MATHEMATICA
Policy Research

REPORT

REVISED REPORT

Design Report: Evaluation of the Tanzania Road and Mafia Island Runway Activities

May 31, 2019

Delia Welsh
Anthony Harris
Anu Rangarajan
Patricia Costa
Ariel Ben Yishay
Sarah Hughes
Laura Meyer
Seth Morgan
Ali Protik

Submitted to:
Millennium Challenge Corporation
875 15th Street, NW, Suite 200
Washington, DC 20005-2227

Project Officer: Cindy Sobieski
Contract Number: MCC-17-CON-0002

Submitted by:
Mathematica Policy Research
1100 1st Street, NE
12th Floor
Washington, DC 20002-4221
Telephone: (202) 484-9220
Facsimile: (202) 863-1763

Project Director: Anu Rangarajan
Reference Number: 50365

This page has been left blank for double-sided copying.

CONTENTS

CONTENTS III

ACRONYMS IX

I. INTRODUCTION 1

A. Key research questions and guiding principles for the evaluation design 1

B. Overview of project activities, program logic, and beneficiaries 3

1. The Mainland Trunk Road and Zanzibar Rural Roads Activities 5
2. The Roads Maintenance Activity 7
3. The Mafia Island Airport Activity 7

C. Literature review 7

1. Impacts on direct beneficiaries of improved roads 8
2. Broader economic effects of infrastructure investments 11
3. Contribution of this evaluation to the literature on roads investments 12

D. Design report overview 13

II. ECONOMIC ASSESSMENT 15

A. Motivation and research questions 15

B. Data sources 17

C. Sampling approach for new data collection 18

D. Analytical approach 20

1. Calibrating HDM-4 models 20
2. HDM-4 Analysis and sensitivity testing 22
3. Road condition data reporting and visualization 23

E. Risks and mitigation strategies 23

III. PERFORMANCE EVALUATION 25

A. Motivation and research questions 25

B. Roads maintenance 25

1. Data sources and outcomes 28
2. Analytical approach 28

C. Road users 30

1. Data sources and outcomes 32
2. Sampling for the expanded vehicle intercept survey 33
3. Analytical approach 34

D. Transportation cost savings 35
1. Data sources and outcomes 36
2. Analytical approach 36
E. Mafia Airport activity 37
1. Data sources and outcomes 37
2. Analytical approach 38
F. Risks and mitigation strategies 38
IV. IMPACT EVALUATION 41
A. Motivation and research questions 41
B. Background and overview of evaluation 41
C. Data Sources 43
1. Nighttime lights imagery 43
2. Trunk Road Network Information 44
3. Daytime imagery 44
4. Data for developing and testing the poverty measures 44
D. Analytical approach 46
1. Mainland trunk roads 46
2. Pemba rural roads 47
3. Steps involved in carrying out the evaluation 48
E. Simulation-based power calculations 48
1. Simulation specification and data sources 49
2. Results 50
F. Optional validation of impact estimates using survey data 51
G. Risks and mitigation strategies 53
V. DATA COLLECTION 55
A. Primary quantitative data 55
1. Road condition data 56
2. Road user data 57
3. Process for collecting primary data 60
B. Secondary data 61
1. Types of secondary data 61
2. Process for collecting secondary data 62
C. Primary qualitative data 62

D. Data quality plan 64

VI. EVALUATION ADMINISTRATION AND MANAGEMENT 67

A. Institutional review board 67

B. Data access, privacy, and documentation 67

C. Dissemination 68

D. Evaluation team 68

E. Evaluation timeline and reporting schedule 68

REFERENCES 71

APPENDIX A: SUMMARY OF STUDIES IN THE LITERATURE REVIEW 77

APPENDIX B: LIST OF VARIABLES FOR HDM-4 ANALYSIS 83

APPENDIX C: SAMPLE SIZE ESTIMATION METHODOLOGY FOR THE EXPANDED VEHICLE INTERCEPT SURVEY 89

APPENDIX D: DATA COLLECTION COMPARISON TABLE 95

APPENDIX E: QUANTITATIVE DATA COLLECTION STANDARDS TABLE 101

APPENDIX F: SUMMARY OF HDM-4 SENSITIVITY TESTING 107

APPENDIX G: RESPONSES TO MCC COMMENTS ON THE DRAFT EVALUATION DESIGN REPORT 119

TABLES

II.1. Overview of research questions, key outcomes, and analytical approach 16

II.2. Data sources for the economic assessment, by category 18

III.1. Roads maintenance research questions, data sources, and outcomes 27

III.2. Traffic and road users research questions, data sources, and outcomes 31

III.3. Illustrative O-D matrix 34

III.4. Transportation cost savings research questions, data sources, and outcomes 36

IV.1. Key data sources 46

IV.1. Quantitative data sample units, sample size, and instruments 56

IV.2. Traffic count stations per road segment 59

IV.3. Secondary data requests 62

IV.4. Qualitative data collection focus areas 63

V.1. Projected timeline for the evaluation 69

A.1. Studies examining impacts on beneficiaries close to improved roads 79

A.2. Studies examining broader economic effects 81

B.1. HDM-4 variables and data sources 85

C.1. Proportion of vehicles by O-D and purpose of travel 92

C.2. Estimated sampling error by trunk road and outcome 93

D.1. Road Data Collection Comparison 97

D.2. Airport Runway Data Collection Comparison 100

E.1. Required data standards 103

F.1. Ex-ante ERR estimates and key model assumptions extracted from Due Dilligence report 111

F.2. Ex-ante ERRs, initial traffic and key outcomes 111

F.3. Sensitivity of economic outcomes to different SN for Tanga – Horohoro 113

F.4. Sensitivity of economic outcomes to different SN for Songea – Namtumbo road 114

F.5. Sensitivity of economic outcomes to different SN for Tanga – Horohoro 115

G.1. Evaluator comments on design report 121

FIGURES

I.1. Program logic for the Transport Sector Project 4

I.2. Locations of MCC Transport Project activities 6

IV.1 Example simulated consumption levels 51

F.1. Location of trunk roads in Mainland Tanzania 110

This page has been left blank for double-sided copying.

ACRONYMS

AADT	Annual average daily traffic
ANN	Artificial neural network
CAPI	Computer-assisted personal interviewing
CBR	California bearing ratio
CNN	Convolutional neural network
CRM	Commercialized Road Management
DID	Difference-in-differences
DMSP	Defense Meteorological Satellite Program
DN	Digital number
DNB	Day/Night Band
EDI	Economic Development Initiatives
ERR	Economic rate of return
ESAL	Equivalent standard axle load
FWD	Falling weight deflectometer
GoT	Government of Tanzania
HDM-3	Highway Design and Maintenance Standards Model 3
HDM-4	Highway Development and Management 4
IRB	Institutional review board
IRI	International Roughness Index
KII	Key informant interviews
LTPP	Long-term Pavement Performance
M&E	Monitoring and evaluation
MCA-T	Millennium Challenge Account-Tanzania
MCC	Millennium Challenge Corporation
MoIC	Ministry of Infrastructure and Communications
NTL	Nighttime lights
NOAA	National Oceanic and Atmospheric Administration
NPP	Suomi National Polar-Orbiting Partnership
O-D	Origin-destination
OLS	Operational Linescan Sensor
RDWE	Road deterioration and works effects
RED	Roads Economic Decision
RMMS	Road Maintenance and Management System
RTCS	Road traffic count survey
RUE	Road user effects

SUMATRA	Surface and Marine Transport Regulatory Authority
TAA	Tanzania Airport Authority
TANROADS	Tanzania National Roads Agency
ToRs	Terms of reference
TRL	Transport Research Laboratory
VIIRS	Visible Infrared Imaging Radiometer Suite
VOCs	Vehicle operating costs

I. INTRODUCTION

Nearly 50 percent of Tanzanians live on less than $2 per day (World Bank 2016). The vast majority of the poor in Tanzania live in rural areas, relying on agriculture for their livelihood. Although the country has made great strides in terms of economic growth and poverty reduction, the agriculture sector, accounting for over 70 percent of employment, has seen slower growth (World Bank 2016; GoT 2016). Rural populations living below the poverty line in Tanzania are not connected to the growth process, and the lack of an adequate transportation network has been a key constraint to economic growth (MCA-T 2013). Improved roads and infrastructure are critical tools for reducing poverty and promoting economic growth because they connect populations to markets and promote investment opportunities (Foster and Briceño-Garmendia 2010). At the time of the compact, the Millennium Challenge Corporation (MCC) estimated that less than 7 percent of Tanzania's road network was paved.

To address these issues, in 2008, MCC and the government of Tanzania (GoT) signed a five-year compact designed to promote economic growth and reduce poverty by investing $698 million in three key sectors: transport, energy, and water. The GoT set up MCA-Tanzania (MCA-T) to oversee the implementation of all projects funded through the compact. MCC disbursed more than $405 million for the transport sector project, which aimed to improve trunk and rural roads in order to reduce transport costs, increase cash crop revenue for farmers, and facilitate access to social services through improvements in the country's road network. The project also upgraded the Mafia Island Airport runway to allow for easier, more efficient, and safer access to the island to facilitate travel and increase tourism. The estimated economic rate of return (ERR) for the transport sector project over 20 years was 17.2 percent (MCC n.d.).

MCC has contracted Mathematica to conduct an evaluation of the Transport Sector Project that will estimate ERRs for the completed roads projects. The evaluation will also assess whether maintenance practices are likely to sustain the roads investments, whether road use conforms to assumptions in the program logic, and the likelihood of transportation cost savings being passed on to consumers. We will also assess any benefits from upgrades to the Mafia Island Airport runway. This design report draws on our evaluability assessment as well as our initial design trip and discusses our proposed approach to the evaluation.

In the next section, we summarize the high-level evaluation questions and guiding principles that we used to develop the evaluation design. We follow this with an overview of the program logic for the Transport Sector Project and a brief literature review to place the Tanzania investments in the context of the evidence to date on the impact of roads improvement projects, maintenance, and transportation costs. We end this chapter with a roadmap to the rest of the design report.

A. Key research questions and guiding principles for the evaluation design

Improved roads benefit economic growth and poverty reduction by lowering vehicle operating costs (VOCs) and travel times for people using the roads resulting in lower transportation costs (MCC forthcoming). These reduced costs benefit the economy in a variety of ways; for example they can promote access to health and social services, trade linkages, or tourism (Bryceson et al. 2008; Escobal and Ponce 2004). Reduced transport costs can also lower costs of inputs for both agricultural and non-agricultural production, benefiting productivity and increasing household income and consumption (Khandker et al. 2009).

Because of these potentially widespread benefits, improvement of transportation infrastructure has been a high priority for many developing countries, and many countries eligible for MCC compacts have requested infrastructure investments. Over the past 10 years, MCC has made significant investments in roads projects, totaling nearly $3 billion via 18 compacts (MCC forthcoming). Along with investing in roads, MCC has invested in evaluations to measure the impacts of the roads projects. However, many of the early evaluations of such projects were not able to measure interim or longer-term outcomes on prices and incomes. Peer reviewers for these evaluations have emphasized that impact evaluation designs are not always based on a solid understanding of the theory of change for specific roads investments. As a leading donor in the sector, MCC continues to learn lessons from roads investments and to develop reliable ways to evaluate them, looking more closely at the intermediate objectives of the roads improvements according to the program logic.

This design report drew upon the Transport Sector program logic and MCC's learning objectives for this and future roads evaluations to develop the evaluation approach. The evaluation will reestimate the economic rate of return for the roads and assess the performance of maintenance practices, road users, transportation costs, and the Mafia Island Airport runway through the following main research questions:

1. What is the estimated economic rate of return for the completed roads projects?
2. How have maintenance practices affected the projected life of the roads?
3. Who is traveling on the roads and why? Have travel times decreased?
4. Are consumers benefiting from reduced transportation costs?
5. How accurate were the assumptions in the ERR for the Mafia Island Airport?

The evaluation will address these questions through a mixed-methods approach that analyzes roads engineering and road user data, administrative records, and qualitative data. In addition to the main research questions posed here, Mathematica and MCC are developing an impact evaluation design to measure the impact of the roads investments on economic activity, using new data technologies including a combination of satellite and field-level/household data collection.

In developing the design report and impact evaluation design, the Mathematica team has applied the following guiding principles to the evaluation data collection and analysis:

- **Clearly identify the theory of change for the roads investment in order to develop the analytical approach for the evaluation.** Grounding the evaluation in the theory of change will help identify intermediate outcomes that can help predict the longer-term impacts of the roads. As noted in the evaluation assessment memo, the expected long-term impacts were not directly tied to a specific theory of change for trunk roads and rural roads. For example, the initial program logic for the Transport Project in Tanzania tied roads investments to increases in agricultural production of cash crops and price increases for those crops without a complete assessment of the linkages between improved transportation networks and prices. In addition, the pre-compact ERRs may have overestimated projected impacts because they estimated access to social services and growth in agriculture separately from the standard variables in the Highway Development and Management 4 (HDM-4) model.

- **Focus data collection activities on the variables that have the highest impact on ERR estimates, and draw lessons for future investments in roads projects.** As part of the evaluation design, we are trying to identify areas where data collection activities could be streamlined, while prioritizing those variables that are key to generating robust ERR estimates (which will rely on HDM-4 outputs); for example, sampling road segments for some measures of road conditions, which may not be needed for every road segment or at the frequency outlined in the data collection specifications. Similarly, we will rely on existing secondary data sources for HDM-4 parameters where data quality can be verified or for parameters to which HDM-4 is not very sensitive.

Develop lessons from the Tanzania (and Ghana) roads evaluations to make future evaluations of infrastructure projects more efficient and useful. In addition to the approaches above, the Ghana and Tanzania evaluation will aim to develop an approach for lower-cost evaluation methods to assess the impact of roads projects, such as by using satellite imaging data. In addition, we will attempt to keep track of the various lessons learned from the design of these evaluations to inform MCC’s future roads evaluation efforts.

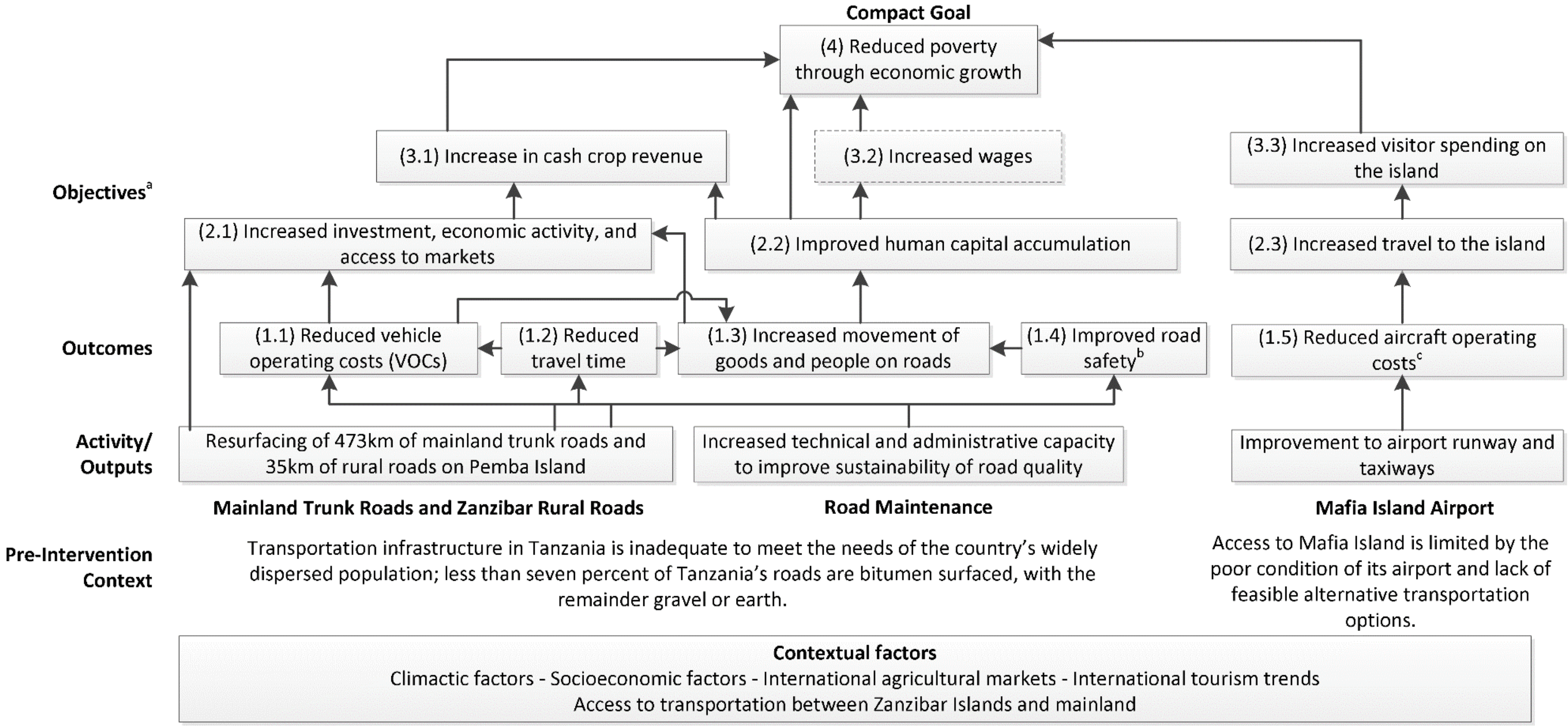
B. Overview of project activities, program logic, and beneficiaries

The $405 million Transport Sector Project, which comprised more than half of total compact investments, aimed to increase cash crop[1] revenue and aggregate tourism spending through four activities. Using the compact’s monitoring and evaluation (M&E) plan and other documents, we created a program logic model, depicted in Figure I.1. Although the program logic evolved over time, this logic model illustrates the pre-compact understanding of the causal path between program inputs, outputs, outcomes, and the ultimate goal of poverty reduction through economic growth (MCA-T 2013). Figure I.1 also describes the implicit assumptions in the program logic.

The project activities, as illustrated in the logic model, are briefly summarized below.

[1] Cash crops identified in compact documents include cashew nuts, groundnuts (peanuts), sunflowers, coffee beans, tobacco, coconuts, and wheat.

Figure I.1. Program logic for the Transport Sector Project



Notes: A number of key assumptions underlie the links between the project outputs, outcomes, and overall objectives. These include (1) upgrading the roads will reduce travel times, and (2) will reduce the cost of operating and maintaining vehicles used along improved segments; (3) reduced VOCs and travel time will lower transportation costs, and these savings will pass on to consumers through competitive transport and agriculture markets; (4) road improvements will lead to increases in cash crop revenue without additional investments in agricultural practices or capacity; and (5) the mainland and Zanzibar roads authorities have sufficient capacity to adequately maintain the roads

[a] Increased wages were not explicitly included in the compact's M&E plan, but are included in ERR benefits streams and are an implicit objective necessary to link the lower-level objectives of aggregate tourist spending and improved human capital accumulation with the compact's goal. Objectives 2.1 through 2.3 were not explicitly included as objectives in the compact's M&E plan until 2012, although they were included in initial ERR benefit streams.

[b] Improved road safety was added as an output of the project mid-compact.

[c] The program logic for the Mafia Island Airport reflects the benefit streams from the compact closeout ERR estimates. See the discussion below.

