807
           | 95203.337  80436.992 |  89246.25
           |       904        696 |      1600
-----------+----------------------+----------
         1 | 99871.775  96489.826 | 98430.604
           |  127221.7  112454.97 | 121119.37
           |       808        600 |      1408
-----------+----------------------+----------
     Total | 90529.529   82474.25 | 87058.903
           | 111782.85  97423.245 | 105893.38
           |      1712       1296 |      3008
```

From this table it is seen that the raw double difference estimate of impact for primary roads is

$$DD_{raw} = (96490 – 70392) – (99872 – 82179) = 8,405.$$

This estimate does not take into account features of the design, such as *caserío* or household. These features affect not only the value of the estimate, but also the estimate of its standard error. The salient feature of the design is that most households are interviewed in both survey rounds. The regression estimate shown in the following table (Figure A1) takes into account the design features.

Instead of the general linear model formula presented above, the following simpler version is used to estimate the double difference, taking into account the design features (but no covariates):

$$y_t = \beta_0 + \beta_1 \text{ Round} + \beta_2 \text{ Treated} + \beta_3 \text{ RoundTreated} + e_t.$$

"Treated" refers to the treatment variable for a particular type of road, e.g., it is Trtprit for primary roads, Trtsect for secondary roads, and Trtrurt for rural roads. The variable RoundTreated stands for the interaction of Round with the treatment variable. The results are somewhat easier to compare if the variables are demeaned. In the data analysis, the symbols RTP, RTS and RTR were used for these three road types (i.e., RTP for primary, RTS for secondary, and RTR for rural). The estimated design-adjusted double difference is the coefficient of the RoundTreated interaction term, β_3 (i.e., the interaction effect of treatment and time).

With respect to model specification and identification, all of the explanatory variables included in the preceding model are fixed effects, and therefore not correlated with the model error term (e_t). The key assumption required for use of the general linear statistical model estimator (viz., uncorrelatedness of the explanatory variables with the model error terms) is hence satisfied. Unobserved variables are assumed to be time-invariant (in which case they drop out of the two-round panel-data fixed-effects model). Alternative variance estimators were examined (e.g., Stata "vce(robust)" option) to address the issue of heteroskedasticity of variances.) (Note: In the analysis, a number of regression models were run under alternative assumptions about the variances. In what follows, this explanation will not be repeated, if it applies.)

In the case of primary roads, the variable RTP is equal to RoundStd * TrtpritStd, where RoundStd denotes the demeaned Round and TrtpritStd denotes the demeaned Trtprit (the "R" in RTP stands for Round, the "T" stands for "treatment" (or "Trtprit") and the"P" stands for "primary"). The regression analysis used to estimate the design-adjusted double difference is shown in Figure A1. The impact estimate is the RTP effect (the interaction effect of Round and the binary treatment variable), which (for the fixed-effects model) is 7,711, with a standard error of 6433. The Student's t value associated with this estimate is 1.20, indicating that the impact is not statistically significant. (Since the sample size is large, the critical values for the Student's t statistic are approximately equal to those for the normal distribution, viz., 1.96 for two-tailed tests and 1.645 for one-tailed tests.

Note that taking the design features into account makes a very great difference in the size of the standard error of the estimate. It the design is ignored and the observations are considered independent (i.e., it is not recognized that in most cases households are interviewed in both survey rounds), the standard error of the preceding estimate (8,405) would be sqrt $(127222^2/808 + 95203^2/904 + 112455^2/600 + 80437^2/696) = 60{,}431$, whereas the standard error for the estimate (7,711) that accounts for the design is 6,433 – one tenth as large!

Regression model with no covariates

The model for this example is:

$$y_t = \text{NetHHInc} = \beta_0 + \beta_1 \text{ Round} + \beta_2 \text{ Trtprit} + \beta_3 \text{ RTP} + e_t,$$

where RTP = RoundStd*TrtpritStd, RoundStd is demeaned Round, and TrtpritStd is demeaned Trtprit.

With respect to model specification, all explanatory variables represent fixed effects, and hence they are not correlated with the model error terms.

Figure A1. Double-Difference Estimate of Impact, Taking into Account Design Features but No Covariates, for NetHHInc, for Primary Roads

```
. xtreg NetHHInc Round Trtprit RTP, fe

Fixed-effects (within) regression               Number of obs      =      3008
Group variable: idhh                            Number of groups   =      1600

R-sq:  within  = 0.0284                         Obs per group: min =         1
       between = 0.0066                                        avg =       1.9
       overall = 0.0105                                        max =         2

                                                F(2,1406)          =     20.51
corr(u_i, Xb)  = 0.0078                         Prob > F           =    0.0000

------------------------------------------------------------------------------
    NetHHInc |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       Round |   20053.59   3181.534     6.30   0.000     13812.53    26294.66
     Trtprit |  (dropped)
         RTP |   7711.064   6433.365     1.20   0.231    -4908.963    20331.09
       _cons |   77689.13   2141.588    36.28   0.000     73488.08    81890.19
-------------+----------------------------------------------------------------
     sigma_u |  87957.678
     sigma_e |    84411.8
         rho |  .52056268   (fraction of variance due to u_i)
------------------------------------------------------------------------------
F test that all u_i=0:     F(1599, 1406) =     2.05          Prob > F = 0.0000
```

The regression model produces estimates that are similar in magnitude to the raw double difference estimate, but they provide an estimate of the standard error of the estimate that is based on the design. As mentioned, the value of the Student's t statistic for this estimate is 1.20, indicating that this estimate is not statistically significantly different from zero.

Regression model with covariates

The next model is one that includes covariates. A number of different selections of covariates were examined, and the one that worked best for income and expense response variables included HouseholdSize, MeanEduc, TotHaOwnFarm and town1000tt0. For the other response variables (access, use of health services, school attendance, employment), the ones that worked best were MeanEduc and town1000tt0. In all cases, the covariates are included in the model as interactions of the demeaned RoundStd, demeaned treatment variable (e.g., TrtpritStd), and demeaned covariate (e.g., MeanEducStd). For NetHHInc, the output from the regression program for that model is shown in Figure A2.

The model for this example is:

$$y_t = \text{NetHHInc} = \beta_0 + \beta_1\ \text{Round} + \beta_2\ \text{Trtprit} + \beta_3\ \text{RTP} + \beta_4\ \text{RTPHouseholdSize} + \beta_5\ \text{RTPMeanEduc} + \beta_6\ \text{RTPTotHaOwnFarm} + \beta_7\ \text{RTPTown1000tt0} + e_t,$$

where RTP = RoundStd*TrtpritStd, RoundStd is demeaned Round, and TrtpritStd is demeaned Trtprit.

With respect to model specification, Round and Trtprit are fixed, and the covariates included in the model are considered to be exogenous (sequentially exogenous, predetermined) with respect to NetHHInc, given the baseline conditions and various fixed-effects assumptions about the project and the household survey. The covariates may be considered as random variables, i.e., the model is a "mixed-effects" model, containing both fixed and random effects. Over a considerable period of time, the covariates may be considered to be mutually causally related to NetHHInc, and this relationship would affect estimation of the model parameters (i.e., OLS would not be appropriate). Over the relatively short period of the project term, however, the three covariates are considered to be exogenous with respect to NetHHInc. If this assumption is viewed as unpalatable, then this "covariates included" model and estimator should not be used. Instead, use the "no covariates" estimator given earlier. It is noted, however, that the effect of including covariates is small. The covariates are weak, and there is little difference between the "covariates included" and the "no covariates" models. Unobserved variables are assumed to be time-invariant.

To summarize, none of the explanatory variables included in the model is endogenous – the model does not include any questionnaire travel times, the GIS-model travel time is exogenous (under the fixed-effects framework), and the three covariates are considered to be little affected by the dependent variable over the term of the study. Unobserved variables are assumed to be time-invariant. Conditional on the baseline and the fixed-effect framework, the explanatory variables of the model satisfy the condition of uncorrelatedness with the model error term, and the method of ordinary least squares (OLS) may be used to construct the estimates. (Over a long period of time, many household variables could be considered endogenous relative to income. This project spans a relatively short period of time and has relatively little impact on income, however, and it is unlikely that variables such as household size, education, and farm size would be much affected by changes in income. Given the baseline conditions, they are assumed to be uncorrelated with the model error terms.) Note that the preceding discussion about endogeneity refers only to the "covariates included" models. If the assumptions made are unacceptable, then the "no covariates" models (which are similar) should be used.

Note that although the treatment variable is binary, nonbinary (continuous) travel-time variables (such as the Round 0 travel times to points of interest) may be included in the model as explanatory variables.

Figure A2. Double-Difference Estimate of Impact, Taking into Account Design Features and Covariates, for NetHHInc, for Primary Roads

```
. xtreg NetHHInc Round Trtprit RTP RTPHouseholdSize RTPMeanEduc RTPTotHaOwnFarm
RTPtown1000tt0, fe

Fixed-effects (within) regression               Number of obs      =      2924
Group variable: idhh                            Number of groups   =      1557

R-sq:  within  = 0.0369                         Obs per group: min =         1
       between = 0.0034                                        avg =       1.9
       overall = 0.0124                                        max =         2

                                                F(6,1361)          =      8.70
corr(u_i, Xb)  = -0.0015                        Prob > F           =    0.0000

------------------------------------------------------------------------------
    NetHHInc |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       Round |    19199.4   3329.189     5.77   0.000      12668.5     25730.3
     Trtprit |  (dropped)
         RTP |   6856.083   6573.794     1.04   0.297    -6039.785    19751.95
RTPHouseho~e |   6139.936   2972.313     2.07   0.039     309.1235    11970.75
 RTPMeanEduc |  -5846.949   3212.433    -1.82   0.069    -12148.81    454.9071
RTPTotHaOw~m |   2054.215   1308.463     1.57   0.117    -512.6067    4621.038
RTPtown100~0 |  -1218.506    555.134    -2.19   0.028    -2307.517   -129.4953
       _cons |    78679.8   2214.251    35.53   0.000     74336.08    83023.51
-------------+----------------------------------------------------------------
     sigma_u |  88623.171
     sigma_e |  85229.789
         rho |  .51951126   (fraction of variance due to u_i)
------------------------------------------------------------------------------
F test that all u_i=0:     F(1556, 1361) =     2.03          Prob > F = 0.0000
```

It is seen that, in the case of NetHHInc for primary roads, the inclusion of covariates in the binary-treatment-variable model did not make a substantial difference in the size of the estimate of impact. The estimated impact for this model is 6,856, with an estimated standard error of 6,574 and a Student's t statistic of 1.04. This compares to the estimate 7,711 with standard error 6,433 for the no-covariate model. In this example, there was no advantage to including covariates, relative to the precision of the impact estimate (even though some of the covariates were statistically significant). The impact is still not statistically significant (Student's t = 1.04). For most response variables, including covariates in the model did not make a significant difference, but in a few it did.

The reason why covariates are of little significance is that, within the same household, many variables have the same values for both survey rounds, i.e., are time-invariant. For a two-round panel survey, the fixed-effects estimator may be represented as a first-difference model, in which case such (time-invariant) variables drop out of the model. This situation holds for the models that follow, and this explanation will not be repeated. (An advantage of the fixed-effects models is that unobserved ("hidden") time-invariant variables also drop out of the (estimation) model.)

In summary, the analysis of NetHHInc for primary roads did not show a significant impact. Although including covariates in the model did not result in a significant impact as measured by the RTP interaction effect, it is noted that some of the covariates are statistically significant. While this may seem to be contradictory, it is not – some variables may have a statistically significant relationship to a response variable in a model, without the treatment effect being significant, overall (on average). This is the situation for NetHHInc. For some response variables, statistically significant results were obtained both for the overall treatment effect and one or more covariates.

Impact Estimation for the Full Set of Outcome Variables, for All Three Road Types

Having completed the example using NetHHInc with primary roads, we now present the impact estimates for all response variables and all three road types (primary, secondary and rural). Each estimate of impact requires a separate regression analysis. Since there are a total of 29 outcome variables of interest and three road types, this is a total of 87 separate models (of which the NetHHInc / primary road example just discussed is a single example). In every case, the average treatment effect is the coefficient of the interaction effect of RoundStd and the binary treatment variable, Trtprit, Trtsect, or Trtrurt (which, in the program output contained in the Do12 file, is denoted as RTP for primary road, RTS for secondary roads, and RTR for rural roads).

Note that the estimator considered in this section is a single coefficient in a regression model, viz., the coefficient of the interaction effect of RoundStd with a binary treatment variable.

Note that "household" is a design variable that is implicitly included in all regression models. It is specified to the Stata *xtreg* procedure as a "panel" variable. Regressions were run with and without *caserío* indicator variables, and no significant differences were observed, so *caserío* is not included in the model.

Table A.11 presents estimates of ATE for the three road types, using the OLS regression estimation procedure, for all of the outcome variables. For this table, the covariates-included regression models are used (although it does not matter much (over all models) whether covariates are included, either for the value of the impact estimate (coefficient of RTP, RTS, or RTR) or its standard error).

Each cell of the table is an estimate derived from a separate regression model. (The full regression output is presented in Figure A2 above just for primary roads and the outcome variable NetHHInc.) For each regression, the regressand (explained variable) is the "Outcome Variable" specified in column 1 of Table A.11, and the regressors (explanatory variables) are the binary treatment variable (one of Trtprit, Trtsect or Trtrurt) and various design and questionnaire variables.

The estimated average treatment effect (ATE, impact) is the coefficient of RoundTreated (denoted in the .log file as RTP for primary roads, RTS for secondary roads, and RTR for rural roads), the interaction effect of treatment and time. Note that the components of IncEmp (IncEmpAg and IncEmpNonAg) do not sum to IncEmp (as they did for the raw data). This is not an error, but a result of the separate estimation of the three effects.

Table A.11. Ordinary-Least-Squares (OLS) Regression Estimate of Average Treatment Effect (ATE) based on Binary Treatment Variables, for Selected Outcome Measures

Outcome Variable	Project Component					
	Primary Roads		Secondary Roads		Rural Roads	
	Estimate	Standard Error	Estimate	Standard Error	Estimate	Standard Error
Household Income and Expenditure						
IncEmp (monthly)	689	448	131	464	351	471
IncEmpAg (monthly)	-101	194	163	202	430*	204
IncEmpNonAg (monthly)	849*	327	119	339	-215	345
TotHHExp (monthly)	180	249	-5103	257	-487	261
NetHHInc (annualized)	6856	6574	-416	6810	1638	6814
Access						
CostToSchool	.132	.248	.059	.256	-.0056	.260
TimeToSchool	-.317	.776	.318	.802	-.788	.813
CostToCollege	-.381	3.92	5.74	4.05	-4.87	4.11
TimeToCollege	-3.44	2.29	11.9	2.35	4.09	2.40
CostToHospital	14.06	7.87	-16.2*	8.13	5.16	8.27
TimeToHospital	5.48	2.86	5.50*	2.94	-1.92	3.02
CostToHealthCtr	.790	1.05	-1.91*	1.09	1.71	1.10
TimeToHealthCtr	-3.41*	1.82	7.98*	1.86	-6.20*	1.91
CostToMarket	-6.35*	2.59	.513	2.69	-.473	2.72
TimeToMarket	-2.35	2.60	1.43	2.69	-4.54	2.73
CostToPulp	1.22	.665	-.597	.689	.0314	.699
TimeToPulp	1.19	.763	-.612	.790	.821	.801
TimeToTegus	4.88	4.99	2.44	5.12	1.44	5.17
TimeToSPS	9.37	8.99	21.7*	9.19	-21.26*	9.33
TimeToDepCap	6.98	5.58	3.39	5.78	-.973	5.75
TimeToMunCap	-3.73	4.12	11.9*	4.30	-8.58*	4.32
School Attendance						
ChldInSch712	.114*	.0509	-.0754	.0527	.0235	.0536
ChldInSch1318	.0208	.0407	-.0791	.0420	.0634	.0427
Use of HealthCare Services						
VisHospital	-.0786	.0758	.132	.0781	-.0531	.0796
VisPrHlthCtr	.0554	.0670	-.0381	.0694	.0024	.0707
VisPubClinic	-.220*	.115	-.235*	.115	-.0514	.130
VisNonProf	.0569	.0785	-.0162	.0813	-.104	.0825
VisPharm	-.115*	.0519	.0266	.0537	-.106*	.0545
Employment						
WrkdPrevWk	..0846	.0705	-.0411	.0730	.0224	.0740

Note: Income, expense and travel costs measured in Honduran lempiras; travel times measured in minutes

The results in Table A11 fail to reveal a strong the impact of the road improvement interventions. Statistically significant effects (of the anticipated sign) are indicated with an asterisk (*). (Two sided tests are used for effects that may be of either sign (positive or negative); for two-sided test, effects that exceed their standard errors by a factor of 1.96 are considered to be statistically significant. One-sided tests are used for effects that are expected to be of a particular sign; for two-sided test, effects that exceed their standard errors by a factor of 1.645 are considered to be statistically significant. As mentioned, on the indication of statistical significance, the interpretation is that over many independent investigations, the probability that the confidence interval includes the true value of the parameter is approximately .95. Confidence intervals within the same investigation are correlated.) A few travel-cost and travel-time effects are statistically significant and of the expected sign, but are small in magnitude. Some are statistically significant but not of the expected sign (e.g., time and cost to hospital for primary roads). Variables relating to use of health care services showed a few statistically significant decreases. The income components IncEmp and IncEmpNonAg are statistically significantly positive for primary roads, but small. IncEmpAg shows a statistically significant positive impact for rural roads.

There are a number of possible reasons why the project impacts, as measured by the binary treatment variable, are small in magnitude and, in some cases, not statistically significant. These include the following:

- The boundary of the zone of influence, separating “treatment” households from “control” households is “fuzzy.” Because of the interconnectedness of the road network, a road improvement has a “ripple effect” throughout the network, and affects all households to some extent. For this reason, there is not a clear separation of treatment and control households, and the effect of the program is diluted. (Note that it is somewhat inaccurate to refer to “treatment” and “control” households – the unit of treatment is the road segment, not the household. Defining those households that are located within a zone of influence as “treatment” households is a “rough approximation” associated with the binary-treatment-variable approach. This is an inherent problem associated with using the “buffer zone” approach to assessing road projects. The continuous-variable-approach (to be examined) overcomes this limitation.)
- The endline household survey for the evaluation was completed prior to the completion of all road improvements. Some of the road segments were still under construction when the endline survey was conducted and, could conceivably have had negative impact on network travel times.
- Improvements to rural roads are of short duration, generally on the order of six months, until the next rainy season. Rural road improvements under the MCA program were completed early in on in the project; as such, many of these roads had succumbed to the detrimental effects of rain storms and landslides by the time the endline data collection occurred.
- A significant lapse of time may be required before noticeable results would manifest themselves in some of the indicators of interest. For example, while we would expect to see immediate changes in travel times and costs following the completion of a road

improvement intervention, changes in income, employment, and health-seeking behavior are likely to take several months or a year or more to manifest. Given the timing of the endline survey, which was dictated by the end date for the MCA Honduras compact, this evaluation did not allow for such a lag time.

- The Transportation Project was implemented in conjunction with and, partly, in support of the MCA - Honduras Farmer Training and Development Assistance (FTDA) Project, to provide improved means of moving horticultural products to market. Since the two project components were conducted in some of the same area over the same time period, the effects of the projects may be confounded. The continuous-treatment-value estimates of impact are conditional on travel times, and therefore not sensitive to confounding with the FTDA project, which had, apart from the rural-road component that was incorporated into this evaluation, nothing to do with travel time.
- The recent global recession began about when the Transportation Project and FTDA Project began. If the recession was more severe for people living near these projects or served by these projects, the impact of the Transportation Project would be diluted.

The CTV estimator to be addressed overcomes the preceding limitations.

Ex Post Statistical Power Analysis

One of the issues to address with respect to Table A.11 is whether the small number of statistically significant results is an indication of low power. That is, the sample size may not be sufficiently large to detect effects of anticipated or realized size. Statistical power analysis was done at the beginning of the project to estimate sample size. That *ex ante* power analysis was complicated by the fact that the standard error of the impact estimates was not known (at that time, prior to the survey). For that reason, the power analysis was based on a model that involved a number of parameters about the test, the population under study, and the sample design. Now that the data analysis has been completed, estimates are available for the standard errors of the impact estimates, and an *ex post* (or *post hoc*) power analysis may be conducted much more easily than the *ex ante* power analysis. It depends on just the test parameters (significance level; test direction (one-sided or two-sided)) and the standard error of the impact estimate. The significance level, α, of the test is the probability of a Type I error of making a decision that the effect (impact) is present (different from zero) when it is not. The probability of a Type II error of making a decision that the effect is not present when it is, is β. The power is 1 – β.

There are a number of indicators that may be examined in an *ex post* power analysis. Two standard indicators are the power of the test to detect a true effect equal in magnitude to the observed effect, and the minimum detectable effect (MDE) that can be detected for a specified level of power, which we shall set at 90%. The formula for the first indicator is:

$$Power = prob\left(\hat{t} \geq \hat{t}_{1-\beta}\right), where\ \beta\ is\ defined\ by\ \hat{t}_{1-\beta} = t_{critical:\alpha/2} - \frac{\hat{\Delta}}{\hat{\sigma}_{\Delta}}$$

for a two-sided test and

$$Power = prob\left(\hat{t} \geq \hat{t}_{1-\beta}\right), where\ \beta\ is\ defined\ by\ \hat{t}_{1-\beta} = t_{critical:\alpha} - \frac{\hat{\Delta}}{\hat{\sigma}_{\Delta}}$$

for a one-sided test, where $\hat{\Delta}$ denotes the impact estimator and $\hat{\sigma}_{\Delta}$ denotes the standard error of this estimate. (The power formulas and notation presented here are from David M. Murray, *Design and Analysis of Group-Randomized Trials*, Oxford University Press, 1998.)

The formula for the second indicator is:

$$\hat{\Delta} = \hat{\sigma}_{\Delta}(t_{critical:\,\alpha/2} + t_{critical:\,\beta})$$

for a two-sided test, where α = .05 and β = 1 – power = .1, and

$$\hat{\Delta} = \hat{\sigma}_{\Delta}(t_{critical:\,\alpha} + t_{critical:\,\beta})$$

for a one-sided test.

For α = .05 and β = .1, the critical t values are

$$t_{critical:\,\alpha/2} = 1.96;\ t_{critical:\,\alpha/2} = 1.645;\ and\ t_{critical:\,\beta} = .84.$$

Two other indicators of interest in an *ex post* power analysis are the power to detect an effect equal to 10 percent of the mean of an outcome variable of interest and the power to detect an effect equal to 10 percent of the standard deviation of an outcome variable of interest. These are standard cases often considered in *ex ante* power analysis, and it is of interest to estimate the power for these two cases after the data have been analyzed and values are known for the various parameters that were unknown at the beginning of the study. The power is calculated for these two indicators from the same formula given above (for the first indicator) , simply by substituting the effect size (ten percent of the mean or standard deviation) in place of $\hat{\Delta}$.

It is also of interest to calculate the ratio of the standard error of the estimate to the mean and to the standard deviation. These indicators are related to the two just described. The latter one is of interest for estimating the (Kish) design effect of the study.

The following table (Table A.12) presents the indicators just described, for a selection of the outcome variables, for primary roads. The table is constructed using one-sided tests, in which case, for α = .05 and β= .1 the value of $t_{critical:\alpha} + t_{critical:\beta} = 1.6449 + 1.2816 = 2.9265$.

The power to detect an effect equal in magnitude to the observed effect is shown in column 4 of the table. This indicator is of interest only for the larger effects, since if a true effect is small, the power to detect it will be, too. The minimum detectable effect for a test of power 90% is shown in column 5. The most interesting indicators are shown in columns 8 and 9 – the power to detect effects equal in magnitude to ten percent of the variable mean and standard deviation.

The results presented in the table reveal that the power of the design to detect effects of the size observed is not high (column 4), for many outcome variables. The sample size was estimated (in the *ex ante* power analysis done at the beginning of the project) to provide high power (.9) for effects that were on the order of ten percent of the standard deviation (the standard deviation was used instead of the mean, since suitable information on the variable means was not available). From column 11 of the table, it is seen that few of the observed effects were that large. In summary, the evaluation design provides low power for detecting effects for many variables, based on BTV models. High power (90%) was present only for detecting effects related to InEmpNonAg.

A revealing indicator of the power of the design is the ratio of the standard error of the estimated impact to the variable mean. This is shown in the penultimate column of the table. For an effect to be statistically significant, it has to be about twice as large as the entry in this column, as a fraction of the mean. This means that, for primary roads, impacts would have to be a substantial proportion of the mean, in order to detect them with high power. It is not expected that a road improvement project would result in impacts equal to twice the entry in the penultimate column, as a fraction of the mean, for some outcome variables. (For example, the relative standard error of the estimate of impact for NetHHInc is .0853. Twice this is .1706. This means that for the impact of NetHHInc to be statistically significant, the effect would have to be about 17 per cent of the mean NetHHInc. This magnitude change may be expected from a road improvement project for NetHHInc (since NetHHInc is probably moderately correlated with ERR, and the pre-project economic analysis estimated that economic rates of returns (ERRs) of this magnitude would occur as a result of the project (17% for CA-5 and at least 12% for rural roads), but it is unreasonable to expect this magnitude change for all indicators. As another example, the relative standard error of IncEmpAg is .2576. Twice this is .5152. It does not seem reasonable to expect an impact of this magnitude (relative to the mean, i.e., 52% of the mean) for IncEmpAg.)

Table A.12, for primary roads, shows that, for a number of outcome variables, the power associated with binary-treatment-variable estimates of impact is too low to detect the small changes that would be expected from road improvement projects. The results are even weaker for secondary and rural roads, and are not presented here.

The last column of Table A.12 is useful for estimating the design effect of the study. For estimation of double differences, the standard deviation of the double difference estimator if simple random sampling is used for all four design groups (if of equal size) is 4σ/sqrt(n), where σ denotes the standard deviation and n denotes the total sample size for all four groups. The value of the design effect is deff = (standard error of estimate) / (4σ/sqrt(n)) = (sqrt(n)/4) (standard error of estimate) / (Round 0 sd). From the last column, it is seen that the average value (over the outcome variables) of the ratio of the standard error of the estimate to the Round 0 standard deviation is about .07. The value of n is 3008, so deff is approximately equal to .07 sqrt(3008)/4 = .96, or about 1. This indicates that the interviewing of households in both rounds compensated for the loss in precision associated with multistage sample (of *caseríos*).

Table A.12: *Ex Post* Statistical Power Analysis for Selected Impact Estimates Based on Binary Treatment Variables, for Primary Roads

Outcome Variable	Estimate of Impact (observed effect)	Standard error (se) of estimate	Power of test to detect a true effect equal in magnitude to the estimate of impact	Minimum detectable effect (MDE) for a test of power 90%	Round 0 mean	Round 0 standard deviation (sd)	Power to detect a true effect equal to .1 of Round 0 mean	Power to detect a true effect equal to .1 of Round 0 sd	Estimate of impact relative to Round 0 mean	Estimate of impact relative to Round 0 sd	Standard error of estimate relative to Round 0 mean	Standard error of estimate relative to Round 0 standard deviation
IncEmp	689.	448.	.457	1311.	4632.	6033	.277	.391	.148	.114	.0967	.0743
IncEmpAg	-101.	194.	.132	568.	753.	2218	.106	.312	-.134	-.040	.2576	.0875
IncEmpNonAg	849.*	327.	.828	957.	3779.	5058	.319	.471	.235	.168	.0865	.0647
TotHHExp	180.	249.	.179	729.	5245.	4066	.687	.503	.034	.044	.0475	.0672
NetHHInc	6836.	6574.	.274	19239.	77052.	89246	.325	.395	.089	.077	.0853	.0737
TimeToSchool	-.317	.776	.110	2.27	14.28	13.56	.589	.552	-.022	-.023	.0543	.0574
TimeToMarket	-2.35	2.60	.231	9.61	82.17	58.53	.950	.757	-.029	-.040	.0316	.0444
ChldInSch712	.114*	.0509	.723	.149	.874	.997	.529	.623	.130	.114	.0582	.0511
VisPubClinic	-.220*	.115	.606	.337	.64	1.55	.142	.397	-.344	-.142	.1797	.0742
WrkdPrevWk	.0846	.0705	.329	.206	1.80	1.23	.817	.540	.047	,069	.0392	.0573

What is readily apparent from this table is that the power associated with estimates of impact based on binary treatment variables is low. From one point of view, this is not very surprising, since the evaluation design was intended to construct impact estimates based on *continuous* treatment variables, not on *binary* treatment variables. Had the study been designed to provide high precision for BTV impact estimates, the sample would have been concentrated in and balanced among buffer zones near improved roads and in comparable control areas (buffer zones just beyond the treatment buffer zones and around comparable non-improved roads), rather than marginally stratified with respect to explanatory variables. While this would have increased power somewhat, however, it would not have made a large difference.

A more important fact leading to low power for many of the BTV-based estimates and to high power for the CTV-based estimates to be discussed in the next section, is that the BTV-based estimates are based only on the household survey data available from both survey rounds, whereas the CTV-based estimates are based not only on the household survey data from both rounds, but also on the pre- and post-project traffic survey data and on the GIS model used to estimate project-induced changes in travel times (from the traffic survey data and from road characteristics such as elevation variation). The GIS-based model estimates of travel times and travel-time changes reflect the travel times and travel-time changes that *would have occurred* had all road improvements been completed and maintained to the level observed in the final traffic survey. This can make a large difference in the estimation of impact. The BTV-based estimates are based solely on the household survey data, and hence reflect the fact that the road improvements were not completed for some treatment roads, and that the improvements may have not lasted (e.g., in the case of rural roads, for which improvements are known to last in some cases for just on the order of six months (prior to the next rainy season)).

Since so few BTV-based impact estimates are statistically significant, and since the paucity of significant results is considered to stem from low power, these estimates are not presented in the main text. It is noted that the low power and subsequent lack of utility of the BTV impact estimates is not necessarily considered to represent an intrinsic characteristic of the evaluation design, but more likely of its implementation, specifically, the timing of the second survey round. The *ex ante* statistical power analysis estimated a sample size based on detectable effects equal to 10-20 percent of the standard deviation of an outcome variable (e.g., a 10-20 percent change in the value of a proportion). While the design was focused on construction of CTV estimates, not BTV estimates, that fact is not the reason why the BTV estimates are of low power. Use of the CTV model should improve power somewhat (over that of the BTV model), but not a lot (since a regression coefficient is similar to a difference, and the loss in information associated with replacing continuous treatment variables with binary treatment variables is not expected to be great). It is considered, instead, that the low power of the BTV estimate is the result of the facts that for some roads the road improvements were not completed in sufficient time to observe effects, and that some of the road improvements did not last (e.g., unpaved rural roads, which often "wash out" when the next rainy season arrives). Without incorporating the effects of the traffic surveys (done for completed treatment roads and a comparison sample) and the GIS model (which estimated travel times and travel-time changes conditional on completion and maintenance of the road improvements), the power of the BTV and CTV estimates would have been similar (since, as noted, the regression coefficients of the CTV model are similar to differences, and the process for estimating impact in the CTV model is similar to estimation of a double difference, as used in the BTV model).

It is noted that the Transportation Project was done in conjunction with and in support of the MCC-funded Farmer Training and Development Assistance (FTDA) Project, which was implemented in some of the same areas and over the same time period as the Transportation Project. The estimated effects (impacts) of the Transportation Project and the FTDA Project are confounded. The estimated impacts may reflect the influence of both projects. Because the road improvement projects were conducted in geographically limited areas and the FTDA Project was nationwide, it is considered that the degree of confounding may be modest. It is expected that the confounding of the income outcome variables would result in an increase in the observed impact (i.e., a positive bias), since the Transportation and FTDA projects are complementary (i.e., synergistic effects are anticipated).

It was possible to use the sample data to construct useful BTV estimates, and this was therefore investigated, but it was not a goal of the evaluation or a requirement for the evaluation design. It might have provided useful results, but it turned out that it did not. It is emphasized that the fact that the BTV estimates of impact are of low power and the CTV estimates (to be discussed in the next section) are of high power has little to do with the lower efficiency of the BTV estimates, and much to do with the fact that the CTV estimates are conditional on the GIS-model estimates of travel-time changes induced by the undertaken road improvements, taking into account the pre- and post-project traffic survey data (for completed road-improvement projects) and making the assumptions that the travel-time changes for uncompleted improvements would have been similar to those for completed improvements and that the effect of the road improvements lasts (i.e., the roads are maintained after being improved). *The CTV estimates (of the next section) are based on the assumption that the road improvements would have been completed for all project roads, and would have lasted.*

Intra-unit correlation coefficients

In the *ex ante* statistical power analysis that was done at the beginning of this project to estimate sample size, one of the key parameters involved in the calculations was the intra-unit correlation coefficient for outcome variables of interest, at two levels of sampling (*caserío* and household). That parameter was not known for any specific outcome variable, and “nominal” values of .1 and .5 were assumed, for *caseríos* and households, respectively. Once the survey data are available, the intra-unit correlation coefficient can be calculated for various levels of aggregation. These values are not of direct interest to the analysis presented in this report, but they would be of interest to assist power analysis and sample size estimation for future studies. Here follows a table of the intra-unit correlation coefficients for the outcome variables of this study, for various levels of aggregation (household, *caserío, aldea*, municipality, and department). The lower levels of aggregation (household, *caserío, aldea*) are the ones of interest for use as sampling units in multistage sampling. (The intra-unit correlation was not calculated for all levels of sampling for all variables).

The intra-unit correlation coefficients are estimated by using the Stata procedure *loneway*. Here follows a sample output (for variable NetHHInc at the household level). (Both rounds of survey data were used to calculate the intra-unit correlations for household, and the Round 0 (baseline) data were used to calculate the intra-unit correlations for the higher levels.) The program output included the estimated intra-unit correlation and its standard error. The standard errors are not included in the table shown below, but are included in the .log file. As a general rule, intra-unit correlations are positive and increase as the size of the sample unit increases. The intra-unit correlations are estimated from an analysis of variance procedure, and are restricted to be positive (the zero entries in the table represent truncated estimates).

In the *ex ante* power analysis done at the beginning of the project, "nominal" values were assumed for the intra-unit correlations. The intra-unit correlation associated with households was assumed to be .5, and the intra-unit correlation associated with *caseríos* was assumed to be .1. It is seen from the table that the intra-unit correlations at a particular level of sampling vary substantially over the various outcome variables, but that these assumed values were reasonable "nominal" values for households and *caseríos*.

```
. loneway NetHHInc Caserío if Round==0

                  One-way Analysis of Variance for NetHHInc:

                                                Number of obs =      1600
                                                    R-squared =    0.0325

    Source                 SS         df       MS            F      Prob > F
-------------------------------------------------------------------------
Between Caserío       4.144e+11      23     1.802e+10      2.30      0.0004
Within Caserío        1.232e+13    1576    7.818e+09
-------------------------------------------------------------------------
Total                 1.274e+13    1599    7.965e+09

         Intraclass       Asy.
         correlation      S.E.        [95% Conf. Interval]
         ------------------------------------------------
            0.02129     0.01376       0.00000     0.04826

         Estimated SD of Caserío effect          13041.86
         Estimated SD within Caserío             88420.59
         Est. reliability of a Caserío mean       0.56605
              (evaluated at n=59.96)
```

Table A.13. Intra-unit correlation coefficients for sampling units of various sizes

Variable	Sampling unit (level of sampling in multistage sampling)				
	Household	*Caserío*	*Aldea*	Municipality	Department
IncEmp (monthly)	.324	.025			
IncEmpAg (monthly)	.311	.027			
IncEmpNonAg (monthly)	.427	.045			
TotHHExp (monthly)	.465	.069			
NetHHInc (annualized)	.347	.021	.040	.033	.048
CostToSchool	0	.013			
TimeToSchool	.462	.109			
CostToCollege	0	0			
TimeToCollege	.524	.27449			
CostToHospital	.020	.053			
TimeToHospital	.639	.215			
CostToHealthCtr	.355	.020			
TimeToHealthCtr	.523	.228			

Table A.13. Intra-unit correlation coefficients for sampling units of various sizes					
Variable	Sampling unit (level of sampling in multistage sampling)				
	Household	*Caserío*	*Aldea*	Municipality	Department
CostToMarket	.254	.027	.053	.100	.083
TimeToMarket	.619	.318	.217	.262	.323
CostToPulp	0	.010			
TimeToPulp	.288	.067			
TimeToTegus	.861	.188			
TimeToSPS	.810	.316			
TimeToDepCap	.382	.162			
TimeToMunCap	.366	.362			
ChldInSch712	.544	.012	.012	.020	.027
ChldInSch1318	.397	.007			
VisHospital	.199	.002			
VisPrHlthCtr	.132	.012			
VisPubClinic	.154	.013	.009	.018	.033
VisNonProf	0	.007			
VisPharm	.012	.003			
WrkdPrevWk	.470	.007	.007	.011	.007

2. Regression Estimator (OLS) of Partial Treatment Effect (PTE) Based on Continuous Treatment Variables (OLS/PTE/Continuous-Treatment-Variable Model)

[The material in this section on model specification and effect identification and estimation is similar to that presented for the preceding model. Some paragraphs are repeated, so that each section may be read independently.]

The CTV estimation of impact consists of two parts: (1) a regression model of the PTE and (2) the estimation of impact from a linear combination (vector inner product) involving the regression model coefficients and changes in the means of treatment-related variables between the two survey rounds (i.e., before and after the program intervention). The PTE regression model is of the same form as the BTV model (equation (1)), except that the treatment variables are continuous, not binary. In this case, the coefficient of a treatment variable is an estimate of the *partial treatment effect* (PTE) of that variable. The partial treatment effect is the marginal change in the outcome variable associated with a unit change in an explanatory variable. (This is true also for the BTV estimates, but in a model with a single binary treatment variable, the "partial" treatment effect represented by this single variable is in fact the total treatment effect.) The continuous travel-time variables in this model are travel times to nearest points on nearest MCA primary road (mcapritt), nearest MCA secondary road (mcasectt), nearest MCA rural road (mcarurtt), nearest primary road (pritt), nearest secondary road (sectt), nearest rural road (rurtt), travel time to Tegucigalpa (tegustt), travel time to Tegucigalpa (sanpedtt), travel time to nearest *caserío* of population 1,000 or more (town1000tt) and travel time to top 10 Honduran cities, ranked by population (top10tt). The estimated partial treatment effects are the coefficients of these variables.

A distinguishing feature of the OLS/PTE/continuous-treatment-variable model is that, in contrast to the OLS/ATE/ binary-treatment-variable model, it makes use of the GIS-model travel-time estimates for *both* survey rounds. The partial treatment effects are with respect to the travel-time changes (variations) estimated by the GIS travel-time model. Hence, the validity of this model is vested not only in the validity of the sample-survey data, but also in the validity of the GIS travel-time model, which we believe to be high. The GIS travel-time model is based on a carefully constructed road network model for Honduras, which was based on data from a series of traffic surveys conducted by MCA - Honduras for this evaluation, conducted before and after the program intervention for treatment and comparison roads. The GIS road network travel-time model that was used in this project was based on GIS road data obtained from the Government of Honduras and private sources, and on traffic surveys conducted before and after the project road improvements.

A problem that arises immediately with the continuous-treatment-variable model is that it is not practical to explicitly include all ten GIS-model travel-time variables, or even just seven of them (after dropping the three mca* variables). As discussed earlier when identifying covariates for the BTV models, the set of ten or seven travel-time variables are highly intercorrelated, and models based on all of them are not stable (i.e., have coefficients that are intercorrelated and have large standard errors). The problem is compounded even more for the CTV models, since if covariates are included in the model, it is necessary to include a demeaned interaction term (by round) for each treatment variable and each covariate. There is hence a veritable explosion in the number of model parameters, and the results are rather useless (unstable, confounded / correlated effects, large standard errors).

To address this problem, the same approach is used as in the case of covariates for the BTV models, viz., the effect of the road improvements are represented by a single continuous treatment variable. The logic for selecting this single variable is the same as presented earlier – the single treatment variable is town1000tt. (Note that the treatment variable is town1000tt, the travel time in either survey round, not just town1000tt0, the travel time in Round 0, considered as a covariate in the BTV model.) With this representation, the estimation model includes (as explanatory variables) the continuous treatment variable (town1000tt) and (demeaned) interactions of Round, town1000tt and covariates of interest, if any. For the economic response variables (income and expense), the covariates are HouseholdSize, MeanEduc, and TotHaOwnFarm. For the access and other response variables, the covariates are just MeanEduc (since there is no reason to include the other two, from a causal-model viewpoint).

After conducting the analysis, it was observed that inclusion of the covariates made little difference in the magnitudes of the estimates of impact or their standard errors. Nevertheless, the covariates-included models are of interest to identify covariates that have a significant relationship to impact. *For estimation of impact, it is advantageous to use the no-covariate models since the standard errors of the impact estimates tended to be somewhat lower (and the magnitudes of the estimates were similar).* The reason for this is that, unlike the case of the BTV, where the impact estimate is (mainly) represented in a single model coefficient, the impact estimate for the CTV case is a linear combination of all treatment-related coefficients (times the difference in means between rounds for each treatment-related coefficient), and the variance of these linear combinations is quite sensitive to how many nontreatment-related variables are included in the linear combination. The linear combinations for the covariates-included models contained several more terms, and had substantially larger variances (squares of the standard errors).

This situation should be kept in mind in the planning of future road-evaluation projects. For CTV models, efforts should be made to keep the number of treatment-related variables low, to promote small standard errors for the impact estimates. This was not a serious issue with the BTV models, because the impact was reflected in a single regression coefficient, rather than in a linear combination of several regression coefficients times differences in means between rounds.

Figure A3 below presents the regression models for NetHHInc, for the covariates-included and no-covariate cases. In the model output, Round denotes survey round (0 or 1), trt denotes the treatment variable (= town1000tt), and Rtrt denotes the Round by treatment interaction term (Rtrt=RoundStd*trtStd, where RoundStd is the demeaned Round variable and trtStd is the demeaned treatment variable). In the course of developing these models, a number of other models were examined. For example, we developed regression models using just the Round 0 data, and we developed "first difference" models based on the difference in variables (both response variables and explanatory variables) between the two survey rounds. These models (which are not shown here, but are included in the Stata output (.log) file) were used as "checks" on the models presented in Figure A3, which include the (undifferenced) data from both survey rounds. (The simpler models were developed using the Stata *regress* procedure; the models that used data from both rounds were developed using the Stata *xtreg* procedure.)

In the BTV models considered earlier, the main estimate of impact was one of the regression-model coefficients (an interaction term of treatment and time). For the CTV models, the estimate of impact is not represented in a single model coefficient. Instead, the model coefficients represent *partial treatment effects*. The estimate of impact is obtained as a linear combination (sum) of those coefficients multiplied by the changes in means of the treatment-related variables between the two survey rounds. How this estimate is constructed, and how its standard error is estimated, will now be explained. That estimate is labeled "Impact" in the regression-model output. (More than one model coefficient reflects treatment in the CTV model, even if there is but a single treatment variable, because the interaction of Round by that treatment variable is also a treatment-related variable. Hence the minimum number of treatment-related variables in the CTV model is two.) Note that the coefficient of the Round by treatment interaction is *not* an estimate of *total* impact (as it was in the BTV model), but an estimate of the *partial* treatment effect.

The model for the examples shown in Figure A3 are as follows.

Regression model with no covariates:

$$y_t = \text{NetHHInc} = \beta_1 \text{ trt} + \beta_2 \text{ Rtrt} + \beta_3 \text{ Round} + \beta_0 + e_t.$$

where y_t denotes an output variable; trt is the treatment variable, equal to town1000tt; Round = survey round (0 or 1); Rtrt = RoundStd*trtStd, where RoundStd and trtStd denote the demeaned Round and trt variables; e_t denotes the model error term, and the β's are the model parameters (regression coefficients).

With respect to model specification and identification, all of the explanatory variables included in the preceding model are fixed effects, and therefore not correlated with the model error term

(e_t). The key assumption required for use of the general linear statistical model estimator (viz., uncorrelatedness of the explanatory variables with the model error terms) is hence satisfied. Unobserved variables are assumed to be time-invariant (in which case they drop out of the two-round panel-data fixed-effects model).

Regression model with covariates:

$$y_t = \text{NetHHInc} = \beta_1\ \text{trt} + \beta_2\ \text{Rtrt} + \beta_3\ \text{RTPTown1000tt0} + \beta_4\ \text{RTPHouseholdSize} + \beta_5\ \text{RTPMeanEduc} + \beta_6\ \text{RTPTotHaOwnFarm} + \beta_7\ \text{Round} + \beta_0 + e_t,$$

where HouseholdSize, MeanEduc and TotHaOwn farm are as defined earlier.

With respect to model specification, Round and Trtprit are fixed, and the covariates included in the model are considered to be exogenous (sequentially exogenous, predetermined) with respect to NetHHInc, given the baseline conditions and various fixed-effects assumptions about the project and the household survey. The covariates may be considered as random variables, i.e., the model is a "mixed-effects" model, containing both fixed and random effects. Over a considerable period of time, the covariates may be considered to be mutually causally related to NetHHInc, and this relationship would affect estimation of the model parameters (i.e., OLS would not be appropriate). Over the relatively short period of the project term, however, the three covariates are considered to be exogenous with respect to NetHHInc. If this assumption is viewed as unpalatable, then this "covariates included" model and estimator should not be used. Instead, use the "no covariates" estimator given earlier. It is noted, however, that the effect of including covariates is small. The covariates are weak, and there is little difference between the "covariates included" and the "no covariates" models. Unobserved variables are assumed to be time-invariant.

To summarize, none of the explanatory variables included in the model is endogenous – the model does not include any questionnaire travel times, the GIS-model travel time is exogenous (under the fixed-effects framework), and the three covariates are considered to be little affected by the dependent variable over the term of the study. Unobserved variables are assumed to be time-invariant. Conditional on the baseline and the fixed-effect framework, the explanatory variables of the model satisfy the condition of uncorrelatedness with the model error term, and the method of ordinary least squares (OLS) may be used to construct the estimates. (Over a long period of time, many household variables could be considered endogenous relative to income. This project spans a relatively short period of time and has relatively little impact on income, however, and it is unlikely that variables such as household size, education, and farm size would be much affected by changes in income. Given the baseline conditions, they are assumed to be uncorrelated with the model error terms.) Note that the preceding discussion about endogeneity refers only to the "covariates included" models. If the assumptions made are unacceptable, then the "no covariates" models (which are similar) should be used.

To investigate the distributional aspects of the program intervention, regression models were constructed that included distance to nearest project road as a regressor. These models would show whether households closer to the project roads realized greater impacts. These regressions were not significantly different from the models that omitted this explanatory variable. Such effects are considered likely. The fact that they were not detected may be due to low power for detection of this particular effect. The survey design was a national sample of households.

While the distribution of the sample was somewhat heavier in areas closer to project roads, detection of this effect was not a major consideration of the sample design.

Figure A3. Ordinary-Least-Squares (OLS) Regression Estimate of Partial Treatment Effect (PTE) for NetHHInc, Based on Continuous Treatment Variables.

Regression model with no covariates:

```
. xtreg NetHHInc trt Rtrt Round, fe

Fixed-effects (within) regression               Number of obs      =      2924
Group variable: idhh                            Number of groups   =      1557

R-sq:  within  = 0.0296                         Obs per group: min =         1
       between = 0.0063                                        avg =       1.9
       overall = 0.0007                                        max =         2

                                                F(3,1364)          =     13.86
corr(u_i, Xb)  = -0.3025                        Prob > F           =    0.0000

------------------------------------------------------------------------------
    NetHHInc |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         trt |   1746.904   3561.863     0.49   0.624     -5240.42    8734.227
        Rtrt |   420.6705   250.7172     1.68   0.094    -71.16256    912.5036
       Round |   20793.72   3412.582     6.09   0.000     14099.24     27488.2
       _cons |    54951.6   47589.65     1.15   0.248    -38405.24    148308.4
-------------+----------------------------------------------------------------
     sigma_u |  93573.398
     sigma_e |  85460.206
         rho |  .54522372   (fraction of variance due to u_i)
------------------------------------------------------------------------------
F test that all u_i=0:     F(1556, 1364) =     2.01          Prob > F = 0.0000

. matrix b=e(b)

. matrix list b

b[1,4]
          trt        Rtrt       Round       _cons
y1  1746.9038   420.67051   20793.717   54951.603

. matrix c=e(V)

. matrix list c

symmetric c[4,4]
              trt        Rtrt       Round       _cons
  trt    12686869
 Rtrt   186316.72   62859.104
Round   3489875.3   46565.685    11645717
_cons  -1.693e+08  -2483064.6   -51573847   2.265e+09

. EstimateTotalEffectForLastModel Delta2_2
Impact = 422.28396; StdErrorImpact = 608.61871; t = .69383994
```

Regression model with covariates:

```
. xtreg NetHHInc trt Rtrt RTtown1000tt0 RTHouseholdSize RTMeanEduc RTTotHaOwnFarm
Round, fe
```

```
Fixed-effects (within) regression               Number of obs      =      2924
Group variable: idhh                            Number of groups   =      1557

R-sq:  within  = 0.0306                         Obs per group: min =         1
       between = 0.0062                                        avg =       1.9
       overall = 0.0007                                        max =         2

                                                F(7,1360)          =      6.13
corr(u_i, Xb)  = -0.3094                        Prob > F           =    0.0000

------------------------------------------------------------------------------
    NetHHInc |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         trt |   1790.308   3610.151     0.50   0.620    -5291.761    8872.376
        Rtrt |   437.1715    332.397     1.32   0.189     -214.895    1089.238
RTtown1000~0 |  -1.986771   16.13739    -0.12   0.902    -33.64366    29.67011
RTHousehol~e |    113.366   107.6674     1.05   0.293    -97.84626    324.5782
  RTMeanEduc |   57.37012   122.9912     0.47   0.641     -183.903    298.6432
RTTotHaOwn~m |   6.784115   31.53852     0.22   0.830    -55.08532    68.65355
       Round |   21241.72   4490.934     4.73   0.000     12431.81    30051.62
       _cons |    54208.1   48438.81     1.12   0.263     -40814.8      149231
-------------+----------------------------------------------------------------
     sigma_u |  93806.989
     sigma_e |  85541.714
         rho |  .54598728   (fraction of variance due to u_i)
------------------------------------------------------------------------------
F test that all u_i=0:     F(1556, 1360) =     2.00          Prob > F = 0.0000

. matrix b=e(b)

. matrix list b

b[1,8]
                trt            Rtrt    RTtown1000tt0  RTHouseholdSize
RTMeanEduc
y1        1790.3076       437.17151       -1.9867711        113.36597
57.370118

     RTTotHaOwnFarm           Round            _cons
y1        6.7841153       21241.716          54208.1

. matrix c=e(V)

. matrix list c

symmetric c[8,8]
                             trt             Rtrt    RTtown1000tt0  RTHouseholdSize
            trt         13033189
           Rtrt        299121.08        110487.77
  RTtown1000tt0       -8354.7361       -3474.7519         260.4155
RTHouseholdSize       -10883.778       -1284.0996       -28.243335        11592.273
     RTMeanEduc        398.26822       -4685.3916         518.0676         14.33956
 RTTotHaOwnFarm         6456.912        140.20578       -68.470254        215.22184
          Round        5043620.9        662991.36       -44059.339       -29239.774
          _cons       -1.746e+08       -4262771.8        131008.47        160844.51

                      RTMeanEduc   RTTotHaOwnFarm            Round            _cons
     RTMeanEduc        15126.842
 RTTotHaOwnFarm       -205.50876        994.67854
          Round        23785.314        2644.1738         20168487
          _cons       -15410.723       -85909.181        -76098628        2.346e+09
```