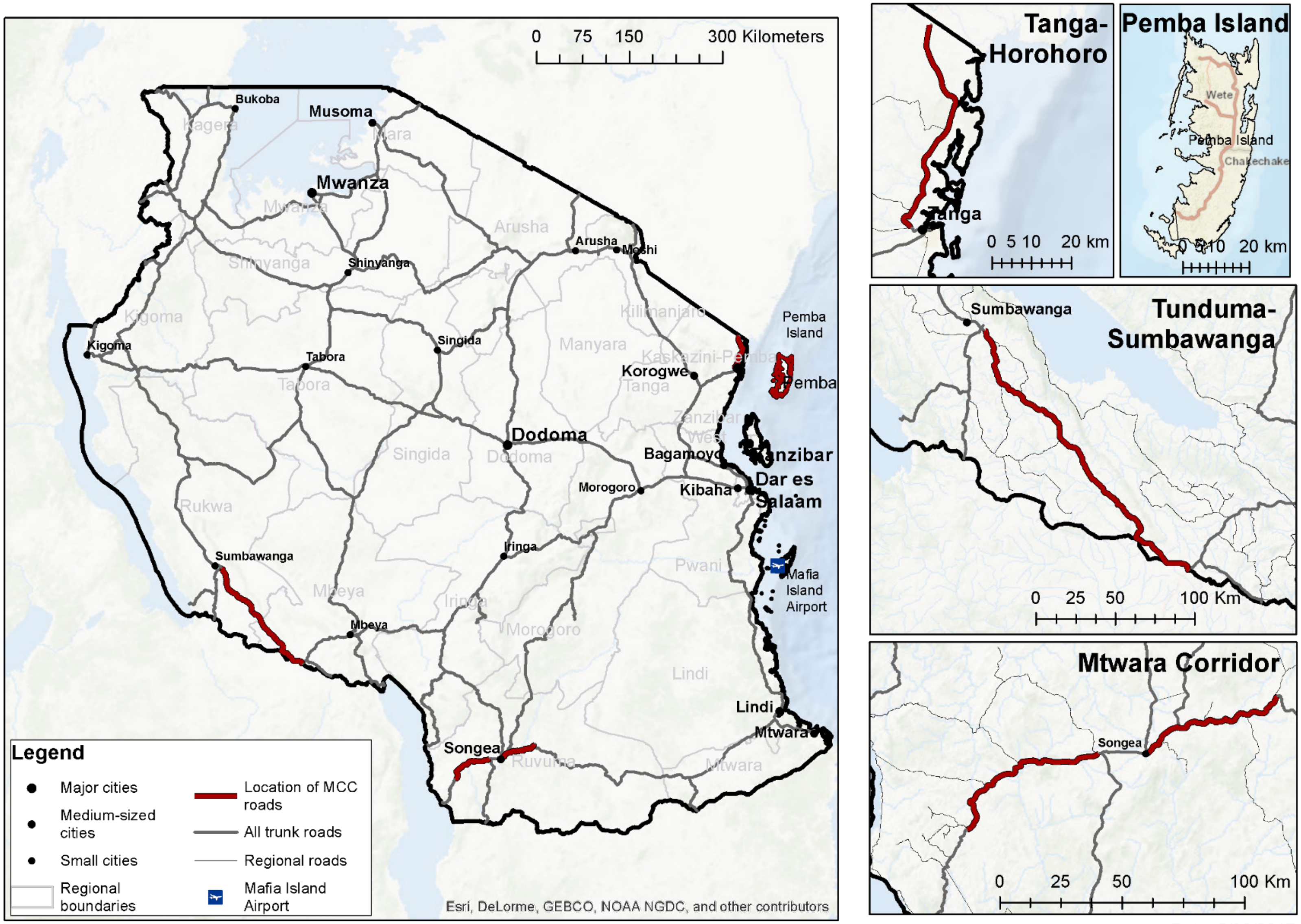
1. The Mainland Trunk Road and Zanzibar Rural Roads Activities

The trunk roads activity, implemented by the Tanzania National Roads Agency (TANROADS) and overseen by MCA-T, aimed to increase cash crop revenue on the Tanzanian mainland by upgrading 473 km of trunk roads that connect different regions within Tanzania and connect Tanzania with other countries. These roads segments include 224 km between Tundumba and Sumbawanga in the west, 68 km between Tanga and Horohoro in the east, and 139 km between Mbinga and Nambtumbo in the south. The Tanga-Horohoro road was completed in 2012, the Tunduma-Sumbawanga road was completed in 2013, and the two segments of the Mtwara Corridor road were completed in 2013 and 2014 (MCA-T 2015). All roads were upgraded from gravel to bituminous (asphalt) surfaces. Figure 1.2, below, depicts the location of all MCC-funded trunk roads and provides detailed views of the trunk roads and their placement within the network of regional roads; we have also mapped the location of Pemba Island and the Mafia Island Airport for reference.

The Zanzibar Rural Roads Activity, implemented by the Zanzibar Ministry of Infrastructure and Communications (MoIC) and overseen by MCA-T, aimed to increase revenue and reduce travel time and VOCs by upgrading 35 km of rural feeder roads on Pemba Island from dirt to pavement. Most of the roads targeted for upgrades serve as terminal nodes that connect isolated villages to the main road network. Prior to being upgraded, many of the road sections were impassable by motor vehicles during the wet season.

The program logic expected that roads improvements would reduce travel time on the roads and reduce the cost of operating and maintaining vehicles used along improved segments. Over the longer term, the savings from the reduced VOCs and travel time would be passed on to farmers and consumers through competitive transport and agriculture markets, and farmers would increase their cash crop revenue. In addition, the program logic assumed that improved roads would increase access to social services such as schooling and health care for those living close to the roads, which would in turn increase human capital and, eventually, wages. The expected beneficiaries were road users (including transport operators and vehicle passengers) as well as farmers, who would receive increased cash crop revenue, and households living near the roads. The initial program logic did not distinguish between beneficiaries of trunk roads—who include inhabitants living at the ends of improved road segments and beyond, largely in urban areas—and those living in rural areas along the roads. These trunk roads can also reduce barriers to interregional trade and facilitate cross-border trade, whereas beneficiaries of rural roads are more likely to be households living in the vicinity of the upgraded roads who now have better access to the transportation network.

Figure I.2. Locations of MCC Transport Project activities



Sources: TANROADS; Esri, The General Bathymetric Chart of the Oceans, the National Oceanic and Atmospheric Administration of the U.S. Department of Commerce, and other contributors. The regional boundaries within Tanzania are from 2011. Many of these boundaries have since moved.

Notes: Detailed maps of the MCC-funded roads on Pemba Island were not available.

2. The Roads Maintenance Activity

Under the Compact, MCC also made investments in the road maintenance capacity on the mainland and in Zanzibar. MCC funded technical assistance for conducting a road inventory and road condition survey, and funded TANROADS to purchase equipment to measure road conditions. TANROADS staff then trained MoIC on the establishment of a Road Maintenance and Management System (RMMS) in Zanzibar. These maintenance investments were designed to complement both direct roads investments by MCC and other donors and previous institutional reforms by the GoT. The intention was that, as a result of these activities, TANROADS and MoIC would be able to better maintain the road networks (including, but not limited to the roads upgraded by MCC). The initial program logic assumed that the improved capacity in the two agencies would be sufficient for maintaining the improved road conditions after completion of the roads upgrades.

3. The Mafia Island Airport Activity

Under the Mafia Island Airport Activity, MCC funded the paving of the runway and taxiways of the airport located on Mafia Island in Zanzibar; the Tanzania Airports Authority conducted implementation and oversight for the activity. The program logic expected that these upgrades would increase travel to the island, thereby increasing tourism and tourist and leading to higher wages for workers on the island. The program logic for the activity evolved over time. Initially, the ERR estimates predicted that the runway would accommodate larger planes and higher traffic volumes; however, revised ERRs adjusted the beneficiary assumptions to include mainly airplane operators, who benefit from reduced maintenance, operating, and insurance costs, and increased visitor spending as a result of increased travel to the island.

C. Literature review

In this section, we review the literature on impacts of roads improvement programs and similar transportation infrastructure. As described above, roads improvements benefit economic growth and poverty reduction primarily through reduced transportation costs, which can lead to other intermediate and longer-term benefits, such as improved local economic conditions and greater interregional trade and cross-border trade.

It is especially challenging to measure the impact of roads projects on income and poverty reduction when improving the movement of goods and labor affects a diverse group of people including people traveling on the roads, populations living near the road, or consumers of products transported on the road. Measuring the specific ways that roads improvements affect a broad group of beneficiaries is especially challenging because the mechanisms are complex and harder to track. For instance, institutional factors, such as the competitiveness of the transport and agriculture markets in a country, can play a considerable role in determining who benefits and to what extent. In addition, because roads improvements have small impacts on a very broad set of beneficiaries, measuring these impacts is especially challenging.

In the next section, we review the literature on the impacts of roads improvements for two groups: (1) populations located close to improved roads, and (2) other beneficiaries who use the roads or benefit from goods traveling on the roads but may not live near them.

1. Impacts on direct beneficiaries of improved roads

We organize the literature review starting with impacts of roads improvements on intermediate outcomes such as reduced travel times and greater access to markets and social services, followed by longer-term impacts on economic activity and income. The literature review draws on the findings of completed MCC roads evaluations, as well as other roads evaluations that used impact evaluation methods to estimate a counterfactual.

a. Travel time

MCC's roads evaluations support the conclusion that the most immediate benefits of roads improvements come from reductions in VOCs and travel times. For example, the upgrading of secondary and rural roads as part of MCC's compact in Honduras slightly reduced travel time to health centers and to municipal capitals; it also reduced travel costs to hospitals, health centers, and basic markets (NORC 2013a). In Georgia, rehabilitation of a regional highway that connects the economically isolated Samtskhe-Javakheti region to the capital and the rest of the country, increased traffic and average speed of vehicles leading to decreased travel times to the capital and basic markets (NORC 2013b).2 Similarly, the upgrading of rural roads in Armenia led to a large, over 30 percent decrease, in travel time by households to sell agricultural products, and increased use of roads by households for shopping or visiting relatives (Fortson et al. 2015). These studies employed a difference-in-differences (DID) approach to estimate impacts on travel times and other outcomes compared to a comparison group. Other studies examining the impacts of road investments also found evidence of decreased VOCs and travel times. For example, a study that used a DID approach to examining the construction of concrete bridges and culverts for roads crossing rivers and streams, thus allowing year-round passage in Brazil, found significantly reduced travel time to population centers and increased use of public transportation (Iimi et al. 2015). A panel data study examining similar interventions in Bangladesh also decreased travel costs by more than a third (Khandker et al. 2009).

b. Access to services, markets, and use of social services

Although roads improvements can increase access to markets and social services such as schools and health centers, the empirical evidence for this is mixed, likely depending on the availability of these services in the local context. The Armenia roads evaluation found that households near rehabilitated roads reported improved market access after completion of the roads. However, the roads improvements did not have an impact on access to social services, perhaps because community members did not use the regional roads to travel to social services located nearby (Fortson et al. 2015).

In the Brazil study, researchers found mixed results on perceived access to social services such as health and schooling. The study reported improved access in two regions, but no consistent increase in two other regions (Iimi et al. 2015). These results may be related to travel distances to the services: the average distance to the nearest hospital was 27 km in the two regions where perceived access increased, compared to 39 km in the two regions where

[2] The Samtskhe-Javakheti road reduced self-reported travel times to the capital by 40 minutes and to the local markets by 44 minutes. We could not calculate the percentage changes for these reductions because the authors did not report average travel time to these destinations at baseline.

perceived access did not. Similarly, distance to schools in communities that reported improved access was half of the distance to schools that reported no improvement, suggesting that households in the regions with closer access were able to take advantage of the roads improvements.

Studies looking at the use of social services also found mixed results. In particular, studies with shorter follow-up windows were less likely to see any changes in utilization. For example, the MCC-funded Honduras study did not detect increased utilization of social services in terms of number of visits to hospitals, health care facilities, or pharmacies by household members in Honduras (NORC 2013a). MCC's evaluation of the Georgia roads project also did not find impacts on the use of health care centers (NORC 2013b) within one to two years of exposure to the improved roads.

Impacts of roads on schooling follow a similar mixed pattern. Road rehabilitation in Brazil increased school attendance of girls in two regions, but attendance decreased in two other regions after one year of exposure (Iimi et al. 2015).3 MCC's Honduras project also did not find increases in the number of children attending school (NORC 2013a). The lack of measurable impact could have resulted from an evaluation period that did not allow enough time to observe changes in household behavior caused by the improved access to services. The evaluation design and program design also may not have adequately considered how the roads would promote the use of social services such as schools or health clinics (MCC forthcoming).

Studies that have longer exposure windows, however, have detected improvements in school outcomes, particularly enrollment. An evaluation of a World Bank-financed rural road rehabilitation project in Vietnam using a DID approach found increases in both primary and secondary school enrollment seven years after the completion of the project (Mu and van de Walle 2011). School enrollment for both boys and girls between ages 5 and 17 also increased in Bangladesh approximately four years after the completion of the project (Khandker et al. 2009).

c. Economic activity in rural areas

Improvements in roads can connect isolated rural areas to larger roads networks and provide households an opportunity to shift away from agricultural production to the non-agricultural wage sector as a result of increased access to regional and national labor markets. MCC's evaluation in Georgia showed an increase in the number of industrial facilities located in the intervention communities near the improved roads (NORC 2013b). Another study of roads projects in Georgia, using a DID approach, found that improvement of rural roads led to increases in the number of non-agricultural small and medium enterprises, and reductions in barter trade (Lokshin and Yemtsov 2005).

[3] Girls' attendance in these two regions decreased for both the intervention and the comparison groups, but the reduction was larger in the intervention group. The authors noted that the decline in attendance is likely related to the fact that the total number of children ages 5 to 9 years has been declining in the state of Tocantins, where the study was conducted. However, that does not explain why the decrease in attendance was significantly larger among the intervention group.

The Vietnam study discussed earlier also found small increases in off-farm activities, mostly in the service sector, as a result of roads improvements (Mu and van de Walle 2011). A study, using a DID design, examined the effects of a newly built road that connected formerly inaccessible rural areas to more urban areas in the Republic of Palau and found increases in non-agricultural wage-sector employment, reduced self-employment in agriculture, and decreases in the number of immigrants sent abroad (Akee 2006). Finally, the roads improvement project in Bangladesh cited above also found increases in total labor supply of both males and females in terms of monthly employment hours (Khandker et al. 2009).4

In contrast, the roads improvement project in Brazil did not change the composition of labor across sectors (Iimi et al. 2015). Agriculture was the dominant sector in all four regions, which could be one reason for not observing significant changes in labor composition. MCC's projects in Honduras also did not find any impacts on employment (NORC 2013a).

Increased market access through roads improvements should, in theory, affect agricultural productivity but here again, the evidence is mixed. Although the highway improvements in the MCC evaluation in Georgia project found an increase in industrial facilities along the improved road segments, there was no change in land use or cropping patterns (NORC 2013b). The study did not allow for measure impacts on agricultural productivity in areas farther from the improved road segments. MCC's Armenia project also did not find impacts on agricultural production during the short exposure period after completion of the road (Fortson et al. 2015). On the other hand, the roads project in Bangladesh, complemented by interventions to improve secondary markets, lowered fertilizer prices by 5 percent, increased aggregate crop prices by 4 to 5 percent, and increased agricultural output by 30 to 39 percent (Khandker et al. 2009).

d. Income and consumption

Many of the studies in this literature review measured impacts on income and poverty with results varying depending on the context. For example, some studies of tertiary roads have shown that these projects improve outcomes related to income and wealth. For example, in Mexico, the upgrading of feeder roads connecting peri-urban neighborhoods to the larger roads network increased property values by 17 percent (Gonzalez-Navarro and Quintana-Domeque 2015). The roads paving program in Mexico led to increased ownership of vehicles and home appliances and increased home improvements (Gonzalez-Navarro and Quintana-Domeque 2015). Ownership of cars also increased in the Republic of Palau and Brazil (Akee 2006; Iimi et al. 2015). It is possible that these increases represent a shift in consumption, as motorized vehicles are more useful with improved road conditions.

In Peru, rural roads rehabilitation in districts with high poverty rates led to a 35 percent increase in per-capita income for households near the rehabilitated rural roads (Escobal and Ponce 2004). Roads improvement in Bangladesh led to increased household expenditure (Khandker et al. 2009), and access to all-weather roads increased consumption growth in Ethiopia (Dercon et al. 2006. A study that examined density of roads and consumption found that higher density of roads was positively correlated with higher rates of consumption (Jalan and Ravallion 2002). These studies used modeling approaches to estimating effects, as opposed to

[4] The study did not specify whether there were changes in the composition of labor hours across different sectors.

using a counterfactual approach. (The Khandker et al. [2009] and Dercon et al. [2006 studies used household fixed-effect regression models, whereas the Jalan and Ravallion [2002] study used a regression model controlling for initial conditions.

Other studies report no significant income effects in the one to two-year post completion evaluation timeline (Akee 2006; NORC 2013b; Fortson et al. 2015; Iimi et al. 2015). MCC's Nicaragua project did not find impacts on the cost and availability of a basket of basic consumer goods at the establishments selling these products, but found increases in household consumption of perishable (dairy) and fragile (eggs) goods, indicating improved access to these products from improved roads conditions (Alevy 2014).

2. Broader economic effects of infrastructure investments

Measuring the effects of improving trunk roads and highways is challenging because the impacts are more diffuse, making the identification of credible counterfactuals difficult. At the same time, recent methodological advancements are making headway in addressing the challenge. In a seminal paper exploring the impacts of building India's railroad network in the 19th and early 20th century, Donaldson (2017) analyzes a new data set from archival sources to show that the rail network reduced trade costs and interregional price gaps, increased interregional and international trade, and improved income levels. In another study, using a combination of U.S. county-level data and Geographic Information Systems (GIS) data, Donaldson and Hornbeck (2015) also found that railroad access in the U.S. had a positive impact on gross national product.

Other studies confirm that transportation costs in Africa are an important constraint to economic growth. Atkin and Donaldson (2015) investigated the cost of distance in two African countries—Ethiopia and Nigeria, where transportation costs are high because of sparse and poorly maintained road networks compared to countries with more developed networks. They find that the costs of intra-national trade are approximately four to five times larger in these two countries than in the U.S. Similarly, Storeygard (2016) studied 15 countries in sub-Saharan Africa whose largest city is a port, and found that transportation costs have important impacts on the income of cities located near these ports. With night time lights satellite data, the study uses increases in oil prices from 2002 to 2008 as a proxy for per-kilometer transport costs, and finds that cities located close to a port benefited from lower transportation costs during this period, experiencing a 7 percent increase in income compared to similar cities located farther away.

Several studies examining the impacts of the entire roads network report shifts in production and labor from the agricultural sector to the non-agricultural wage sector. For instance, Banerjee and colleagues (2012) found that although new and improved roads in China improved agricultural productivity and raised income and consumption, they also took capital and skilled labor away from rural regions through migration. A study of the effect of a national rural roads construction program in India, which connected previously unconnected villages to the regional transportation network, decreased the share of households and workers in agriculture and increased the share participating in wage labor (Asher and Novosad 2016). Gollin and Rogerson (2014) also report similar findings in a multisector multiregion model, where higher transport cost contributed to a larger agricultural workforce and a larger fraction in subsistence.

Finally, the success of roads infrastructure projects is likely to be dependent on the country contexts in which the investments are made. Inefficiencies in local and regional markets can reduce or eliminate the potential longer-run effects of infrastructure investments for the intended beneficiaries in remote locations. For example, Casaburi and Reed (2013) investigated the impacts of a rural roads rehabilitation program in Sierra Leone and found that it lowered the price of rice and cassava. The impacts were stronger in areas farther away from large urban centers and with lower productivity, likely because improved road quality decreases the cost of reaching local markets where farmers and traders conduct business.[5] However, the price effects were lower in areas with better cell phone coverage, because reductions in travel times likely were less relevant in places where farmers and traders could exchange information using the cell phone network. Atkin and Donaldson (2015) investigated the extent to which the benefits of import price reductions were passed on to consumers in two African countries—Ethiopia and Nigeria. They found that intra-national trade costs rise substantially with distance in these two countries, suggesting that consumers in remote locations did not benefit from reductions in international prices. Because of higher transport costs, it is less profitable to sell products in remote locations, which results in less competition in the local markets. As a result, intermediaries are able to capture a greater fraction of the benefits of price reductions, leaving little to pass on to communities in remote locations.