```
. EstimateTotalEffectForLastModel Delta2_6
Impact = 39.753187; StdErrorImpact = 1980.6082; t = .0200712
```

The approach to estimating impact is somewhat different for the continuous-response-variable models than for the binary-response-variable models, and we elaborate this approach in some detail for the case of NetHHInc. The procedure starts with the construction of the partial-treatment-effect regression model describing the relationship of NetHHInc to the treatment-related explanatory variables (trt and the various demeaned interaction terms), as shown in the preceding figure. The estimate of impact (total treatment effect) and its standard error are derived from this model.

The total treatment effect of the program intervention is obtained from the regression model (which describes the partial treatment effect) by substituting mean values in the regression equation for each round, and differencing. For travel-time-related variables, these means are the separate means for each round (representing the different travel times in the two rounds). For the other variables, these means are the means over *both* survey rounds. Since the regression model in this case is linear in the explanatory variables, this is equivalent to substituting the changes in means between survey rounds for all travel-time-related variables and zero values for all non-travel-time-related variables, including the constant-term indicator value (which is normally equal to one). (Zero covariate values are entered when the model representation is as shown in Figure A3, in the form of a set of demeaned variables and a constant term. Had the model been represented in terms of non-demeaned explanatory variables, the procedure would be equivalent to substituting the mean value for each covariate. *The total effect is conditional on the (both-round) mean values for all covariates (explanatory variables other than treatment-related variables*).)

In symbolic terms, the impact is determined as follows. As discussed in the main text, a simple conceptual representation of the relationship of outcome to travel time may be presented as:

> *Vehicle speed = f(project intervention (road improvement), road characteristics (primary, secondary, rural; elevation variation), given vehicle type, season, day of week, time of day, weather)*
>
> *Travel time (for pickup truck) = g(road characteristics, mean vehicle speed given road characteristics)*
>
> *Outcome measure = h(travel-time variables and other variables ("covariates"))*
>
> *Impact = Expected value (mean) of outcome measure conditional on completion and maintenance of road-improvement project and on covariate means for both survey rounds – Expected value of outcome measure at beginning of project, conditional on covariate means for both survey rounds*

where f(.), g(.), and h(.) denote continuous functional relationships.

The functional relationship was estimated using the general linear statistical model. Using this approach, an unbiased estimate of the average treatment effect (ATE) may be obtained from a regression model that expresses outcome as a function of explanatory variables. The ATE may be estimated conditional on the values of explanatory variables (**x**), or as an unconditional average (over those variables). The basic regression model on which outcomes of interest are related to explanatory variables, and from which impact estimates are derived, is the following:

$$y_t = \mathbf{x}'_t\boldsymbol{\beta} + e_t,$$

where

t = survey round index (0 for Round 0 or baseline, and 1 for Round 1 or endline)
y_t = outcome variable
$\mathbf{x}_t$ = vector of explanatory variables (design variables, treatment variables, and covariates; one component of which is "1", representing a constant term)
$\boldsymbol{\beta}$ = vector of parameters (one parameter is a constant term)
e_t = model error term.

To obtain unbiased estimates of the model parameters using the ordinary-least-squares (Gauss-Markov) estimation procedure, it is necessary to make assumptions about the model error terms and the relationship of the explanatory variables to the error terms, such as that the model error terms have zero mean, constant variance and are uncorrelated with each other and the explanatory variables. Such assumptions are reviewed for each model constructed in the course of the analysis (since each model has a different specification (set of explanatory variables and functional form)).[18]

For the preceding outcome model, the regression coefficients corresponding to treatment-related variables (i.e., to travel-time-related variables) are partial treatment effects (PTE). This model will be referred to either as an "outcome" model or a "partial treatment effects" model. For this application, the main feature of the model is that all of the explanatory variables outside of the household are considered to be fixed effects, in which case they have zero correlation with the model error term.

Differences in the relationship between rounds are addressed by the inclusion of interaction terms.

The expected value of y_t is given by the same equation as above, omitting the error term:

$$E\ y_t = \mathbf{x}'_t\boldsymbol{\beta}.$$

Identification of Parameters and Effects Related to Impact

An estimate of impact is obtained from the preceding equation by taking the difference of the expectation equation between survey rounds:

---

[18] The models used for this analysis are described in the book, *Econometric Analysis of Cross Section and Panel Data*, 2nd edition, by Jeffrey M. Wooldridge (Massachusetts Institute of Technology Press, 2010, 2002). For additional information, see *Mostly Harmless Econometrics* by Joshua D. Angrist and Jörn-Steffen Pischke (Princeton University Press, 2009); *Counterfactuals and Causal Inference: Methods and Principles for Social Research* by Stephen L. Morgan and Christopher Winship (Cambridge University Press, 2007); *Micro-Econometrics for Policy, Program, and Treatment Effects* by Myoung-Jae Lee (Oxford University Press, 2005; *Analysis of Panel Data* 2nd edition by Cheng Hsaio (Cambridge University Press, 1986, 2003); *Econometric Analysis of Panel Data* by Baki Baltagi, 4th ed., (Wiley, 2008); and *Econometric Analysis* 7th ed. by William H. Greene (Prentice Hall, 2011).

$$Impact = \bar{\boldsymbol{x}}_1'\boldsymbol{\beta} - \bar{\boldsymbol{x}}_0'\boldsymbol{\beta}$$

where each component of $\bar{\boldsymbol{x}}_t$, denoted as $\bar{x}_{ti}$, is as follows: for travel-time-related (treatment) variables the component is the GIS-model mean for round t (t = 0 or 1); and for other variables (covariates) the component is the mean over *both* survey rounds (t = 0 and 1).

The preceding general linear statistical model is linear in the parameters (**β**). It is also linear in the explanatory variables (**x**), so that (in this special case) the formula for impact may be represented as:

$$Impact = \Delta\bar{\boldsymbol{x}}_1'\boldsymbol{\beta}$$

where the subscript on $\bar{\boldsymbol{x}}$ denotes survey round, and where Δ denotes the backward difference operator (defined by $\Delta x_t = x_t - x_{t-1}$). In this representation, the components of $\Delta\bar{\boldsymbol{x}}_1$ are zero for all non-travel-time-related variables and equal to the difference in means between survey rounds for the travel-time-related variables. (Although the preceding model is linear in the explanatory variables, nonlinear representations are allowed in terms of the observed variables, e.g., one explanatory variable may be x and another may be x^2.)

As mentioned, we developed regression models with and without covariates. In the case of the BTV models, the standard error of the estimate did not differ much between the no-covariate and covariates-included models. In the case of the CTV models, the standard error of the estimate is somewhat different for these two models, and tends to be less for the no-covariate model, and so *the no-covariate model is used to estimate impact.* Note that it is only the treatment-related coefficients and their variances/covariances that are involved in estimating the total effect. The impact estimate is obtained by substituting zero values for all of the (demeaned) covariates (including the variable associated with the constant term), so there is no effect from including them. The standard error of this estimator is obtained from the formula for the standard error of a linear estimator based on a general linear statistical model. (The covariates-included model is used to identify covariates that may have a significant relationship to output, since knowing what covariates are related to outcomes was of some interest.)

In quantitative terms, the procedure for estimating the total treatment effect (average treatment effect) from the partial treatment effects (coefficients of treatment values in the regression model) is as follows. Let **Δx** represent the (column) vector of changes in travel times between the two survey rounds, where the last component of the vector represents the constant term and has the value zero. Let **b** denote the (column) vector of parameters of the regression model, where the last component represents the constant term and has the value 1. Let **V** denote the variance-covariance matrix of the estimated parameters of the regression model. Then the estimated impact is **Δx**' **b** and its variance is **Δx**' **V** **Δx**. (Note the ambiguity and confusion in terminology. The "total treatment effect" refers to what is normally called the "average treatment effect," or ATE. Later, we shall refer to the total impact for the country, also called a "total treatment effect," in contrast to the impact estimate **Δx**' **b**, which is the total treatment effect for a randomly selected household. Which meaning is intended for "total treatment effect" should be clear from context.)

In **Δx**, we are concerned only with changes in the treatment-related variables, since those are the only components that are nonzero. Since we are "adjusting" the estimate to the same values of the covariates in both rounds, all of the components corresponding to non-treatment terms are zero. In the no-covariate models, there are just two treatment-related terms: trt and Rtrt. For Round 0, the means of trt and Rtrt are 14.93652 and -.8042389, and for Round 1 the means are 14.78404 and .832805, so that the difference in means between the two rounds for trt and Rtrt are -.15248133 and 1.6370394.

The value of **Δx** is hence

> **Δx**' = (trtd, Rtrtd, 0, 0)
> = (-.1525, 1.6370, 0, 0)

where the "d" suffixed variables are the differences in means for the explanatory variables of the regression model (e.g., trtd = trt1 – trt0 where trt0 denotes the mean of trt in Round 0 and trt1 denotes the value of trt in Round 1.

Note that for estimation of the components of **Δx**, it is the *weighted* means that are used (where the weights are proportional to the reciprocals of the probabilities of selection of the units). The regression model was developed without weights, since if the model is correctly specified the use of weights makes little or no difference in bias and would simply reduce precision. The components of **Δx**, however, represent the changes in the *mean levels* of the explanatory variables over the population under study. It is for that reason that weights are used in estimating the means.

The preceding **Δx** had two non-zero components, corresponding to the treatment-related variables trt and Rtrt. In general, the number of nonzero components in **Δx** is equal to the number of treatment-related variables in the model. Had the covariates-included models been used, there were six treatment-related variables in the income and expense response variables: trt, Rtrt, RHouseholdSize, RMeanEduc, RTotHaOwnFarm and Rtown1000tt0 and four for the other response variables (trt, Rtrt, RMeanEduc and Rtown1000tt0).

From Figure A3, we see (in the case of NetHHInc) that the vector of estimated parameters (**b**, also denoted as $\hat{\boldsymbol{\beta}}$ in the main text) is

> **b**' = (1746.904, 420. 420.6704, 20793.72, 54951.6).

Taking the inner product (**Δx**' **b**) of **Δx** and **b**, we obtain the impact estimate 422.28396 (the large number of significant digits is retained so that the reader may find this number in the regression model output). The variance matrix, V, of the estimated parameters is given in Figure A3 (at the end of the figure). The value of the quadratic form is **Δx**' **V** **Δx** is seen to be 370416.7342, so that the standard error of the impact estimate is 608.61871. The Student's t value of 422.3/608.6 = .69 indicates that this effect is not statistically significant (one-sided test at the .05 level of significance).

Note that the estimate **Δx**' **b** is considered to be *conditional* on the value of **Δx**. The error variance **Δx**' **V** **Δx** includes variation in **b**, but not in **Δx**.

The interpretation of the meaning of the impact estimate for the CTV models is quite different from the interpretation for the BTV models. For the BTV models, the impact estimate is the estimate of the mean change (difference) in the response variable (e.g., NetHHInc) over the two-year period of the project, between households within the treatment zone and those outside the zone. An estimate of the total impact of the program is obtained by multiplying the estimated *number of people within the treatment zone* by this estimate. For the CTV models, the impact estimate is the estimate of the mean change in the response variable (over the project term) for a randomly selected household in the entire country. The total impact of the program is obtained by multiplying the estimated *number of households in the country* by this estimate.

To summarize, the model used as a basis for estimating the relationship of outcome to explanatory variables (treatment variable, design variables, and covariates, if any) is:

$$y_t = \beta_1 \text{ trt} + \beta_2 \text{ Rtrt} + \beta_3 \text{ Round} + \beta_0 + e_t,$$

where y_t denotes an output variable; trt is the treatment variable, equal to town1000tt; Round = survey round (0 or 1); Rtrt = RoundStd*trtStd, where RoundStd and trtStd denote the demeaned Round and trt variables; e_t denotes the model error term, and the β's are the model parameters (regression coefficients).

With respect to model specification, all explanatory variables in the preceding model are fixed. Hence the requirement for uncorrelatedness of the model explanatory variables with the model error term is satisfied, and OLS may be used to construct estimates of the model parameters. It is assumed that unobserved variables are time-invariant (so that , as discussed before, they do not introduce a bias into the estimated coefficients).

The estimate of impact is obtained by substituting the difference in means (between survey rounds) for the treatment-related variables in this equation. That is, if trtd and Rtrtd denote the differences in means of the treatment-related variables between survey rounds, and if we denote the regression equation as $y = f(\mathbf{x})$, then the impact (for this model, linear in the explanatory variables) is given by $\Delta y = f(\mathbf{\Delta x})$ (where Δ denotes the backward difference operator), where

$$\mathbf{\Delta x}' = (\text{trtd, Rtrtd}, 0, 0)$$
$$= (-.1525, 1.6370, 0, 0).$$

Estimated Impact for All Outcome Variables

There is a separate model for each outcome variable, each with its own set of β's. The full regression output for those models is not presented here, but are included in the Stata .log file that accompanies the project documentation. Here follows a table (Table A.14) of key model parameters (coefficients of treatment-related parameters). The table presents the values of β_1 and β_2 for each outcome variable, along with their standard errors. Two partial-effects parameters are involved, both involving trt (i.e., trt and Rtrt). The "main effect" of treatment is β_1, with β_2 representing an adjustment (associated with round). The coefficient β_1 (trt) varies in sign over the various outcomes. The trt parameter represents the partial effect of treatment, as measured by town1000tt. The coefficient is the estimated change in outcome per unit change in town1000tt. That coefficient along does not represent impact – impact is estimated by estimate $\mathbf{\Delta x}'\ \mathbf{b}$.

Table A.14. Key Model Parameters for Outcome (Partial-Treatment-Effect) Regression Model with No Covariates				
Outcome Variable	β_1 (trt)		β_2 (Rtrt)	
	Estimate	Standard Error	Estimate	Standard Error
IncEmp (monthly)	135	243	15.6	17.1
IncEmpAg (monthly)	-293*	105	16.6*	7.39
IncEmpNonAg (monthly)	610*	176	-9.85	12.4
TotHHExp (monthly)	218	135	4.89	9.47
NetHHInc (annualized)	1747	3561	421	251
CostToSchool	.087	.134	-.0061	.0094
TimeToSchool	.628	.420	-.0143	.0245
CostToCollege	5.59*	2.11	.262	.149
TimeToCollege	2.38*	1.34	.112	.088
CostToHospital	4.39	4.20	-1.74*	.296
TimeToHospital	.352	1.54	-.397*	.108
CostToHealthCtr	.614	.568	-.062	.040
TimeToHealthCtr	3.65*	.979	-.0044	.0692
CostToMarket	2.88*	1.40	-.102	.099
TimeToMarket	-1.35	1.40	-.176	.099
CostToPulp	.905*	.359	-.0073	.0254
TimeToPulp	.040	.412	-.020	.029
TimeToTegus	-1.59	2.88	-.991*	.187
TimeToSPS	-.032	5.91	.459	.381
TimeToDepCap	7.18*	3.21	-.950*	.214
TimeToMunCap	3.42	1.80	-.357*	.158
ChldInSch712	.022	.0275	-.0045	.0019
ChldInSch1318	-.030	.0220	-.0033*	.0015
VisHospital	-.004	.041	.0045	.0029
VisPrHlthCtr	-.011	.036	-.0024	.0025
VisPubClinic	-.078	.052	-.0104*	.0044
VisNonProf	.086*	.042	-.0076*	.0030
VisPharm	.010	.028	-.0033	.0020
WrkdPrevWk	.066	.038	.0088*	.0027

The estimated impact, presented in Table A.15, is given by $\mathbf{\Delta x' b}$ = trtd β_1 + Rtrtd β_2 = -.1525 β_1 + 1.6370 β_2. For example, in the case of NetHHInc, the impact is -.1525 (1747) + 1.6370 (421) = 422. The standard error of the estimated impact cannot be obtained from just the standard deviations shown in this table – it also depends on the covariance between the coefficient estimates, which is not shown in this table.

Table A.15 presents estimates of the total treatment effect (which, as noted, is the average treatment effect, ATE) for the complete set of response indicators. Impact estimates that are statistically significant (two-sided or one-sided, as appropriate, .05 significance level) are marked with an asterisk (*). Note that the components of IncEmp (IncEmpAg and IncEmpNonAg) do not sum to IncEmp, because they are estimated independently. (As mentioned, on the indication of statistical significance, the interpretation is that over many independent investigations, the probability that the confidence interval includes the true value of the parameter is approximately .95. Confidence intervals within the same investigation are correlated.)

It is clear from Tables A.11 and A.15 that the magnitudes of the impact estimates differ substantially for the BTV and CTV models. This is a reflection of the fact that the populations to which they refer differ very much in size (i.e., those households within a treatment "buffer" zone vs. all of the households in the country). It is also a reflection of the fact that the impact of a road improvement project varies substantially according to the distance (in time or space) of a household from the project. The BTV estimates show the impact for households within .5-1 hours of project roads (or approximately 12-24 km from the project roads, were distance used to define the treatment zones of influence ("buffer zones"). Lowess curves (nonparametric regression curves) were plotted for a number of outcome indicators, but the sample sizes were not sufficiently large for these curves to be useful. Had the sample size been much larger (or the effects larger), an interesting lowess curve would have been impact plotted as a function of distance (time or space) from project roads. Since project roads are geographically distributed over the country, it is not expected that the relationship would be strong (since a household far from one project road may be close to another project road).

Table A.15. Ordinary-Least-Squares (OLS) Regression Estimate of Average Treatment Effect (ATE) based on Continuous Treatment Variables, for Selected Outcome Measures

Outcome Variable	Estimate of Impact	Standard Error of Estimate
Household Income and Expenditure		
IncEmp (monthly)	5.00	41.5
IncEmpAg (monthly)	71.9*	17.9
IncEmpNonAg (monthly)	-109*	30.1
TotHHExp (monthly)	-25.2	23.0
NetHHInc (annualized)	422*	609

Table A.15. Ordinary-Least-Squares (OLS) Regression Estimate of Average Treatment Effect (ATE) based on Continuous Treatment Variables, for Selected Outcome Measures		
Outcome Variable	**Estimate of Impact**	**Standard Error of Estimate**
Access		
CostToSchool	-.0232	.0228
TimeToSchool	-.119	.0718
CostToCollege	-.424	.361
TimeToCollege	-.180	.212
CostToHospital	-3.52*	.718
TimeToHospital	.704*	.263
CostToHealthCtr	-.194*	.097
TimeToHealthCtr	-.549*	.168
CostToMarket	-.606*	.239
TimeToMarket	-.083	.240
CostToPulp	-.126*	.0613
TimeToPulp	-.0394	.0704
TimeToTegus	-1.38*	.476
TimeToSPS	.757	.994
TimeToDepCap	-.459	.557
TimeToMunCap	-1.106*	.387
School Attendance		
ChldInSch712	-.00402	.00470
ChldInSch1318	-.000843	.00375
Use of Health Care Services		
VisHospital	.00675	.00700
VisPrHlthCtr	.00230	.00621
VisPubClinic	-.00513	.01058
VisNonProf	-.0255	.00722
VisPharm	-.00702	.00479
Employment		
WrkdPrevWk	.00436	.00649

Note: Income, expense and travel costs measured in Honduran lempiras; travel times measured in minutes.

The preceding results show that the program intervention had a statistically significant effect (of the expected sign) on many of the access times and costs, on some of the indicators concerned with use of health care facilities, and on some of the income indicators. In all cases, the magnitude of estimated impact is small. The principal reason for small magnitude of impact is that it is the expected impact of the project *for a randomly selected household in the country*. This is quite different than for the case of the BTV estimate, for which the impact estimate is the expected double difference (difference in change over time between the treatment and control groups) relative to the zone-of-influence buffer zones.

To assess the total national impact of the road-improvement program, the impact estimates are multiplied by the number of households in the nation. The population of Honduras is 8.2 million people (2010 est.), and the average household size is approximately 5, so this corresponds to approximately 1.64 million households. Based on these numbers, the estimated total program impact for the nation is approximately as follows:

Table A.16. Estimated National-Level Impact of the Transportation Program (table entries are lempiras)

Outcome Variable	Estimated Impact per Household	Standard error of Estimated Impact per Household	Estimated Total Impact for Nation	Standard Error of Estimated Total Impact for Nation
IncEmp (monthly)	5.00	41.5	8.2M	68M
IncEmpAg (monthly)	71.9*	17.9	119M*	30M
IncEmpNonAg (monthly)	-109*	30.1	-180M*	50M
TotHHExp (monthly)	-25.2	23.0	-42M	38M
NetHHInc (annualized)	422	609	692M	999M

An approximate 95% confidence interval for the estimated national-level project impact is the estimate plus and minus twice its standard error. For example, an approximate 95% confidence interval for income from agricultural employment (IncEmpAg) is (59M lempiras, 179M lempiras).

Note that in addition to the direct impact estimated here, the Transportation Project may have indirect effects, such as providing the improved transport required for the FTDA project.

The preceding estimates of total national impact are comparable in magnitude to, but small compared to, the cost of the road-improvement project (which was, according to the MCC Compact, 125.7 million dollars).

The results presented in Table A.15 for the continuous-treatment-variable models are more "sensitive" than the results presented in Table A.11 for the binary-treatment-variable models, in the sense that a larger proportion of the effects are statistically significant. In technical terms, the CTV approach is much more powerful than the BTV approach. This is to be expected to some extent, since the binary-treatment-variable formulation loses information in transforming the continuous time variables to binary variables. Another reason is the "fuzzy" distinction between "treatment" and "control" households for the BTV model. A more important reason is that the BTV and CTV estimates refer to quite different populations (i.e., in the case of the BTV to the populations in the treatment and control zones of influence, and in the case of the CTV to the entire country). In contrast to the BTV models, the impacts for access times are usually of the expected sign (i.e., negative). Note that the impacts for use of health care services are statistically large (of negative sign) but not of the expected sign.

A number of reasons were cited why the BTV impact estimates presented earlier were weak. Note that none of these reasons pertain to the CTV estimates, which are conditional on completion and maintenance of the project, and are based on the relationship of outcome to travel time as estimated from the partial-treatment-effects model (i.e., there is no "lag" associated with the conditional impact, which is estimated directly from the partial-treatment-effect equation).

Ex Post Statistical Power Analysis

The section dealing with estimation of impact based on binary treatment variables contained a discussion of the power associated with tests of hypothesis based on the BTV-based impact estimates. That section was brief because the evaluation design was concerned primarily with CTV estimates of impact – the BTV estimates were presented simply because it was possible to do this in the analysis, even though the design was not oriented to that type of estimate. This section will present a similar *ex post* statistical power analysis for the CTV estimates, which were the focus of this investigation.

Table A.17 shows the power and minimum detectable effects for all outcomes of interest. This table uses the same test parameter values as in the BTV case (i.e., $\alpha = .05$ and $\beta = .1$, one-sided tests).

The table shows (column 4) that for some of the response indicators the power to detect impacts of the magnitude of those observed is high, for many it is of moderate size, and for some it is low. Columns 9 and 10 show that the power of the design is very high for detecting effects (impacts) that are equal in magnitude to ten percent of the outcome variable mean or standard deviation. Columns 11 and 11 show, however, that (in contrast to the situation for the BTV estimates) the observed effects were much smaller than this (ten percent). The interpretation of these results is that although the impact estimates based on the CTV models are small (since they refer to a randomly selected household in the country), the CTV model is sufficiently precise to detect them with very high power. *What this says about the evaluation design is that the decision, in the project planning phase, to base impact estimation on CTV models based on GIS-model travel times was a very good decision.*

As was discussed in the *ex post* power analysis of the BTV estimates, the data on the ratio of the standard error of the estimate to the Round 0 standard deviation (in the last column of the table) can be used to estimate the (Kish) design effect. The average value of this quantity is seen to be (from the last column) about .006. Using the same formula as was presented in the BTV *ex post* power analysis, the design effect is hence estimated to be deff = (sqrt(n)) (standard error of estimate) / (Round 0 sd) = .006sqrt(3008)/4 = .08. This design effect is extremely good. It is an order of magnitude smaller than that for the BTV models. It reflects the fact that the CTV models based on travel times are very efficient (high precision for the sample size). The efficiency comes mainly from interviewing the same households in both survey rounds, not from the marginal stratification on explanatory variables, which were seen to have a small effect on impact estimates).

Another indicator of interest is the ratio of the standard error of the estimate (of impact) to the mean of the variable. This is presented in the penultimate column of Table A.17. For an impact to be statistically significant, it must be about twice as large as this quantity, as a fraction of the mean. It is clear from this indicator that the study could detect very small changes in most response indicators, relative to the mean. (For example, for NetHHInc, the ratio of the standard error of the impact estimate to the mean is .0051. Twice this is .01. For an impact in NetHHInc to be statistically significant, it must be at least as large as .01 times the mean NetHHInc, or .01 x

77,052 = 770. This is a reasonable-sized impact to expect from a road improvement project. This example illustrates that the power of the design is satisfactory, with respect to NetHHInc.)

Table A.17 includes a column that specifies the coefficient of variation of the outcome variables for Round 0. The coefficient of variation is the standard deviation divided by the mean. It is presented in the column headed "CV (sd/mean)" in the table. In the power calculations done at the beginning of the project, not much was known about the statistical properties of the population with respect to the variables of interest. Data were available from which the CV for income could be estimated, and it was seen to be in the range 1-2. The sample data show that the CV may be much larger than this for small components of income (e.g., IncEmpAg), and for non-income variables. (The value of the CV for travel cost to pulperia (CostToPulp) is very large. The value of this variable is zero for most households, very small for a few (about 30 households), and substantial for a few (about 15 households.)

Table A.17. *Ex Post* Statistical Power Analysis of Ordinary-Least-Squares (OLS) Regression Estimate of Average Treatment Effect (ATE) based on Continuous Treatment Variables, for Selected Outcome Measures

Outcome Variable	Estimate of Impact (observed effect)	Standard error (se) of estimate	Power of test to detect a true effect equal in magnitude to the estimate of impact	Minimum detectable effect (MDE) for a test of power 90%	Round 0 mean	Round 0 standard deviation (sd)	CV (sd/ mean)	Power to detect a true effect equal to .1 of Round 0 mean	Power to detect a true effect equal to .1 of Round 0 sd	Estimate of impact relative to Round 0 mean	Estimate of impact relative to Round 0 sd	Standard error of estimate relative to Round 0 mean	Standard error of estimate relative to Round 0 sd
Household Income and Expenditure													
IncEmp (monthly)	5.00	41.5	.066	121.	4632	6033	1.30	.999+	.999+	.001	.008	.0090	.00688
IncEmpAg (monthly)	71.9*	17.9	.991	52.	753	2218	2.95	.993	.999+	.095	.032	.0237	.00807
IncEmpNonAg (mon)	-109*	30.1	.976	88.	3779	5058	1.34	.999+	.999+	-.029	-.022	.0080	.00595
TotHHExp (monthly)	-25.2	23.0	.292	67.	5245	4066	.78	.999+	.999+	-.005	-.006	.0044	.00566
NetHHInc (annualized)	422	609	.172	1782.	77052	89246	1.16	.999+	.999+	.005	.005	.0079	.00682
Access													
CostToSchool	-.0232	.0228	.266	.0667	.12	1.70	14.17	.212	.328	-.193	-.014	.1900	.0134
TimeToSchool	-.119	.0718	.505	.210	14.3	13.56	.95	.999+	.999+	-.008	-.009	.0050	.00529
CostToCollege	-.424	.361	.319	1.056	10.0	63.81	6.38	.870	.999+	-.042	-.007	.0360	.00566
TimeToCollege	-.180	.212	.214	.620	53.3	44.56	.84	.999+	.999+	-.003	-.004	.0040	.00476
CostToHospital	-3.52*	.718	.999	2.101	52.0	96.78	1.86	.999+	.999+	-.068	-.036	.0138	.00742
TimeToHospital	.704*	.263	.849	.770	112.	65.08	.58	.999+	.999+	.006	.011	.0023	.00404
CostToHealthCtr	-.194*	.097	.639	.284	5.39	16.49	3.06	.999+	.999+	-.036	-.012	.0180	.00588
TimeToHealthCtr	-.549*	.168	.947	.492	43.5	36.52	.84	.999+	.999+	-.013	-.015	.0039	.00460
CostToMarket	-.606*	.239	.813	.699	25.2	36.77	1.46	.999+	.999+	-.024	-.016	.0095	.00650
TimeToMarket	-.083	.240	.098	.702	82.2	58.53	.71	.999+	.999+	-.001	-.001	.0029	.00410
CostToPulp	-.126*	.0613	.659	.179	.28	5.85	20.89	.999+	.999+	-.450	-.022	.2189	.0105
TimeToPulp	-.0394	.0704	.140	206	7.91	12.01	1.52	.999+	.999+	-.005	-.003	.0089	.00586
TimeToTegus	-1.38*	.476	.894	1.39	241.	137.6	.57	.999+	.999+	-.006	-.010	.0020	.00346

Table A.17. *Ex Post* Statistical Power Analysis of Ordinary-Least-Squares (OLS) Regression Estimate of Average Treatment Effect (ATE) based on Continuous Treatment Variables, for Selected Outcome Measures

Outcome Variable	Estimate of Impact (observed effect)	Standard error (se) of estimate	Power of test to detect a true effect equal in magnitude to the estimate of impact	Minimum detectable effect (MDE) for a test of power 90%	Round 0 mean	Round 0 standard deviation (sd)	CV (sd/ mean)	Power to detect a true effect equal to .1 of Round 0 mean	Power to detect a true effect equal to .1 of Round 0 sd	Estimate of impact relative to Round 0 mean	Estimate of impact relative to Round 0 sd	Standard error of estimate relative to Round 0 mean	Standard error of estimate relative to Round 0 sd
TimeToSPS	.757	.994	.189	2.91	269.	159.3	.59	.999+	.999+	.003	.005	.0037	.00624
TimeToDepCap	-.459	.557	.206	1.63	112.	67.89	.61	.999+	.999+	-.004	-.007	.0050	.00820
TimeToMunCap	-1.106*	.387	.886	1.074	68.8	53.96	.78	.999+	.999+	0.016	-.020	.0056	.00717
School Attendance													
ChldInSch712	-.00402	.00470	.216	.0138	.874	.997	1.14	.999+	.999+	-.004	-.004	.0054	.00471
ChldInSch1318	-.000843	.00375	.081	.0110	.371	.666	1.80	.999+	.999+	-.023	-.001	.0101	.00563
Use of Health Care Services													
VisHospital	.00675	.00700	.248	.020	.306	1.11	3.63	.999+	.999+	.022	.006	.0228	.00631
VisPrHlthCtr	.00230	.00621	.101	.018	.303	.977	3.22	.999+	.999+	.076	.002	.0205	.00636
VisPubClinic	-.0513	.01058	.999	.031	.64	1.55	2.42	.999+	.999+	-.080	-.033	.0165	.00683
VisNonProf	-.0255	.00722	.969	.021	.222	1.29	5.81	.999+	.999+	-.115	-.020	.0325	.00560
VisPharm	-.00702	.00479	.429	.014	.13	.631	4.85	.993	.999+	-.054	-.011	.0368	.00759
Employment													
WrkdPrevWk	.00436	.00649	.166	.0190	1.80	1.23	.68	.999+	.999+	.002	.004	.0036	.00528

Note: Income, expense and travel costs measured in Honduran lempiras; travel times measured in minutes.

Comparing the impact effect sizes to the means and standard deviations of the variables (in the Round 0 summary tables presented earlier), the magnitudes of the observed effects (impacts) are seen to be very small. The evaluation design was able to detect these very small effects in the response variables. For many outcome variables, the standard error of the estimate is less than one percent of the variable mean, and yet the impact is statistically significant.

While the power of the CTV estimates is high, there is a disturbing fact that emerges from Table A.17. This is the fact that the power is high because the CTV model was able to detect very small effects – on the order of one percent of the means and standard deviations of outcome variables – not because the effects are large. This high power owes to use of the GIS-model travel times (which estimates the travel-time change under the assumption that all project road improvements are completed and maintained). The estimation of sample size that was done at the beginning of the project assumed that minimum detectable effects (MDEs) would be in the range of 10-20% of the standard deviation of selected outcome variables (or of the mean, for variables with coefficient of variation approximately equal to one). That size MDE was considered because that was the magnitude of the impacts that was considered by the MCC Compact *Monitoring and Evaluation (M&E) Plan* for other indicators (for example, an Economic Rate of Return (ERR) of 17-21% for the primary-road improvements and at least 12-15% for rural roads). For many of the income variables, the magnitudes of the impacts that resulted from the road improvement projects were substantially less than ten percent (of the mean or of the standard deviation). Had it turned out that the impact evaluation models been able to detect MDEs of on the order of 10-20 percent, as planned, few statistically significant results would have been observed. The study would have been underpowered. The reason why the evaluation was in fact able to detect impacts of the small size that were realized was because of the very high power of the model when it incorporated the traffic survey / GIS-model information, and estimated impacts conditional on completion and maintenance of all planned road improvements. The implication of these results is that if a roads-improvement evaluation is planned, the second round of the evaluation survey should be conducted after sufficient time has elapsed for the economic benefits to manifest, and the improved roads should be maintained.

Summary of Ex Post Statistical Power Analysis

Based on this *ex post* statistical power analysis, it is concluded that the power of the evaluation design was quite adequate to detect effects of the anticipated size, and is reasonably characterized as high. The reason for the high power for the CTV estimates is the fact that they are conditional on GIS-model travel-time estimates, which are based on the assumption that all planned road improvements were completed and maintained.

A major recommendation for future road improvement evaluations is that if they are to be based solely on household survey data (and not traffic surveys/GIS-model travel-time estimates) and evaluated before the results of the project have had a chance to manifest fully, then the sample sizes must be much larger than the sample sizes used in this evaluation. Alternatively, and more reasonably, the evaluation should be delayed until the economic benefits of the project have had time to manifest. Since improvements to rural roads do not last unless the roads maintained, there is little point to evaluating such projects if they are not maintained.

As was noted in the discussion of the BTV-based estimates, the estimated effects (impacts) of the Transportation Project and the FTDA Project are confounded (because the projects were conducted in some of the same areas over the same time). Because the CTV model is based on the GIS travel-time model (and not simply on a double difference, as was the case for the BTV model), it is believed that the degree of confounding is low.

The Impact Estimates Are Estimates of Causal Relationships

The impact estimates presented in this section are estimates of the total effect of the program intervention, conditional on setting the non-treatment-related (non-travel-time) explanatory variables to their mean values, and setting the treatment-related variables equal to the difference in their mean values between the baseline and endline rounds. The validity of these results depends on the validity of the household survey data and the GIS network travel-time model, which depends in turn on the traffic survey data. The statistical models developed in this analysis were based on underlying causal models, and the effect estimates are estimates of the *causal relationship* of the outcomes to the program intervention.

Note that the term "setting" refers to substituting variable values (for treatment-related variables (travel times) and covariates) in an estimated model. (We are "fixing" the values of the treatment-related variables, in the sense of Judea Pearl's "do" calculus – the project is "fixed.") In the case of the treatment variables, the substitution corresponds to the forced changes *in the treatment variables* caused by the program intervention. It does not refer to making forced changes *in the covariates*. The "forced change" in this program was made at the project level by the program implementers. Although these changes were not made by randomization, they were definitely forced changes. The data used in this evaluation are "forced-change" data, not "observational data." The data analysis measures statistical relationships associated with these forced changes, and are estimates of the change that would be observed if similar forced changes were made under similar circumstances, again.

This type of evaluation is called an experiment, since forced changes were made to the treatment levels (by the program implementers). Since randomization was not used to assign treatment (i.e., decide which roads would be improved), the experiment is not referred to as a "designed experiment," but it is nonetheless an experiment involving forced changes to treatment variables. (Because the assignment to treatment was not determined by randomization, the analysis must be concerned with selection effects, which may introduce biases. This was done through the use of causal modeling and the use of analytical statistical models related to the causal models. The impact estimates are "model-based" estimates, in contrast to the "design-based" estimates from a designed experiment.)

It should be kept in mind that the term "statistical significance" refers to associational relationships, quite independently of whether the relationships are causal or simply associational ("correlational"). Whether those relationships represent causal relationships is not determined by the statistical model, but by a causal model and the relationship of the statistical model to it. *The estimates resulting from this analysis are estimates of the causal effect of the program intervention on the outcome variables.* As stated, they represent the impacts that would be expected if similar forced changes were made under similar circumstances.

It is recognized that some researchers refuse to attribute causality to program interventions unless randomization has been used to select experimental units and randomization has been used to assign treatment levels to the selected experimental units. There is a large body of scientific opinion that does not support this point of view. Much scientific progress has been made in recent centuries in settings in which randomization was not used. This progress belies the assertion that causal inferences can be made only if treatment levels are assigned using randomization. A very important consideration is the presence of forced changes in treatment variables (whether caused by randomization or other means). (The forced change supports compliance with Judea Pearl's "back door" criterion of valid statistical estimation (identification) of causal relationships (causal effects).) The lack of randomization certainly makes the analysis and interpretation of results more difficult and more subject to threats to validity, but it does not alter the fact that the analysis presented here is based on causal modeling, and the estimates are estimates of the causal impact of the program intervention. (Even if forced changes are not made, causal inference is possible, e.g., the US government project on Smoking and Health, which established a causal relationship of health to smoking, even though smoking levels were never determined by randomization.)

3. Instrumental-Variable Estimator of Partial Treatment Effect (PTE) Based on Continuous Treatment Variables (IV/PTE/Continuous-Treatment-Variable)

The use of instrumental variables is a popular method for estimating impact for non-randomized designs. This approach was not used in this study, for reasons discussed in this subsection. This subsection is included because it may be asked why the instrumental-variable approach was not used. Apart from consideration of instrumental variables, the subsection includes results of a canonical-correlation analysis of the questionnaire and GIS-model travel times.

As has been mentioned, it is not correct to include the questionnaire travel times as explanatory variables in ordinary-least-squares regression models for income, because those variables are endogenous (i.e., they may affect each other, over the time frame of the project). This possibility introduces correlations among the explanatory variables and the model errors terms, with the result that the regression coefficient estimates may be biased and inconsistent. One way of including such variables in a regression model is to apply the method of instrumental variables. In this project, the GIS-model travel times may be used as instruments for the questionnaire travel times, since the GIS-model travel times are exogenous (i.e., they were produced by the GIS model and traffic surveys, and are independent of the questionnaire travel-time variables).

The use of the instrumental variable method requires instrumental variables that have substantial correlation with the endogenous variables, but are uncorrelated with the model error term, given the full set of covariates. The GIS-model travel times satisfy these conditions.

To show the correlation of the GIS-model travel times with the questionnaire travel times, a canonical correlation analysis was performed. The results of this analysis are shown in the following table.

```
. canon (TimeToSchool TimeToCollege TimeToHospital TimeToHealthCtr TimeToMarket
TimeToPulp TimeToTeg
> us TimeToSPS TimeToDepCap TimeToMunCap) (mcapritt mcasectt mcarurtt pritt sectt
rurtt tegustt sanp
> edtt town1000tt top10tt) if Round==0

Canonical correlations:
  0.9415  0.8719  0.6010  0.4093  0.3362  0.2825  0.1777  0.0755  0.0547  0.0303

----------------------------------------------------------------------------
Tests of significance of all canonical correlations

                         Statistic      df1      df2            F     Prob>F
         Wilks' lambda    .0113344      100  4787.39      41.6299     0.0000 a
        Pillai's trace     2.40947      100     6760      21.4584     0.0000 a
Lawley-Hotelling trace     12.0009      100     6652      79.8299     0.0000 a
    Roy's largest root     7.80847       10      676     527.8527     0.0000 u
----------------------------------------------------------------------------
                            e = exact, a = approximate, u = upper bound on F
```

The canonical correlation analysis shows a reasonably high degree of correlation of the GIS-model travel times and the questionnaire travel times (e.g., a value of .9415 for the first canonical correlation). Despite this high correlation, it was decided not to present results based on the method of instrumental variables (IV). The main reason for this decision is that the GIS-model travel times were used directly in the OLS models both as treatment variables (Round 0 and Round 1) and as covariates (Round 0), with statistically significant results associated with the treatment variables and with little benefit associated with the covariates. Instrumental-variable estimates are biased and inefficient, and are prone to be unstable and have large standard errors, even for very large sample sizes. There was no compelling reason to entertain more complex models that used the GIS-model travel times as instrumental variables, when satisfactory results had already been obtained using the GIS-model travel times directly.

III.G GIS Model-Based Estimates in Impact Evaluations

The GIS-model variables were useful in the evaluation survey design and in estimating the OLS regression models presented in the preceding subsection. As mentioned, in the survey design, *caseríos* were stratified by estimated change in travel time associated with the project (calculated by the GIS model) and by estimated travel time (to various places of interest) from project roads, a probability sample of *caseríos* was selected from each stratum, and a probability sample of households was selected from each *caserío*. This procedure (marginal stratification based on a large number of travel-time-related variables) assured that there would be adequate variation in travel-time-related variables to assure high power for impact estimates of interest. (In order to achieve high precision and low correlation among regression coefficients, it is necessary that adequate variation (spread, balance) be represented in the explanatory variables, and that the correlation (degree of orthogonality) among them be low. We were able to achieve adequate variation in explanatory variables by using the GIS travel-time model.)

The validity of the impact estimates based on OLS regression of GIS-model travel times rests on the validity of the household sample survey and the validity of the traffic surveys and GIS travel-time model. This situation differs from many development-program evaluations, which rely mainly on household survey data, focus-group discussions, key-informant interviews, and existing data sources, but do not make use of engineering models to provide additional exogenous variables.

Overall, the impact estimates based on the CTV GIS-model travel times were more sensitive than the impact estimates based on the BTV models. In summary, the conceptual framework of assessing the impact of road improvement using continuous treatment variables – travel-time estimates produced by a GIS road network model – proved successful.

The development of the GIS model was resource-intensive. It involved assembly of a significant amount of GIS geographic data, the construction of a complete GIS road-network model, collection of vehicle speed data from traffic surveys conducted about the same time as the household sample surveys (i.e., near the beginning and end of the program), for treatment and comparison roads, and entry of the speed data into the GIS model (so that it could be used to estimate travel times from all *caseríos* to places of interest). It hence required substantial investment of time and money and cooperation from Honduran agencies having GIS data.

Perhaps the most challenging aspect of using GIS technology is the requirement for verification and validation (V&V) of GIS-based models. The V&V is necessary to establish the validity of the impact estimates. In order to verify a GIS model, it is necessary to construct command files (or "script files," in programming languages such as ESRI's Avenue or Visual Basic) that enable the analysis to be replicated, after its completion by the GIS analyst. "Replication" refers to the ability to completely reconstruct (duplicate) the GIS model results, starting with the traffic survey data and the GIS road network data, and ending with the travel-time estimates to be used in conjunction with the household survey data analysis. (This is the same as the requirement to reconstruct the output of Stata analyses, starting with the cleaned survey data files and executing one or more Stata command ("do") files. There is no justification for a lower standard of documentation for the GIS model than is required for the data processing and statistical analysis.) Unfortunately, most GIS analysts use GIS programs (such as ArcView or ArcGIS) interactively, but do not have the interest or capability to construct the command files necessary to reproduce the GIS-model results starting with the cleaned traffic data files and ending with the estimated travel times. Under these conditions, much of the GIS model remains undocumented, and it becomes difficult or impossible either to verify (replicate, duplicate) the GIS model results or to validate them. What results is documented and validated household survey data and analysis and an undocumented GIS model that cannot be verified or validated. If the evaluation contract and evaluation firm do not possess the GIS skill resources, funding resources, and time to implement a GIS model approach that can be documented at a professional level, the GIS approach should not be used.

A competently executed GIS-model-based evaluation can require substantially more programming resources than one that does not include GIS modeling. A major factor is whether the GIS is used simply to create maps and construct "buffer" zones, vs. whether it is used to perform spatial analysis (e.g., estimate travel times).

In summary, the CTV estimates are conditional on the GIS-model estimates of travel-time changes between rounds (which are based on the traffic survey data and the assumption that all of the undertaken road improvements are completed and maintained). The use of the GIS approach was useful, but very resource intensive and time consuming. For these reasons, the costs and benefits of using GIS-model-based data should be carefully assessed before deciding to adopt this approach for future road-improvement program evaluations. The use of GIS spatial-analysis technology adds substantially to the cost of the evaluation, and if the GIS model is not adequately documented (with script (e.g., Visual Basic) files similar to the Stata command ("do") files), the validity of the results is suspect, since it cannot be validated or verified.

The significant advantage of using GIS technology for evaluation of roads projects over the conventional "zone of influence" approach based on binary treatment variables is that the former approach is a more sensitive representation of the relationship of outcome to road-improvement impact. For the present evaluation, many effects of interest were statistically significant for the CTV model but not for the BTV model. The statistical power of tests of significance associated with the CTV model is substantially higher than for the BTV model.

III.H Summary of Program Impact

Discussion

The preceding sections have presented estimates of program impact for a variety of outcome measures of interest. Impact was estimated using two classes of models, one based solely on the household survey data and binary treatment variables (BTVs), and the other, a continuous-treatment-variable (CTV) model, based on the household survey data plus travel times estimated from traffic surveys and a geographic information system (GIS) road network model. Impacts were estimated for a large number of outcome variables of interest, including income and expenses; travel times to points of interest; access to health-care and educational facilities; and employment. The impact estimates were less precise and the ability to detect impacts was less powerful for the BTV models than for the CTV models, for a number of reasons. One factor contributing to the lower-than-expected impact for the BTV models may be the fact that some of the road improvements were not completed when the second household survey round was conducted. Another factor is that rural-road improvements are considered to last for only a short time (e.g., six months). This factor could have contributed to the fact that statistically significant results were not observed for rural roads. Additional reasons for the lower power of the BTV model are that the binary treatment variables contain less information than the CTV variables, and that the distinction between "treatment" and "control" households is "fuzzy" (imprecise).

If impact is estimated using only the household survey data, then the estimates of impact (the binary-treatment-variable estimates) in most cases are not statistically significant. Statistically significant results were observed for many outcome variables using the continuous-treatment-variable model, conditional on the travel-time estimates obtained from the GIS road network model and traffic survey data, and on the relationship of outcomes to travel times that are represented in the partial-treatment-effects model. The GIS-model travel-time estimates were based on pre- and post-project traffic surveys done on a sample of roads for which the improvements had been completed and a sample of (untreated) comparison roads. These estimates are based on the assumption that the estimated travel times observed in the traffic surveys would apply to all project road segments had the improvements been completed (before the second round survey of the evaluation project), and that the improved roads would have been maintained. The CTV estimate is conditional on manifestation of long-term effects as estimated by the PTE model.

Based on this analysis, there is strong statistical evidence that Transportation Project road improvements are associated with decreased travel times to some places of interest, but the per-household effects are small and they are not associated with substantial increases in income or improvements in school attendance, use of health care services, or employment. Statistically significant impacts were observed for income and access to places of interest. The total national impact of the project appears to be comparable in magnitude to, but small compared to, the total cost of the road improvement program, but the statistical power associated with a test of the hypothesis that the program intervention is associated with an increase in net household income is not high.

An *ex post* statistical power analysis was conducted, to assess the power of the evaluation design to detect impacts of reasonable size. The *ex post* power analysis was done for both the BTV and CTV models. The analysis showed that the power of the CTV models to detect impacts of meaningful size (relative to the variable means) is very high for many outcome variables of interest. For the BTV models, the power to detect impacts of meaningful size was usually low. This low power is a reflection of the facts that the magnitudes of the observed program effects are small and that the evaluation design was oriented toward providing high power for the CTV-based estimates, not the BTV-based estimates. The small magnitudes of the program effects observed for the BTV models are associated with a number of factors, including the fact that some road improvement projects were not completed when the second round of the household survey was completed (so that the economic effects of the program did not have time to fully manifest). For the CTV models, the small magnitudes of the impact estimates are a reflection of small per-household program impacts, not of an underpowered study.

The impact results presented in this report are based on an evaluation design, causal models, and analytical models that are considered to have high validity. Based on an *ex post* statistical power analysis, the power associated with the impact estimates is high. The analysis results are considered to be an accurate (valid, reliable and high-power) assessment of the impact of the Transportation Project. The program impact is statistically significant for many outcome variables, but not as large as might have been anticipated. The largest income impact, 9.5 percent of the baseline mean, was observed for income from agricultural employment (variable IncEmpAg). This is somewhat less in magnitude than the economic rates of returns anticipated for the program (17-21% for CA-5 and at least 12-15% for rural roads). The estimated impact for net household income (variable NetHHInc) was just one percent of the baseline mean. These estimates are from the CTV model, and represent the per-household impact for a randomly selected household in the nation.

Summary of Impact of the Transportation Project

The principal finding of this evaluation is that although the Transportation Project shows some statistically significant effects on a variety of indicators (income and travel times to places of interest), those impacts are very small. A detailed statistical power analysis was conducted, which showed that the small number and size of the statistically significant results is not the result of an underpowered survey, but a result of the small magnitudes of the project effects.

The evaluation design adopted for this impact evaluation was to estimate impact from household-survey data, conditional on project-caused changes in travel time. The travel times are determined by a GIS road network model, using mean vehicle speeds estimated from traffic-survey data. The estimates are conditional on completion and maintenance of the road improvement project. The analysis produced estimates of the mean impact expected for a randomly selected household in Honduras. It was determined that the mean household-level impact of the project, averaged over the nation, is low.

Analysis of the traffic-survey data showed substantial changes in travel speeds and travel times over the project roads, compared to similar non-project roads. While the per-household impact averaged over the nation is low, the effect of the project on the speed of vehicles using the project roads is substantial. The disadvantage of using traffic-survey data alone to assess project impact is that it provides “intermediate” outcomes, not “higher level” impacts such as income, employment, and access to health, medical and other facilities.

The approach used in this report has a number of strengths that argue for the validity of the findings. The results are based on causal modeling, and the assumptions required of the statistical estimation models used to estimate impact are not in doubt. The per-household impact of the Transportation Project is low. While the economic impact of new roads is known to be substantial, the impact of the road improvements implemented in the Transportation Project, on a national level, are not high. On a national level, they represent marginal improvements to the road system and to household access, and they produce marginal impacts. The noticeable effect of the road improvements is on speeds and travel times of users of the improved roads.

In summary, the impact results presented in this report are based on an evaluation design, causal models, and analytical models that are considered to have high validity. An *ex post* statistical power analysis demonstrated that the power associated with the impact estimates is high. Conditional on the soundness of the assumptions described, above, we consider the results of this analysis to be an accurate (valid, reliable and high-power) assessment of the impact of the Transportation Project.

Summary of Assumptions and Limitations

With respect to the “macro-level” causal model and associated statistical model used to estimate impact, the assumptions underlying the estimates of impact are the following:

1. The stable unit treatment value assumption (SUTVA, no-macro-effects assumption, partial equilibrium assumption) is made. Among other things, this assumption implies that the project is not so large that it changes the basic relationship of outcomes of interest to travel time.

2. The estimates of travel time are based on the GIS model of the Honduran road network. This model includes all official Honduran roads, as well as others. The GIS model is highly detailed, and considered to be up-to-date and of high accuracy. The quality of the GIS model is not considered to be a limitation on the quality of the evaluation.

3. The impact estimates are conditional on completion and maintenance of the Transportation Project as finally configured. Under this assumption, the impact estimates refer to this particular project, not to the mean impacts associated with a conceptually infinite population of similar projects in other locations or at other times. The estimate of impact is the average treatment effect of this particular project on a randomly selected household in Honduras, not the average treatment effect associated with improving a randomly selected eligible-for-treatment road segment.

4. Although the unit of treatment was the road segment, the unit of analysis was the household. In this case there could be two types of unobserved (hidden) variables, which may introduce biases into the estimates of the model parameters. First, there may be unobserved variables that are time invariant. In a fixed-effects model, these, however, "drop out" for the two-round panel specification. Second, there could be unobserved variables that are not time invariant (though no such variables were identified). It is assumed that such variables, if any exist, are uncorrelated with the explanatory variables.

5. The continuous-treatment-value impact estimates are conditional on the travel speed table derived from the traffic surveys. This table presents estimates of the average speed of a pickup truck over a route by route type (primary, secondary, rural), elevation variation and program intervention status (improved or not improved). The speeds are conditional on the season of the year, day of the week, time of day, and weather conditions under which the traffic surveys were conducted. The speeds are used to calculate travel times to places of interest (using the GIS model). These travel times are used (as described above) to estimate impact. A number of travel times were available (i.e., travel times to various places of interest from a *caserío* such as to a municipal or department capital, or the nearest town having population 1,000 or more. Attention focused on the one that had the highest relationship to outcomes of interest (i.e., travel time to the nearest town of population 1,000 or more). The continuous-treatment-value estimates are also conditional on manifestation of long-term benefits, as estimated from the partial-treatment-effects model.

Other more specific model assumptions are listed for particular estimation equations in the detailed analysis presented in Annex 1.

The limitations of the evaluation are:

1. The impact estimates constructed in this analysis pertain solely to the Transportation Project as it was finally configured, and when and where it was implemented, not to similar projects in other settings (locations or times). The estimated standard errors reflect sampling variation associated with estimation of characteristics of this project, and do not include variation associated with hypothetical variations in the project location or time, or the higher-level sample units of the household sample survey (*caseríos* and households) (as would be the case for a "random effects" approach). It is expected, however, that similar results would be expected for similar projects in similar settings.

2. More of the resources for primary data collection for this evaluation were invested in the household sample survey, and less in the traffic surveys. The traffic surveys were limited in scope and in sample size. The design of the traffic surveys rested on judgment, not on the methods of statistical sample survey. The estimates of speed were conditioned only on elevation variation, and not on other road variables that may have an important effect on average speed, such as number of lanes, access, road roughness, and curviness. The vehicle speed on which travel times were based for analysis of impact was a pickup truck. The traffic survey data were used as input to the GIS road-network model for estimation of travel times, not for estimation of direct impact (such as via a double-difference estimator applied to the traffic data). The travel-time estimates pertain to a particular vehicle type, season, day of week, time of day, and local weather conditions. The level

of correlation among the available travel times was relatively high, so that the limited scope of the traffic surveys is not considered to be a major weakness. In the future, however, if this approach is used, it is recommended that a substantially greater level of resources be allocated to the traffic surveys (at least comparable to that of the household surveys).

3. The study focuses on a variety of indicators to assess impact of the road improvements. Some of them are direct effects, such as travel times to points of interest (education and health facilities), and others are indirect effects such as income and employment. Additional direct effects might have been of interest, such as number of trips or length of trips.

4. As described above, an assumption of the analysis is that are no time-varying unobserved variables that are correlated with any of the explanatory variables, which in this case refers to the treatment variables. While we have investigated and confirmed the soundness of this assumption to the extent feasible, it is not possible to conclusively rule out the possibility that such variables exist and may be influencing the results. For example, if economic conditions such as labor market characteristics or the level of private investment are changing in systematically different ways that are correlated with travel time,our approach may mistake the impact of the road improvement for the impact of changes in these conditions. This possibility must be considered a limitation of the analysis.

ANNEX 2: DESCRIPTION OF METHODOLOGY USED TO CALCULATE TRAVEL-TIMES USING A GIS ROAD NETWORK MODEL

I. INTRODUCTION

This chapter describes the process we followed to calculate the travel-time measures that were used in the Honduras Transportation Project impact evaluation models. The report is divided in two sections. First, we describe the overall process for calculating the Honduras travel-times; and second, we provide a brief description of a Geographic Information System (GIS) and how travel-times are estimated in GIS using a road network database, descriptive data on road segments, and utilizing a least-cost path algorithm for finding optimal routes through the network.

II. ESTIMATION OF TRAVEL-TIMES FOR THE IMPACT EVALUATION

Estimation of travel-times in Honduras was performed using a highly accurate and detailed GIS dataset of the complete Honduran road network that was calibrated with measured travel speeds from the NORC traffic surveys conducted in Honduras. All calculations were made using ESRI ArcGIS software. A graphical representation of this complete network is shown in Figure 1, with a map of the MCA road improvement locations in Figure 2.

The data for this GIS network came from two sources. The vast majority of the GIS road network data was provided by the Honduran road maintenance agency SOPTRAVI[19], including all data for primary and secondary roads, and for some rural roads. SOPTRAVI has compiled detailed spatial data the "official" Honduran road network for which it has maintenance responsibility. However, this "official" network does not include a large number of "unofficial" rural roads which are not officially maintained. Consequently, NORC merged into this dataset a extensive GIS database of Honduran rural roads that had been compiled by a special Consultant to MCA - Honduras, Alden Rivera. Alden told NORC that he had systematically compiled extensive GIS data on the spatial locations of unofficial rural roads through integration of various Honduran government GIS database, maps obtained from logging companies, and satellite imagery, and he estimated that he had captured 70-80% of all significant rural roads in the country. NORC merged Alden Rivera's rural GIS road database into the SOPTRAVI official road network database. Overlaying this GIS Honduran road database on selected satellite images and digital airphotos indicated that the spatial data of the road segments was highly accurate.

This GIS road network was then converted from a standard GIS file to a specialized GIS database object known as a GIS "network" object. This network object allows for the delineation of travel-routes through the network using a "least-cost path" algorithm (Djikstra 1959[20]; Weibull 1976) that will find the shortest distance between any two points on the network by minimizing a specified criterion. For example, if the specified criterion is geographic distance, the least-cost path algorithm will find the optimal route through the network that minimizes geographic distance from any point to any other point. However, since different types of roads

[19] http://soptravi.gob.hn/dr/

[20] http://en.wikipedia.org/wiki/Dijkstra's_algorithm#CITEREFDijkstra1959

have varying travel speeds, the shortest path in terms of distance is not always the shortest path in terms of time. Since road users will almost always opt to select the route that minimized their travel time, the least-cost path algorithm can also be used to find the optimal route that minimizes travel-time if the travel-time for each individual road segment in the larger network is recorded. This is a more accurate measure of the actual "economic distance" (in terms of what users actually pay, and seek to minimize) than geographic "Euclidean" distance. The technical details of the calculation of optimal travel routes using the GIS network dataset are described in more detail below in Section 2.

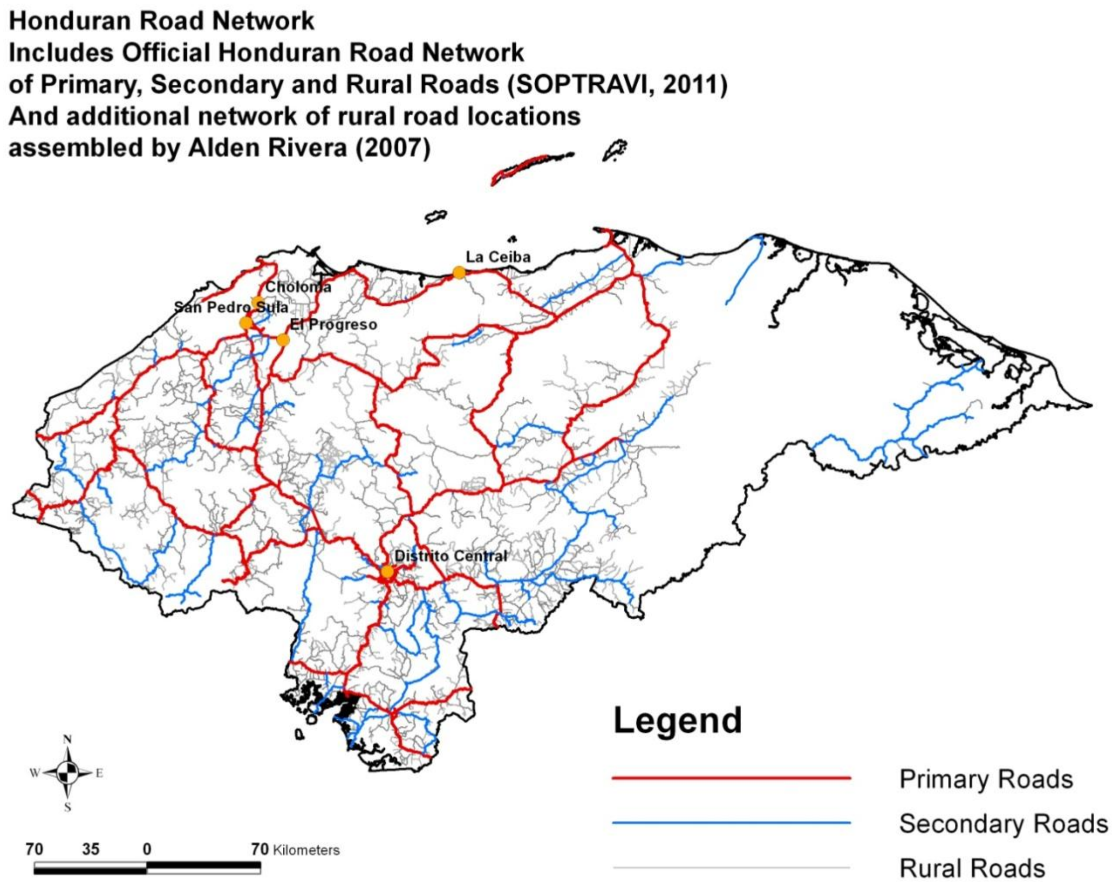


Figure 1: Honduran GIS Road Network (SOPTRAVI & MCA Honduras)

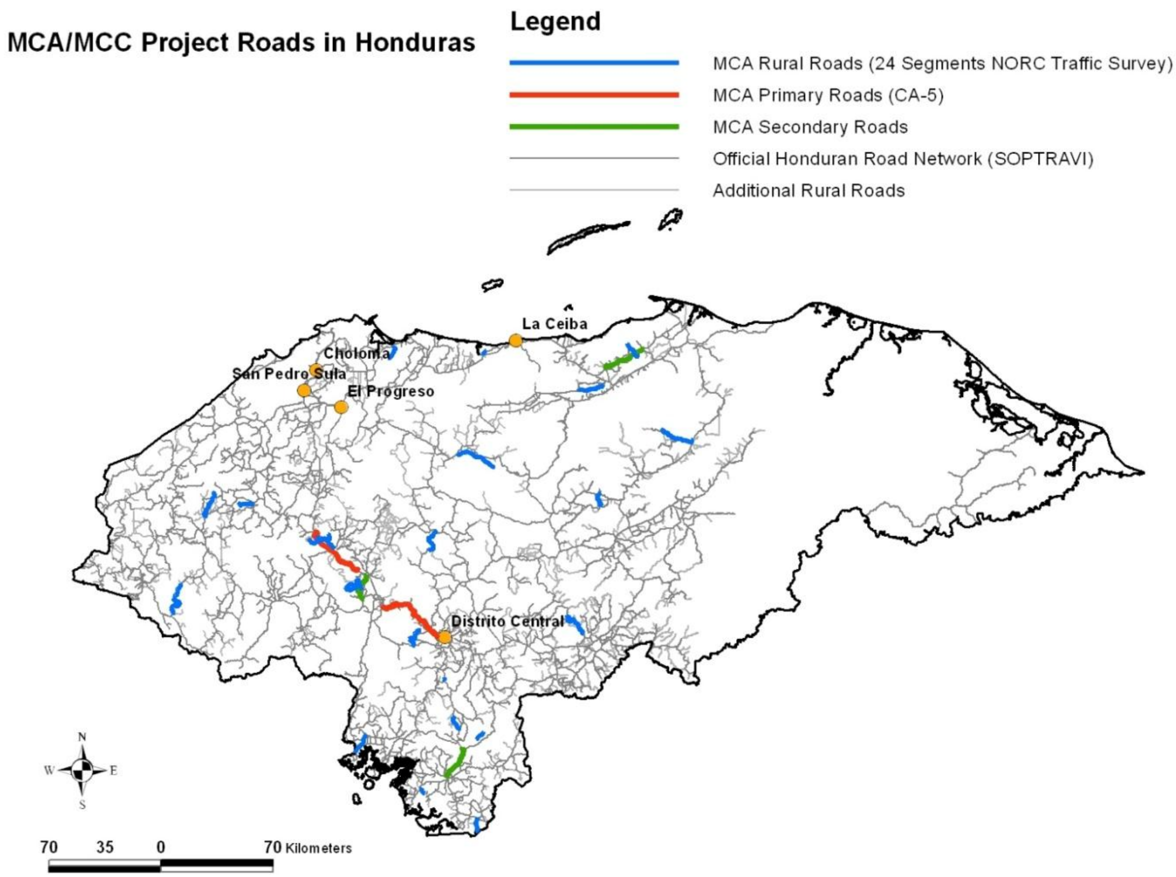


Figure 2: MCA Road Improvements in Honduras

The calculation of travel-times for the Honduran impact evaluation proceeded with the following steps:

(1) **Classification of Honduran Roads by Topographic Variation.** All Honduran road segments in the NORC Honduran road GIS network, including all MCC project and comparison roads, were classified into three categories by their degree of topographic variation, to classify all project and comparison roads into categories of "high topographic variation", "some elevation variation", and "low elevation variation." This was achieved by using a detailed GIS Digital Elevation Model (DEM) which allowed for point extraction of Honduran elevation values at any geographic location. The GIS was used to obtain an elevation measurement at 1 kilometer distances along all MCC project and comparison roads. The standard deviation of these elevation measures were then calculated for each individual GIS road segment. These standard deviation measurements were then binned into terciles to obtain the three categories of elevation variation, to classify all roads into the three categories of elevation variation. This process also sub-divided all MCC project and comparison roads into three categories of elevation variation.

(2) **Measured Vehicle Speeds from Traffic Surveys.** Traffic surveys were conducted in Honduras in cooperation with MCA - Honduras and NORC, and vehicle travel speeds were measured on all MCC primary, secondary and rural road improvements, as well as on a selection of non-improved comparison roads. The surveys were conducted pre and post-improvement, with a pre-improvement round conducted in 2009 and a post-improvement round conducted in 2011.

(3) **Linking Measured Speeds to Project and Comparison Road Segments.** Average measured travel speeds for pick-up trucks (the most common vehicle type on Honduran roads as measured by the traffic surveys) for all MCC project and comparison roads were obtained from the traffic surveys and then linked in the Honduran road GIS network to the respective MCC project or comparison where they were measured, for MCC project and comparison primary, secondary and rural roads. Average speeds on all project primary, secondary and rural roads, as well as on all comparison primary, secondary and rural roads, were linked to their respective GIS road segments. Because all Honduran roads, including the MCC project and comparison roads, were sub-divided into three categories of elevation variation, this resulted in 18 categories of pick-up truck average speeds linked to MCC project and comparison roads: project primary, secondary and rural categories; comparison primary, secondary and rural categories; and then three categories of elevation variation for each of those categories. Further, average speeds for pick-up trucks were linked for pre-improvement (traffic survey Round 1, 2009) and then also post-improvement (traffic survey Round 3, 2011). Table 1 displays the specific measured speeds from the NORC traffic surveys for the 36 categories that the Honduran road segments were binned into, for both treatment and comparison roads: 18 categories for pre-improvement (traffic Round 1) speed measures, and 18 categories for post-improvement (traffic Round 3) measures.

Table A.18: Measured Honduran Pickup-Truck Mean Travel Speeds

Measured Pick-Up Truck Travel Speeds		
Honduras Road Network	Pre-Improvement	Post-Improvement
	(Round 1: 2009)	(Round 3: 2011)
Primary Treatment Roads, Greatest Elevation Variation	**53.2**	**54.7**
Primary Comparison Roads, Greatest Elevation Variation	**50.4**	**50.9**
Primary Treatment Roads, Some Elevation Variation	**61.2**	**62.4**
Primary Comparison Roads, Some Elevation Variation	**57.6**	**58.1**
Primary Treatment Roads, Low Elevation Variation	**67.9**	**69.7**
Primary Comparison Roads, Low Elevation Variation	**64.2**	**64.6**
Secondary Treatment Roads, Greatest Elevation Variation	**41.2**	**70.3**
Secondary Comparison Roads, Greatest Elevation Variation	**54.3**	**55.2**