Our literature review highlights the complexity of evaluating investments in transportation projects and the need for having a clear, context-specific theory of change. Not surprisingly, studies have been more likely to find effects on the proximal outcomes, such as increased traffic, reduced travel times, and reduced VOCs. However, impacts on other intermediate and longer-term outcomes depend on the context in which these investments are made. For example, improvements in access to health and education depend on the availability of related infrastructure. The timing of the evaluation also matters, particularly for longer-term outcomes such as income and consumption, when it can take several years after a project is completed for impacts to materialize; shorter-term studies have failed to detect impacts on these outcomes. Further, tying the nature of roads improvements to the relevant beneficiary population is very important. Beneficiaries, particularly of highway or trunk roads, are likely to be dispersed across the entire transportation network and not just along the improved road segments; it is important for the evaluation designs to capture benefits where the beneficiaries are located. Studies that examined upgrading of regional highways but measured impacts only for households along the improved road segments were unlikely to find improvements in economic activity and income. Finally, improvements in roads may fail to benefit the intended beneficiaries if there are market inefficiencies, in which case complementary context-specific investments or policy changes could improve the effectiveness of roads investments.

3. Contribution of this evaluation to the literature on roads investments

This evaluation of Tanzania roads activities will contribute to the literature on the impact of infrastructure by examining factors that affect the intermediate outcomes for roads investments. The traditional impact evaluation approach of identifying treatment and comparison groups has

---

[5] Areas with higher productivity have stronger demand from traders because traders can buy more in these markets for given transportation cost. A reduction in transportation costs increases demand from traders more in the lower productivity areas.

proved to be problematic and appears not to capture the full impacts of roads, in particular trunk roads and highways. Given the potential problems with these approaches that focus on a narrow segment of the beneficiary population, developing alternative ways to assess the performance of roads projects could guide governments and funders in determining roads investments.

Although studies of the impacts of roads projects have not used HDM-4 or ERR assessments, ERRs are commonly used to determine which roads projects to fund, because they allow decision makers to compare road segments to one another in terms of anticipated benefits and to conduct sensitivity testing to look at changes in outcomes under different assumptions. HDM-4 is a consumer surplus model, capturing a much fuller set of impacts of the roads than the quasi-experimental approaches but without disaggregating benefits by type and beneficiary. Refining the application of HDM-4 for ex-post performance assessment, as we discuss in more detail in Chapter II, could benefit future decision making on roads investments. The evaluation will also take a closer look at maintenance practices and could help MCC consider how to incorporate assessments of maintenance practices into due diligence process and pre-implementation ERR estimates. Finally, the impact evaluation design will attempt to develop a broader set of outcome indicators using satellite data that could add to methodological tools for measuring the impacts of specific infrastructure projects.

D. Design report overview

The remaining chapters of the evaluation design report describe our approach to addressing the key research questions described earlier in this chapter. In Chapter II, we describe our approach to the HDM-4 analysis and its use in calculating the ERRs for the completed project. We describe our plans for conducting a performance assessment of roads maintenance, road users, and transportation costs in Chapter III. This chapter also covers our proposed design for assessing the likely economic impact of the Mafia Island Airport runway. Chapter IV describes the impact evaluation design using satellite data and other data sources. Chapter V describes our data collection approach. Chapter VI describes evaluation administration and presents the updated work plan.

This page has been left blank for double-sided copying.

II. ECONOMIC ASSESSMENT

A. Motivation and research questions

The economic assessment of the Tanzania Transport Project reestimates the ERRs for the Mainland Trunk Roads and Zanzibar Roads Activities and compares them to the pre-implementation ERRs. This will involve reestimating the direct benefit streams from the roads improvements and assessing the main assumptions underlying the model used to estimate these benefits. As discussed above, the pre-implementation ERRs for the roads estimate both direct and indirect benefits. Direct benefits include reduced travel time for road users and reduced costs for vehicle operators; these benefits are scaled as traffic flows increase along improved roads. The indirect benefits come from growth in agriculture and better access to social services through improved market integration. Our recalculation of the ERRs will be based on collecting new data on road users and current road conditions in order to estimate the direct benefits of road improvements. We will not reestimate the indirect benefits of increases in agricultural production and improved access to social services, as we believe that accounting for traffic growth captures these benefits already (see discussion below). The recalculation will also use data on actual construction costs and maintenance expenses and will allow us to produce a side-by-side comparison of predicted and realized direct benefit streams and costs.

The economic assessment will use the HDM-4 model as its framework, which is the same model used in the pre-implementation ERRs.[6] HDM-4 is a program that analyzes the economic viability of road transportation projects over their lifetime by simulating pavement deterioration, road user costs, and traffic over time and trading this off against the costs of construction and periodic and routine maintenance. At its core, HDM-4 models the interaction between road users and road deterioration. Improvements or upgrades to a road allow for smoother driving, reducing damage to vehicles and allowing for faster travel. This, in turn, increases traffic along improved roads, as traffic may be diverted from slower, poorer roads, and as economic growth generates new traffic. However, traffic and general weathering lead to deterioration in the road over time, which again impacts road users. The model also estimates the effects of regular maintenance and upgrading on road condition. Using all these data, HDM-4 calculates the economic benefit of a road by multiplying the reduction in VOCs and travel time by the traffic volume along the road. These benefits are compared to the costs of construction and any maintenance programs to arrive at an ERR for specific roads or road segments.

The HDM-4 model is very adaptable, in that its parameters can be customized to reflect the local context. This process of tailoring the model to local conditions is referred to as "calibration" and is a critical first step in using the model to estimate economic impacts. Road conditions deteriorate differently around the world, depending in part on traffic loads, traffic volumes, climate, building standards, and construction materials. Likewise, important determinants of VOCs, such as the cost of new and used vehicles, the composition of a country's

[6] HDM-4 is a software program used to model the engineering and economic viability of roads projects. It is based on the Highway Design and Maintenance Standards Model (HDM-3), which was originally developed by the World Bank. The current version of the model is HDM-4 version 2, which was released in August 2005. HDM-4 is considered the industry standard for project appraisal and evaluation and is the product of extensive collaboration between international donors, research groups, and practitioners.

vehicle fleet, and the costs of fuel, replacement parts, and maintenance labor, can vary considerably. Selecting appropriate parameters ensures that the benefit streams from reduced VOCs and travel times are accurately estimated. The process of calibrating the model entails setting many of the assumptions that need to go into estimating the direct benefit streams from road improvements. The economic assessment will involve analyzing the calibration choices used in the pre-implementation HDM-4 model and comparing these pre-implementation assumptions with assumptions based on our newly collected data.

In addition to assessing the main assumptions underlying the ERRs for the Tanzania roads investments, this evaluation will aim to develop useful lessons for future implementation and evaluation of roads activities. We will use the HDM-4 model to address a number of other important subquestions related to comparing key outcomes from the pre-implementation ERRs to actual outcomes and the expected life of the road. We outline these questions and outcomes in Table II.1.

Table II.1. Overview of research questions, key outcomes, and analytical approach

Research questions	Key outcomes	Analytical approach
1. How do the reestimated ERRs compare to the pre-implementation ERRs for the road segments? 2. Using actual measures of traffic, axle load, and realistic assumptions about road maintenance, what is the expected remaining life of the roads? 3. How sensitive are the updated ERRs to errors in key input data, such as traffic or maintenance, and what lessons can be learned for future ERR calculation?	• Traffic (annual average daily traffic [AADT]) • Road roughness (International Roughness Index [IRI]) • Vehicle operating costs • Estimated lifetime maintenance expenditures • Capital costs	• HDM-4 project analysis: road condition and traffic forecasting and sensitivity analysis • Comparison of the key outcomes, construction costs, time savings

The pre-implementation ERR models map closely to the elements of the program logic outlined for both the Mainland Trunk Roads and Zanzibar Roads Activities, modeling both direct and indirect benefit streams. As discussed in the Evaluation Assessment report, we believe that the direct benefit streams modeled with HDM-4 already implicitly capture any indirect benefits of road upgrading. HDM-4 is a consumer surplus model of the transport sector, so any indirect benefits are already captured by the increased number of trips taken following an improvement in the road and a reduction in transport costs. Including both direct and indirect benefit streams therefore likely overestimates the benefit of roads improvements and inflates the ERRs. Although the pre-implementation ERR models scale down the direct benefits in order to avoid double-counting, the percentage is chosen arbitrarily. For this reason, we will use the benefits projected by the HDM-4 model without adding any additional benefit streams that may already be captured by the model. The most valuable lessons for future ERR analysis will come from comparing the parts of the pre-implementation ERRs that are based on HDM-4 outputs with the reestimated model.

The following sections describe how we will use the HDM-4 model to reestimate the ERR for the roads activities and address the research questions described above. First, we discuss the data sources for the analysis and our approach to sampling. Then we describe our analytical approach, which includes the calibration of the HDM-4 model to the Tanzanian context, conducting the economic analysis, and conducting sensitivity testing using the calibrated model.

B. Data sources

The economic assessment will draw on a broad range of primary and secondary data sources to reestimate the ERRs for the roads activities and assess the validity of the initial assumptions of the pre-implementation ERRs around growth in traffic volumes and road deterioration over time.

The primary and secondary data sources for our economic analysis fall under three broad categories: road condition and inventory data, road user data, and other. Table II.2, below, shows the main variable types used in the HDM-4 model and the related data sources. Appendix B includes a more detailed list of calibration and input variables within each variable type, as well as our expected data sources. Chapter V provides greater detail on the data collection methods for the key variables.

The HDM-4 model uses many input variables and allows for numerous parameters of the model to be adjusted based on the local context. It also includes default values for the model parameters that could be appropriate depending on the purpose and use of the analysis. Making a judgment about when to use default parameters and when to adapt the parameters is a critical part of model development and affects the reliability of the model output. Our approach will be to calibrate those parameters to which the ERRs are most sensitive and those that vary most by context or across time, such as fuel costs or vehicle prices. HDM-4 defines these variables as the Level 1 parameters, which come from a combination of existing secondary data sources or from primary data (Bennett and Paterson 2000). In determining which data and how much data to collect, we will also consider whether a representative sample of road conditions data can replace data collection along the full length of all the built roads for some variables. In addition to focusing new data collection on the variables to which the ERRs are most sensitive, we will rely on existing HDM-4 models used in Tanzania or similar countries that are available to us. We propose to collect the following data as part of our primary data collection:

- **Roughness** will be measured using the International Roughness Index (IRI), which is a standard measure of road smoothness.
- **Surface condition** measures characteristics of the road surface, including the number and area of potholes, rutting, raveling, and other pavement defects. Surface condition data will collected using GPS-linked video and through manual surface condition assessments.
- **Modified structural number** summarizes the structural property of each layer of pavement and pavement strength. It may be derived from deflection data combined with either data from core sampling or data from the construction specifications in the as-built drawings, or if unavailable, the design drawings. We will explore whether deflection data is available from TANROADs.
- **Axle load** measures the average weight carried on a single vehicle axle; together with the traffic volume and composition, this is used to calculate the overall traffic loading on the road.

- **Traffic** measures the volume of traffic along the roads and the composition of traffic by vehicle type.
- **Vehicle unit costs** measure the costs faced by road users for various goods and services, including maintenance, spare parts, and fuel.
- **Vehicle fleet characteristics** measure the physical characteristics of the vehicle fleet, vehicle speeds, and prices.

Table II.2. Data sources for the economic assessment, by category

Variable category	Primary data sources and illustrative variables	Secondary data sources and illustrative variables
Road condition		
Pavement characteristics	**Engineering surveys** • Roughness • Percentage of road area that is potholed • Road thickness	**As-built drawings, road agency** • Shoulder and carriageway width • Structural number
Road inventory	n.a.	**Road agency** • Road segment length • Surface type
Road works and activities	n.a.	**Road agency maintenance department** • Maintenance strategy (e.g., the area of road with potholes that triggers road agency repairs)
Road users		
Road user data	**Vehicle operator surveys** • Cost of vehicle • Cost of fuel	**Transport sector reports, road agency** • Vehicle fleet characteristics (e.g. number of axles, engine speed, power)
Traffic data	**Traffic counts** • Annual average daily traffic (AADT)	**Historic data collection, feasibility reports** • Traffic growth rate
Other		
Project finance data	n.a.	**MCC project documentation, road agency maintenance departments** • Interest rates • Construction costs
Climate data	n.a.	**World Meteorological Association** • Average annual rainfall

C. Sampling approach for new data collection

As described in the guiding principles for the evaluation, we will seek to streamline data collection activities while prioritizing those variables that are key to generating robust ERR estimates. The sampling approach applies this principle to the data collection for road condition. Some of the primary data required for HDM-4 will be collected along the full length of the roads. However, some of the road condition variables for which we will collect data require expensive and time-intensive methods of collection. For these variables, we propose to collect data from a sample of segments drawn from each of the improved roads that are representative of the overall road condition, consistent with World Bank Information Quality Level standards for roads projects (Bennett and Paterson 2000) and TRL Overseas Road Note 18, Section 4 (TRL and DFID 1999). These segments will be defined using the existing reference system used by TANROADS and the MoIC to identify road segments so that data collected can be linked back to their systems.

Based on our design trip and field visits to the Tanga-Horohoro road and Pemba roads, as well as reports from the TANROADS staff, we believe that the roads built under the MCC Transport Sector Project seem to be in generally good condition. As a result, we expect that we will not find an unreasonably high degree of variation in road condition variables and that a sampling rate of 10 percent for surface condition will be sufficient. However, we will adapt our sampling strategy depending on the roughness profile of the road, observation of road surface condition from video footage, and internationally recommended practices. The sampling approach will follow an iterative process. First, we will collect data on road roughness for the full length of the improved roads and use video footage to create a coarse classification of surface condition[7]. Then we will combine these with secondary data related to terrain and road geometry available in engineering drawings and define a sample selection grid that stratifies road segments by the following attributes: road class, average roughness, coarse classification of surface condition, traffic class, region, construction lot number, and road terrain (flat, hilly, or rolling). Segments of 1 km for detailed study will be randomly selected from this grid to ensure that the sample covers the different types of road characteristics. We will assign a higher probability of selection for problematic segments, defined as those with greater average roughness or worse surface condition. In each of these 1km detailed study segments, we will conduct manual surface condition assessments.

We will collect primary data for the variables noted below:

- **Roughness** will be collected along the full length of the road network being analyzed.
- **Traffic** will be collected along multiple points on the road network to ensure that local traffic is excluded from the count and that any segments defined by important exits or entries onto the road are covered.[8]
- **Axle load** will be collected at mobile weigh stations on each of the roads. Axle load stations will be located near to traffic count stations.
- **Surface condition data** will be collected on the detailed study segments. Coarser measures of surface condition will be collected for the road network using GPS-linked video.
- **Modified structural number** will be taken from the as-built drawings or the design drawings. If deflection data is available from TANROADs we will explore using this instead.

[7] Coarse classification refers to a simple classification of road condition, such as good, fair, poor, based on an engineer's subjective assessment of road surface condition observed in the windshield video.

[8] We exclude local traffic because we want to estimate the number of trips along the full road segments. Local trips may only use a small part of the improved road.

We will work with TANROADS and MoIC to establish whether reliable secondary sources exist that can provide information on vehicle fleet characteristics, prices, and unit costs. In addition to these secondary data, we propose to collect primary data on vehicle prices and unit costs of maintenance, replacement parts, and fuel from a sample of transport companies and repair shops. Where possible, we will use primary data to validate information on road users drawn from other sources.

D. Analytical approach

The evaluation will cover three large segments of the road network in mainland Tanzania, covering of 470 km of trunk roads, and five segments of the network on Pemba Island, covering 35 km of rural roads. Although it is possible to use the Roads Economic Decision (RED) model to assess the ERRs for the rural roads, we propose using HDM-4 to better compare ERRs for rural and trunk roads. The cost difference for running the two models is relatively low because we are already collecting primary data to calibrate the HDM-4 model to analyze the mainland trunk roads. Many of the additional data requirements for running HDM-4 on the Pemba Island roads, such as roughness and traffic, are also required for running the RED model.

As described above, HDM-4 analysis is based on the concept of road pavement life cycle analysis, which is made up of two models that interact with each other: the road user effects (RUE) submodel and the road deterioration and works effects (RDWE) submodel (Odoki and Kerali 2000):

- **RUE** models the impacts of the road quality on road users, which are measured in terms of road user costs and other social and environmental effects.
- **RDWE** models road deterioration and condition over time; it is predicted as a function of traffic volume and congestion, road pavement type and strength, maintenance standards, and environment and climate.

A first step in using these two models is to ensure that they are appropriately calibrated to the Tanzanian context.

1. Calibrating HDM-4 models

Calibration of HDM-4 ensures that the forecasts generated by the model forecasts are as accurate as appropriate for the required application. In general, this requires high quality data so the outputs generated by the model reflect the actual conditions and benefits of the roads.

Based on conversations with officials at TANROADS, we understand that no major calibration exercise has been conducted for Tanzania and that there is no standardized process for calibrating the model in Tanzania. Although HDM-4 is widely used by TANROADS for prioritization of road upgrades and evaluation of potential projects, there is no standardized approach to calibration and no country-specific set of calibration parameters.[9] We will explore

[9] During our design trip, we were given a sample HDM-4 workspace used by TANROADS. We examined the files to see the extent to which the model has been calibrated and found that the model primarily uses default calibration parameters complemented by calibration of some vehicle fleet information, unit costs, maintenance practices, and climate zones.

further whether the original HDM-4 data files used to generate the pre-implementation ERRs can be found, but will work on the assumption that we will not have access to these files.

HDM-4 sets out three levels of calibration, wherein each level reflects a greater degree of customization to the local context but also a greater burden in terms of data collection and analysis. As noted earlier, we will conduct Level 1 calibration, because the ERRs are the most sensitive to these parameters and because the data collection effort matches MCC's expected timeline for this evaluation. Data required for Level 1 calibration can typically be collected over the course of weeks. In contrast, Levels 2 and 3 calibration require months and years of data collection, experimentation, and analysis to determine calibration parameters more accurately, and are typically used for roads planning and program purposes (Bennett and Paterson 2000).

Our proposed data collection for Level 1 calibration will help us determine the values of basic input parameters, adopt default values where appropriate, and calibrate the most sensitive parameters with best estimates from desk studies or field surveys. Appendix B includes the Level 1 calibration variables and indicates whether they will come from primary or secondary sources. The list in the table expands on the variables outlined above, including specifics on the variables to be collected from secondary sources. Because there is no standardized approach to calibration in Tanzania, we will study how previous HDM-4 models used in Tanzania have been calibrated to assess the most appropriate calibration values.

The choice of calibration parameters determines the outputs of the RUE and RDWE submodels and how the submodels interact. A large number of parameters can be adjusted in both the RUE and RDWE submodels; as part of our design, we identify the level of detail needed for the analysis.

- The RUE submodel predicts VOCs, travel times, safety, and environment. Level 1 calibration of these parameters are set using best estimates, desk studies, or minimal field surveys. Appendix B provides a detailed list of the Level 1 parameters that enter the RUE submodel and where we expect to collect the necessary data. For data that change frequently over time and/or are not up-to-date, we will collect the needed information via a vehicle operator survey. This survey will cover unit costs for vehicle operators and information on vehicle fleet costs if the secondary data we get from TANROADS are not up-to-date.
- The RDWE submodel predicts the road pavement strength through the modified structural number, the deterioration of the road surface over time in response to use, as well as the effect of ongoing maintenance on road condition. As with the calibration of the RUE submodel, we will primarily use a combination of secondary data sources or default model parameters, such as data from previous HDM-4 models; data derived from as-built drawings, desk studies, and literature reviews; and stated maintenance practices drawn from TANROADS and MoIC. We will also collect primary data on the road segments on key variables needed for the evaluation, namely on roughness and surface condition. The calibration choices for the RDWE submodel will be refined in an iterative way by comparing the pavement condition predicted by the model to the observed data collected on road roughness and pavement surface condition. We will continue to adjust our assumptions until the model predicts the observed road condition data well.

In addition to primary data collected used as part of the calibration of the RDWE and RUE submodels, we will collect data on traffic composition and volume, which we will compare with previous estimates of traffic volume derived from feasibility studies.