Secondary Treatment Roads, Some Elevation Variation	**44.2**	**74.1**
Secondary Comparison Roads, Some Elevation Variation	**58.9**	**59.6**
Secondary Treatment Roads, Low Elevation Variation	**48.1**	**77.5**
Secondary Comparison Roads, Low Elevation Variation	**62.7**	**63.1**
Rural Treatment Roads, Greatest Elevation Variation	**23.7**	**34.2**
Rural Comparison Roads, Greatest Elevation Variation	**18.3**	**18.7**
Rural Treatment Roads, Some Elevation Variation	**30.2**	**39.9**
Rural Comparison Roads, Some Elevation Variation	**20.4**	**20.6**
Rural Treatment Roads, Low Elevation Variation	**33.8**	**43.5**
Rural Comparison Roads, Low Elevation Variation	**23.1**	**23.4**

1. **Extrapolating Measured Speeds to All Honduran Road Segments.** Measured speeds from MCC comparison roads were then extrapolated to all road segments in Honduras in the GIS by category of road: primary, secondary and rural, and then further by high, medium and low elevation variation. This resulted in measured traffic speeds for all Honduran road segments. For example, the average measured pick-up truck speed on comparison Honduran secondary roads pre-improvements with a "medium" degree of elevation variation was 58.9 kph (see Table 1). This measure was then used to specify the travel-time speed for all non-treatment secondary roads in the larger Honduran road network pre-improvements.

2. **Calculation of Travel-Times for Each Honduran Road Segment.** Using these estimated travel speeds, travel-times were calculated for each Honduran road segment as a function of geographic length.

3. **Calculation of Pan-Honduran Travel-Times.** Travel-times were then calculated using the GIS to find optimal routes that minimized travel-time between the specified origin and a selection of destination locations, using a GIS least-cost path algorithm through the road network. Total accumulated travel-times of the optimal least-cost path route were calculated as an output of this process. For example, to calculate estimated travel-time from each household in the NORC household surveys to Tegucigalpa, the GIS was used

to find the optimal least-cost travel route from the Caserío location of each household to Tegucigalpa that minimized total travel time (considering estimated travel speeds on all Honduran primary, secondary and rural road segments), and to calculate accumulated total travel time for each of those routes.

4. **Calculation of Pre- and Post-Impact Travel-Time Estimates.** We calculated two sets of travel-times: travel-times pre-improvement (using traffic Round 1 measured speeds) and travel-times post-improvement (using traffic Round 3 measured speeds). For the former, the pre-improvement measured speeds were used, and for the latter the post-improvement measured speeds were used.

5. **Calculation of Euclidean Distances.** In addition to the travel-times calculated using the process delineated above, we also calculated straight-line Euclidean ("as the crow flies") distances from each household caserío to the set of specified destinations. These Euclidean distances were used as a robustness check in the impact evaluation models.

6. **Travel times and Euclidean distances were calculated to the following set of destination locations:**

   - to Tegucigalpa;
   - to the nearest point on CA-5 MCA improvements road segments;
   - to the nearest point on MCA secondary road improvements road segments;
   - to the nearest point on MCA rural road improvements road segments;
   - to the nearest point on nearest major highway (primary road) road segment;
   - to the nearest point on nearest secondary road segment;
   - to the nearest point on nearest rural road segment;
   - to San Pedro Sula;
   - to the nearest Caserío of population 1000 or greater;
   - to the nearest of the top 10 Honduran cities (Tegucigalpa, La Ceiba, Comayagua, San Pedro Sula, Choloma, Puerto Cortes, La Lima, Choluteca, Danli, El Progreso).

7. **Notes.** It is important to note the following regarding these estimated travel-times:

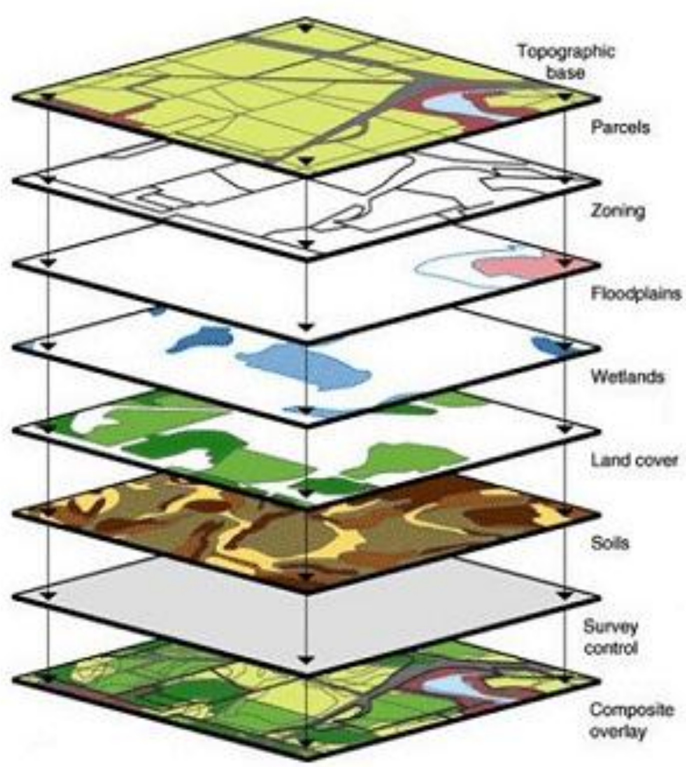
   - For the calculation of these travel-times, the actual travel speeds measured in the NORC traffic survey were used to provide improved calibrated estimates of likely approximate travel speeds for all roads in Honduras. However, the accumulated travel-times for the least-cost path routes that were estimated are not the actual travel-times for those routes, as those were not measured. While they are likely a close approximation of those actual route travel-times, their real purpose for the Honduras impact evaluation models is to represent an *index of relative travel-times* to specified destinations for households and Caseríos modeled in the Honduran impact evaluation models.

- As a highly-precise relative index of travel-times for the individual households in the impact evaluation models, they provide an excellent instrument against the self-reported travel-times provided by respondents in the NORC transportation household surveys, considering that those self-reported travel-times could be endogenous with the MCA road improvements. The GIS relative travel-time index, however, is completely independent of changes or impacts from the MCA road improvements; it represents an excellent instrument to control for any endogeneity in self-reported travel-times. They were used explicitly for that purpose, for example, in the 3rd impact evaluation model.
- Further, using the GIS and the Honduran road network GIS dataset to estimate this relative travel-time index provides us with the flexibility of generating a relative travel-time index for any set of origin locations to any set of destination locations in Honduras at very low cost, without the time and expenditure of conducting a survey measure them.
- We have attached to this report the following files, in a separate compressed "zip" archive:
  1. "tt1_dis1.xls": this excel file provides the complete set of travel-times and Euclidean distances calculated for all Honduran Caseríos (identified both by name and by Honduran Caserío ID codes) to all of the destination locations listed above. A separate worksheet in this file provides a data dictionary for all the travel-time and Euclidean distance calculations.
  2. "tt1_dis1.dta": provides the travel-time and Euclidean distance calculation data in Stata format (same data as "tt1_dis1.xls").
  3. "Hon_road_GIS.zip": this separate zip archive provides the GIS shapefile of the complete, merged Honduran road network used for these calculations.

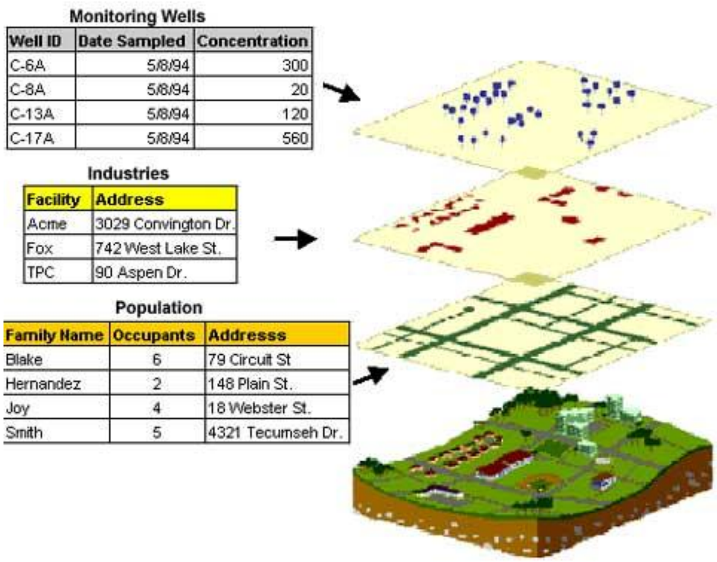
III. GIS AND ACCESSIBILITY

Geographic Information Systems (GIS)

A Geographic Information System (GIS) is a computer geo-database system for capturing, storing, checking, integrating, manipulating, analyzing and displaying data related to positions on the Earth's surface. In essence, a GIS is simply a standard database (running on Oracle or SQL Server platforms, for example) that also has the functionality of incorporating digital geo-spatial data (for query, display and data storage purposes). These geo-spatial data variables and inputs might be represented as several different "layers", where each layer holds data about a



particular kind of feature. Each feature is linked to a position on the graphical image on a map and a record in an attribute table. By layering information such as road networks, village or community locations, and population, spatial relationships among the objects being mapped can be emphasized. A GIS differs from other information systems because it combines common database operations such as query and statistical analysis with the benefits of visual and geographic analysis offered by maps.



GIS can relate otherwise disparate data on the basis of common geography, revealing hidden patterns, relationships, and trends that are not readily apparent in spreadsheets or statistical packages, often creating new information from existing data resources.

Although quite complex and robust GIS data formats exist and are increasingly used in an ever-growing range of industries and sectors, in a simplified representation GIS represents real-world objects in four major categories: points, arcs (straight or curved line segments), polygons, or in a grid/pixel system. The point, line and polygon data model is referred to use the umbrella term “vector”, while the grid/pixel model (such as with a digital airphoto or satellite image that is “geo-rectified” to a cartographic projection system, such as latitude/longitude) is referred to as the “raster” data model. In each case, the GIS computer database keeps track of the geo-location of each object, and recognizes each point, line, polygon or raster pixel as a distinct object, having its own properties. Thus, for example, the computer keeps track of the geo-location, length and starting and ending points of each individual curvilinear line segment in an integrated road network. In addition, tabular data can be attached to each GIS object, and is incorporated into the central GIS database in “attribute tables”. Thus, for example, if cities or towns are represented by spatial point objects, then any of a host of socioeconomic variables describing those cities or towns can be included in the database and integrated (such as city/town population, number of households, date of incorporation, etc.). In the case of a curvilinear line segment in a GIS road network, variables describing the pavement type, approximate road quality, legal speed limit, or date of last maintenance can be included in the GIS database.

Travel-Time, Travel-Cost and Accessibility Indices

There is extensive spatial economic theory (agglomeration theory) that describes the fact that spatial access to markets, controlled by transportation costs, is crucial in economic development.

GIS can be used to assess accessibility as a function of road quality, the time of road building, geography, topography, and other factors that aid or hinder access (“quality of access”), including political or administrative policies or traffic congestion.[21] Using GIS to give continuous accessibility values to observational units, regressions between continuous indices and selected impact variables could be run, to illuminate trends and patterns, establish correlations, and bolster and support conclusions. Furthermore, multivariate regression models could be constructed with the inclusion of controls that might influence or mitigate true accessibility, such as tax policies or after-effects of natural disasters.

Spatial economic theory as far back as the famous Von Thunen (1826) land rent model has been based on the assumption that spatial access to markets, controlled by transportation costs, is crucial for economic development. In principle, improved access to consumer markets (including inter-industry buyers and suppliers) will increase the demand for a firm's products, thereby providing the incentive to increase scale and invest in cost-reducing technologies. Marshall (1890) showed that the geographical concentration of economic activities can result in a “snowball” effect, where new entrants tend to agglomerate to benefit from higher diversity and specialization in production processes. Workers and firms would benefit from gaining access to an agglomeration as they could expect higher wages and to have access to a larger set of employers. Furthermore, access to markets or economic city/town agglomerations can determine if a household is able to afford the cost of shipping products for sale, earning potentially higher wages in agglomeration centers, or gaining access to information spillovers or technology advances, further reducing costs.

There is a rich body of literature on the benefits to firms from gaining improved access or proximity to other firms in the same industry (Henderson, 1974 and 1988; Carlino, 1978). Theoretical and empirical work on urban economics and economic geography (see review by Henderson et al., 2001) suggests that the net benefits of industry concentration and location in dense urban areas are disproportionately accrued by technology intensive and innovative sectors. This is because the benefits of knowledge sharing (ideas) and access to producer services (e.g., venture capital) are considerably higher in these sectors than in low-end manufacturing that employs standardized production processes. As a result, these innovative sectors can afford the high wages and rents in dense urban locations and industry clusters. Paradoxically however, we find a considerable range of standardized industrial activity in most developing country countries. One explanation for this is the lack of inter-regional transport infrastructure linking small centers to large urban areas, thereby reducing the opportunities for efficient location decisions and de-concentration of large urban areas. In a recent empirical study, Henderson (2000) documents the linkages between improvements in inter-regional infrastructure and growth of smaller agglomerations outside of larger city centers.

In general, “access” to markets is determined by the household’s or village’s true cost of traveling to or accessing market centers. This could include the cost of transporting goods for sale, transporting (back to the village) key inputs for production or consumption, or the cost of transporting people for migratory or more permanent employment. Thus, effective access to urban markets also depends on the willingness and ability to afford transport costs, and these in turn are directly a function of road quality as well as actual measured road distance, topography,

[21] See, for example, Harvey J. Miller and Yi-Hwa Wu (2000), "GIS Software for Measuring Space-Time Accessibility in Transportation Planning and Analysis," *GeoInformatica* , 4, pp. 141-159 and Luis Rosero-Bixby, “Spatial Access to Health Care in Costa Rica and its Equity: A GIS-based Study,” *Social Science & Medicine*, 58, pp. 1271–1284

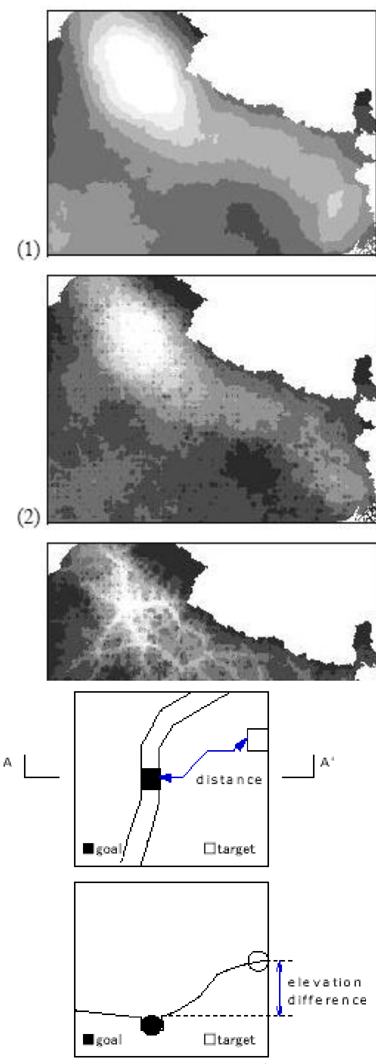
climate, rivers or any other potentially inhibiting (and thus more costly) exogenous geo-physical barriers.

The classic gravity model which is commonly used in the analysis of trade between regions and countries states that the interaction between two places is proportional to the size of the two places as measured by population, employment or some other index of social or economic activity, and inversely proportional to some measure of separation such as distance. Following Hansen (1959)

$$I_i = \sum_j \frac{S_j}{d_{ij}^b}$$

where I is the "classical" accessibility indicator estimated for location i (for example, a village), S is a size indicator at a market destination *j* (for example, population, purchasing power or employment), and *d* is a measure of distance (or more generally, *friction*) between origin i and destination *j*, while *b* describes how increasing distance reduces the expected level of interaction. Empirical research suggests that simple inverse distance weighting describes a more rapid decline of interaction with increasing distance than is often observed in the real world (Weibull, 1976), and thus a negative exponential function is often used.

There are several options for developing accessibility indicators depending on the choice of distance variables used in the computation. These include:
(a) indicators based on "straight-line" or Euclidean distance;
(b) indicators incorporating topography; (c) indicators incorporating the availability of transport networks;
(d) indicators incorporating the quality of transport networks;
and (e) movement across a "cost surface".

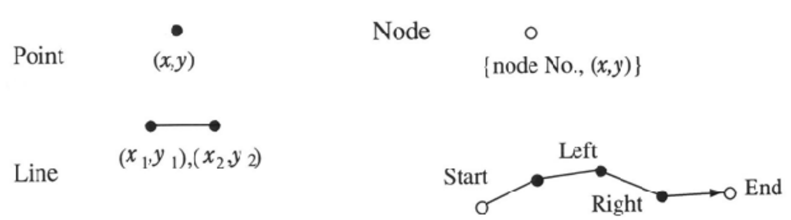


A better alternative is to use actual measured distance along road networks as the basis of the inverse weighting parameter and to incorporate information on the quality of different transportation links. Feasible travel speed and thus travel times will vary depending on each type of network link. A place located near a national highway will be more accessible than one on a rural, secondary road. The choice of the friction parameter of the access measure will therefore strongly influence the shape of the catchment area for a given point—i.e., the area that can be reached within a given travel time. This, in turn, determines the size of potential market demand as measured by the population within the catchment area.

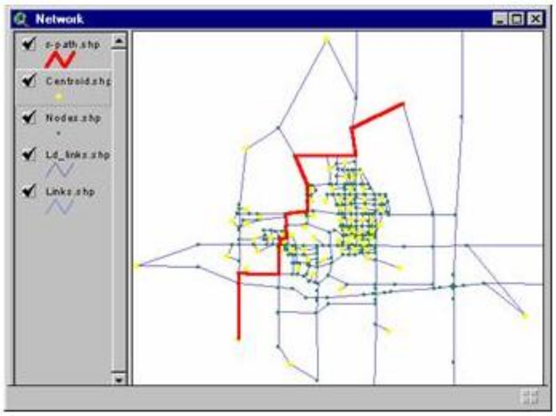
The figure here illustrates these points using an accessibility surface for the Northern Indian Gangetic plain using three measures of market access: (1) based on Euclidean distance, (2) network distance, and (3) network travel time. It is clear that indicators based on (1) and (2) overestimate potential market area. The variation in infrastructure quality (3) between regions leads to a more realistic representation of the structure of market areas. Thus, incorporating the quality of the transport network is important in assessing the potential market integration.

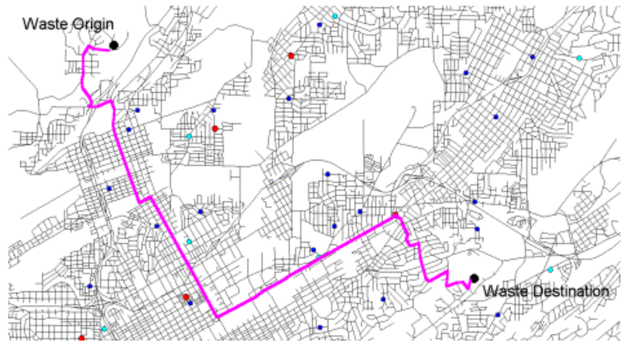
In studies related to agglomeration economies and economic geography (e.g., Hanson, 1998), the distance measure of choice is usually the straight-line (Euclidean, or "as-the-crow-flies") distance, which has the advantage of computational simplicity. However, this assumption of uniform (isotropic) plane is clearly unrealistic, particularly in countries where topography and sparse transport networks of uneven quality greatly affect the effort required to move between different parts of the country. Such an access index takes no account of the fact that hills and mountains greatly reduce travel times and greatly increase travel costs. Nor does it take into account the fact that people and goods move along road networks – not across a uniform plane. Topographic data (such as from contour lines digitized from paper maps, or from spot samples taken on the ground by surveyors, or from airborne or satellite instruments) can be converted using GIS algorithms to a continuous elevation *surface*. In that case, distance *across topography* can be calculated: the GIS calculates Euclidean distance, but then further calculates the actual distance on the ground considering topographic variation. This is partially illustrated by this graphic at right, where distance is measured both across the two-dimensional x,y surface, but also across the topographic z surface, calculating actual distance traveled (in meters, kilometers, etc.).

A far better alternative, however, is to use actual measured distance along existing road networks, considering the fact that goods and people move predominantly along infrastructure networks. This can be accomplished by obtaining an accurate digital GIS road network. Such a road network has all roads digitized into GIS digital "vector" objects. That is, rather than simply a graphical image of the roads, the road network is actually made up of many individual line segments, connected to each other at the end points, which are called "nodes". Each individual road segment in the larger network is "seen" by the GIS as an individual digital object. The GIS can calculate the exact length, direction and curvature of each line segment (just as it can for polygonal objects). These graphics illustrates the underlying road network structure, which is technically referred to as "vector line topological structure":

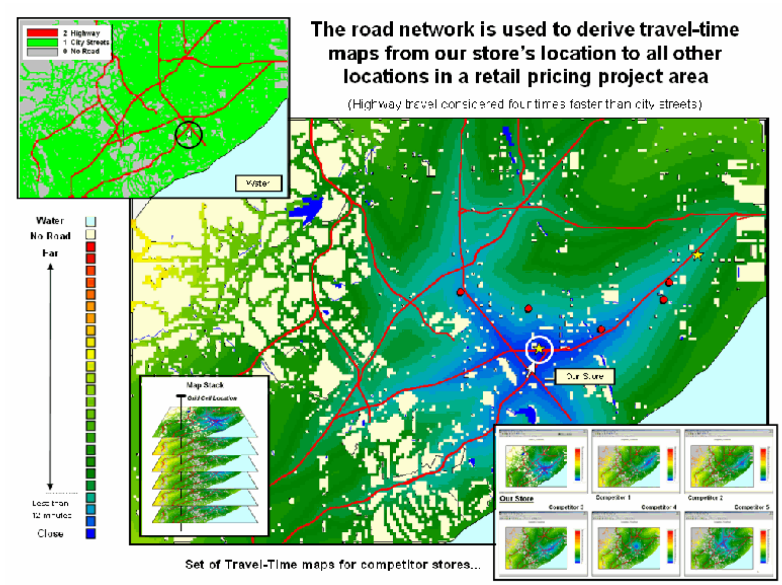


The GIS keeps track of the exact geo-location of each node connecting linear road segments, as well as the exact curve of each line segment. Thus the GIS can calculate precisely the exact distance along each segment in any desired unit (such as meters, kilometers, etc.). Thus, using advanced algorithms, the GIS can calculate travel distance *through* the road network from any node to any other node. Other algorithms (such as the Djikstra algorithm or variants) will pick a "shortest path" through the network to get from node X to node Y, minimizing travel distance, as in this graphic at right. Here the GIS has simply found the shortest path through the road network assuming that all road network segments are equal in terms of road quality or road speed. However, data on road quality or road speed of each individual road segment is often available,





and can be entered into the GIS database and attached to each road segment (in fact any amount of information on road segments – or any other object in the GIS – can be entered into the database, such as data on road segment names, date of paving, cost per segment, number bridges per segment, etc – all of this information is kept track of in the GIS database). If data on road quality is available, then approximate road speeds can be estimated. Typically, road maps categorize roads into categories.



For example, if a road is categorized as "one-lane paved", then an approximate road speed of 45 miles per hour could be assigned to all road segments with that categorization. Once categories of roads are assigned approximate road speeds, then travel times through the road network *considering road speed/road quality* can be calculated. This is a simple calculation: road length divided by road speed. For example, if the road segment is 50 kilometers long, and the road speed of that segment is 25 kilometers per hour, then the travel time (or travel "cost" if the definition of cost here is time) would be 50/25 or 2 hours of travel time. Often, this results in a different "least cost" or "least time" (if the "cost" is in terms of speed) pathway than the minimum distance pathway along all road networks. For example, it may be quicker in terms of time/cost to drive onto a highway and then exit to get to a destination than to travel along intermediate roads even though they provide a more direct link. Thus, the pathway of minimum distance may not always be the same as the pathway of minimum time or cost. This graphic shows the *fastest* route through a network from one destination to another, rather than the minimum distance route:

Topographic information could be combined with road network speed information, so that the road network segments are weighted by elevation or slope. For example, one might burn less gas or put less stress on a truck (lower "cost") to drive around a mountain than across it, even though the minimum distance pathway is across the mountain. In this case, the path of "accessibility" would likely be around the mountain.

Physiographic Data to Weight Travel Cost Estimations

While measuring distance along road networks incorporating data on varying road quality or varying road speed is a far superior method than measuring access "as the crow flies" (Euclidian distance) or even along road networks without considering road quality, the accuracy of the computed access indices can be further enhanced by incorporating weights that reflect further variable that impede travel, adding travel cost and time. For example, topography (as well as slope angle) is an extremely important variable that could dramatically alter travel times and costs, but might not be considered at all if only road network distance and road quality were considered. While digital data on a road network might indicate that a particular stretch of road was paved at high quality, with an official speed limit of 80 kilometers per hour, nonetheless in reality that stretch might involve movement up and down steep hills, in effect slowing travel time and increasing travel cost beyond what is measured simply by the road network data. Furthermore, a flat stretch of road in a low-lying area that rarely encounters debilitating weather such as snowstorms might overall be much easier (and cheaper) to travel than a similar flat stretch of identical road quality located at high elevations. On the latter, travel may frequently be inhibited by severe snow or ice, thus dramatically increasing travel costs.

Other important physiographic factors can affect actual travel costs and times, including land cover, climate, rainfall amounts, and the presence of lakes, rivers, streams and glaciers, which may periodically overflow, or swell during certain times of the year. Furthermore, a road network map may not indicate that certain areas are restricted because they are protected – either for conservation or military purposes, for example – and thus travel through them is impractical. In that case, the road network will need to be digitally altered to reflect the actual travel routes.

By the same token, certain physiographic factors can provide exogenous drivers of village economic productivity, such as inherently fertile soils that would result in higher agricultural productivity, or favorable rainfall patterns or climate, etc. Villages located in areas with good access to clean water, or with less intimidating (and costly) topography for villagers to drive and navigate, might have an inherent (exogenous) advantage over other villages with very similar socioeconomic measures. In their absence, for example, an economic increase in one village over another might be falsely attributed to superior road access, rather than to superior soil fertility, which may be the true driver. Or the reverse could occur, blunting the effective measurement of true, positive road benefits. Ignoring such physiographic conditions for villages could also ignore another key element of "accessibility": the fact that market access may be more valuable for some communities than for others. For example, a community with inherently poor soil fertility may benefit more from access to a fertilizer market than a community with inherently rich soils, but having the same level of access as measured by road distance, quality and even topography.

Once these data are assembled in the GIS, along with geo-locations of impacted communities (such as villages), then the GIS can quickly "map" to each community variables describing the respective physiographic conditions for each. Also, these data inputs can be used to weight the road network segments, as well as the areas of land leading to the nearest road network (in the case of villages that have no road network connection, if these exist).

References

Carlino, G. A. 1978. "Economies of scale in manufacturing location: theory and measure." Leiden: Martinus Nijhoff Social Sciences Division.

Dijkstra, E. W. (1959). "A note on two problems in connexion with graphs". *Numerische Mathematik*, vol. 1, pp. 269–271. doi:10.1007/BF01386390.

Hanson, Gordon H. 1998. "Regional adjustment to trade liberalization," *Regional Science and Urban Economics*, vol. 28, no. 4, pp. 419-444.

Henderson, J. V., Z. Shalizi, and A. J Venables. 2001. "Geography and development," *Journal of Economic Geography*, vol. 1, no. 1, p. 81.

Henderson, J. V. 2000. "The effects of urban concentration on economic growth." Cambridge: National Bureau of Economic Research, processed.

Henderson, J. V. 1988. "Locational pattern of heavy industries: Decentralization is more efficient," *Journal of Policy Modeling*, vol. 10, no. 4, pp. 569-580.

Henderson, J. V. 1974. "The sizes and types of cities." *The American Economic Review*, vol. 64, no. 4, pp. 640-656.

Marshall, A. 1890. "Some Aspects of Competition," *Journal of the Royal Statistical Society*, vol. 53, no. 4, pp. 612-643.

Weibull, J.W. 1976. "An axiomatic approach to the measurement of accessibility," *Regional Science and Urban Economics*, vol. 6, no. 4, pp. 357-379.

Von Thünen, J.H. and P. G Hall. 1826. "The Isolated State." Pergamon Press.

ANNEX 3: HOUSEHOLD SAMPLE SURVEY DESIGN

I. INTRODUCTION

This annex is extracted from the project report, *Sample Selection for MCA - Honduras Program Evaluation*, 30 January 2008. The extract presented here is a verbal description of the sample design and sample selection process. That report also contained the sample design for the evaluation of the FTDA project, which is not relevant to this report. It also contained a list of all of the selected sample units and a number of tables describing the population (frame) and selected sample. Those tables and the sample list are not reproduced here. (This extract is an exact copy of material from the cited report, and it contains references to the omitted tables.) The Microsoft Excel file, *SurvDes4TransportCaseríos.mdb*, contains a complete listing of the sampling frame and selected sample, along with sample selection data such as the selection probabilities (used to generate the weights used in the analysis presented in this report).

II. HOUSEHOLD SURVEY DESIGN FOR TRANSPORTATION PROGRAM EVALUATION

The primary objective of the survey of households for the transportation-program evaluation is to estimate the relationship of program (economic) impact to program treatment effect, where the primary measure of program treatment effect is reduction in travel time. For the transportation program, randomization was not used as a basis for selection of the road segments to be upgraded under the program. At one point, there was consideration of using randomization to select rural roads, but this was not implemented. For this reason, the estimation of program impact will rely solely on estimation of an analytical model of the relationship of impact to program effect as reflected through changes in travel time caused by the program.

At the beginning of this project, it was planned to use the Census segment as the primary sampling unit, because at that time a sample frame was available for that area sampling unit and data were available from a previous (2001) Census for that unit. After an initial visit to Honduras, it became clear that the government of Honduras possessed a large amount of geographic information system (GIS) data, which could definitely be used in the data analysis phase of the study, and possibly could be available in time to use in the survey-design phase of the study. As things turned out, it was possible to make use of the GIS data for the survey design. Since the GIS data were available down to the level of *caserío*, but not Census segment, it was decided to use the *caserío* as the primary sampling unit. It was viewed that this change was actually desirable from a statistical viewpoint, since the *caseríos* tend to be larger than Census segments (causing a reduction in loss of precision from intracluster correlation, which tends to be lower for larger area sampling units).

The sample frame was a list of 22,816 *caseríos* stored in the GIS (and the same as those available from the Census). Caseríos in the Islas de la Bahia and Gracias a Dios departments, and *caseríos* in protected status, were excluded as out of scope, reducing the sample frame size to 20,467. These are the primary sampling units (PSUs) for the study.

The design process began by identifying all known variables that might have had an effect on selection of roads for the program, and that may have a significant relationship on impact. (The word "known" means that data on these variables are available to assist survey design, i.e., are available prior to implementation of the survey data collection.) Three sets of variables were identified: (1) basic demographic and administrative variables, such as population size, agricultural region and urban/rural status; (2) data from the 2001 Census of households; and (3) GIS data.

The purpose of the planned survey is to collect data in support of the development of an analytical model, and the approach to designing the survey followed established procedures for constructing an analytical survey design (see the memorandum, "Sample Survey Design Details" for some background on this methodology).

The first step in the process was to examine the Cramer coefficient of correlation (a nonparametric measure of correlation) among all variables. The variables fall into two categories – those that are known to be closely related to the dependent variables of interest, and those that are simply candidate explanatory (independent) variables. Examples of the former are various measures of travel time and changes in travel time anticipated to be caused by the program intervention, from *caseríos* to various points of interest. Examples of the latter are demographic and administrative variables, Census variables, and physiographic variables produced by the GIS. The purpose of examining the Cramer coefficients is obtain information to guide combining explanatory variables that are highly correlated with each other (or delete some of them), and (to a lesser extent) to eliminate variables that show little relationship to the variables that are considered to be closely related to the dependent variable.

After conducting the analysis of the Cramer coefficients, it was decided that the large number of variables from the Census could be replaced by the Basic Necessities Index (NBI). Of the other variables, it was decided to retain all of the remaining demographic / administrative variables (agricultural region, urban/rural status, and population size), and most of the GIS variables. The sample selection was to be done in such a way as to ensure substantial variation on these variables, and low correlation among them. The procedures described in the "Sample Survey Design Details" memorandum were employed to select the sample.

The following is the final list of variables used in the survey design for the sample of *caseríos*. These variable names will appear in a number of tables to be presented to describe the sample design and sample selection process. The list presented below contains the definition of the variable as extracted from the GIS, and the definition of the recoding used in the sample-selection process. In the following, all travel times are in minutes, and all distances are in meters.

POP: population (individuals, not *viviendas* (households)
AUTO: [GIS definition?]; recoded in quintiles.
NRDIST2: distance (in meters) to nearest road (of any type); recoded in quintiles.
NRDSTT: travel time (in minutes) to nearest road (of any type); recoded in quintiles.
TTTEGUS1: travel time (in minutes) to Tegucigalpa before MCA improvements; recoded in quintiles.
TTEGUS1: same, after MCA improvements; recoded in quintiles.
TTEGUSD: reduction in travel time to Tegucigalpa after MCA improvements; recoded in quintiles.

TT133PP1: travel time to the nearest of the 133 largest Honduran cities and towns, before MCA improvements; recoded in quintiles.
TT133PP2: same, after MCA improvements; recoded in quintiles.
TT133PPD: reduction in travel time (to nearest 133 largest Honduran towns), due to MCA improvements; recoded in quintiles.
TT10CITY1: travel time to nearest of 10 largest Honduran cities, before MCA improvements; recoded in quintiles.
TT10CITY2: same, after MCA improvements; recoded in quintiles.
TT10CITYD: reduction in travel time due to MCA improvements; recoded in quintiles.
TTPC1: travel time to Puerto Cortes before MCA improvements; recoded in quintiles.
TTPC2: same, after MCA improvements; recoded in quintiles.
TTPCD: reduction in travel time due to MCA improvements; recoded in quintiles.
Region: agricultural region code (0-9); not recoded.
Urban: urban classification (1 = urban, 0 = rural); was recoded from 1 = urban and 2 = rural.
NBI; index of basic necessities; recoded in quintiles.
PRDIST: distance to nearest primary road; recoded in quintiles.
SRDIST: distance to nearest secondary road; recoded in quintiles.
CA5DIST: distance to nearest point on CA-5 highway; recoded in quintiles.
CITY5DIST: distance to the nearest of the five largest Honduran cities; recoded in quintiles.
MCAPSDIST: distance to the nearest point on MCA primary or secondary road improvement location; recoded in quintiles.
ELEVATION: elevation of the *caserío*, in meters; recoded in quintiles.
CLIMZONE: climate zone code; 0-9 unrecoded, values above 9 recoded as 9.
SOILCAP: soil capacity for the soil of the *caserío*; recoded in quintiles.
RAINREG: major rainfall rain code; evidently ordinal, so recoded in quintiles.
PREC_MM: median annual rainfall precipitation in millimeters; recoded in quintiles.
TEMP: median annual temperature in degrees Celsius; recoded in quintiles.
VEGCOVER: code for different types of major vegetation cover across Honduras; not recoded.
PROTAREAS: 1 if the *caserío* is in a nationally protected area or national park, 0 if not; not recoded.
HYDRODIS: distance in meters to the nearest major river (for all *caseríos*); recoded in quintiles.
TTMCAP1: travel time to nearest point on MCA-improvement primary road, before MCA improvements; recoded in quintiles.
TTMCAP2: same, after MCA improvements; recoded in quintiles.
TTMCAPD: reduction in travel time due to MCA improvements; recoded in quintiles.
TTMCAS1: travel time to nearest point on MCA-improvement secondary road, before MCA improvements; recoded in quantiles.
TTMCAS2: same, after MCA improvements; recoded in quantiles.
TTMCASD: reduction in travel time due to MCA improvements; recoded in quintiles.
TTC1000_1: travel time to nearest *caserío* with a population greater than 1,00 people (there are about 500 of these), before MCA improvements; recoded in quintiles.
TTC1000_2: same, after MCA improvements, recoded in quintiles.

TTC1000_D: reduction in travel time due to MCA improvements; recoded in quintiles.
TTMCAR1: travel time to nearest point on MCA tertiary improvement segment, before all road improvements; recoded in quintiles.
TTMCAR2: same, after MCA improvements; recoded in quintiles.
TTMCARD: reduction in travel time due to MCA improvements; recoded in quintiles.

All of the variables listed above were used as stratification variables in the design. The stratum categories are defined by the coding specified in the above list (e.g., the quintiles (values 0-4) or the agricultural region code (values 0-9)). In most cases, the stratum boundaries were determined by quantiles, so that the population frequencies are equal or similar for the various categories (they would normally be the same, but some variables contain large numbers of identical values, causing the numbers in the various quantile categories to differ from uniform). The use of quantiles to define the stratum boundaries is justified in cases where the nature of the relationship of impact to the variable is not well known. Since this is a "groundbreaking" study, the nature of the relationships is not known. The use of quantiles to define stratum boundaries is oriented toward "nonparametric" representations of relationships between impact and explanatory variables (since too little is known about the relationships to assume particular parametric forms (e.g., a linear relationship) at the present time). Another advantage of using quantile-defined stratum boundaries is that the stratum allocation results are unaffected by the particular measurement scales used for the variables. It is therefore particularly convenient when considering nonparametric relationships (it is also appropriate for parametric analyses, such as maximum likelihood estimation, which are invariant with respect to reparameterization (e.g., a scale transformation)). (Note that the methodology allows for completely arbitrary definitions of stratum boundaries. The choice of quantiles (e.g., quintiles) was guided strictly by conceptual desirability, and had nothing to do with any technical or programmatic constraints. The boundaries could have been specified using any points on any measurement scale. The use of quantiles is indicated because the nature of the relationship of impact to the design variables is not known. In a later study, the stratum boundaries would likely be set using natural-scale values.)

Most of the stratum cells were defined by quintile quantiles, and the codes for the five quintile categories (0-20%, 20-40%, 40-60%, 60-80% and 80-100%) are 0-4 (shown in the column headings of the tables to be presented later). For the stratum cells that are not defined by quintiles (or other quantiles), the following criteria defined the stratum boundaries. Each of the columns of the tables corresponds to a different stratum category, or "cell," and the code value of the cell is the table heading.

Prior to beginning the analysis, all *caseríos* from the Islas de la Bahia and Gracias Dios departments were deleted, since they are of little relevance to this program. All *caseríos* in a protected status were also deleted.

Agricultural Region ("Region")

0. Islas de la Bahia (all units deleted)
1. Sur
2. Centro Occidental
3. Norte
4. Litoral Atlantico

5. Norte Oriental
6. Centro Oriental
7. Occidental

Urban / Rural Status ("Urban")

Rural
Urban

Climatic Zone ("Climzone")

Values 1-8 as stored in the GIS; all values greater than 8 coded as 9.

Vegetation Cover ("Vegcover")

Code values 1-9 taken directly from GIS

Protected Areas ("Protareas")

Unprotected
Protected (all deleted from the sample frame)

Table 2, entitled "Caserío Population Frequencies," shows the number of population units in each stratum cell. The stratum identification codes (0-9 in all cases) are the numeric column headings in the last 10 columns of the table. This table reveals that, apart from stratification variable having quintile-defined stratum boundaries, the proportion of the population falling in the various cells varies tremendously. As discussed in the Design Report, a sample of 100 primary sampling units (PSUs, here *caseríos*) is to be selected. If simple random sampling were employed, the design would not be at all satisfactory. While many of the strata having quintile-defined boundaries show good variation, in some cases the stratum allocations are far from desirable. For example, the expected number of urban *caseríos* would be less than one, and the number of sample *caseríos* near project roads would be very small (e.g., less than 5 within 30 minutes of a primary project road, less than 10 within 30 minutes of a secondary road, and less than one-third within 30 minutes of a rural project road). Also, some of the distributions of the sample for travel-times and travel-time differences would be quite "unbalanced," such as variation by a factor of three in allocation of the sample to agricultural regions, and a single sample unit in the stratum representing the highest reductions in travel times from *caseríos* to Tegucigalpa.

Table 3, entitled "Caserío Desired Sample Allocation" shows the sample sizes that are desired for each stratum cell. The rationale for these allocations is as follows.

1. The total sample size is 100.
2. An equal distribution of the sample was specified over the seven regional strata.
3. The distribution of the sample over the quintile-defined stratum cells is uniform when the population distribution is uniform.
4. If the distribution of the population over quintile-defined categories is highly unbalanced, the desired allocation is set so that a minimum of 20 sample units occurs in extreme categories, and a minimum of 10 for inside categories. It is noted that changing the

stratum allocations in these cases can introduce substantial variation (or distortion from a uniform distribution) in the probabilities of selection (or cell sampling fractions), and the setting of the stratum cell sizes is done with a view to keeping this distortion to a low level.

5. A total of 80 sample units is desired for cells corresponding to travel times of less than 30 minutes to a project road (CA5, secondary, or rural), and 20 to cells corresponding to travel times of greater than 30 minutes. This is done to ensure that a substantial portion of the sample is relatively close to project roads.
6. No stratum allocations are specified for any of the "time 2" travel times, since these are highly correlated with "time 1" travel times.
7. Except for travel times to project roads, no stratum allocations are specified for any "time 1" travel times, since no reason was seen for doing this.
8. No stratum allocation was specified for climatic zone or vegetation cover, since those variables were non-ordinal (measurement scale) with many categories, and it was not apparent how to reasonably group them into a smaller number of categories that might relate well to impact.

Once the desired stratum allocation had been specified, the algorithm described in the "Sample Survey Design Details" memorandum was applied to construct a set of sample selection probabilities corresponding to this specification. A random probability sample of 100 units was then selected, using these selection probabilities. Table 4, entitled, "Caserío Actual Sample Frequencies," shows the distribution of this sample of 100 units over the stratum cells.

The selected sample shows reasonably good variation in all important stratification variables except region. It is important to realize that the actual stratum allocations do not have to be exactly equal to the "desired" – what is important is that the amount of variation is adequate to support the development of analytical models, and the obtained sample will do so.

(It is noted that the design produced here does not make full use of the methodological procedures described in the "Sample Survey Design Details" memo. In particular, the procedure of stratifying over product variables to promote orthogonality was not employed. That procedure is time-consuming to employ, and no instances were identified where it would have been highly desirable to apply it. Instead, highly correlated variables were removed from the design (e.g., no stratum allocations were specified for any "time 2" variables, and all of the approximately 200 variables of the Census were collapsed into a single variable (viz., the NBI)).)

A map (in the Sample5_maps1.pdf file referenced earlier) was prepared showing the location of each sample *caserío*. The important thing to observe on the map is that the sample *caseríos* have wide geographic distribution and that many are located near program roads. It is seen from the map that there are a number of rural roads (mainly in remote areas) that have no sample *caseríos* located nearby. This is an artifact of the design, which placed emphasis on many other design variables: in order to satisfy all of the desired constraints (i.e., the desired variation on all of the travel-time, physiographic, demographic and administrative variables), remote *caseríos* do little to improve the desired sample variation, and hence they tend not to be included in the sample. This is a little disturbing, however, since travel time is not the only variable of interest in the survey. It is the one, however, that we have most information about, through the GIS model. Much additional information will be collected during the course of the survey, from households, communities and traffic studies. *For this reason, it is recommended that the sample presented* *here be augmented with additional caseríos – one or two near each of the roads that presently have no sample caseríos near them.*

This may be done in two ways – (1) either add about a dozen additional *caseríos*, randomly selected from near each rural road that presently has no sample *caseríos* nearby, to the sample; or (2) reduce the sample size from 100 to 90, design and draw the sample again, and add about a dozen *caseríos*, randomly selected from near each rural road that has no sample *caseríos* nearby, to the sample. From a statistical viewpoint, these additional *caseríos* introduce a technical difficulty when constructing unbiased estimates that take the probabilities of selection into account. (The device of considering them part of a "certainty stratum" does not work here, because they are included after the sample has been selected.) This is not much of a difficulty, however, since these units, small in number, may be omitted from the calculation of unbiased estimates (relative to the total population). Much of the analysis, however, will disregard the selection probabilities and construct estimates that are *conditional on the particular sample selected*, not on the total population. This is because the selection probabilities are of much less importance in the estimation of an analytical model than in estimation of population means or totals; for estimation of an analytical model, the bias associated with basing the model on the selection probabilities is considered negligible compared to the increase in precision associated with assuming the probability of selection of each unit to be the same. The addition of the additional *caseríos* will improve the precision of these estimates – especially of estimates other than travel times (which they will not affect much at all, in view of their irrelevance to the analytical-model design) – and will introduce no significant additional bias.

Based on review of the sample characteristics, especially those available from plotting the sample on a map, the following recommendation is made. (It was not possible to foresee this recommendation, since it depends on the nature of the population, not on the sample design and selection methodology.)

Recommendation: For rural roads having no nearby sample units, randomly select a nearby *caserío* for the household survey. *Action item:* Decide whether to add these additional *caseríos* (about 10 of them) to the already-selected sample of 100, or to reduce the sample size (e.g., to 90) and then add the additional *caseríos*, so that the final sample size is about 100.

Table 5, entitled, "Caserío Sample for the Transportation Program Evaluation," presents a list of the 100 *caseríos* selected for the sample.

As mentioned in the main report, and discussed in detail in the next annex (on determination of sample size), the plan was to select 20 households from each sample *caserío*. This would have resulted in a sample of size 2,000, which was recommended based on power calculations. Because many of the *caseríos* were small, the sample size of 2,000 was not realized. The final sample size for the baseline sample was 1,600 households in the 100 sample *caseríos*.

ANNEX 4: FORMULAS USED FOR STATISTICAL POWER ANALYSIS

I. INTRODUCTION

The sample sizes (numbers of sample *caseríos* and number of sample households per sample *caserío*) were determined in consideration of the statistical power desired to detect impacts of specified magnitude. The sample sizes that were decided on were100 *caseríos* and 20 sample households per *caserío*, for a total household sample size of 2,000. Because many *caseríos* were small, the average number of households selected per *caserío* was less than planned, and the total number of households interviewed in the baseline was 1,600.

As discussed in the main text of this report, the evaluation design for estimation of impact was an analytical model based on travel-time-related continuous treatment variables. The estimate of impact involves a difference in means (of outcomes variables) between two survey rounds. The means are estimated from the regression models, conditional on the changes in the means of travel-time-related variables in both rounds (and adjusting to the mean values of covariates).

To conduct a statistical power analysis, information is required on a number of items: the impact estimator being used; the test parameters (power level, significance level); the minimum detectable effect; characteristics of the sampled (target) population (means, standard deviations, intra-unit correlation coefficients (if multistage sampling is used); and the sample design to be used for the sample survey. In the design phase, not a lot was known about the statistical properties of many of the outcome variables of interest in this study. For income, analysis of available data showed that the coefficient of variation (ratio of standard deviation to mean) of rural incomes in Honduras was about one. For estimation of proportions in the vicinity of .5, the coefficient of variation is also about one. This value (coefficient of variation equal to one) was assumed for the power calculations presented below.

When the project began, a decision had not been made on the choice of the primary sampling unit. Based on considerations of the GIS model, it was decided to use the *caserío* as the primary sampling unit. The Kish design effect (deff) used in the following formulas includes loss of precision associated with multistage sampling. The formula for deff is deff = 1 + (m-1)icc, where m denotes the number of sample households selected per sample PSU and icc denotes the intra-unit correlation coefficient. The value of icc depends on the outcome variable. For example, for m = 20 and icc=.2, the value of deff is 4.8. In the following, the value of deff is modified by a factor to reflect design effects additional to multistage sampling, such as the effects of stratification and matching of treatment and comparison roads.

In the course of the project planning, a number of different cases were examined, corresponding to a range of values of the parameters that affect power and sample size. The table shown below represents part of the "sensitivity analysis" that was conducted to determine the *caserío* and household sample sizes.

The material shown below is extracted from the *Design Report* dated October 30, 2007, with some simplification. When the *Design Report* was written, the choice of sampling unit had not yet been made, and consideration was given to use of Census segments as the primary sampling unit (PSU). After a thorough review of the available GIS data, it was decided to use the *caserío* as the primary sampling unit. The term "cluster" used below hence refers to *caserío* (and "Census segment" is to be replaced by *caserío*). In the course of the power analysis used to estimate sample size, a number of different cases were examined. The table presented below examines one selection of parameters.

All *caseríos* and all households in the country are subject to sampling. The sample frame from which the *caserío* sample was selected was the GIS database. The household sample was selected by systematic random sampling from within sample *caseríos*.

The next section deals with determination of the household sample size within *caseríos*, and the section after that deals with determination of the *caserío* sample size.

II. DETERMINATION OF HOUSEHOLD SAMPLE SIZE WITHIN CASERÍOS

The standard approach to determining sample size (for clusters and households within clusters) and allocation (to strata) are: (1) to specify a total budget for the survey and then configure the design to maximize precision of certain estimates or power of certain tests of hypothesis; or (2) to specify desired or required levels of precision (of certain estimates) or power (of certain tests of hypothesis), and configure the design to minimize cost. Since a survey budget has not been specified, but it is anticipated that it would be sufficient to fund a survey of several hundred clusters and several thousand households, the approach used in this case will be mainly the second one, with some iterations expected if the total cost becomes "too large."

To make sample-size estimates, information is needed about the relative cost of sampling clusters (Census segments) and elements (households) within clusters; about the variances of estimates of interest; about the intracluster correlation coefficient for estimates of interest; and about the intraclass correlation coefficient of strata for estimates of interest. Information is known from previous similar surveys about sampling costs but, as noted earlier, this evaluation design is a new one, and little information is known about estimate variances or the intracluster or intrastratum correlation coefficients for the variables of primary interest (e.g., estimates of change in economic impact as a function of travel time). There is certainly *some* prior information available on variability, however, since the proposed survey will in fact include many of the same variables that have been included in previous surveys – just not on the primary phenomenon of interest (the relationship of change in impact (income, employment, access) to program interventions or its surrogate (latent / endogenous variable), change in travel time.

To assist the survey design, a statistical analysis was conducted to estimate the value of the intracluster correlation coefficient (icc) for a selection of about a dozen variables of the 2001 Honduran Census, and for a general measure of socioeconomic well-being (households lacking three or more basic necessities ("Necesidades Básicas Insatisfechas" (NBI)). The analysis was conducted using formulas presented by Kish (*Survey Sampling*, Wiley, 1965) and using the Stata statistical analysis program. The intracluster correlation coefficient was estimated not only for Census segments, but also for two other area sampling units (aldeas and municipios). The results are as follows:

Conglomerado	Rho (formula Kish)	Rho (Stata)

Segmento censal	0.1949	0.1941
Aldea	0.1184	0.1471
Municipio	0.0599	0.0804

It is noted that the icc's presented in the table are for a single composite variable, NBI, and that the icc varies depending on the variable. For the selection of other (raw) variables taken from the Census (e.g., presence or absence of a refrigerator, presence of a specified type of water, attainment of 4^{th} grade education), the icc varied for Census segments varied from close to zero to as high as .8. The value for NBI, about .2, is fairly typical, and, in the absence of icc estimates for the variables of primary interest in the present survey (change in impact measures associated with program interventions (as reflected in change in travel time) over two years), it will be used to suggest reasonable sample sizes. (The results presented in the table vary a little by estimation type (Kish, Stata) because of different estimation approaches. The Kish formulas assume a fixed cluster size, which is true for segments but not so for aldeas and municipios, and so they were calculated for subsample of clusters of similar size.)

It is expected that an interviewer could conduct two to four household interviews (lasting about an hour) per day, once present in the Census segment. In this case, the ratio of cluster sampling cost to household-within-cluster sampling cost varies from approximately 10:1 to 100:1, depending on how long the questionnaire is. If travel costs between clusters are not very large, then the following formula specifies the "optimal" within-cluster sample size, as a function of the sampling cost ratio and the icc:

$$m_{opt} = \frac{S_2}{\sqrt{S_1^{\ 2} - S_2^{\ 2} / M}} \sqrt{c_1 / c_2} \approx \sqrt{\frac{c_1(1 - icc)}{c_2 icc}}$$

where

S_1^2 = variance among primary (cluster) means
S_2^2 = variance among subunits (households) within primary units
M = cluster size (number of households per cluster)
n = number of clusters in sample
m = number of households sampled per cluster
c_1 = variable cost of sampling per cluster
c_2 = variable cost of sampling per household
C = total variable sampling cost = $c_1n + c_2nm$
icc = intracluster correlation coefficient.

Substituting $c_1/c_2 = 10$ and icc = .2, we obtain $m_{opt} = 6$. If $c_1/c_2 = 100$, then $m_{opt} = 20$. The preceding estimates are based on a number of assumptions, and the results vary according to the variable (since the icc varies according to the variable). Prior survey experience in Honduras suggests that the value of m = 20 is a reasonable one for the within-cluster sample size, and that is what is proposed for the present survey.

It remains to specify the number of clusters to select. As in the case of determining a reasonable intracluster sample size, prior information can, along with a number of assumptions, suggest a reasonable value or range of values. Since this evaluation design is unlike any other, however, it should be recognized that the sample size estimates that follow are simply rough guidelines, making use of best available prior information.

The objective of the evaluation research design is to provide estimates of adequate precision for the relationship of change in impact (income, employment, access) to program interventions, as reflected in change in travel time. There are two standard approaches to sample-size estimation, specifying either the precision of an estimate (e.g., by specifying the width of a confidence interval) or the statistical power of a test of hypothesis. The present study is more concerned with estimation rather than tests of hypotheses (e.g., determining whether results are different for different subpopulations, or at different times (e.g., before-and-after an intervention)), and so the "power" method will be emphasized. Sample sizes will be estimated, however, using both methods (since the survey data will be used both to make estimates of means and to conduct tests of hypotheses (e.g., about differences among subpopulations, such as comparisons by gender, level of education, urban/rural status, sector, and region).

Since we have little prior information about the variance of the estimates of primary interest (change in impact as a function of change in travel time), we shall limit consideration to estimation of proportions, for which the variance is a function of the mean. It is recognized that the estimates of primary interest in this project are *not* proportions, but if the survey is designed to efficiently produce estimates of adequate precision for proportions for a variety of socioeconomic variables in the population of interest, it is reasonable to expect that it would to provide adequate precision for a the socioeconomic variables of interest in this evaluation. This cannot be affirmed with certainty, but it is the best that we can do with the information that is already available, without undertaking a costly and time-consuming preliminary ("first-phase") survey to collect preliminary data which would enable a better full-scale survey design to be constructed. In any event, the planned survey will collect data on many socioeconomic variables that are similar to those collected in the Census, and the results presented here will certainly pertain quite well to those variables.

It is noted that the formula for determining sample size using statistical power analysis can be expressed in terms of the coefficient of variation (CV) of the outcome variable. The coefficient of variation for a proportion equal to .5 is 1.0. The coefficient of variation of income in rural areas of developing countries is in the range .5-2. For Honduras, an analysis of Census data showed that the CV for income in Honduras was about 2. The CV varies for each outcome variable. The table presented below applies to any variable for which the coefficient of variation is 1.0.