2. HDM-4 Analysis and sensitivity testing

Once calibrated, the HDM-4 model can be used to analyze a variety of scenarios. The model will report predictions for the evolution of VOCs, road condition, traffic, maintenance expenditures, and other key variables over the lifetime of the project. These outputs are compared to the base case alternative, in which road improvements are not conducted, to arrive at ERRs (Morosiuk et al. 2006). We will run these analyses to predict road condition and traffic on the MCC roads, and the costs and benefits incurred over time. We will then conduct sensitivity testing to illustrate how benefit streams change depending on the underlying modeling assumptions. Finally, we will compare the ERR from our ex-post model to the pre-implementation ERR, and adjust economic variables for inflation to the same year that pre-implementation models begins.

Each road will be divided into sections that will be analyzed separately.[10] These sections, referred to as homogeneous sections, will be selected so that they have uniform characteristics in terms of their road class, average roughness, traffic volume, region and construction lot number. The homogeneous sections should be generated using information from the pre-implementation period so that the analysis of sections reflects how the ex-ante analysis would have been done before the road was upgraded. We will do this using pre-implementation traffic volume and roughness measurements from feasibility studies. We will compare the homogeneous sections we generate to the homogeneous sections used in the pre-intervention ERRs if they are available.

HDM-4 generates a number of reports as part of the analysis, including important outputs that we will present in our analysis. In particular, we will predict and plot annually, for the lifetime of the road, the following outcomes:

- Traffic volume
- VOCs and travel time
- Road roughness, measured in terms of roughness (IRI)
- Cost streams, measured as expected costs of maintenance

We will compare these predicted values to the original ERR analysis. In addition, we will compute new ERRs based on the updated HDM-4 models. We will plot the original predictions about traffic growth, road roughness, and VOCs against the updated estimates.

Sensitivity analysis is conducted in HDM-4 by rerunning the model using different basic assumptions. For example, we might rerun the model assuming that traffic levels are 20 percent lower than predicted or that construction costs are 50 percent higher than planned. We propose to test the sensitivity of ERRs to assumptions about traffic growth, maintenance, and construction costs, since these variables often change over the course of a project or differ from initial

[10] Although analysis is done section by section, we will generate an ERR for each project.

predictions. We will present the results of the sensitivity analysis as a variety of different scenarios, and report the predicted ERRs, road roughness, traffic volume, and cost streams for each run of the model. We will use the sensitivity analysis to derive lessons for future applications of HDM-4 for MCC projects.

Currently, we do not have access to the original pre-implementation HDM-4 models, so we are limited in the comparisons that can be made across pre-implementation and ex-post ERR models. In the absence of detailed documentation about how the original pre-implementation HDM-4 models were calibrated, we will have difficulty constructing a baseline version of the HDM-4 model using our data that exactly mimics the predicted traffic, road deterioration, and rates of return reported in project documentation and ERRs. Instead, we will use sensitivity analysis to compare the model outputs using "corrected" post-implementation assumptions to the pre-implementation assumptions about traffic growth, road condition, construction costs, and maintenance. We will conduct a similar sensitivity analysis to test differences in ERRs when actual construction costs are used instead of projected construction costs in the cost-benefit analysis. We will also use data on roughness and surface condition to check whether assumptions about maintenance made during the initial modelling exercise are correct. We will assess the ERRs, VOCs, and IRI over time for a variety of different scenarios, including comparing actuals to pre-implementation assumptions for specific variables.

3. Road condition data reporting and visualization

We will develop itinerary diagrams to display road condition, integrating GIS data, available pre-construction data, and primary data collected for the evaluation. These diagrams will include pavement characteristics, traffic count stations and traffic data.

E. Risks and mitigation strategies

There are a number of potential risks to our proposed economic assessment, which we describe below, along with our proposed risk mitigation strategies.

- **Access to administrative data.** The HDM-4 model has large data requirements and we will rely on TANROADS and MoIC to provide us with secondary data. Their cooperation is essential and we will collaborate closely with our points of contact in these agencies.
- **Data quality.** The HDM-4 model relies on good quality data, especially for the most sensitive parameters. We will work with primary data collectors to ensure that data quality is a key priority, and will implement a variety of quality control protocols. Because we rely on secondary data sources for a number of variables, we will also need to establish that these data are reliable. Wherever we have primary data that relate to the secondary data, we will conduct cross-checks. We will also look for secondary data collected in other similar contexts to validate the secondary data we receive. The sensitivity analysis described in Section D, above, may mitigate against data quality issues.
- **Access to initial ERR data.** An important part of our analysis involves comparing current ERR estimates with previous ones. The degree to which such comparisons can be made accurately depends in large part on the availability of previous models used to calculate returns. Having the HDM-4 models that were used at the time of the initial ERR generation

will be an important element of this analysis. Without information on the HDM-4 models used to establish pre-implementation ERRs, we can only compare ex-post outputs with predicted pre-implementation outputs that are reported in the project documentation, such as VOCs and travel time savings, traffic growth, ERRs, and projected roughness. This may limit our ability to identify problematic assumptions in the initial models.

III. PERFORMANCE EVALUATION

A. Motivation and research questions

The performance evaluation of the Tanzania Transport Sector Project will complement the economic analysis of the roads projects by addressing important elements of the program logic that are not fully captured by the HDM-4 model. The pre-compact ERR estimates for the roads assumed that the maintenance capacities of the roads authorities in Tanzania would be sufficient to maintain adequate road conditions over the predicted lifetime of the roads. The program logic also assumed that reduced VOCs and travel time savings would especially benefit agricultural production through competitive transport and agriculture markets, leading to increased traffic related to agricultural production and increased cash crop revenue. If the roads are not adequately maintained, roads condition would deteriorate sooner than expected and the economic benefits would not materialize. Similarly, if the transport and agricultural markets are not competitive, benefits of the project may not accrue evenly to consumers and producers. To assess program logic assumptions about the roads and Mafia Island runway investments and to generate lessons for future roads projects, the performance evaluation will address the remaining following four broad research questions outlined in Chapter I:

1. What are the current maintenance practices and how do they affect the ERRs of the improved roads?
2. Who is traveling on the road, and why? What are they transporting?
3. Are there VOC savings, and are they being passed on to consumers of transportation services?
4. How accurate were the program logic and ERR assumptions for Mafia Island Airport runway?

B. Roads maintenance

Effective maintenance of roads investments is critical to ensuring that they realize their full economic benefits. Lapses in maintenance or inadequate maintenance can inflict a heavy cost burden on public agencies responsible for the roads and on road users. For example, the South African National Road Agency estimated that repair costs increase by six times the initial maintenance costs when maintenance is neglected for three years. This estimate rises to 18 times the initial maintenance costs after five years without maintenance (Burningham and Stankevich 2005). At the same time, maintenance is often delayed in developing countries struggling to fill multiple funding gaps. For every kilometer of road rehabilitated in Sub-Saharan Africa, an estimated three kilometers of road falls into disrepair, leading to a net loss of the overall road network (World Bank 2003). As the cost to repair under-maintained roads increases, the benefits of the roads to road users, who have to deal with slower speeds and higher wear and tear on their vehicles, decline rapidly.

Road maintenance covers three main types of activities: routine maintenance, periodic maintenance, and emergency maintenance. Routine maintenance of paved roads includes ensuring proper drainage and making spot repairs, such as fixing potholes or sealing cracks. Routine maintenance is typically performed at least annually. Periodic maintenance is planned work and involves more intensive roads work, such as resealing or applying asphalt overlay for paved roads. Although the timing of periodic maintenance depends on road use, climate, and the quality of the original construction, periodic maintenance is typically needed every few years (Schroeder 1990). Emergency maintenance covers repairing roads after unexpected events such as flooding. When a road has deteriorated past the point of periodic maintenance, it will likely need to be upgraded or reconstructed (Robinson and Thagesen 2004). All roads, from highways to unpaved rural roads, will likely require each type of maintenance over the course of their lives.

Roads maintenance is a multistage process requiring significant management and oversight. Governments need to monitor road deterioration throughout the network, ensure that there are sufficient funds for maintenance, allocate those funds effectively across the network, and, finally, manage specific maintenance projects so that they are implemented as expected. Donors supporting roads investment projects have attempted to identify and document standards and best practices for public maintenance activities (Gwilliam et al. 2008; Harral et al. 2011; Heggie 1997; Pinard 2015; Robinson and Thagesen 2004), including the following:

- Adequate, dedicated funding for maintenance generated through direct charges to road users or other sustainable revenue sources
- Political autonomy of those who oversee maintenance funds, such as through independent boards
- Prioritization of cost-saving routine and periodic maintenance activities on high-traffic roads that have the greatest effects on network-wide VOCs
- Implementation of maintenance projects by private contractors selected through competitive bidding processes
- Engagement of a variety of stakeholders in the planning, implementation, and oversight of maintenance (for example, through regional road user associations) to promote government accountability and public ownership of the roads network

Roads maintenance in both mainland Tanzania and Zanzibar are managed using funds that are independently managed by separate Roads Funds Boards. The Roads Funds are primarily financed from the levying of fuel taxes; in addition, they receive a small share of funding from vehicle-related taxes and fees. The boards are responsible for the collection and oversight of revenue, including disbursements to all governmental entities responsible for maintenance activities. On the mainland, TANROADS is responsible for maintaining approximately 3,000 kilometers of trunk roads and 21,000 kilometers of regional roads. The Ministry of Infrastructure and Communications (MoIC) is responsible for the maintenance of approximately 1,200 km of roads in Zanzibar. TANROADS and MoIC rely on road condition data that are captured and analyzed in a Roads Maintenance and Management System (RMMS) database. The database includes a series of analytical modules that can be used to identify and prioritize routine and periodic maintenance needs within budget constraints and also tracks maintenance implementation and contract milestones (Mwaipungu and Allopi 2012).

Pinard (2012) previously applied the Commercialized Road Management (CRM) framework, which integrates the best practices and provides a method to evaluate road management, to road maintenance institutions in Tanzania. The assessment found that the

Tanzanian road authorities were characterized by good institutional structure and management approaches. However, it rated Tanzania lower for the autonomy of the Roads Fund, asset management policy and planning, risk assessment, and quality control.

Our assessment of maintenance practices addresses the first research question of the performance evaluation: What are the current maintenance practices and how will they affect the projected life of the road? The evaluation will use a mixed-methods approach that measures the current condition of the roads, road agencies' maintenance practices for the network, and MCC's program logic assumptions about maintenance outcomes. Table III.1 lays out more detailed subquestions for this part of the performance evaluation, as well as the data sources and key outcomes. The subquestions break down how the evaluation will assess maintenance funding and implementation using information from the HDM-4 data collection, secondary data, and stakeholder interviews.

Table III.1. Roads maintenance research questions, data sources, and outcomes

Research questions	Data source and type	Outcomes
1a. To what extent have government entities been able to raise adequate maintenance funding and predict funding flows?	• Roads Fund revenue and documentation—secondary • Roads Fund policy documentation—secondary • Stakeholder interviews—primary	• Roads Fund revenue and outlays
1b. To what extent do available maintenance funds meet the maintenance needs? How are these needs identified and how are funds allocated to meet those needs?	• RMMS data—secondary • Maintenance policy documentation—secondary • Stakeholder interviews—primary	• Annual maintenance expenditures • Percentage of kilometers in the road network receiving routine/periodic maintenance per year • Percentage of kilometers in the network, by pavement condition over time • Percentage of projects prioritized by the RMMS that receive maintenance
1c. How is maintenance implemented? To what extent is it implemented according to plan? Have maintenance planning processes changed since the start of the compact?	• Stakeholder interviews—primary • RMMS data—secondary • Maintenance policy documentation—secondary	• Planned versus actual expenditures • Percentage of maintenance projects completed on time/within budget • Examination of maintenance practices on roads similar to the improved roads.
1d. What are the ERRs of the roads under different maintenance scenarios and how sensitive are they to different maintenance practices? Are trucks traveling on the roads overloaded?	• Road condition data (HDM-4 inputs)—primary/secondary • Sources for maintenance scenarios - Stakeholder interviews—primary - RMMS and Roads Fund data—secondary - Maintenance policy documentation—secondary • Axle loading data—primary	• Adjusted ERRs under different maintenance scenarios • Variance in adjusted ERRs • Percentage of large commercial vehicle axles that are overloaded • Equivalent Standard Axle Load (ESAL)

1. Data sources and outcomes

The data sources for the maintenance questions outlined above will include a combination of primary and secondary data from a number of sources. Outcomes for the performance evaluation of roads maintenance will be based on analysis of (1) administrative data, (2) interviews with maintenance officials and specialists in Tanzania, (3) axle loading data, and (4) road condition engineering data collected as part of the economic analysis.

- **Official records of maintenance budgets and expenditures,** including data on Roads Fund revenue and extracts from the national and Zanzibar regional RMMS databases. We will request RMMS data for the past five years, focusing our analysis on the period from compact end to the present. We will review all secondary data received for completeness and quality. We will use Roads Fund data to develop outcome measures for maintenance funding, such as annual Roads Fund revenue, and use RMMS data to develop outcomes for maintenance implementation, such as the percentage of kilometers of road in the network in need of maintenance and receiving maintenance and the percentage of kilometers of road in the network that are in good condition.
- **Government and Roads Fund documents** on maintenance policies and practices: we will request and review documents such as annual maintenance plans, annual road works reports, road fund performance reports, legislation governing the Roads Fund and road management agencies, strategic plans for the transport sector, guidance for road works contractors, and internal or donor-generated assessments of road management agency performance.
- **Key informant interviews (KIIs)** with staff from the roads maintenance agencies and other specialists in Tanzania. These will include representatives of TANROADS, MoIC, the Roads Fund Boards for the mainland and Zanzibar, and those involved in maintenance contracting and implementation. We will develop protocols for the interviews based on recommendations from the maintenance literature on best practices in managing and maintaining national and regional roads networks. In particular, the interviews will capture information on the MCC-funded maintenance activities and their influence on any current roads maintenance policies and practices; how RMMS is maintained and the quality of the information maintained; how roads maintenance allocations and decisions are made, particularly in a resource constrained environment; how maintenance contracts are issued; oversight of maintenance works, including the quality control and auditing processes are used for maintenance projects; and so on.
- **Axle loading data** will come from the axle load survey conducted on the compact-financed road segments. The survey will measure outcomes such as the weight and the number of axles of a sample of commercial freight vehicles traveling along the MCC-funded roads.
- **Road condition data.** These data will come from the primary and secondary engineering data sources used to run the HDM-4 model and estimate the ERRs, as described in the previous chapter.

2. Analytical approach

Our evaluation approach for maintenance will largely consist of an assessment of maintenance funding, planning, and expenditures. We will also assess the extent to which the Tanzania maintenance system is aligned with best practices for road management, and will consider differences between trunk and rural roads, as appropriate. Our analysis will examine the following:

- **Maintenance funding**. The Roads Funds are the primary source of funding for maintenance activities on mainland Tanzania and Zanzibar, and adequate funding is an essential precondition for maintenance planning and implementation. We will compare revenue from the Roads Funds over time, focusing on projected and actual revenues collected since the end of the compact. We will incorporate inflation estimates to assess if maintenance funds have remained stable in both real and nominal terms. We will also track indirect measures of demand for maintenance, such as expansion of the overall roads network. These comparisons will inform whether the availability of funding over time has remained stable and kept pace with demand.

  We will supplement this with an analysis of the qualitative data gathered from stakeholders, identifying commonalities and differences in the responses of different stakeholders, and compare the situations on the mainland versus Zanzibar. We will also seek insights from these stakeholders on the role of any significant donor contributions to maintenance activities and the implications for sustainability.

- **Maintenance planning and expenditures**. Using maintenance needs and prioritization data from the RMMS, we will compare differences in reported maintenance needs and planned maintenance, the distribution of maintenance expenditures on routine and periodic maintenance, as well as the maintenance backlog over time (measured in terms of the share of roads in poor, fair, or good condition). We will review changes in average roads network conditions over time and whether the share of roads falling into poor condition or needing rehabilitation is increasing. We will also analyze stakeholder perspectives on decision making around the allocation of maintenance funds, assessing how differences perspectives align with information in the administrative data.

- **Maintenance implementation and oversight.** To assess how effectively the roads maintenance funds are used, we will start by comparing planned spending on maintenance to actual spending. KIIs will supplement the comparison and provide insights into the reasons for differences in planned and actual spending. We will compare actual spending to the RMMS prioritization in order to understand the extent to which other factors affect actual spending allocations. We will compare the maintenance policies described in documents and stakeholder interviews to best practices. We will also try to obtain an understanding of the contracting process, the extent to which projects experience overruns or delays, and how maintenance authorities check programs through audits or other checks.

  We will also assess whether the roads authorities have put in place any new practices as a result of the compact-funded Maintenance Activity and whether the equipment and software systems funded by the activity are used to monitor road condition, and solicit perspectives on the current quality of those equipment and whether the activity led to increased capacity in the agencies. Throughout our analysis, we will identify any differences in stakeholder perspectives, assess the extent to which the opinions expressed are consistent with administrative data, and compare the situations on the mainland versus Zanzibar.

- **Axle load analysis.** The amount of pressure that large commercial vehicles exert on the road and how it is distributed across a vehicle’s axles has important implications for the degree of maintenance required and the life of the road. We will analyze the axle load survey data and compute two key measures of truck overloading—the percentage of vehicles that are overloaded, and the degree of overloading in terms of weight per axle. We will compare these with policies on axle loading to see if loading policies are being enforced. Where available, we will compare reports of axle loads from the feasibility studies that MCC and MCA-Tanzania commissioned prior to construction of the roads to those collected in our survey. This comparison will inform whether current use patterns are in line with the anticipated capacity of the road.

- **Road condition.** We will identify other causes of deterioration through the assessment of visual data collected from the field (e.g., photographs/video of drainage systems), and structural analysis. Visual data from surface condition assessments will be used to identify potential damage caused by traffic loading and environmental factors.
- **HDM-4 simulation.** HDM-4 includes assumptions about road maintenance practices, which in turn influence the predicted rate of deterioration over time. These assumptions specify maintenance work triggered by road deterioration or periodic and routine maintenance practices. The model then factors any costs incurred or maintenance into the ERRs. We will use HDM-4 to simulate the effect of different assumptions about maintenance practices in Tanzania, comparing the stated practices from TANROADS and MoIC with data from our analysis of the RMMS, road agency expenditure, and backlog analysis.

  As we test the sensitivity of the HDM-4 model, we will define a set of maintenance scenarios based on observed road conditions and maintenance requirements over the last five years. These scenarios could include (1) a fully funded, “best case” scenario where all maintenance can be completed; (2) a “realistic” maintenance scenario, reflecting actual maintenance as indicated in administrative maintenance records; and (3) a “worst case” scenario that might occur if the fuel taxes underwriting the Roads Fund were cut substantially. We will use these different maintenance scenarios to modify the maintenance-related inputs to the Maintenance Standards in our HDM-4 model, rerun the model, and report the ERRs that result for each road segment.

Finally, the output of these analyses will help the evaluation team identify potential lessons for future roads performance assessments and for developing assumptions about maintenance in the due diligence process of MCC roads investments.

C. Road users

This component of the performance evaluation will address the second main research question: Who is traveling on the road and why? What are they transporting? The program logic assumes that cheaper and increased movement of people and goods as a result of project activities will lead to greater economic activity and growth, particularly in the agricultural sector, via greater access to markets. Roads improvements that facilitate transportation and goods markets that are cheaper and more efficient benefit both producers and consumers. In particular, the program logic expected that expanded access to agricultural markets would generate greater aggregate sales for cash crop producers as more buyers are incentivized to access markets.

The performance assessment for this question will involve collecting data on traffic volume and traffic composition and from road users on their travel purpose and patterns through the expanded vehicle intercept survey. We will also conduct focus groups in communities along improved and unimproved roads to discuss changes in transportation since completion of the roads activities. We will select unimproved comparison roads from the roads that were identified in the baseline report completed by Economic Development Initiatives (EDI). The report documents an impact evaluation design and baseline survey that was completed in 2009 (EDI 2009).