Note that the total household sample size depends on the intra-unit correlation coefficient (icc) for the first-stage sample units (PSUs). The icc affects the within-PSU household sample size. Once the within-PSU sample size, m, has been specified, the PSU sample size depends only on the icc and m through the design effect (deff). The sample size formula to be presented is in terms of the deff, not icc and m. The design effect is the ratio of the variance of an estimator for a specified design to the variance for a simple random sample of the same sample size (of second-stage sample units). For a two-stage sample design, the value of deff is given by deff = 1 + (m-1) icc.

[The material from the *Design Report* dealing with sample size estimation based on specification of precision is omitted here, since it was the power-based estimates that were used.]

III. ESTIMATION OF SAMPLE SIZE BASED ON SPECIFICATION OF STATISTICAL POWER

The formula for the power of a statistical test of an hypothesis of the equality of two group means (i.e., the probability of rejecting a null hypothesis) is:

$$\Pr\left(\frac{\hat{\mu}_1 - \hat{\mu}_2}{[deff\,(\sigma_1{}^2/n_1 + \sigma_2{}^2/n_2 - 2\rho(\sigma_1/\sqrt{n_1})(\sigma_2/\sqrt{n_2}))]^{1/2}}\right) > z_{1-\alpha} \left| \mu_1 - \mu_2 = D \right) = 1 - \beta$$

where

μ_1 = mean for group 1
μ_2 = mean for group 2
n_1 = sample size for group 1
n_2 = sample size for group 2
σ_1 = standard deviation for group 1
σ_2 = standard deviation for group 2
ρ = correlation between the group estimated means
α = significance level of *one-sided* test of hypothesis of equality of group means (the probability of Type I error, i.e., the probability of rejecting the hypothesis of equality of group means, when it is in fact true) (e.g., .05)
β = the probability of making a Type II error, i.e., the probability of accepting the hypothesis of equality of the group means, when it is in fact false) (e.g., .1)
$1 - \beta$ = power of the test (e.g., .9)
$z_{1-\alpha}$ = 1-α percentile point of normal distribution (e.g., 1.6449 for α=.05, or 1.2816 for α=.1)
deff = design effect
D = (true) size of difference between group means,
and a caret (ˆ) over a parameter (symbol) denotes a sample estimate.

(Note that α refers to the significance level of a *one-sided* test of hypothesis. This corresponds to the case in which there is little doubt about which of the two group means will be larger. For situations in which it is not known which of the two group means might be larger, a *two-sided* test would be appropriate. In that case, the value $z_{1-\alpha}$ in the preceding formula (and the one given below) should be replaced by $z_{1-\alpha/2}$. Whether to use a one-sided test or a two-sided test depends on the circumstances. Assuming a one-sided test corresponds to a smaller value of z (i.e., $z_{1-\alpha}$ is less than $z_{1-\alpha/2}$), and to a smaller sample size estimate (a smaller sample is required since there is some *a priori* information about which group mean might be larger).)

The corresponding formula for the sample size is:

$$n_1 = \frac{deff\,(z_{1-\alpha} + z_{1-\beta})^2(\sigma_1{}^2 + (1/ratio)\sigma_2{}^2 - 2\rho\sigma_1\sigma_2/\sqrt{ratio})}{D^2}$$

where

ratio = ratio of the group sizes = n_2/n_1 (i.e., n_2 = ratio x n_1).

The power calculations are somewhat simpler to discuss and present if done in terms of the coefficient of variation (CV), or standard deviation divided by the mean, and the effect size, D, is also expressed relative to the mean. This is done for the table presented below.

In the present application, we shall assume the following values for the parameters of the above formula:

deff = 1 + (m-1)icc = 1 + (20-1)(.2) = 4.8
σ_1 = 1.0 (i.e., a CV of one)
σ_2 = 1.0
α = .05 (so $z_{1-\alpha}$ = 1.6449)
β = .1 (so power = .9, and $z_{1-\beta}$ = 1.2816)
ρ = .5 (corresponding to interviewing of the same households in both survey rounds)
ratio = 1.

With these values, the formula for n_1, as a function of D, is:

$$n_1 = 41.10913/D^2 .$$

Let us denote the total sample size (of both groups) as $n = n_1 + n_2$.

It remains to specify values for D. In the economic analysis that preceded the Transportation Project, it was estimated that the economic rate of return (ERR) for the project would be in the range 12-21 percent. One of the main outcomes of interest for this project is income, which is correlated with ERR. We hence consider values of D in this same range. If we set D=.1, .15 and .2, we obtain the following values for n: 8,222, 3,654, and 2,055. These are household sample sizes, so that, with 20 households selected from each cluster, the corresponding cluster sample sizes are 411, 183, and 103. (Note that these are the total number of households and clusters for *both* waves of the panel survey.)

The following table summarizes the preceding discussion.

Estimate of Household-Survey Sample Size (Roads Evaluation) Based on Specification of Power (.9)

D	Estimated Sample Size (based on 20 households per cluster)	
	Households	**Clusters**
.1	8,222	411
.15	3,654	183
.2	2,055	103

Based on the preceding considerations (estimating sample size by specifying precision and power), the size of the NORC evaluation contract, and the perceived level of funds available for data collection, it is viewed that a minimum acceptable sample size for the household survey is about 4,000 households, selected from 200 clusters (in both survey rounds). Since this is a panel survey, the same households are interviewed in both survey waves, each wave involving 2,000 interviews in 100 clusters. (For costing purposes, it does not matter that the same households and clusters are sampled in the second wave – the cost is the same as if they were sampled independently.)

ANNEX 5: DOCUMENTATION OF COMMENTS FROM PARTICIPANTS AT THE PRESENTATION OF INDEPENDENT EVALUATION RESULTS OF THE TRANSPORTATION PROJECT, MAY 2, 2013, AND NORC RESPONSES

Based on the discussions and comments that surfaced during the May 2, 2013, dissemination workshop, NORC proposed to MCC the following changes to the existing Transportation Project Evaluation Report. In the following, a summary of the workshop discussion is presented, followed by associated changes made to the Final Report.

1. Discuss the short-term and long-term effects of the intervention separately. The estimated changes in travel time and cost were short term, and they were noticeable. The change in income and employment are longer term, and it is not surprising that the results were weak, since the project was hardly completed when we conducted the endline. We will strongly emphasize this point in the report.

NORC Response: *The revised report addresses this issue by pointing out that there are several factors associated with the small magnitude of impact observed in the study. They include the following: the second round of the household survey was completed prior to the completion of all road improvements; that improvements to rural roads are of short duration (on the order of six months, until the next rainy season); and that for many of the indicators of interest, we would expect a considerable lag before we detect significant changes. For example, we expect that changes in access times would occur immediately after completion of the improvement, whereas changes in income, employment, or behavior related to health and education might take several months or a year or more to manifest. Note that the CTV estimate of impact is conditional on completion and maintenance of the road improvements, whereas the BTV estimate reflects the situation as of the time of the second round of the household survey.*

2. Make clear that the impact estimate refers to a randomly selected household in the country, and that the national-total impact of the program is obtained by multiplying the household-total impact estimate by the number of households in the country. There should be more discussion of the fact that the design estimates the "ripple" effect of the program intervention throughout the entire road network and hence, for the entire country. That the target population for which we are estimating impact is all households in the country is a key point, and it is not sufficiently discussed in the report.

   NORC Response: *We have now included this discussion n Section G.2 of the report, which presents national-level impact estimates that attempt to capture the ripple effects of the intervention.*

3. Include a discussion of the validity of the continuous-treatment-variable (CTV) GIS model for estimation of impact.

   NORC Response: *Additional discussion of external validity has been added, in Section C.2.*

4. Disaggregate estimation of effects by road type. At present, the impact is presented for all road improvements combined.

   NORC Response: *The impact estimates based on the CTV model refer to all road improvement projects – primary, secondary and rural – combined. The BTV model was configured to provide separate estimates of impact by road type, but the power associated with that approach was inadequate to show statistically significant income effects for secondary and rural roads. The CTV model could have been used to estimate impact by road type, but this would have required separate estimation of travel times (using the GIS road network model) by road type, and the project analysis plan did not include this extension. In the project planning, it was decided not to disaggregate results by road time under the CTV model. (The decision was made to focus resources, not because it was thought that impact would be low, so that little benefit would occur from disaggregation.)*

5. Jack Molyneaux made the statement that the sample size should have been based on a power analysis. It was, as described in the *Design Report*. Note that statistical power analysis was used as the basis for sample size estimation for both FTDA and Transportation studies. We will emphasize this in the report.

   NORC Response: *A detailed description of the statistical power analysis used to estimate sample sizes (caserío and household) has been added, as Annex 4.*

6. Jack Molyneaux also stated that a power analysis based on a more specific population, for which we might expect stronger effects, may have been preferable for this evaluation. Unfortunately, this was not the evaluation design proposed to MCC in our *Design Report*. The purpose of presenting the evaluation methodology and statistical power analysis in the *Design Report* and submitting it to MCC for review and comment early in the project was to obtain agreement on the sample size, and agreement that it was sufficient to detect effects of an anticipated size with high power. Conducting statistical power analysis for a more restricted population (for which greater impact might be expected) was never part of the design. It is made very clear in the report that a detailed statistical power analysis was conducted in the design phase of the project, to ensure a sufficient sample and that the evaluation was not underpowered. We will also conduct an *ex post* statistical power analysis to determine whether this was in fact the case. The *ex post* power analysis should ideally be disaggregated by road type and by distance from the project roads to show impact as a function of distance from the project road. We will determine whether sufficient data exist to do this and, if so, we will conduct the analysis and include it in the report (Note: this is a time-consuming exercise, but one that NORC would like to undertake).

   NORC Response: *We conducted an ex post statistical power analysis for both the CTV and BTV models. Results are referenced in the main body of the report and presented and discussed in detail in Annex 1. The relationship of impact was investigated by means of regression models that included travel times to project roads as explanatory variables. No statistically significant relationship was observed. Such a relationship is to be expected. A likely reason why it was not observed is that the project roads were distributed throughout the country, so that if a household is far from one project road, it has an increased chance of being close to another project road.*

7. We would like to replace the three impact estimators with a single estimate, given that the presentation of multiple estimators has given rise to much consternation.

   NORC Response: *Two different approaches were used to estimate impact, one based on the use of binary treatment variables and the other based on the use of continuous treatment variables. The CTV approach was the basic approach recommended in the design phase of the study. Discussion of the BTV approach was included to facilitate understanding of the CTV approach, which is more complex (based on estimation of partial treatment effects). We used a single estimator for each approach. Only the CTV estimates are presented in the main text. It is considered important to include results for both the BTV and CTV models, since the BTV approach (the conventional approach) was suggested in the M&E Plan, and the evaluation project recommended the CTV approach instead. As a "learning exercise," a detailed comparison of both approaches is included in the Final Report.*

8. Include OLS regressions on questionnaire travel times, and estimate the correlation of the questionnaire travel times with the GIS-model travel times.

   NORC Response: *We conducted a canonical correlation analysis to compare the set of questionnaire travel times to the set of GIS-model travel times. The results of this analysis are included in Annex 1. The complete set of individual correlations of every travel time with every other travel time is included in the .log file, but is not copied into the report.*

9. Include detailed descriptions of the GIS travel-time model as an annex.

   NORC Response: *Include as Annex 2*

ANNEX 6: SUMMARY OF RESPONSE TO REVIEWER COMMENTS FOR MCC ROADS EVALUATIONS PEER REVIEW LEARNING WORKSHOP, SEPTEMBER 20, 2013

On September 20, 2013, MCC conducted a Roads Evaluations Peer Review Learning Workshop at its office in Washington, DC. In preparation for this workshop, the Draft Final Report dated July 31, 2013, was reviewed. The following tables summarize the reviewer's comments in the left-hand column and describe NORC's proposed response in the right-hand column. The table does not include positive reviewer comments, which do not require a response.

In addition, we have also responded to comments received from MCC reviewers at an earlier date.

Comments from Alexander Rothenberg	
Comment	**Response**
2.1. Identification	
I felt like the discussion of identification, while mentioned in the appendix, was not in the primary report and needed to be, particularly in Section C and / or Section G.	A description of the causal model, statistical model specification, parameter and effect identification, and estimation procedure that is common to all models has been added to the main text. Additional description beyond this is added for specific models, in Annex 1.
The discussion of identification, where it appears, is in Annex 1, and it needs to be much tighter and sharpened.	Same response as above.
I could not find a clear, succinct statement of what the identifying assumptions are, really anywhere in the text.	Same response as above.
2.1.1. Difference-in-Differences	
The key identifying assumption required for using difference-in-differences is the common trends assumption. Is this identifying assumption credible? This is hard to assess, but one approach commonly used in the literature is to see if the pre-treatment trends in outcomes for treated and control sites are similar.	The double-difference measure of impact was used only for the binary-treatment-variable (BTV) model, not for the continuous-treatment-variable (CTV) model. The BTV model was used only for comparison purposes. It is the estimates of impact from the CTV model that are presented in the main report and for that model it is not necessary to make a "common trend" assumption.
2.1.2. Continuous Treatment Effect Estimates	
The authors seem to think that using GIS travel time data effectively solves the identification problem, which is basically that the road improvements were not randomly assigned across locations. On page 56, the authors state: "The travel times estimated by the GIS travel-time model are exogenous variables." This is absolutely not correct and makes me suspicious about the authors' understanding of endogeneity problems. What assumptions are the authors making to achieve identification? The key assumption required for	Discussion of selection effects has been included. Fixed-effects models and fixed-effects estimators are used to estimate impact. The project is fixed – impact estimates are constructed for the particular Transportation Project. *Caseríos* and households are selected using a nationwide probability sample in which every *caserío* and every household in the country is subject to sampling. Selection effects are not an issue, with respect to the project roads or with respect to the sample units (*caseríos* and households) of the household sample survey. The

Comments from Alexander Rothenberg	
Comment	**Response**
identification is the following…. More work needs to be done to argue the credibility of this assumption.	GIS-model travel times are in fact exogenous, given the model specification (a fixed project, and a nationwide probability sample of households).
2.1.3. Specification Problems	
I would strongly encourage the authors to use a specification that is closer in spirit to (1).	Additional discussion and justification of model specification has been added. Note that the model specifications for the binary-treatment-value (BTV) approach and the continuous-treatment-value (CTV) approach are quite different. In the BTV model, impact is reflected in the coefficient of the round by treatment interaction term. In the CTV approach, an equation like (1) was in fact used to estimate the partial treatment effect. The text describes the difference: "In the BTV models …, the main estimate of impact was one of the regression-model coefficients (an interaction term of treatment and time). For the CTV models, the estimate of impact is not represented in a single model coefficient. Instead, the model coefficients represent *partial treatment effects*. The estimate of impact is obtained as a linear combination (sum) of those coefficients multiplied by the changes in means of the treatment-related variables between the two survey rounds."
2.1.4. Other Sources of Identification	
Finally, the authors make note that several proposed segments of the Honduran roads were scaled back during implementation, due to budget pressures and rescoping. In particular, 4 segments were listed as "terminated" in Table 1. Is there any way to make use of these terminated segments for identification?	The estimates presented in the report are fixed-effects estimators for the project as finally configured. Under this assumption, there is no selection effect with respect to the project roads.
2.1.5. Complementarities Between Transportation and Rural Development Projects	
Also, the authors discuss that the transportation project was taking place at the same time as a rural development project (Annex I, Section III.F, p. 83). It would be really interesting to see if the program data could be used to shed light on any interesting complementarities between the different projects. Are there areas where the transport project took place by itself, without the FTDA project? Are there areas where the FTDA project took place, but road improvements did not? If so, then it seems like it would be possible to estimate the independent impact of both the road and FTDA projects separately, as well as their interaction, would be possible.	The continuous-treatment-value estimates of impact are conditional on travel times, and therefore not sensitive to confounding with the FTDA project, which had, apart from the rural-road component that was incorporated into this evaluation, nothing to do with travel time.

Comments from Alexander Rothenberg	
Comment	**Response**
2.2.1. Random Effects vs. Fixed Effects.	
Random effects has really fallen out of favor in the econometric literature, and I'd strongly recommend avoiding random effects in any treatment effect estimation going forward, especially given the absence of random assignment.	For the earlier draft of the report, a Hausman test was performed for all models. For estimation of impact, the fixed-effects estimator was used if the Hausman test was significant (i.e., indicated a statistically significant difference in the parameters of the fixed-effects and random-effects models); otherwise the random-effects estimator was used. For estimation of relationships to time-invariant variables, the random-effects estimator was used. There was little difference in the estimate of impact, whichever assumption (fixed-effects or random-effects) was made. For simplification, <u>in the current version of the report</u>, fixed-effects estimates are used to estimate impact, in every case (i.e., for every outcome variable).
2.2.2. Measurement of Binary Treatment	
The authors focus on a specification where settlements within 30 minutes travel time to improved roads are treated, and those that are within 30 minutes of a comparison road are controls. The choice of 30 minutes seems to be pretty arbitrary, and it might be useful to investigate the sensitivity of these binary treatment response effects to differences in the 30 minute buffer (i.e. report effects for 20 minutes, 10 minutes, etc.). This would be a useful robustness check, and it may lead to interesting results on how the effects of improved roads vary with distance to the treated site.	The zone of influence was defined by 15 minutes, 30 minutes, 45 minutes, and 60 minutes. Impact had maximum sensitivity for 30 minutes, and that is why that value was selected.
2.2.3. Measurement of Continuous Treatment	
In the continuous treatment specifications, the authors focus on a measure that is the travel time to the nearest town of a population of 1,000 or more. This is fine, but it may also be useful to consider estimating effects of a broader index, measuring market access, which is defined as follows….	The rationale for focusing on travel time to the nearest town of a population of 1,000 or more was presented in the report. A principal components analysis was conducted of all travel-time variables. They are highly intercorrelated, and any one of them explains much of the variation explained by all of them. The travel time to the nearest town of population 1,000 or more had the strongest relationship to outcomes of greatest interest. Using a composite measure would change things little, while increasing complexity and reducing face validity.
2.2.4. Mean Effects Analysis	
In the paper, the authors present estimates for a huge number of outcomes. With so many outcomes, hypothesis testing becomes difficult, and the exercise borders on data mining. One approach	The face validity associated with the individual outcomes is considered valuable. As noted above, a principal components analysis of travel times was conducted. As a result of this analysis, the number

Comments from Alexander Rothenberg	
Comment	**Response**
to resolve this would be to focus on a smaller number of more aggregate outcomes. Another approach would be to conduct a mean effects analysis (Kling et al., 2007). This involves grouping related outcomes and estimating the impact of changes in infrastructure access on an index of the dependent variables. This has been shown to be useful in situations with lots of outcome variables, especially for improving power and summarizing effects.	of treatment-related variables was reduced from travel times to ten points of interest to a single one (travel time to nearest town of 1,000 or more).
2.2.5. Results Discussion	
Generally, I think the discussion of the results could be greatly improved. One approach would be to convert the continuous results into elasticities, so that we can say things like a 10% reduction in distance to treated roads is associated with an X% increase in investment. Right now, the results discussions focus on pointing out only significance, and while statistical significance is doubtless important, economic significance is also crucial.	The impacts are very small in absolute magnitude. Conversion to elasticities would result in very small numbers and, as such, we do not consider it a useful exercise in this situation. The impact estimates are presented along with baseline means in the tables dealing with statistical power analysis. Both statistical and economic significance were commented on.
2.3. Other Outcomes	
2.3.1. Sustainability of Road Investments	
Sustainability of Road Investments: There's quite a bit of discussion in the document about how the road improvement projects may not be long lasting, due to problems with construction and poor quality materials. I think that in any evaluation of road improvement projects, understanding why poor road construction outcomes occurred is really important. Could there be better ways of selecting the road builders, or incentivizing them to build roads properly? Would better monitoring, or financial threats for poor performance, have lead to better outcomes? While this is clearly outside the scope of this study, this seems like a really important area for future research.	Agreed.
2.3.2. Non-Linear Response	
An advantage of the continuous specifications is that we can explore potential non-linearity in the relationship between changes in transport costs and outcomes. So far, the authors have focused entirely on estimating linear treatment response. It would be useful to see more flexible specifications, either semi-parametric estimation, or a flexible polynomial in transport costs, to determine whether there are any interesting and important non-linearities in the data.	Response was graphed by distance from project roads, and no discernible nonlinear relationship was observed. Regression models that included travel time to nearest project road (in addition to the other model variables) showed no statistically significant effect.
2.3.3. Treatment Effect Heterogeneity	

Comments from Alexander Rothenberg	
Comment	**Response**
The estimates so far focus on average treatment effects, but I'd be really interested to see whether the road improvements had larger effects for households that were poorer, more agricultural, and spent more on transport in the last few months. In some specifications, these effects are estimated, but the coefficients are not separately reported. I think this sort of work would be really useful and would provide a much richer story for the distributional consequences of the program.	Interaction terms were included to estimate the relationship of impact to some covariates, but not those specifically mentioned. Since the overall effects are so small, the value of investigating these relationships is small. However, we agree that additional efforts to investigate heterogeneity of the effects may be interesting to see.
3. Smaller Comments	
It would be great to see a map of the locations of the road projects, which are listed in Table 1. That would give us a good sense of the spatial variation in the treatment.	We have added three maps, showing locations of project roads and comparison roads, and locations of the sample *caseríos* of household sample survey.
The report needs some carefully editing and rewriting before it is posted on the MCC website. There are many typos, sentences that need to be rewritten, and grammar mistakes that need to be addressed.	We have corrected the few typographical errors that we found.
The authors cite a number of papers from the literature on transport improvements and economic development, but they do not include a bibliography (it only appears for a few sections in the appendix). It would be great if the final report had this information.	We included references (in footnotes) for texts and articles directly relevant to the analysis, at points where they were relevant.
It would be nice to see tables reporting all of the coefficients in an appendix.	All coefficients for all models are included in the Stata .log file. This file is hundreds of pages long. Regression coefficients and standard errors were reported for key models.
The raw STATA output that appears in the appendices isn't particularly attractive, and should be cleaned up if possible.	The courier font retails the vertical alignment of the program output.
How much imputation of missing data took place? When imputation occurred, why did you use means to impute? It would be great to discuss how much of an impact the censoring and imputation had on the final results; ideally, you would report separate estimates for the non-imputed and imputed data.	Counts were included for all response and explanatory variables in tables in Annex 1. We do not believe the regression models that would have been obtained if incomplete observations were dropped are useful.
Did you take logs of most of (or any of) the dependent variables? It seems that this would be really useful to do before estimating effects.	Logs were taken for income variables, in some regression models.
4.1. Technical Rigor and Factual Accuracy	
Generally, the report communicates its hypotheses and methods clearly, but more work needs to be done on identification and estimation issues. Right now, the technical rigor of this is lacking, but I think with a bit more work and a better	As stated, we have included more discussion has on model specification, parameter/effect identification and estimation. The panel-data regression procedures employed are appropriate and correct.

Comments from Alexander Rothenberg	
Comment	**Response**
understanding of panel data regression techniques, this could be seriously improved.	
4.2. Follow the Program Logic	
Overall, I think the report does a good job of following the MCC program logic, focusing on estimating the impacts of intermediate and final outcomes associated with the transportation improvement project. I'm not entirely sure what the value of the GIS data, specifically for estimating travel times, was for the project. To the extent that it wasn't useful (because it does not solve the endogeneity problems), this would represent a waste of resources.	The impact estimates presented in the report are conditional on completion and maintenance of project activities, as reflected in the GIS-model travel times. These conditional estimates were based on the GIS-model travel times. The GIS model was an essential part of the continuous-treatment-value model used to estimate impact. Please see Annex 7, which we have added to this version of the report, for additional discussion in Annex 7 in support of this view.
4.3. Clarity and Expected Contribution	
I think the clarity and exposition could really be improved, particularly on the identification assumptions, and I've highlighted this above. Overall, I think more work here would make the report much better, and I think this is an area where the report is relatively weak.	Substantially more explanation has been included on model specification and effect identification.
4.4. Peer Reviewer Recommendation	
I think with more work, this study could be a nice contribution to the literature on transport infrastructure and economic development. The identification issues, however, are pretty important and may be difficult to overcome. Absent better identification, working with a structural model would help place the results of the study in a good journal.	More explanation has been included on model specification and effect identification. We believe that the model specification is well-founded, and have presented detailed reasons for this view in Section C.2 and in Annex 7.