These data will allow us to compare the present-day traffic with assumptions about economic activity and market access that the roads investments were intended to produce. As feasible, we will assess changes in traffic composition and volume by comparing the new traffic counts to those collected prior to implementation. Data from the road users survey will allow us to conduct a descriptive analysis of who is using the road and how, including information on the origin and destination of individual travelers and their trips, purpose, and frequency. In addition, the information acquired from the focus groups discussions provide perspectives on how transportation options and roads use have evolved since the completion of the roads. Below, we describe the main outcomes and data sources for this aspect of the evaluation, as well as the analytical approach we will use.

Table III.2 lists the outcomes and data sources for the road users assessment and links the outcomes to the research subquestions.

Table III.2. Traffic and road users research questions, data sources, and outcomes

Research questions	Data source and type	Key outcomes
2a. How many and what types of vehicles are traveling on project road segments?	• Classified traffic counts— primary • Traffic counts from the feasibility study—secondary	• Traffic counts • Traffic composition • Estimate of AADT
2b. Who are the road users, where are they traveling, and why are they using project road segments? Have their travel patterns changed over past 10 years?	• Expanded vehicle intercept survey—primary • Focus group discussions with communities near improved and unimproved roads	• Road user demographic characteristics • Origin • Origin activity categories • Destination • Trip purpose categories • Travel direction • Trip frequency categories
2c. How long does it take road users to move along key routes? What alternative routes or modes of transportation do road users use?	• Expanded vehicle intercept survey—primary	• Time per unit distance traveled • Alternative routes • Alternative modes of transportation categories
2d. What are freight carriers (trucks) transporting and what is its value? How many passengers and what kinds of goods are passenger transportation services carrying? What is the capacity for each?	• Expanded vehicle intercept survey—primary • Axle load survey • HDM-4 estimates of time savings	**Freight carriers**: • Cargo type, weight • Vehicle capacity • Number of crew **Passenger transportation services:** • Number of passengers, seats • Number of crew
2e. Where are commercial passengers traveling, and what are they paying for transport?	• Expanded vehicle intercept survey—primary	• Passenger origin and destination • Fare charged (Tsh.)

1. Data sources and outcomes

The outcomes data related to road users will come from three sources: (1) traffic counts, (2) expanded vehicle intercept surveys, and (3) focus group discussions.

- **Traffic count data,** an important input to the ERR analysis, will also inform the performance evaluation. Traffic counts provide data on traffic volume and traffic composition that will enable the performance evaluation to assess how many people and vehicles are currently using the improved road segments.
  - **Traffic volume.** Data on volume of traffic on upgraded project roads will come from the primary classified traffic counts. We will convert traffic counts into standard volume measures including estimates of ADT (the average number of vehicles passing a specific point during a 24-hour period) and AADT (the total volume of vehicle traffic of a road segment for a year divided by 365 days). Total volume in the calculation of AADT is typically estimated using seasonal adjustment factors. The traffic counts will take into account market days and seasonality considerations. They will also disaggregate traffic by time of day and day of the week (for example, weekday versus weekend), and will indicate directional movement of traffic volume.
  - **Traffic composition**. The traffic count survey will assign a class to each vehicle using the categories from the road traffic count survey (RTCS) of the pre-implementation feasibility studies to facilitate any potential pre-post analyses of traffic composition. Vehicles in the RTCS were disaggregated into motorized and nonmotorized vehicles; motorized vehicles were further disaggregated into passenger and goods vehicles. Passenger vehicle categories include cars, utility vehicles, two-wheel vehicles, and buses. Goods vehicle categories include light trucks, medium trucks, heavy trucks, semi/full trailers, and tractors. Bicycles and carts make up the two main nonmotorized categories.
- The **Expanded Vehicle Intercept Survey** will interview drivers and/or passengers of commercial and private vehicles traveling on the improved road segments. It will also capture pedestrians and non-vehicular use of the road. The survey will include two modules as described below:
  - **Origin-destination (O-D) module** will collect details on the origin and destination of each respondent’s trip, including the trip purpose, such as travel to home, work, school, leisure, or commercial activities. The survey will also collect data on travel direction, trip frequency, travel time between the trip’s origin and the interview point, and basic driver and vehicle characteristics.[11]
  - **Road users module** will collect data on an expanded set of questions about the road users, including information on cargo, costs, alternative routes, and household or business travel patterns. For trucks and other freight vehicles, we will gather information such as cargo type. For passenger vehicles, we will gather information on the number of passengers, capacity, and passenger destination and fares charged. We will also gather descriptive information from road users to obtain their perspectives and opinions about

[11] Details on the categories of information collected are included in Chapter V.

the road and their travel options, including the availability of public transportation services.

- We will pre-test the road users module to understand the level of detail we can collect from drivers and passengers over various periods of stoppage time (we understand that 10–15 minutes is a tolerable amount of time for a vehicle intercept survey in Tanzania, but this will need to be confirmed in a pre-test). We will use this information to determine the optimal length of both modules given the information we hope to obtain from the survey.

- **Focus group discussions** will be conducted in select communities along the improved roads as well as along selected comparison roads segments from the EDI baseline report. The discussions will involve interviewing community members about transportation in their area since the upgrading of project roads. In particular, we are interested in learning respondents' perceptions of changes to transportation availability and options, transportation use and costs, access to markets and health and social services, and travel times to access markets and services. We will use the EDI baseline data to guide the selection of communities within which to conduct the focus group discussions, and classify responses into subgroups such as where focus group participants live (for example, close to a commercial center or more remote).
- **Axle load surveys** will include administering the freight module from the expanded vehicle intercept survey in order to collect additional information on cargo, origin and destination, and other route options.

2. Sampling for the expanded vehicle intercept survey

The expanded vehicle intercept survey will be carried out on three days per road, including both a market and nonmarket day. On the mainland trunk roads, we will sample at least 20 percent (approximately every fifth vehicle) of all traffic moving in both directions along the roads. This proportion of traffic falls between the 10 percent rate that is considered adequate for descriptive studies and the 50 percent that is recommended for surveys used to predict future travel patterns; it has also been used for previous traffic surveys in Sub-Saharan Africa (Hajek 1977; Damsere-Derry et al. 2016). This will result in a total sample size of approximately 230 vehicles on the Mtwara Corridor, 760 vehicles on the Tanga-Horohoro trunk road, and 675 vehicles on the Tunduma Sumbawanga trunk road. To ensure driver and enumerator safety, we will conduct the vehicle intercept survey only during the day. Pemba Island roads have a lower traffic volume, so we plan to sample a larger proportion of vehicles, potentially on fewer days.

We anticipate that this sample size will allow us to estimate O-D outcomes, including the share of inter- and intraregional traffic, with a sampling error of 4 percent or less (see Appendix C for methodology).These sample sizes will also allow us to estimate outcomes based on the other survey modules, such as the share of vehicles traveling for noncommercial purposes (school, shopping, or medical), and the share of trucks carrying agricultural inputs or outputs, with similar levels of precision. We will use data from traffic counts and survey pre-tests to finalize our selection of traffic stations, survey days, and sampling rates to ensure that our final sample will permit us to estimate O-D and other key outcomes with adequate precision. The interview rate will be expanded to daily totals using the volumetric count totals.

3. Analytical approach

The analytical approach to using information from the various data related to road use will involve developing several comparisons related to understanding the movement of goods and people on the road segments and how the roads improvements may have contributed to it. These comparisons include the following:

- **Pre-post analysis of traffic volume and composition.** As described in Chapter V, where possible, we will conduct traffic counts in similar locations as the counts conducted for the feasibility studies. This will allow us to examine changes in traffic volume and composition over time. Of course, we cannot attribute all observed changes in traffic volume or composition to the improved roads, but these comparisons will provide some insights as to whether the changes are in the right direction and align with the program logic. Although these comparisons would not determine the impact of the project on traffic volume or composition, we would be able to measure whether traffic volume has significantly increased since the roads project was completed.
- **Origin-destination matrix.** Using data from the vehicle intercept surveys, we will construct an O-D matrix to analyze current road usage patterns. An O-D matrix illustrates travel patterns and demands by mapping trip origins and/or destinations located within a region (see the example in Table III.3). In practice, it would start with local market towns, moving up through regional centers and then to the greater Dar es Salaam area, and then to international destinations, indicating the routes that are the most heavily used. These travel patterns within and across subregions can reveal information about the types of economic activity in the region. For example, if there are increases over time in travel on a project road originating in the interior of the region and ending outside a subregion contiguous to an international border, it could suggest that international trade was facilitated by the project. The O-D matrix also allows for an analysis of trip length and intra- and interzonal traffic patterns.

Table III.3. Illustrative O-D matrix

	Destination				
Origin	Within region	Dar es Salaam	Other regions	International	Total
Within region (%)	16	16	27	0	58
Dar es Salaam (%)	17	NA	0	0	17
Other regions (%)	23	0	1	0	24
International (%)	0	0	0	NA	0
Total (%)	56	16	28	0	100

Source: Authors' calculations based on BCEOM and Inter-Consult Ltd. (2008).

- **Descriptive analyses of road users.** Using data from the expanded vehicle intercept survey, we will conduct descriptive analyses of road users, for example, of how they are using the road and what they are transporting. To understand economic activity of road users, we will use the O-D data to estimate the percentage of traffic devoted to agricultural production, travel to social services, other productive activities such as wage labor, and leisure activities, and the frequencies with which these types of trips are made. We will also describe results from qualitative questions about alternative routes and road users' opinions about transportation options and changes in road user travel patterns since the improvements of the roads. The analysis will look at direction of travel on project roads, alternate routes and modes of transportation, and (as applicable) cargo and passenger capacities and freight charges and fare costs. These will be presented in graphical illustrations, disaggregated by key subgroups such as road segment, vehicle type, and economic activity.

- **Access to markets and social services for communities along the roads**. Conducting focus group discussions along improved and comparison roads will allow us to solicit perspectives on how travel patterns have changed over time and whether road users have more access to transportation options and/or markets. The purposeful selection of communities from locations near improved roads and comparison segments will allow us to get a better understanding of who is benefiting from the roads and changes over time, and to what extent. We will use the EDI baseline data to validate the changes respondents perceive to have transpired over time.

D. Transportation cost savings

High transportation prices and costs are a major obstacle to economic growth in Sub-Saharan Africa (Teravaninthorn and Raballand 2009). The program logic assumes improved roads will lower transportation costs in Tanzania by reducing travel times and VOCs, particularly maintenance and fuel. In a competitive transportation market, reduced costs and the opportunity to earn larger profits should encourage more providers to offer transportation services. The increased competition would then put downward pressure on consumer prices and encourage transportation operators to pass on cost savings to consumers.

However, market imperfections such as barriers to entry often obstruct competition in African transportation markets, and policies and regulations such as subsidies and price ceilings can distort markets, preventing prices from adjusting to changes in costs. Such price rigidity can prevent cost savings from being passed on to consumers. Even if VOC savings are effectively passed on to consumers by traditional operators or by new competitors entering the market, measuring these cost savings poses a considerable challenge. There are likely multiple factors that influence freight costs and passenger fares other than VOCs. This aspect of the performance evaluation will address the third main research question: Are there VOC savings, and are they being passed on to consumers of transportation services? Table III.4 below outlines our related subquestions for assessing what happens to transportation costs saving, with the data sources and key outcomes.

The transportation cost section of the performance evaluation will gather information on transportation service prices and fares from transportation operators, as well as qualitative information on the Tanzanian transportation market structure from key informants within agencies and organizations including TANROADS, MoIC, the Surface and Marine Transport Regulatory Authority (SUMATRA), and the transport sector union. Data from the expanded vehicle intercept surveys will help us check cost data from these sources and inform whether there are more transportation options available to consumers.

Table III.4. Transportation cost savings research questions, data sources, and outcomes

Research questions	Data source and type	Key outcomes
3a. Have vehicle operator costs been reduced since the upgrades to project road segments?	• Administrative transportation fare data—secondary • Transportation service operator interview data—primary • Expanded vehicle survey (passenger module)—primary • Vehicle unit costs data—TBD	• Historical transportation fares • Service operator characteristics • Passenger fares charged (Tsh.)
3b. Are vehicle operator cost savings being passed on to consumers? Are more transportation services available to consumers?	• Administrative data on the Tanzanian transportation market structure—secondary • KIIs with government ministries and regional transportation departments—primary • Interviews with transport company officials and operators—primary • Focus group discussions	• Qualitative administrative information on transportation rules, regulations, and subsidies • Qualitative information on how fares and tariffs are determined, and opinions on price fluctuations and price elasticity • Operator costs and fares • Perspectives on transportation options

1. Data sources and outcomes

We will draw on the following data sources to create the key outcomes needed for the analysis of transport costs savings:

- **Administrative data.** These data from TANROADS, MoIC, and other entities such as SUMATRA will allow us to document historical and current fares along improved roads. These data should include information on transportation fare rules, regulations, and subsidies.
- **Transportation service operator interviews.** We will conduct KIIs with officials from government ministries and regional transportation departments to understand how fares and tariffs are determined, and develop questions that examine transport price fluctuations and elasticity. We will also interview staff at major transport and cargo companies and collect data on their operation costs and fares.
- **Outputs from HDM-4 analysis.** We will use VOCs generated from the HDM-4 analysis to assess changes in the costs incurred by drivers.
- **Expanded vehicle intercept survey.** The expanded vehicle intercept survey, described in the previous section, will also provide data on passenger origin and destination and fares charged.

2. Analytical approach

As noted above, improved roads should lead to costs savings for transport companies and private vehicle owners, which are ideally passed on to customers in the form of lower fares and freight charges. The analytical approach for this piece of the performance evaluation will first start with whether there were indeed cost savings to transport companies and other vehicle owners; if so, we will try to assess whether some of these savings resulted in lower costs for consumers given the current transportation market structure in Tanzania.

- **Vehicle operator costs.** HDM-4 measures expected reductions in VOCs given the current road conditions. We will use the output from these analyses to understand and analyze the extent to which operators of different vehicle types benefit from lower costs. This will help us determine whether it is likely that costs have gone down along the improved roads segments.
- **Transportation market structure.** Our analysis will focus on understanding the influence that existing policies and market imperfections may have in limiting observed transportation cost savings from being passed on to consumers. Data collected through document reviews and interviews with key transport sector staff on transportation fare rules, regulations, and subsidies will help us understand the regulatory environment, gather perspectives on the main cost drivers in the market, and identify which channels of cost savings might be passed on to consumers and what market imperfections may limit such pass-ons. The focus group discussions in communities near the roads may also provide additional insights into how the transportation market has changed over time.
- **Transportation fares.** We will analyze trends in historical transportation fare data for evidence of fare changes, and map them against external factors identified above. Examples of external factors we may identify include changes in fuel prices as well as other large-scale infrastructure upgrades. The elasticity or rigidity of fares to past external shocks may suggest how responsive fares might be to the upgrades made to project roads. We will supplement this analysis with the passenger fare data collected by the expanded vehicle intercept survey and the transportation fare information we collect from the interviews with TANROADS staff and fleet operators.

E. Mafia Airport activity

The performance evaluation will also include an assessment of the Mafia Island Airport runway upgrade. The initial program logic and ERR estimates anticipated that improvements to the runway and taxiways would allow larger planes to land on the island, making travel to the island easier and cheaper. As a result, tourist and business travel would increase. However, the landing of larger aircraft also requires improvements to fire and communications equipment for these types of planes, and these complementary investments were not made due to a lack of interest from airlines.

As a result, the revised ERRs calculated at compact closeout conceptualized the activity's benefit streams more narrowly. The projected increase in travelers to the island was scaled down significantly; the majority of estimated economic returns were expected to accrue to airplane operators in the form of reduced insurance costs from a safer runway and reduced maintenance and operating costs. Given the changes to the program logic, our analysis of the runway investment will be guided by the revised ERR estimates and will focus on the following question (research question 4): How accurate were the revised ERR assumptions for the Mafia Island Airport in terms of increased traffic and reduced aircraft operating costs?

1. Data sources and outcomes

For this analysis, we will draw on two main data sources: (1) administrative data that we will request from the Tanzania Airport Authority (TAA) on the number of flights arriving and departing and the types of aircraft, the number of passengers, and the amount of cargo and freight handled at the Mafia Island Airport and comparison airports, and (2) brief, semi-structured interviews with TAA officials and airline operators eliciting their perspectives on trends in air traffic and operating costs. We will also supplement the analysis with a visual assessment of the pavement condition of the upgraded runway and taxiways.

2. Analytical approach

Our analysis will involve a pre-post comparison of trends in air traffic to and from Mafia Island prior to the runway improvements and since the completion of these improvements. We will examine the number of flights, the types of aircraft over time, the number of passengers, and the quantity of cargo, and will calculate outcomes such as the number of passengers and quantity of cargo per flight, comparing levels before and after the project improvements. We will also compare actual traffic flow to traffic estimates from the compact closeout ERRs. We will ground our findings by comparing the trends in traffic flow on Mafia Island to overall trends in air traffic in Tanzania and on other, similar islands in Zanzibar.

Engineers from our data collection team will visit Mafia Island Airport to assess if the runway pavement condition is a significant improvement on its pre-compact condition and, as a result, is likely to result in reduced airplane operating costs. They will assess the overall runway pavement condition, divide it into sections based on overall condition, and rate each section on a standard scale based on the Federal Airline Administration (FAA) Asphalt Airfield Pavements: Pavement Surface Evaluation and Rating (PASER) Manual (Walker, Entine, and Kummer 2014). If we find evidence of improved pavement condition *and* increased air traffic, we can infer that the program logic is plausible enough to merit further investigation. In particular, if we find significant evidence of improved traffic, we will then conduct KIIs with Mafia Island Airport officials to learn about changes in tourism and other traffic to the island.

F. Risks and mitigation strategies

There are a number of risks to our proposed design, resulting mainly from our reliance on administrative data and potential challenges with vehicle intercept surveys. We can address these challenges through alternative analysis approaches and additional data collection.

The primary risk to the maintenance analysis is that administrative records are not available or not of sufficient quality and level of detail to generate a clear picture of maintenance. We will mitigate this risk to the quality of the evaluation by supplementing the administrative data with additional interviews with key stakeholders to systematically gather information on maintenance expenditures.

A risk to the road users analysis is that we will be unable to use the data from feasibility studies to conduct a pre-post analysis of traffic volume and composition because sites and data do not meet the needs of the current evaluation. Where possible, we will mitigate this risk by using the same count sites, and will collect similar primary data using similar methods. Where pre-implementation data either do not exist or cannot be replicated, we will still conduct descriptive analyses of traffic characteristics and try to benchmark them with other existing studies in Tanzania.

Another risk is that the challenging logistics of the expanded vehicle survey could affect the quality of the data. In particular, stopped vehicles will be anxious to move and we cannot ask them the number of questions required to get a complete, detailed picture of their transport patterns. We will mitigate this risk by carefully pre-testing the survey and getting road user input as we develop the survey instrument and test it before it is fully rolled out. We also plan on limiting the duration of the survey to under 15 minutes, which seems feasible based on input from potential subcontractors in Tanzania. We will conduct extensive enumerator training, develop a clear, effective protocol with police for flagging down cars to prevent traffic disruption, and develop effective sampling strategy (for example, flagging down groups of vehicles).

The transportation cost analysis is vulnerable to poor quality or limited data on transportation fares and vehicle operator cost data. If high quality fare data are unavailable, we will rely more heavily on interviews with agency staff and companies, acknowledging the limitations of our findings using these data in any deliverables. We will also triangulate across data sources when overlapping information is available.

This page has been left blank for double-sided copying.

IV. IMPACT EVALUATION

A. Motivation and research questions

In addition to estimating the economic return of the roads, MCC is interested in learning about the impacts of the investments on economic growth and poverty reduction. However estimating the impact of roads is challenging because the decision to invest in a road relates to location-specific characteristics, such as economic potential or political importance, which can make it difficult to identify a credible counterfactual. Another complicating factor is that roads improvements often have small individual impacts spread across a broad set of beneficiaries, making traditional impact measurement a significant challenge. Despite a strong economic rationale, many studies find little or no impact of roads projects on income and poverty in communities and households located near the roads.

The increase in availability of new data technologies—namely, remote sensing—has the potential to overcome some of the challenges in previous studies. Satellite-based measures of economic activity and poverty can be powerful, inexpensive alternatives in cases where survey data are either unavailable, costly (Jean et al. 2016; Henderson et al. 2012). Satellite data are also collected over a long period of time and can be used generate a baseline in cases where surveys or traditional data collection have not been done. This evaluation of the Tanzania roads investments presents an opportunity to conduct an impact assessment that could help develop evaluation approaches for future infrastructure projects.

This chapter presents a proposed impact evaluation design for the Tanzania roads improvement activities, which addresses the following research questions:

- Did the improvements in trunk roads lead to increased economic activity and poverty reduction in urban and rural communities connected to the roads?
- Did the improvements in rural roads lead to increased economic activity and poverty reduction in Pemba Island communities?

To address these questions, we propose developing two distinct satellite-based measures of poverty. These measures, which vary across location and time, will be used as outcome variables in a regression analysis to estimate project impacts. Applying this approach in Tanzania will enable MCC to understand the impact of road investments on poverty reduction while also advancing the use of satellite data methods to evaluate infrastructure projects. The Tanzania evaluation also presents a unique opportunity to validate findings from the satellite-based impact estimates using a survey-based impact evaluation design developed under the Tanzania Compact. Economic Development Initiatives, Ltd. (EDI) developed the design, which identified treatment and comparison road segments and collected baseline survey data prior to the upgrading of the MCC roads.

B. Background and overview of evaluation

Recent work in measuring the impacts of transportation infrastructure on economic activity has shown that the availability of satellite data can help overcome the difficulty in identifying a counterfactual, through the availability of data over a wide geographic area and across time. The long time span enables analysis of changing outcomes before and after infrastructure investments are made, and the wide geographic coverage provides counterfactual comparisons across space. In mostly urban and suburban areas, nighttime light imagery is used to derive light intensity which serves as proxy for economic activity. Several studies have used NTL images to measure changes in economic activity and poverty in developing countries (Chen and Nordhaus 2011; Henderson et al. 2012; Storeygard 2016). However, nighttime lights data is of limited use in measuring economic activity in rural areas, especially when these areas are not sufficiently electrified. A potential solution is to use daytime satellite imagery to estimate the income and poverty in rural areas. Jean et al (2016) combine survey and satellite data from five African countries (Tanzania, Malawi, Nigeria, Rwanda, and Uganda) to create a proxy measure for economic activity that can explain up to 75 percent of the variation in survey-based economic outcomes.

Building on these methods, we propose two distinct approaches for the impact evaluation of the Tanzania trunk and rural road investments: first, using nighttime light data for urban and sub-urban areas, and second, using daytime satellite imagery for rural areas. As described in greater detail below, we propose to estimate the impact of MCC's trunk and rural road upgrades on economic activity using a variety of panel data methods.

1. For urban areas, the evaluation will use nighttime lights data to assess the impact of the trunk road improvements on urban areas because urban areas are more likely to be electrified. We with link data on the average annual light intensity derived from nighttime lights imagery across time to the location of each urban pixel in Tanzania for the period before and after the upgrading of the trunk roads. The panel of light intensity for urban pixels will be linked to a proxy measure for transport costs and market access that will be derived from each pixel's position in the Tanzania trunk road network and the surface type for each road. The analysis of this data will use a regression model that relates improvements in travel times arising from road surface upgrades to economic activity as proxied by nighttime light intensity, controlling for fixed characteristics of urban pixels and flexible time trends. This will allow us to capture benefits accruing to urban areas immediately along the improved roads, as well as to more distant urban areas that are connected to these improved segments via the broader road network.
2. Because rural locations are not typically electrified, we will employ newly developed machine learning methods to derive poverty measures from daytime satellite imagery. These methods have been used to estimate decade-scale poverty measures for Tanzania, and this evaluation will use several advances to produce annual-level measures for locations near the improved roads. We will then estimate regression-adjusted differences in poverty levels between locations near improved roads and those near similar but unimproved roads, flexibly controlling for time variation via year fixed effects. The set of similar roads will be drawn from the comparison roads identified in the initial EDI design (validating and potentially refining these using the satellite imagery).

The rest of this chapter describes in more detail the data sources for the evaluation, the construction of satellite-based measures, the proposed framework for analysis and methodological approach, and an optional proposal to further validate the evaluation approach by following up on the EDI impact evaluation design via the collection and analysis of survey data.

C. Data Sources

We propose several data sources for this impact evaluation, and below outline outline how we plan to generate proxy measures of poverty and economic activity using satellite imagery.

1. Nighttime lights imagery

Several researchers have used nighttime lights data to estimate national income (Henderson et al. 2012) as well as income in sub-national regions (Baum-Snow et al. 2015). Previous studies have shown that nighttime lights can explain cross-country difference in income growth (Henderson et al. 2012), as well as differences in income at the sub-national level for less developed countries (Baum-Snow et al. 2015).[12] Recent work by Storeygard (2016) estimated the relationship between changing transport costs within countries and economic growth by using nighttime lights as a proxy for economic activity in urban areas and estimating the impact increasing transportation costs have on income growth.

Economic activity in a given location will be estimated using satellite imagery of light emitted at night from human settlements. We will use imagery collected by the United States Air Force Defense Meteorological Satellite Program (DMSP) and distributed by the US National Oceanic and Atmospheric Administration (NOAA). The DMSP satellites use an Operational Linescan Sensor to measure light at night every night across the globe. Light intensity (from cloud-free composites) is reported on an annual average basis for each pixel after non-stable light sources, such as forest fires or moonlight reflecting on clouds or water, have been removed. Each pixel represents 0.86 square km at the equator (pixel size decreases further from the equator) and is assigned a digital number (DN) representing light intensity.

We propose to use DMSP-OLS data from 1992 until 2012 and will extend the DMSP-OLS time series to 2016 by using nighttime light data collected from a different sensor that has been available since 2012. This data is collected using the Visible Infrared Imaging Radiometer Suite (VIIRS) Day/Night Band (DNB), which is carried by a satellite operated by the Suomi National Polar-Orbiting Partnership (Suomi NPP). The NPP-VIIRS data is processed and distributed by NOAA. The NPP-VIIRS data is available on a monthly basis, is more granular, and can measure a greater range of light intensities.

We will use nighttime lights data both in our analysis of the impact of trunk road upgrades on urban areas and in our machine learning approach to measuring poverty with daytime imagery. For our analysis of urban areas, we will identify the location of urban areas in Tanzania using nighttime light imagery to create a panel data set of light intensity for each urban pixel in Tanzania over the study period so that we can establish growth trends before and after the roads were upgraded. For our machine learning approach, we will use the panel of nighttime light data for the full country.

[12] Nighttime light (NTL) data does not perform well in explaining differences in GDP growth across American cities because most cities have reached the highest level of intensity that can be measured by the DMSP-OLS sensor.

2. Trunk Road Network Information

In addition to data on nighttime lights, we will collect data on the trunk road network in Tanzania in order to estimate transport costs and travel times within the country. The network spans regions within the country, connecting urban areas and linking Tanzania to neighboring countries. We will work with TANROADS to compile data on the trunk road network, including the year in which unpaved roads were upgraded to paved roads. Using this data, we will create a proxy measure of overall transport costs within Tanzania by assuming a fixed travel speed for paved and unpaved roads. We will use this data to estimate the transport costs faced by rural and urban areas along the trunk network in two ways. First, transport costs for each urban area will be estimated similar to Storeygard (2016) and by calculating the time it takes to travel to Dar es Salaam using the least-cost path through the trunk network. Dar es Salaam is the biggest economic center in the country and also Tanzania's most important port, so this is a key measure of market integration. However, this measure is imperfect, since market integration between locations within Tanzania is also important for economic growth. To account for this, we will develop a second measure of connectedness for each location that is a weighted average of the travel time between a given location and all major urban centers. Travel times will be estimated by making an assumption about the travel speeds that can be achieved on different types of road surfaces and estimated using the least cost path to travel between origins and destinations, weighted by the size of urban areas. Using both methods, travel times for each location will change as the trunk road network is upgraded.

3. Daytime imagery

We will use daytime imagery to construct measures of assets and consumption, in rural and peri-urban areas (both on the mainland and on Pemba Island), which are unlikely to be electrified and their light emissions too faint to be robustly detected. Promising recent work by Jean et al. (2016) combining survey and satellite data demonstrates that daytime imagery can generate measures of economic activity which explain up to 75 percent of the variation in survey-based economic outcomes, and future applications of the method could improve measurement accuracy. The evaluation will develop daytime satellite-based asset and consumption measures using a similar methodology as employed by Jean et al. (2016).

Our approach involves training a machine learning algorithm to recognize patterns in the Landsat 30m resolution daytime imagery that correlate with location-based measures of income and poverty derived from traditional surveys. The Landsat program is run by NASA and the US Geological Survey, and provides a long-running time-series of daytime imagery collected using visible light sensors. We will use patterns in the daytime imagery to predict poverty in locations and time periods where we have daytime imagery but not survey-based measures. The surveys we will use to train and test the algorithm include the EDI baseline survey and the 2008–2009 nationally representative Living Standard Measurement Survey (LSMS).

4. Data for developing and testing the poverty measures

Our proposed approach will involve testing poverty prediction methods using a combination of data sources, including nighttime lights measures, consumption and poverty indices from a pooled sample of traditional surveys (the MCC/EDI baseline survey and the 2008/2009 LSMS nationally representative survey), Landsat 30m (or comparable) imagery, and a variety of potentially relevant covariates (selected to serve as control variables, e.g., elevation). The machine learning approach designed to predict assets and consumption can use satellite and survey data from both rural and urban areas, including data on nighttime lights.

In rural areas, we plan to explore and contrast different parameterizations for the construction of poverty measures using daytime imagery, building on the approach used in Jean et al. (2016) which uses convolutional neural network (CNN) approaches to the spatial estimation of poverty. This method has been successfully implemented for Tanzania, but has only estimated levels of assets over a coarse (decadal) time-span. By integrating additional sources of daytime imagery, alongside additional relevant information and local surveys, we will identify the best parameterization for the construction of poverty measures in low-light areas, as well as assess the spatial and temporal granularity at which such estimates can be generated with high degrees of accuracy.

We will first identify the best model parameterization (based on our ability to predict in-sample data through bootstrapped validation) to estimate annual-level poverty measures for 2008/9 for locations surveyed in the LSMS survey. Based on the findings from the cross-sectional 2008/2009 model, we will conduct an out-of-sample prediction to locations that were not surveyed in the LSMS survey and for years through 2016.[13] As LSMS panel data is also available for 2010/2011 and 2012/2013, we may also seek to train and validate algorithms on the full set of all years for which data is available (2008/2009, 2010/2011, 2012/2013), subsampling to only include locations far from the treated roads to ensure our prediction is not biased by the actual intervention. We will also explore the feasibility of generating measures for a set of preceding years to test for comparability in pre-trends.

The primary model we will explore parameterizing will be a specific flavor of artificial neural network (ANN), which is convolutional neural networks (CNN). This model, part of a suite of "transfer learning" approaches (Pan and Yang 2010), identifies features in the Landsat daytime imagery that are predictive of nighttime lights—with the latter being an available but noisy proxy for economic well-being in rural areas—and then uses these daytime features to predict sparser consumption and asset measures that are available from LSMS and MCC/EDI baseline surveys. This approach outperforms nighttime lights imagery in predicting variation in economic well-being in Africa in rural area, and Jean et al. (2016) show evidence that a model built using data from one African country can make accurate predictions when applied to other African countries where the model had not been trained. This suggests that our approach can be used to reasonably predict poverty and well-being in countries and regions where validation data are unavailable and within regions of a country where we do not have survey data available.

As a part of this analysis, we will prepare and integrate a large set of candidate ancillary datasets in advance of the data analysis steps. Table IV.1 below lists several pieces of data to be assessed for inclusion. These data sources will (a) enable the identification of candidate control

[13] In conjunction with ongoing work being carried out by researchers at Stanford University, we may opt to train the algorithm to identify changes in poverty levels (i.e. 2008 - 2010) rather than baseline levels. We would then use the algorithm to predict the changes in poverty between 2008 and 2016 (and narrower windows) directly, rather than calculating these differences from annual-level predictions of the baseline poverty level. This would base the estimation of changes in poverty on additional temporal data, although it would not use the MCC baseline (as no comparable follow-up data is available). The determination of which approach to take will be based on their predictive performance.

and treatment locations near the improved roads by comparing pre-investment characteristics in the data, (b) enable the estimation of travel-time, as a proxy for access-to-markets, and (c) provide ancillary variables to promote more accurate estimates of impact by controlling for confounding factors.

Table IV.1. Key data sources

Domain	Source	Topic	# of Obs.	Current coverage: Temporal	Current coverage: Spatial	Spatial resolution
Human Development	VIIRS	Nighttime lights	N/A	1992-2016	Global	Grid cell (1km; 250m)
	AidData	International Aid	>2 Million	~1990-2015	Global	Variable
	gROADS	Road networks	N/A	1980-2010	Global	Grid cell (~1km)
Politics	GeoEPR	Political exclusion of ethnic groups	>4 million	1946-2014	Global	Ethnic group boundaries
	GADM	Administrative units	>200,000	1990-2014	Global	Admin unit
	AfroBarometer	Multiple survey questions regarding perception	>30,000	Variable rounds	Africa	Variable
	WDPA	WDPA Environmental protection areas	220,453	2015	Global	Variable
Demography	GPW	Population	N/A	1990-2020 every 5 years	Global	Grid cell (5km / 1km)
	GRUMP	Urban/rural settlements	N/A	1990-2010	Global	Grid cell (1km)
	SRTM	Elevation / Slope	N/A	2000	Global	Grid cell (500m)
	MODIS	Land cover and Fire	N/A	2001-2012	Global	Grid cell (1km)
	UDel	Air temperature & Precipitation	N/A	1900-2014	Global	Grid cell (50km)

D. Analytical approach

1. Mainland trunk roads

We will estimate the impact of upgrading the trunk road network on urban areas by estimating a regression model, which relates improvements in travel times to economic activity as proxied by nighttime light intensity. Our model controls for fixed characteristics of urban areas and flexible time trends by including year and urban area fixed effects. We will estimate the following model, using pixels in urban areas as the primary unit of analysis:

$$Y_{it} = \alpha + \beta TravelTime_{it} + D_i + D_t + \epsilon_{irt}$$

Where Y_{it} nighttime light intensity of pixel *i* and time *t* and $TravelTime_{it}$ is estimated as above, as the mean time required to reach a set of major urban destinations using the existing network and vehicle speeds based on road surface quality. D_i and D_t are fixed effects that account for fixed characteristics of each pixel (and thus of each urban area) and Tanzania-wide trends in nighttime light intensity. This model identifies the impact of reduced travel time, β , for all urban areas across Tanzania for the full sample period. For each urban area, we will calculate the change in $TravelTime_{it}$ that is attributable to MCC's trunk road investments and use, β , to estimate the resulting change in economic activity.

2. Pemba rural roads

Because feeder roads usually provide access for rural areas, we adopt a specification that focuses on estimating impacts on communities that live near to the road. As mentioned above, our base specification will estimate regression-adjusted differences in poverty levels between locations near improved roads and those near similar but unimproved roads, flexibly controlling for time variation via year fixed effects. We will use the comparison roads identified in the initial EDI design and the pixel-level as our primary unit of analysis to estimate the following specification in levels:

$$Y_{irt} = \alpha + \beta Post_t * T_r + D_i + D_t + \epsilon_{irt}$$

Where Y_{irt} is the outcome for pixel *i* near road *r* in year *t*, $Post_t$ denotes whether the road segment improvements have been completed, T_r indicates whether the nearest road is in the treatment group, and D_i and D_t are fixed effects for pixel and year that control for confounding cross-sectional and common time factors.

While we adopt this as our primary specification, we also recognize that it is possible that the rural roads improvements may affect more distant populations as well. We will therefore also extend this model to address two key (and related) questions:

1. Do the benefits of the road improvements accrue primarily to those living closest to the roads (say, within a few km), or do they reach farther populations as well?
2. If the benefits reach farther populations, how should we adjust our sample construction to avoid bias from spillovers?

To address the first question, we will explore the heterogeneity in treatment effects within the treatment group. To do so, we will add an interaction between our main treatment measure and the distance from each pixel to the nearest road.

$$Y_{irt} = \alpha + \beta Post_t * T_r + \lambda Post_t * T_r * Dist_{ir} + D_i + D_t + \epsilon_{irt}$$

In this model, $Dist_{ir}$ is the linear distance between the pixel and the road segment. We will thus be able to assess whether, say, only communities within 1km of the road experience gains, or whether more distant ones do as well. If we find large treatment effects still accrue at the 5km distance, we may extend our sample further out to assess at what distance these gains dissipate.

To address the second issue related to spillovers, we will make two adjustments. First, we will estimate a continuous treatment model as we do for the trunk roads, in which our measure of treatment is the change in travel time from each location to a set of urban destinations. We will estimate the following specification:

$$Y_{irt} = \alpha + \beta TravelTime_{irt} + D_i + D_t + \epsilon_{irt}$$

$TravelTime_{irt}$ reflects the mean time required for a vehicle to reach a set of major urban destinations, calculated using the road network and vehicle speeds associated with the road surface quality. This measure captures spillovers in which locations not immediately near the improved road nonetheless experience reductions in travel time because the improved segment serves as their main route to access a key destination. We will describe this measure for an expanded sample that includes both buffers along the matched comparison roads as well as buffers extended in both directions from the improved segments. This will allow us to assess whether spillovers may have biased our primary model both by indirectly treating our comparison units as well as by undercounting indirectly treated but nonsampled areas. Our second correction for spillovers involves adjusting our standard errors to account for the resulting autocorrelation in unobservables. To do so, we will estimate spatially lagged standard errors (a la Conley 1999).

3. Steps involved in carrying out the evaluation

To complete the evaluation, the team take the following steps. We will need to acquire road network data from TANROADS in addition to data on roughness and surface type. If we are not able to obtain electronic records or shape files of the road network, we will need to digitize the maps we have received to date. And if we cannot access data for the entire country, working with regional data may be sufficient. After obtaining the roads data, we will develop measures of market access reflecting each location's weighted average time required to reach other urban markets (weighted by market size) following Donaldson (2016). These measures will capture the changing access due to the MCC-funded road improvements.

We will then develop the poverty measures by obtaining daytime satellite imagery (Landsat 30m time series record) and developing the machine learning algorithm to generate annual poverty estimates. This analysis builds on Jean et al (2016) in developing a convolutional neural network. We will use the Landsat 30m imagery combined with the assets reflected in the geo-referenced DHS and LSMS-ISA to generate annual asset estimates. The final analysis will incorporate additional geospatial data for theory-informed heterogeneity, by merging additional data on agricultural production, climate, and other features and use these to test theoretical predictions about the heterogeneity of treatment effects. Finally we will analyze the main treatment effects due to improved market access and theory-informed heterogeneity tests.

E. Simulation-based power calculations

We use a simulation approach to determine whether the satellite-based consumption measurement strategy are likely to have sufficient statistical power to detect impacts of the road investment projects. First, we develop a probabilistic model of satellite-based consumption data over time that includes a hypothetical treatment effect of road investment on household consumption. The model of satellite-based consumption is based on data from the Tanzania National Panel surveys in 2008 and 2014-15 and factors in additional statistical noise that arises from using satellite-based measures as a proxy for consumption. Second, we use this model to simulate a dataset for 1000 grid cells over 5 years. Using the simulated data, we run a fixed effects regression model to estimate the treatment effect of the road investment and determine whether the treatment effect is statistically significant at p-values of 0.01, 0.05 and 0.1 We repeat this step 1000 times and report summary statistics on how frequently the treatment effect estimate is statistically significant at p-values of 0.01, 0.05 and 0.1. Our simulation approach tests how frequently a hypothetical treatment effect on household consumption of 0.025 (2.5%)

would be detected. This effect size is considerably smaller than the projected increase in daily traffic used in MCC's pre-investment ERR models.