Comments from Máximo Torero	
Comment	**Response**
The authors overstate some aspects of the design, for example the exogeneity of travel times and the GIS model.	As mentioned, a detailed discussion of causal model, statistical model specification, parameter and effect identification, and estimation procedure has been included that pertains to all models, and specific additional material included for specific models. Given the stated assumptions for the analysis (fixed project, nationwide probability sample of households), the GIS-model travel times are in fact exogenous.
The continuous treatment results are difficult to discuss, because of the lack of explicit equations ("model") estimated.	We presented explicit equations for all models (model parameters area presented in tables). Additional explanatory material has been added to the model descriptions.
One would be hardly press to call the methods use here "structural." The authors should improve the presentation of the report, and make the "structure" of their estimation methods explicit, in main text.	Additional description has been added. The descriptor "structural" has been dropped.
The authors should discuss the selection biases that might remain and how they differ in the DD methodology and the continuous treatment methodology.	We now have expanded the discussion of selection bias in the main text. It describes why there is no selection bias at the project or household levels for the CTV model upon which the impact estimates are based. Concerning the selection bias difference between the DD and the CTV methodologies, we do not discuss this since the former was not the basis for our estimates of impact.
The matching exercise was not presented. Since these are not the main results, it is of limited importance. The matching seems to be only done for the traffic survey, in which case the selection issues when using the household survey should be discussed more.	We have included additional description of the traffic surveys. The matching of the treatment and comparison roads was done by judgment (taking into account road type, location, and "connector" status), not by statistical methods. As discussed earlier, selection of roads is not an issue, under the stated assumptions (impact estimated for a fixed project).
The sampling methodology should be clearer in the main text. Specify the strata used and be clear about what "marginal stratification" entails.	We have included an additional description of the sampling methodology to the main text. A detailed description of the stratification process is presented in Annex 3.
Authors do not use the traffic data to run simple regressions to see the improvement in the intermediate outcome.	The traffic surveys were planned and designed to estimate speeds for the GIS travel-time model, not to estimate impact (see Design Report). We have added a recommendation suggesting this additional work for a future study.
The authors provide little context for the specific evaluation.	Context was provided in the Executive Summary and Introduction.
The timing of data collection was done before the completion of the projects and arguably after the "useful" life of the rural roads.	Agreed. The BTV estimates are as of the time of the second survey round. The CTV estimates are conditional on completion and maintenance of

Comments from Máximo Torero	
Comment	**Response**
	project road improvements. The road sample size is indeed small, and was a motivation for the fixed-effects approach. The *caserío* sample (100) is modest. The household sample size (3,008 in both survey rounds, with 1,408 interviewed in both rounds) is considered adequate. The *ex-ante* statistical power analysis suggested these sample sizes, and the *ex post* statistical power analysis showed that they were adequate (for tests of hypotheses about effects relating to impact based on CTV estimates).
The sample size is small and attrition seems to be an important issue to discuss. Originally the sample was to be 2,000 household because of sample frame issues this was 1,600 at baseline and then 1400+ at follow up.	The roads project that existed after several components were terminated was the project evaluated, using fixed-effects estimators conditional on baseline conditions. We conducted an *ex post* statistical power analysis for all outcome measures reported. We also analyzed attrition and examined a selection model for attrition. No correlates with attrition were identified.
It would be recommendable to do a longer term effects study, given that the household data is not very costly and that a community survey focused on smaller areas should be feasible.	Agreed.
Heterogeneity by road type could be useful to relate all the results. Note that simple interactions would be enough for this, in lieu of "running" the GIS model by type of roads (as proposed in pg.28), since doing this would go against the argument that impacts are to be larger since they are reflected across the road system, even if no near the improved roads.	There is no single road type associated with travel to points of interest. Inclusion of interactions is not a feasible way of addressing this issue.
Heterogeneity across gender and socioeconomic status at baseline should be explored.	Detailed tables were presented of all analyzed response and explanatory variables at baseline. Because of the overall weak response, little effort was allocated to estimation of heterogeneity of response. Analysis was done of the relationship of outcome to travel times to project roads.
The presentation of the results is long and convoluted. It seems that the extension of the report is due to previous comments that the authors address. However, while the use of annexes is a good way to do it, the reader should have enough detail in the main body to understand what is being done and the results presented are achieved	Some parts of Annex 1 have been shortened, and additional descriptive material included in the main text.
Some of the outcomes defined were not discussed nor presented in the tables. For example, the basic/grain other crops impacts where not presented. If they are defined in the text one would	This is true. Detailed models were developed for aggregate measures of income, such as NetHHInc. The impact was very weak, and so the various components of income were not analyzed in detail

Comments from Máximo Torero	
Comment	**Response**
expect to see them in the tables.	(as they were in the FTDA report), and are not included in tables. All of the travel time and access indicators were included in tables.
A descriptive map of the roads should be presented in the main text. Having the aldeas in the sample shaded in the map and/or the location of surveyed households would be illustrative of the variation being exploited. (like the ones in pg.112-113)	Three descriptive maps, showing locations of project and comparison roads and sample *caseríos*, have been added to the main text.
The mechanisms through which the impacts are achieved are not discussed thoroughly and no attempt is made to quantitatively trace the mechanisms.	A high-level causal model diagram has been added. More discussion of causal mechanisms has been added to the description of specific models.
There is a gross misrepresentation of the endogeneity problem at hand when evaluating the effect of road improvements.	Additional discussion of causal relationships has been added. Endogeneity is not a problem. The impact estimates are conditional on the fixed project and a nationwide probability sample of households. Under the assumptions made (fixed project, nationwide probability sample of house-holds), the GIS-model travel times are in fact exogenous.
The discussion of the choice between random effects and fixed effects, while largely correct, relies on Hausman test, which have their pros and cons, and do not use any theory/intuition based reasoning to select them.	The impact estimates are fixed-effects estimates. Random effects estimates were used to assess the effects of some time-invariant variables (which are not estimable in ("drop out of") the fixed-effect models.
The way that the sampling design is being accounted for is not very convincing. All regressions should be clustered at the caserío level. When the estimation is fixed effects at the household level, most of the other aspects of the design are accounted for by these, but if one is to use the random effects strata (at least) should be controlled for.	Regression models were run with and without *caserío* indicator variables. Negligible difference in estimates of interest. Since there were 116 *caseríos*, it is not appropriate to leave them in the model, given their unimportance.
Be consistent in how thing are modeled. The outcome equations are sometimes given by differences/changes and sometime by the level. Also using an overly general functional form makes the explanation more confusing and the reader more uncomfortable in pairing the model versus the empirical results. This is especially true if one wants to pass this as somewhat "structural"; be clear about your assumption and what you ARE and NOT doing.	Additional discussion has been added concerning the use of travel times vs. changes and travel times. Conditions under which outcomes may be expressed in terms of changes in travel times are discussed. Models are described in general terms and in complete detail.
Note that with two rounds of surveys the difference in difference and the fixed effects should be numerically equivalent. In the results presented, pg.70-72, they are not. This is because of the demeaning, which the authors say it would be	The double-difference estimate presented was the "raw" double-difference estimate, ignoring the sample design. Taking into account the sample design, the double-difference estimate is in fact the fixed-effects estimate.

Comments from Máximo Torero	
Comment	**Response**
easier to interpret.	
On pag.9, note that “networks” effects are present because of the GIS model, but these are not separately identified; the “allowing for network effects to be estimated” might be misleading.	The word “estimated” has been replaced by “taken into account.”
Seeing how the effects change by type of road would be interesting. As the authors note (pg.28), running the GIS model improving one type at a time would get at this, but entails more than was planned. However, simple interaction with the type of road that the travel time variable refers to, meaning what is the type of road nearest to that household, could shed light on these differential effects.	Travel to points of interest is often over multiple road types. No single road type is associated with a travel time. In this case, the approach of using interaction terms does not work.
Note on pg.29, the vector of mean presented there comes as a surprise to the reader that has not read the annex; the presentation should be self-contained.	Presentation has been made clearer.
Table 5 [on page 30], should present the standard errors. The standard deviation of the “structural” parameter is not a population moment of interest. The standard deviation is just trying to measure the degree of uncertainty of the estimation of the parameter.	The column labels are in error. The column entries are in fact standard errors.
On pg.32, the baseline means should be discussed so that the reader can get a sense of the impact estimates. An extra column in the impact table might prove useful.	The impacts are very small compared to baseline means. The impact estimates and baseline means and their quotient, are included in the tables dealing with statistical power analysis.
Note17 on pg.54, seems to point to the need of discussion attrition. It seems that households that are nearest to the roads in the baseline are less likely to respond to the survey.	Attrition was examined. A selection model for nonresponse was examined. No correlates of nonresponse were identified.
Given the timing of the survey, the results on income are hard to believe. The authors should focus on the intermediate outcomes (like travel time, etc.)	Report focused on estimates for both income and intermediate outcomes. Both are presented in all summary tables. A recommendation has been added to conduct additional analysis of the traffic survey data.

Comments from unnamed MCC reviewers (recorded in file MCCCommentsOnNORCTransportEvaluation.docx)	
Comment	**Response**
Power calculations (Transport Project Evaluation). Please note that the justification given for the power calculations that linked income changes to the level of the ERR appear to suggest a weak understanding of the relevant economic relationships between returns on the roads investments and implications for broad economic growth. MCC will be asking the peer reviewers to critically assess this justification in an upcoming peer review workshop of our Roads evaluation portfolio.	There is no misunderstanding of the difference between an ERR and an increase in a farmer's income. The client did not specify expected changes in income in the *M&E Plan*. The anticipated benefits of the program were stated in terms of ERR. The evaluation design focused on changes in income, not changes in ERR. The *ex-ante* statistical power calculations done to estimate sample size were related to the coefficient of variation of an outcome indicator. This was made clear in the report. The fact that the statistical power analysis was done for arbitrary outcome variables (having specified coefficients of variation) and sample units having specified intra-unit correlations has been emphasized.
2. Exposure to treatment: Given the variation in road completion dates, can NORC clarify the exposure to treatment range for households?	The exposure to treatment is not reflected in questionnaire variables. Because of the differing completion dates, the primary estimates presented in the draft final report were conditional on completion and maintenance of all project components. Impact is conditional on the Transportation Project as finally configured. This is made clear in the report.

ANNEX 7: EXAMPLE ILLUSTRATING FIXED EFFECTS AND ENDOGENEITY

Background

In the several reviews of this report, much confusion reigned concerning the issues of endogeneity, fixed effects and selection effects.

*Endogeneit*y refers to a mutual causal relationship between an outcome variable and an explanatory variable of a model. The presence of endogeneity complicates the estimation of causal effects. *Selection effects* refers to the fact that the road segments or the households included in an analysis may not be simple random samples from identified populations. If not properly taken into account, selection effects may lead to biased estimates of impact. *Fixed effects* (in the present context) refers to constructing estimates for under specified conditions, such as estimating impact for a fixed project, or for the particular sample of households selected for the household sample survey.

The approach of estimating impact for a particular (fixed) transportation project was adopted for several reasons: (1) the uniqueness of the primary road (CA-5); (2) the lack of randomization in the selection of secondary and rural roads; and (3) the small size of the sample of secondary roads. To take selection effects into account, and develop estimates for a broader (statistical) scope of inference, it would have been necessary (because of a lack of randomized assignment to treatment) to develop a selection model. A selection model would relate to selection of road segments (because they were the unit of treatment), and the road-segment sample sizes were very small. For the reasons stated, it was not feasible to develop a selection model for road segments. The approach adopted therefore was to estimate impact for the particular project represented by the Transportation Project.

The issues of endogeneity and selection effects are reflected in the reviewers' comments that road location and household location are determined jointly, when a homeowner decides where to locate his home.

Selection effects may come from two sources – the selection of the road segments (by the government), and the selection of the households (by NORC). By assuming that the project is fixed, selection effects associated with the first source disappear. With respect to the second source, the household sample was a nationwide probability sample. It was stratified by travel time to project roads, but included all of the country (except for some remote regions and tourist areas). In estimating impact, the effect is averaged over all sample households (including those close and far from the project roads). This estimate is the impact (of the fixed project) on a randomly selected household in Honduras.

It is noted that for the analysis, a fixed-effects assumption was made. This means that the locations of the households are considered fixed (once they have been selected). This assumption does not remove the household selection effect. The household selection effect is removed because the household sample is a probability sample from the entire country (excluding certain remote areas and tourist areas), and impact is estimated over the entire probability sample.

Example

A very simple example will be presented to clarify the concepts. Suppose that the Transportation Project consisted of improvement to a single road segment, and that it is desired to assess the impact of improving this single road segment on the households of Honduras. The unit of treatment is the road segment, and the unit of analysis is the household. The objective is to estimate the effect of improving this single road segment on a randomly selected household in Honduras. It is *not* the objective to estimate the impact associated with improving a road segment that is randomly selected from the population of all eligible road segments in Honduras. If this latter were the goal, it would matter very much how the road segment was selected. Given the goal of estimating the impact associated with that particular road segment, however, how it was selected is completely irrelevant. The segment may have been chosen as the very worst segment in the most economically deprived part of the country, with the potential of making the largest improvement in household welfare. The segment may have been chosen as the country's "showcase" road segment, in which case the impact of further improvement may be negligible. Improvements to different road segments will have different effects on households. The effect of the road improvement will doubtless vary, depending on how the road segment is selected (i.e., on which road segment is selected). Irrespective of how the road segment was selected, however, the objective is to estimate the impact of that *particular* road segment. The estimate of impact will not depend on how the road segment was selected. For the purpose of this estimation objective, this road segment – this project – is fixed.

To estimate the effect of improving the road segment, a nationwide sample of households is selected. Selection effects may be introduced either through the means of selection of the treatment units (roads), or the selection of the households. The sample is a probability sample selected from the population of all households in the country (with the exclusion of certain remote or tourist areas). Every household has a known, nonzero probability of selection. The sample might have been a simple random sample, or it may be a highly stratified two-stage sample. What matters is that it is a probability sample from the entire nation (excluding the remote areas and tourist areas), so that the effect of the road segment improvement can be estimated for a randomly selected household in Honduras. Note that if the (*caserío* and household) sample size is large, it does not matter whether a fixed-effects or random-effects assumption is made with respect to the higher-level sample units (*caseríos* and households). That decision may affect the standard error of the estimates of impact a little, but it will not bias the results. Note, in particular, that the sample survey may be stratified by distance or travel time from the project road. That non-uniform allocation of the sample does not bias the sample estimates of impact, since the entire country is stratified, and each household has a known probability of selection. The sample may be a simple random sample or a highly complex sample – because it is a probability sample, it may be used as a basis for estimating the impact of the project.

With respect to the issue of endogeneity, the following comments are made. Within a household, many household variables are mutually causally related. For example, household income and travel times and costs are likely to be mutually causally related. The issue of interest is whether GIS-model travel times are mutually causally related to household characteristics. Prior to selecting the road segment (i.e., if the road segment were not fixed), there could be any sort of relationship between household characteristics and the road segment location, and hence, the GIS-model travel times. This follows since it is possible that selection of the road segment is made on the basis of variables related to household characteristics, such as income. In other words, unconditionally (without fixing the project), the GIS-model travel times and household characteristics may be correlated. Once the road segment is fixed (specified, conditioned on), however, the GIS-model travel times have no stochastic relationship to household characteristics – they are fixed numbers. The key point to remember is that, once the road segment is fixed, then the scope of inference is restricted to that particular road segment.

The same reasoning applies to a project that contains more than one road segment. If the road segments are not fixed, then there may be a stochastic relationship between the road segment locations and household characteristics. For simplicity, let us restrict consideration to projects consisting of a single road type (i.e., a well identified population of eligible road segments, such as secondary road segments). In this case, if the goal is to make inferences about the effect of a randomly selected road segment (of a given type), then it is necessary to take into account the method of selection of the road segments. If they are randomly selected from the eligible population, then the impact estimates represent the average treatment effect of improving a randomly selected road segment. If they are selected in some other way, then it is necessary to take into account the selection model, by means of a sound causal model that describes the relationship of causal variables to selection and/or outcome (e.g., using a Rosenbaum-Rubin ("statistical") approach or a Heckman ("econometric") approach to impact estimation).

If, however, we are interested in estimating impact only for a particular project, however configured, the issue of selection of the road segments becomes irrelevant. When the road segments are considered fixed (i.e., the project is considered fixed), then the estimates are conditional on the selection – in a directed-acyclic-graph causal model (Figure 4e), this conditioning breaks the causal link from selection to all variables affecting outcome. Once the household locations and the project locations are fixed, the GIS-model travel time is simply a number, not a random variable. Once the household sample is available, the relationship of household variables to the GIS-model travel times may be related to GIS-model travel times, but the GIS-model travel time for a particular household is fixed. Once the project and the household are fixed, there is no endogenous (stochastic) relationship between the GIS-model travel time and any household variables. The household variables may be causally related to the GIS-model travel time (as a fixed number), but the GIS-model travel time is not causally dependent on any household variable. There is no *simultaneous* relationship (mutual causal relationship) between these variables.

Note that the household sample survey may be stratified to be more heavily concentrated near project roads, and no bias is introduced by doing this, as long as the sample is a nationwide probability sample. A different sample (or sample design) would produce similar results. If the sample were restricted to a peculiar subnational area, however, (e.g., to a "near" buffer zone and an "outside" buffer zone that did not include the rest of the nation), the impact estimates would not represent the impact expected for a randomly selected household in the country.

In the preceding, the households of the sample survey are considered fixed effects. If this restriction were dropped, the basic result would be the same (although the standard errors of estimates would be greater, and the power of tests of hypothesis less). The important feature of the example is the fixing of the project (and averaging over all households in the country), not adopting a fixed-effects estimation model for the household sample survey.

On the Need for the GIS-Model Travel Times

One reviewer questioned the need for the GIS-model travel times in this analysis. The GIS-model travel times were essential to the CTV approach. Under the assumption of a fixed project (and a nationwide probability sample of households), the GIS-model travel times are exogenous, whereas the questionnaire travel times are highly endogenous with other household variables. Because of this endogeneity, the questionnaire travel times may not be used to develop the partial-treatment-effects model. That model (showing the relationship of outcome variables to travel time) requires the use of an exogenous travel time, or an instrumental variable for the questionnaire travel times (such as the GIS-model travel times).

The questionnaire travel times are endogenous regardless of the fixed status of the project or households. The GIS-model travel times are endogenous if the location of the project is allowed to vary or if the household sample survey is not a nationwide probability sample. Under the assumption of a fixed project and a nationwide probability sample of households, the questionnaire travel times are still endogenous (with household outcome variables), but the GIS-model travel times are not.

Absent the GIS-model travel times, the adopted approach of basing impact estimates on a partial-treatment-effect travel-time model would not have been feasible (because the questionnaire travel times are endogenous with respect to other questionnaire outcome variables of interest, and an instrumental-variable model would have been unavailable for them).

Some reviewers objected to referring to the GIS-model travel times as exogenous, pointing out that the location of the project roads and the locations of the households are jointly determined. The assertion was made that the fact that a given household is at a certain location is jointly determined with income and other household characteristics, the GIS-model travel time is not exogenous with respect to household characteristics. Prior to the determination of the project location (i.e., if the project location is considered as a random variable), the project location and household location may indeed be jointly statistically related (i.e., endogenous). Once the project location and the household location are fixed (or the household sample survey is a nationwide probability sample), however, this joint relationship disappears (i.e., in Pearl's terminology, the causal link is "severed" by the conditioning). The project road locations are fixed, and are no longer correlated with household location or other characteristics. Once the project location is fixed, it is no longer even a random variable, and so it has no correlation with anything (since correlation is defined for two random variables, and if one is fixed (conditioned on), it is no longer a random variable).

The key point here is that the project is considered fixed, and the impact estimates are conditioned on this particular fixed project. The project is "fixed" in the same sense that Tegucigalpa is fixed.

The partial-treatment-effects model must take into account not only the GIS-model travel times but also other variables that may affect outcome variables of interest. For a single cross-sectional survey, constructing such a model represents a formidable task, since travel times may be correlated with many other variables, with the result that the travel-time effects are confounded with these variables. When two rounds of a panel survey are available, however, this task is much simplified, since (in a fixed-effects analysis) time-invariant variables drop out of the model. This applies both to observed time-invariant variables and to unobserved time-invariant variables. More significantly, it applies to variables (observed or unobserved) that may be correlated with travel time, and with which the effect of travel time (on outcomes of interest) may be confounded. By using two survey rounds in a fixed-effects analysis, the confounding influence of all time-invariant variables is reduced or removed. In particular, the effect of unobserved time-invariant variables is removed (so that, under the assumption that all unobserved variables affecting outcome are time invariant, there are no unobserved variable in the model, so that ordinary least squares may be used to construct unbiased estimates of model coefficients. (This point has to do with confounding, not with endogeneity. It is an advantage of the pretest-posttest design having either treatment and control groups (and a binary treatment variable) or a continuous treatment variable.)

The point was made (by a reviewer) that the partial-treatment-effect model coefficients depend on project location. That is absolutely correct, but it does not represent a deficiency of the approach (of estimating impact conditional on the project's being fixed). The partial-treatment-effect model is conditional on the project – there is a different set of model coefficients for every different project. What is being estimated are model coefficients for the particular, fixed project. It is emphasized that *we are not estimating the impact of a randomly selected road segment, but the impact associated with a particular project.*

It is not the case that the GIS travel times were used because they were considered more precise estimates of travel times. They are not. Relative to a particular household, they are much *less* precise than the questionnaire travel times. They are measured from the *caserío* centroid (rather than from the household location), and are conditioned on many variables other than road characteristics and project intervention status (vehicle type (a pickup truck), season, day of week, time of day, weather). They are of interest because, given the project as fixed, they are exogenous, and the questionnaire travel times are not. The main reason why GIS travel times were used was that, under the assumption of a fixed project, they are exogenous variables that enable implementation of the continuous-treatment-variable approach, whereas the questionnaire travel times do not.