1. Simulation specification and data sources

We model four processes as part of the simulation model: i) impacts, ii) initial log annual consumption, iii) annual change in log consumption and iv) satellite-based annual log consumption. Impacts occur along ten road segments, with 100 grid cells associated with each segment, for a total of 1000 grid cells assessed for impact over a five year period.[14] Each road segment is assigned for intervention in a randomly selected year during the first four years of the five year period. A panel dataset is constructed in which each cell receives a treatment binary of 1 during post-intervention years, and a 0 in pre-intervention years. In addition to the treatment data, each cell has data simulated along three additional dimensions.

We distinguish between true measures of consumption, which are unobserved and satellite-based measures, which are observable, but statistically noisy measures of true consumption. True initial log annual consumption is modelled using the mean and standard deviation for the rural subsample of the Tanzania 2008/9 National Panel Survey. Unobserved, true initial log annual consumption is modelled as:

$$Y_{i,t=1} = N_i\left(\mu = 9.8,\ \sigma = 0.6\right) \qquad \text{eq. 1}$$

Where $Y_{i,t=1}$ represents the unobserved true mean annual consumption level for cell *i* and time period 1.

We model subsequent true consumption levels using mean changes estimated in the Tanzania 2014/15 National Panel Survey and distributional parameters on the changes in consumption from Beegle et al. (2011).[15] Unobserved, true annual change in log consumption is modelled as:

$$C_{i,t} = N\left(\mu = 0.05,\ \sigma = 0.15\right) + \rho^* \varepsilon_{z,i} \qquad \text{eq. 2}$$

Where $C_{i,t}$ represents the unobserved change in consumption levels at cell *i* during year *t*, $N\left(\mu = 0.05,\ \sigma = 0.15\right) + \rho^* \varepsilon_{z,i}$ represents a random value drawn from a normal distribution with a mean of 0.05 and standard deviation of 0.15[16], ρ a parameter for the maximum error (set to

[14] The grid cells used in our final analysis will be 30m x 30m, thus providing more than 11,100 cells within 1km of a 10km long road segment. Even if we limit the analysis to only villages along road segments and assume a given village is 1km squared, we will have more than 1000 cells per village, and likely many times that many along a given road segment.

[15] Beegle et al. (2011) use a panel dataset of rural households in Kagera, Tanzania that followed households over a longer duration than the National Panel Survey, making it a better source for the distributional parameters for changes in consumption.

[16] See parameters drawn from Beegle, Kathleen, Joachim De Weerdt, and Stefan Dercon. "Migration and economic mobility in Tanzania: Evidence from a tracking survey." *Review of Economics and Statistics* 93, no. 3 (2011): 1010-1033.

25%), and $\varepsilon_{z,i}$ an error term estimated to approximate a within-road intervention correlation of 0.3:

$$\varepsilon_{z,i} = N_z\left(\mu = 0.0,\ \sigma = (0.3)*0.15\right) + N_i\left(\mu = 0.0,\ \sigma = (0.7)*0.15\right)$$

Where N_z is the component of error that is held constant across all cells *i* which are attributed to road *z,* and N_i the component of error unique to each cell *i*. We then model subsequent levels of consumption based on these changes:

$$Y_{i,t>1} = Y_{i,t-1} * \left(1 + C_{i,t}\right) + \theta * T_{i,t} \qquad \text{eq. 4}$$

Where $Y_{i,t>1}$ denotes log annual consumption in time periods after the first, θ the hypothetical treatment effect (2.5%), and $T_{i,t}$ a binary value indicating if cell *i* was considered treated at year *t*. Parameters for the initial time period are drawn from equation 1.

Finally, we model the observed, satellite-based consumption levels for each period as a function of the true consumption levels using distributional parameters from Jean et al. (2016). The observed satellite-based annual log consumption is modelled as:

$$Y^{*}_{i,t} = Y_{i,t} + \left(Y_{i,t} * U_{i,t}\left(0.25, 0.45\right) * B\left(-1,1\right)\right)$$

where $Y^{*}_{i,t}$ is the value of change in log consumption a researcher might observe when leveraging satellite-based approaches to the detection of consumption[17], $Y_{i,t}$ the simulated log change in poverty calculated following equations 3 and 4, $U_{i,t}(0.25, 0.45)$ a random (uniform) noise term applied based on the reported variance satellite data can explain in log consumption (see footnote 2), and $B(-1,1)$ a binary random (uniform) draw of a -1 or 1 to adjust for the potential negative and positive directionality of errors.

2. Results

After Y^{*} is generated, a linear regression model is fit to estimate the impact of the hypothetical road intervention on each plot *i*.:

$$Y_{i,z,t} = \alpha + \theta T_{i,t} + D_t + D_z + \epsilon_{i,z,t}$$

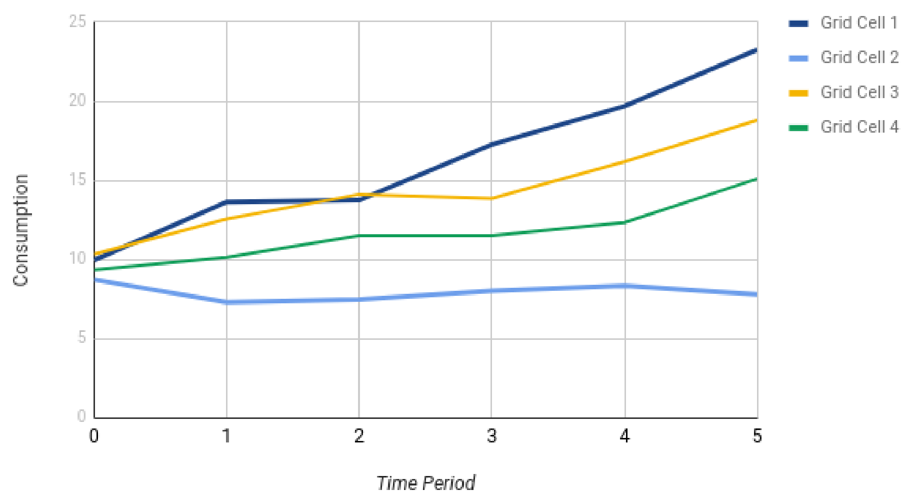
This model reflects the type of model we propose to use in our analysis. A two-step procedure is followed in which first a standard linear model is fit (including an intercept term

[17] Parameters are drawn from Jean, Neal, Marshall Burke, Michael Xie, W. Matthew Davis, David B. Lobell, and Stefano Ermon. "Combining satellite imagery and machine learning to predict poverty." *Science* 353, no. 6301 (2016): 790-794, who found that 55% to 75% of the variation in average household asset wealth across five countries in Africa could be explained by satellite modeling (following a cross-validation procedure). The random noise term applied here represents potential variance due to the remaining 45 to 25% of unexplained variance; a uniform term is chosen as a conservative estimator.

and road-level fixed effects), and then a heteroskedasticity robust covariance (HC1) is used to re-estimate standard errors and concomitant p-values. Each iteration the p-value at which significance would have been found is recorded.

We ran 1,000 simulations following the procedure outlined above. As an illustrative example of the model, Figure IV.1 below shows the outcome of equation 4 for four randomly selected cells (plots). For θ = 0.025, the distribution of p-values at which significance would have been identified was recorded, as well as the percentage of times significance would have been found at the 0.01, 0.05 and 0.01 levels, respectively. At $p <= 0.01$, significance was found in approximately 88% of all simulations. At $p <= 0.05$, significance was found in 90.7% of simulations; at $p <= 0.1$ it was found in 92.5% of simulations.

Figure IV.1 Example simulated consumption levels



Note: Consumption is measured as the log of consumption and time period is measured in one year increments.

These findings suggest that a satellite-based approach to measuring consumptions has a strong likelihood of detecting true effects, even with relatively small treatment effect sizes. An effect could be detected even while accounting for potential errors in measurement, heteroskedasticity, and both year and intervention fixed effects. This provides sufficient confidence that there is adequate power to detect even quite small effects.

F. Optional validation of impact estimates using survey data

Because EDI identified comparison roads and collected baseline data prior to the upgrading of the MCC roads, the Tanzania roads evaluation presents a unique opportunity to confirm and validate our satellite-data based approach, and assess whether these methods offer a low-cost alternative to survey-based impact evaluations. Below we present this optional addition that would allow MCC to compare the impact estimates for the two approaches in order to generate lessons on the future measurement of impacts for roads investments.

To conduct the validation exercise, we would collect follow-up survey data on the treatment and comparison households in the baseline study, and, through a difference in difference approach, estimate the impact of the roads improvements on those households.[18] As described above, the EDI baseline study identified communities living near the upgraded trunk and rural roads that would serve as a comparison group. We propose to complete the initial EDI design on Pemba only because the household surveys conducted on Pemba are more complete than the surveys conducted on the mainland, and the roads that were upgraded were designed largely to benefit these rural communities (unlike the trunk roads, which were also intended to connect other parts of the country).

Completing the EDI design would involve some additional analysis prior to designing and conducting the follow-up survey. Although EDI was careful in selecting comparison communities, their baseline report shows statistically significant differences in baseline characteristics between households in the treatment and comparison communities. To correct for this, we propose to implement a propensity score matching (PSM) design to ensure that the comparison group more accurately reflects the treated group at baseline. A PSM design involves two steps. First, we will compare baseline data on community and household characteristics across treatment and comparison areas, using this data to estimate how baseline characteristics correlate with being near an MCC road. Second, we will use these estimates to generate weights that we apply to follow-up data on communities and households, giving a greater weight to units in the comparison group that are similar to the units in the treated group. This approach will enable us to estimate income and poverty effects of the MCC's roads investments. The impacts of the upgraded roads will be estimated using a matched comparison design, which controls for baseline characteristics of households prior to road construction.

Implementing the initial impact evaluation design requires conducting follow-up surveys with approximately 1,200 households across 80 villages. We propose to collect data on economic activities, assets, and household characteristics and will work to ensure comparability between survey rounds. The follow-up survey would collect data through in-person interviews with household heads or members of the household who are knowledgeable of the household's economic activities. These survey data will be used to generate measures of economic activity, and poverty and wealth, which will be the outcomes used in our impact evaluation.

We propose validating the satellite-based methodology in two ways. First, we will assess the extent to which variation across villages in economic activity and poverty as measured by a survey correlates with similar measures derived from satellite-based methods. We will also assess the extent to which changes over time in economic activity and poverty, as measured by the two approaches, correlate with each other. These comparisons will provide a general sense of whether the daytime imagery approach predicts poverty and economic activity well. Second, we will compare impact estimates of the rural road upgrades that are based on survey data and satellite data. Provided the satellite-based measures predict poverty and economic activity well, we would expect to find similar conclusions about the impacts of the road project.

Because this validation approach is optional and would require additional funding to conduct the household survey, the decision to move forward could be made after completing the satellite-based impact analysis. The successful implementation of impact evaluation and the

[18] Adding a follow-up household survey to the impact evaluation approach would also enable the evaluation to collect a richer data set on beneficiaries and on the outcomes predicted by the program logic.

validation exercise would inform future roads evaluation designs, especially where survey data are unavailable or costly.

G. Risks and mitigation strategies

There are two potential risks to the proposed impact evaluation, which are described below, along with proposed mitigation strategies.

- **Availability of administrative data.** Some of the models we propose to estimate use measures of travel time and distance that rely on road condition and GIS data on trunk and feeder road networks. Obtaining comprehensive data on roughness and surface type for the national road network could be a challenge. It is possible that some of the road condition data is missing and we are unable to credibly estimate travel times. If this is the case, we will have to adapt or drop our analysis that includes travel times. Missing road condition data will not affect our ability to estimate the main model specification.
- **Poverty measures may be noisy.** Our evaluation design builds on a method that has successfully generated accurate consumption and asset measures that can detect differences across low-level administrative units. We propose to estimate consumption and assets at annual rather than 5- and 10-year time spans. There is a risk that our models may produce noisy estimates of consumption and assets, which would prevent us from detecting impacts of the program. We mitigate this risk by including a broader range of survey data in the machine learning algorithm, including two types of surveys (DHS and LSMS-ISA) that range across a number of years. By training the machine learning algorithm to predict consumption and assets across years and survey instruments, we expect to reduce the noisiness of the imagery-based measures and increase the likelihood of detecting a program impact.

This page has been left blank for double-sided copying.

V. DATA COLLECTION

As described in previous chapters, the evaluation of the Tanzania Transport Sector Project will draw on a diverse set of data sources to respond to the key evaluation questions. The guiding principles for this evaluation, described in Chapter I, serve as a blueprint for our data collection plans: we are seeking high quality, cost-efficient data collection approaches that meet both Tanzania and international standards and lead to actionable findings. We are prioritizing primary data collection efforts for variables to which the ERRs are most sensitive and those that vary most by context or across time, in order to obtain robust ERR estimates. We expect to learn lessons from carrying out the proposed data collection approach in Tanzania that will inform future MCC road evaluations and contribute to industry learnings. We will share these lessons on data collection methods and standards in our evaluation findings.

In this chapter, we detail our data collection plans for the economic assessment and the performance assessment, describing both our data collection sources and the proposed approach for obtaining data. The impact evaluation data collection approach will be submitted later, with the impact evaluation design memo.

We organize our proposed data collection approach in three broad sections covering (1) primary quantitative data (roads condition data, and road user data); (2) secondary data, including program administrative data and other data drawn from existing documents; and (3) qualitative data obtained from KIIs and focus groups. For each of these data sources, we start with a description of the data and our measurement approach, followed by our approach to collecting high quality data. We have already started collecting some secondary data, as depicted in Table IV.3, below. We end the chapter with a summary of our data quality plans.

A. Primary quantitative data

To estimate the projected ERRs and conduct the performance evaluation, we will collect information on road performance, including primary data on pavement quality, traffic volume, travel times, and VOCs. Driven by our guiding principles, we will focus our resources on collecting data for variables that are key drivers of the ERRs, and look for opportunities to reduce costs without sacrificing the credibility of the analysis. We separate our description of our primary quantitative data collection approach into engineering roads data, road user data, and VOCs to mirror the inputs for our HDM-4 modeling. Table IV.1 provides a summary of the sample and the associated equipment/instruments needed. We describe each of the data collection approaches in greater detail below.

We end the section with a description of our data collection process, including our future engagement with an engineering subcontractor and our intended approach to working with the road authorities and relevant government entities on data collection tasks.

Table IV.1. Quantitative data sample units, sample size, and instruments

Data type	Sample unit	Sample size	Relevant instruments/ equipment	Anticipated Quality Control Protocols
Roughness	All roads	470 km of trunk roads, 35 km of rural roads	IRI	Link roughness data to GPS readings. Review all raw and treated data. Photographic evidence of field application
Surface condition	**Video capture:** All roads	470 km of trunk roads, 35 km of rural roads	GPS-linked camera	Spot checks by local consultant.
	Manual surface condition assessment: Road segment	A stratified random sample of 1km study segments along all trunk and feeder roads	Trained data collection team, using instrument based on TRL Road Note 18, Section 4 (TRL and DFID 1999) and the 2003 TANROADS field testing manual.	Comparison of GPS linked video capture with results of manual assessment. Spot checks by local consultant.
Axle load survey	Road segment	One per road segment, 3 days for 12-hour counts. For Pemba roads, only one station for all roads	Mobile or fixed weight stations; TANROADS instrument	Calibration and testing of weight stations. Spot checks by local consultant.
Traffic Counts	Select census stations/road segment	A total of 13; see Table IV.2; 7-day counts	MCA-T Classified Traffic Count instrument	Training of enumerators. Pneumatic tubes and/or other technologies to verify accuracy of manual counts. Spot checks by local consultant.
Expanded vehicle intercept survey (including O-D and road user modules)	Select census stations/road segment	20% of traffic/8-hour day * 3 days; larger sample for rural roads	TANROADS; TRL instrument; adapted to this evaluation	Pre-testing and pre-testing of data collection instruments. Training and testing of enumerators. Data checks. Use of Tablets.
Vehicle Operator Survey	TANROADS standard VOC sample	5 major transport operators 5 garage /mechanics 5 freight companies 5 vehicle sellers – second hand and new. Coverage by vehicle type based on common vehicles sold and/or used	TANROADS; TRL VOC instrument, adapted to TZ	Triangulation of information

1. Road condition data

Road engineering data include three main elements: road roughness, deflection, and surface conditions surveys.

a. Data to be collected from the complete road chainage

Road roughness. We will measure road roughness using the International Roughness Index (IRI) for the entire network of MCC-funded roads on both mainland Tanzania and Pemba Island. Our engineering subcontractor will lease specialized equipment such as accelerometers and laser-mounted roughness measures from either the Central Materials Laboratory of TANROADS or the University of Dar es Salaam. We will secure similar equipment from Pemba, or transport it from the mainland if it is not available. Once we obtain the equipment, the subcontractor will calibrate the measuring equipment for use with the measuring vehicle. The road roughness data will be captured for analysis using an IRI software analysis package.

Video-based road condition assessment: To assess overall road condition for all project roads, we will first use a GPS-enabled video camera mounted on the vehicle collecting roughness data. If feasible, we will to collect video prior to the data collection exercise. In this case, we will add a GPS-enabled camera focuses on the road surface designed to capture detailed images of the road surface. A trained engineer will review the video footage and rate the condition of the roads using a coarse classification. We will use a combination of traffic, roughness and surface condition to stratify the road chainage.

Then, we will combine roughness and surface condition data collected via the video camera to select a random sample of road segments from which to collect manual surface condition in a manner consistent with TRL Overseas Road Note 18, Section 4 (TRL and DFID 1999). We will oversample from the roads and segments with high IRI and/or poor condition observed in the video.

b. Data to be collected from a stratified random sample of road segments

Detailed road surface condition surveys. These surveys are visual inspections of the pavement on road segments, carried out by engineers. Our engineering firms will hire TANROADS and MoIC engineers and technicians to conduct the survey for the MCC-funded roads. The survey will identify potholing, cracking, raveling, patching, pumping in cracks, deformation, edge-break, bleeding, surfacing, and drainage, and will be conducted in accordance with the TRL Road Note 18, Section 4 (TRL and DFID 1999) as a minimum standard. We will also use the 2003 TANROADS field testing manual, where it is consistent with the TRL's guidance. We will use similar procedures and equipment for the rural roads in Pemba. Where we find evidence of potholing and cracks, we will take additional measurements as well as photographs.

Mafia Island Airport runway inspection. Our engineering subcontractor will travel to Mafia Island Airport to conduct a visual inspection of the runway; if there is evidence of potholing and cracks, will take additional measurements and photographs. The engineering subcontractor will collect video footage of the airport runway.

2. Road user data

Road user data include traffic counts, axle load surveys, and expanded vehicle intercept surveys, which will be used for HDM-4 modeling as well as for our performance evaluation.

Traffic counts. We will collect classified traffic counts on all the roads constructed by MCC. Classified traffic counts categorize traffic data by vehicle type. Passenger vehicle categories will include cars, utility vehicles, two-wheel vehicles, and buses. Goods vehicle categories will include light trucks, medium trucks, heavy trucks, semi/full trailers, and tractors. Bicycles and carts make up the two main nonmotorized categories. To compare current traffic volumes to previous ones, we will use the RTCS instrument, which was used in the pre-implementation feasibility studies. Where possible, we will place the traffic count stations in the same locations as used in the pre-implementation feasibility studies to allow for maximum comparability. However, we will assess these locations to make sure they are appropriate from a methodological perspective; if they are not, we will select alternate locations. We will also visit the sites prior to confirming their selection to assess safety, the presence of proper shelter for enumerators, and that the locations are a true representation of the traffic on the roads.

Our data collection subcontractor will conduct traffic counts throughout a seven-day period to include three-day, 24-hour counts and four-day, 12-hour counts, with one of the 24-hour count days being a regional market day. The same count will be replicated at each traffic count station. We will employ TANROADS' standardized classified traffic count instrument. Table IV.2, below, includes information on our anticipated traffic count stations per road. To determine the number of traffic count stations per road segment, we will look at the possibility of placing the same traffic count stations as in the baseline to make comparisons, and also ensure count stations are placed along segments where there are no major exits or entries. We will also place traffic count stations at least 5 km from major junctions. In addition, we will place traffic count stations outside of urban centers to ensure that local traffic is not counted. We will collect traffic counts during the dry and the rainy season, although we anticipate having more dry season traffic count stations.

Axle load surveys provide information on traffic loading. Our data collection subcontractor will partner with the road authorities in Tanzania to supply both the technicians and the equipment to weigh vehicles. This survey will require mobile weigh bridges placed on the side of the road where current permanent weigh bridges are located. Permanent weigh bridges offer the necessary and safe space to conduct these surveys; however, the permanent equipment does not offer the possibility of measuring weight per axle, as the mobile weigh bridges do. We will conduct the axle load survey for three days[19] in each of the axle load stations; at least one of the three days will be a market day. We will conduct this survey along each of the road segments in mainland Tanzania using either existing TANROADS permanent weigh stations or mobile weigh station at the sites of existing weigh stations (for safety reasons). Where no permanent weigh stations exist along critical segments, we will consult with MoIC and TANROADs on the safest location to place the mobile weigh station. Through the axle load survey, we will obtain estimates of the average loading per axle by vehicle type, and examine vehicle overloading. We will conduct axle load surveys separately for each lane (direction) of traffic. It is a fairly common occurrence in Africa for trucks to be overloaded in one direction (to markets) and relatively empty in the other direction. Having the two different data points will affect the usable life of the lane. While a commercial vehicle is being weighed, we will use the time to conduct part of our expanded vehicle intercept survey, described below. We anticipate

[19] We will expand the duration of the axle load survey to ensure a sufficient sample of vehicle types, as needed.

conducting this survey at the same time as the rest of the dry season data collection in summer 2019, and are aware that harvest season in Tanzania spans June to August. Therefore, we will discuss with MCC the possibility of conducting the axle load survey earlier than the rest of the engineering data, or incorporating an adjustment factor to calculate possible loads during harvest season.

Table IV.2. Traffic count stations per road segment

Road segments	Traffic count stations (dry)	Traffic count stations(wet)
T-9 Tundumba–Sumbawanga	3	1
Mtwara corridor	4	2
T2 (Tanga–Horohoro)	1	1
Mzambarauni Takao Pandani Finya	2	1
Mzambarauni Karim Mapofu	1	1
Chwale Kojani	1	1
Kipangani Kangagani	1	

Expanded Vehicle Intercept Survey. As described in Chapter III, the expanded vehicle intercept survey will include an O-D module and a road user module. The O-D module will focus on collecting details on the origin and destination of the respondent's trip, including the activity at the origin and destination, as well as trip purpose.[20] This module will also collect data on travel direction, trip frequency, travel time between the trip's origin and the interview point, and basic driver and vehicle characteristics such as name, age, gender, and vehicle type. Additional questions include the following:

- For trucks and other freight vehicles, we will ask questions about cargo type,[21] and number of crew members.
- For passenger vehicles, we will inquire about number of passengers, seats, and crew members, and passenger destination and fares charged (Tsh.).

The road user module will contain additional questions related to road use and travel, such as travel patterns, alternative routes and modes of transportation, availability of travel services and their use, travel costs, benefits, and challenges of the road since its improvement.

To administer the expanded vehicle intercept survey, we will systematically stop traffic on the roads (in accordance with the sampling procedure outlined in Chapter III), and conduct roadside interviews with drivers of each vehicle type, and passengers in commercial vehicles

[20] Activities at origin and destination include (1) home, (2) work, (3) school, (4) shopping, (5) leisure/social, (6) medical facilities, (7) trade (i.e., buying and selling at a market), and (8) other. Trip purpose categories will include (1) to home, (2) to work, (3) to school, (4) business, (5) shopping, (6) government/official, (7) leisure/social, (8) tourism, and (10) other. Trip frequency categories include (1) several times per day, (2) once per day, (3) several times per week, (4) once per week, (5) several times per month, (6) once per month, (7) several times per year, (8) once per year, and (9) rarely or only trip.

[21] Cargo type categories will include (1) livestock, (2) agricultural produce or inputs, (3) market items, (4) machinery, (5) medicines, (6) fuels, (7) hardware and building materials, and (8) other.

such as taxis and buses. We will conduct the survey during the daytime on a market and nonmarket days for each road segment in mainland Tanzania, and a smaller segment of roads on Pemba. We will discuss optimal locations to place enumerators for this survey in Pemba with MoIC, likely in some of the same spots as the traffic count stations and sufficiently outside urban and market areas to ensure adequate representation of different vehicle types. We anticipate administering both modules to stopped vehicles; however, based on our pre-test results and traffic volume, we may administer the road user module to a subset of stopped vehicles. The two modules combined may take up to 15 minutes.

Our local engineering firm will work closely with TANROADS in Dar es Salam, as well as TANROADS regional managers, to secure approvals and to assist in working with the local police commander. The local police will be responsible for pulling over vehicles in line with our sampling plan. We will follow the same procedure with MoIC and local police in Pemba. If feasible, we will program the questionnaire for use on tablets, to streamline the data management process.

Vehicle Operator Survey. To understand changes in costs associated with roads improvements predicted by the HDM-4 model, we will obtain information on vehicle operating costs. It is our understanding that TANROADS and MoIC currently have data on VOCs. Should the data be outdated or of questionable quality, we will survey Tanzanian vehicle fleet operators (bus companies, transport operators, shipping companies) to collect VOCs as well as information on fleet characteristics. With the support of a transport economist hired by our engineering firm, we will survey major transport operators, travel companies, garages, and concessionaires to obtain both fixed and variable transportation costs such as unit costs of trucks, tires, vehicle maintenance, and fuel. We will obtain the list of fleet operators (with five or more vehicles) from the World Bank transport group, TANROADS, MoIC, or another reliable source. We will use electronic data capture programmed with the World Bank or U.K.'s Transport Research Laboratory (TRL) standardized VOC questionnaire, adjusted to the Tanzania context. The information from this vehicle operator survey will be complemented by secondary data from TANROADS and MoIC (see below).

3. Process for collecting primary data

All of the data collection activities will be carried out by experienced subcontractors in Tanzania and Mathematica staff, working in close partnership with local authorities. After the design report is approved, we will hire a road engineering firm to collect the road engineering data, axle load surveys, and to provide logistics support for the traffic counts and vehicle intercept survey. Our terms of reference (ToRs) for the engineering firm will require that the chosen firm has significant experience in Tanzania and Zanzibar in the following:

- Working with the relevant roads authorities to collect engineering data, including roughness, and road condition
- Training enumerators to conduct axle load surveys and a proven track record in conducting or supporting O/D surveys that use safe traffic stops as part of the sampling procedure
- Assembling multidisciplinary teams with knowledge on pavement engineering, transportation economics, and highway and rural road engineering

We will also engage a local engineering expert to serve as the data coordinator for this evaluation. The coordinator will be based in Tanzania, provide oversight on the data collection process, and participate in or conduct some of the KIIs. The coordinator will have prior experience conducting or overseeing roads data collection, an understanding of road evaluations and TANROADS road data collection standards, and a strong network of contacts within Tanzania to facilitate obtaining secondary sources of data and KIIs.

EDI will be responsible for collecting traffic count data, conducting the expanded vehicle intercept survey, and collecting data from focus groups. EDI has extensive experience in collecting high quality survey data in Tanzania and can implement the required data quality protocols required for the evaluation in addition to ensuring local IRB approval for road user data collection.

The data collection activities require working very closely with Tanzanian authorities, such as the government agencies responsible for roads planning and maintenance and the police. TANROADS and the Zanzibar Ministry of Infrastructure and Communications will be instrumental in obtaining the necessary approvals to conduct much of the primary data collection activities. We will partner with the road authorities to lease specific data collection equipment and request secondary data such as vehicle operating costs and roads maintenance data. Road users data collection will require the involvement of local police in order to monitor traffic counts or stop traffic to interview passengers and drivers. Finally, to facilitate our data collection efforts for the Mafia Island Airport, we will liaise with the Tanzanian Airport Authority in order to request permission to conduct a visual inspection of the runway.

B. Secondary data

We will complement the primary data with a number of secondary data sources. Examples of secondary data include construction costs, VOCs, maintenance expenses, and climate. We will request information on a series of documents such as as-built drawings or records, administrative records, and data from the RMMS database.

1. Types of secondary data

Road condition data. To supplement the roads engineering data collected by our engineering firm, we will extract geotechnical data from as-built drawings. The as-built drawings will provide us with information related to the pavement type and pavement age, as well as the functional classification of roads. We will also use as-built drawings and project completion reports to estimate the Structural Number, using the reported pavement thickness and subgrade CBR. As depicted in Table IV.3, below, we have already received the as-built drawings for the mainland roads but not for Pemba Roads. Based on our interviews during the evaluation design trip in February 2017, we believe that TANROADS and MoIC have data on deflection and IRI that we can also access.

Road users and other data. As mentioned previously, we will review the VOC data collected by TANROADS and MoIC and determine whether we can limit or supplement the vehicle operator survey described above. We learned that every year TANROADS collects VOC costs for one of four regions in Tanzania, which means that their VOC data may be recent or out of date, depending on the overlap of MCC roads with their data collection schedule. Finally, we will gather climate data by subregion from the Tanzania National Bureau of Statistics or a reliable international body, and will use it to adjust the HDM-4 model to account for seasonality.

Administrative data. As described in Chapter III, the performance evaluation will rely on a number of administrative data sources. To inform the analysis of maintenance funding, expenditures, and implementation, we will review documented maintenance policies, practices, and financial data. This information will come from the Roads Fund revenue records, the RMMS database, and similar records for Zanzibar. In addition, we will review any regulated fares for passengers on the mainland or in Zanzibar. Finally, we will request air traffic and passenger data for the Mafia Island Airport from the Tanzania Airport Authority. In particular, we will request data on the number of flights arriving and departing, the number of passengers, and the amount of cargo and freight handled at the Mafia Island Airport and comparison airports.

2. Process for collecting secondary data

We will work closely with MCC and our local contacts to request the data described in Table IV.3, below. For example, we will prepare draft data request letters for MCC or other local representatives to put on their letterhead introducing Mathematica and the evaluation. During the evaluation design trip, we established a productive relationship with Eng. Salum Sasillo, formerly the director of Transport Sector Projects MCA-T and currently working for TANROADS. Dr. Sasillo and our local data coordinator can assist as needed in requesting and collecting secondary data. Table IV.3 includes a list of secondary data that we have received, requested, or are planning to request in order to supplement our primary data collection efforts. We expect that as we begin our data collection planning, we will include additional sources of secondary data.

Table IV.3. Secondary data requests

Type of secondary data	Status
Example HDM-4 workspace	Received
GIS shapefiles showing the current road network for mainland trunk road network	Received
As-built drawings and final reports for 4/5 Mainland Roads	Received
As-built drawings and final reports for Pemba	Requested
Maintenance records from TANROADS and MoIC	Requested
Remaining as-built drawings and final reports for Zanzibar Rural Roads and Mainland Roads	Requested
Data on IRI and deflection contained in TANROADS RMMS database	Requested
GIS shapefiles showing current road network for Pemba	Requested
Tanzania and Zanzibar policy documentation (legislation, charters, development plans, regulations)	To be requested
Mafia Island Airport passenger and cargo records	To be requested

C. Primary qualitative data

We will also draw on data from interviews with key informants, including officials from the roads ministry and other stakeholders. These data will primarily inform the performance evaluation research questions on roads maintenance and transportation costs savings. We will conduct the interviews with staff from the following entities:

- TANROADS (both staff in headquarters as well as staff based in regional offices)
- MoIC
- Roads Fund Board (Mainland)
- Zanzibar Roads Fund (Zanzibar)
- Transport operators
- Tanzanian Airports Authority

The semi-structured instruments will include predefined questions, allowing us to gather specific information while also permitting open-ended conversation that may reshape our interview and, potentially, research questions. A preliminary list of data sources and topics for the KIIs is included in Table IV.4, below.

To secure meetings with the relevant representatives, we will first request an introductory letter from MCC to TANROADS, MoIC, and other relevant entities. We will also seek the support of our engineering subcontractor and of stakeholders we met during the design trip to help secure meetings with key informants.

In addition to relying on key information interviews, we will also conduct focus group discussions with community members in areas near the improved roads and in areas near the comparison road segments identified in the EDI baseline. The focus group discussions will provide further information on perceptions of transportation options, costs, and improvement in access to markets and/or other social services.

Table IV.4. Qualitative data collection focus areas

Data source	Areas of focus
Former MCA-T representatives in the roads, contracting, and M&E departments	• Reflections on implementation and sustainability • Perceptions of the implementation of infrastructure interventions • Maintenance practices
TANROADS policy, management, and engineering staff	• Perceptions on any road use change, benefits, unintended consequences • Quality control and auditing process • Maintenance practices of RMMS • Data quality in RMMS and links to decision making • Priority identification of RMMS and links to decision making • Factors driving maintenance decisions • Procurement and road construction processes
MoIC policy, management, and engineering staff	• Reflections on implementation and sustainability • Sustainability of road enhancements and maintenance • Challenges to sustainability • Maintenance practices of RMMS system • Data quality in RMMS and links to decision making • Priority identification of RMMS and links to decision making • Factors driving maintenance decisions
Tanzania Airport Authority officials and airline operators	• Perspectives on trends in air traffic and operating costs.

Data source	Areas of focus
Local regional roads officials	• Reflections on implementation • Involvement in roads management process • Satisfaction with roads maintenance practices and roads upgrades and implementation • Perceptions on road uses regionally • Process to communicate road conditions with headquarters • Degree of involvement in maintenance decisions of local roads
Transportation operators	• Perceived benefits of improved roads and perception of maintenance practices • Perceptions of maintenance savings based on improved roads. • Expanded life of vehicle fleet due to improved road conditions • Increased competition since emergence of roads • Changes in fares over time • Share of travel conducted on MCC-funded roads
Representatives from Road Funds Boards (Mainland and Zanzibar)	• How levels of funding of the Roads Funds have changed over time • Challenges and successes • Independence of the Roads Funds • Revenue and expenditure data for the Roads Funds • Whether the Maintenance Activity funded by MCC influenced road maintenance policies and management practices, and if so, how
Road users (via focus groups)	• Change in travel patterns over time; changes in access to transportation options and markets. • Involvement in maintenance process

D. Data quality plan

Collecting high quality data depends on developing data collection plans around the realities of local capacity, and building data collection skills as needed. It begins with ensuring the right mode of data collection for the context and developing a reasonable process for data quality checks. For example, electronic data collection can reduce human error, but it depends on the training and oversight of enumerators during the data collection process.

Selecting a high quality data collection firm, strong training of enumerators, and periodic checks. Critical to collecting high quality data is the selection of a highly capable subcontractor, and defining practices to ensure data quality as part of the contract. Prior to selecting the engineering subcontractor, we will check references and include a data quality plan as part of the ToRs. We will then include our standard Mathematica data quality control procedures in the subcontract with the engineering firm and require that all surveys include a pre-testing stage and a training stage. As an extra quality assurance measure, our local data coordinator/engineer will provide oversight to data collection efforts and give regular updates to the evaluation team. He will also conduct unannounced visits to traffic count stations and sites where the road users survey takes place at randomly selected collection points. In addition, as we receive engineering data, we will engage our HDM-4 specialist to review the outputs and compare them to previous data available, to assess if there are major discrepancies that could be caused by poor quality data collection practices.

Pairing technology with in-person measurement. To the maximum extent possible, we will use computer tablets for data collection related to road users. Computer-assisted personal interviewing (CAPI) software has the potential to reduce errors through internal programmed validity checks, increase the timeliness of data collection by avoiding a separate data entry process, and more closely monitor data collection in the field. We will explore the feasibility of including tablets for both the vehicle intercept surveys and the traffic count surveys. We will also explore leasing GPS-tracked video cameras from TANROADS to control for road condition data quality and traffic counts. In addition, we will work with our engineering subcontractor to identify additional traffic count technology (such as automatic traffic counters and classifiers) for a sample of traffic count sites to conduct additional data quality reviews. As a data quality assurance step in our measurement of roughness, we will test the accuracy of the pavement roughness data capture by comparing it to the roughness of a test section which either has a known roughness or has roughness measured using a static method.

Comparing information from KIIs When obtaining qualitative data from KIIs, we will compare information from data sources that overlap, assess inconsistencies, and when possible follow up or note the inconsistencies. In addition, for both our qualitative and quantitative research, senior staff review all methods, results, and interpretation via the QA process.

This page has been left blank for double-sided copying.

VI. EVALUATION ADMINISTRATION AND MANAGEMENT

In this chapter, we summarize several administrative and management issues related to the successful implementation of the proposed evaluation design for the roads and airports activities.

A. Institutional review board

Mathematica will prepare and submit institutional review board (IRB) applications for the four primary data collection activities that involve human subjects: (1) the expanded vehicle intercept survey, (2) KIIs with roads stakeholders, (3) focus group discussions, and (4) the vehicle operator survey. We intend to use Health Media Lab as our IRB, based on our positive experience with it on previous MCC projects. For the IRB application, we will submit a set of required documents, including a research protocol providing details of the study and data collection activity, copies of data collection instruments, and a completed IRB questionnaire that summarizes the key elements of the data collection protocols and plans for protecting respondents' confidentiality. The data collection instruments that we prepare and share with the IRB will include consent statements approved by MCC that guarantee the confidentiality of respondents. We will work with local data collection partners to identify and obtain any local approvals needed from Tanzanian authorities for the data collection efforts described above.

We do not anticipate seeking IRB approval for road condition data and classified traffic counts, as they do not involve intervening or interacting with human subjects or identifiable private information. However, we will obtain all necessary permission and cooperation from roads and police authorities in Tanzania to collect these data.

B. Data access, privacy, and documentation

Primary data used for the evaluation will come under three main categories in terms of anticipated access, privacy, and documentation.

1. For **road condition data and classified traffic counts,** we will produce a data collection report that includes the key input data used in modeling road segments in HDM-4, data collection tools and templates, and summary statistics of key variables.
2. For the **qualitative data collection**, including focus group discussions and KIIs with roads stakeholders, we will maintain confidentiality of respondents and follow MCC's latest guidelines regarding access to qualitative data.
3. For the **quantitative surveys**, including surveys of vehicle operators, the expanded vehicle intercept survey, and the axle load survey, we will follow MCC's latest guidelines regarding public use data, and make these data available in a form that is consistent with the confidentiality guaranteed in the respondent consent statements included in the data collection instruments and approved by the IRB.

These products will be de-identified according to the most recent guidelines set forth by MCC. The public use data files will be free of personal or geographic identifiers that would permit unassisted identification of individual respondents or their households. In addition, we will remove or adjust variables that introduce reasonable risks of deductive disclosure of the identity of individual respondents. Mathematica will remove all individual identifiers, including names, addresses, telephone numbers, government-issued identification numbers such as license plates, and any other similar variables. We will also recode unique and rare data by using top and bottom coding or replacing affected observations with missing values. If necessary, we will also collapse into less identifiable categories any variables that make an individual highly visible as a consequence of geographic or other factors (such as ethnic classifications or languages spoken). These measures are designed to retain the usefulness of the data while preserving the privacy of survey respondents.

C. Dissemination

Mathematica will present the findings from our analysis in person to MCC and stakeholders in Tanzania after completing a draft of the analysis report. These presentations will be valuable for both disseminating the findings to relevant stakeholders and gathering feedback from them to revise the draft report. In addition, we will collaborate with MCC and stakeholders to identify a variety of forums—including conferences, workshops, and publications—to share results and encourage donors, implementers, and policymakers to integrate the findings into future programming.

D. Evaluation team

Mathematica's evaluation team brings together strong design, data collection, and HDM-4 expertise. Our core team includes Dr. Anu Rangarajan, Ms. Delia Welsh, and Dr. Harry Evdorides. Dr. Rangarajan will serve as project director, providing leadership and technical support for all aspects of the project. Ms. Welsh, as deputy project director, will oversee the evaluation team on a day-to-day basis and lead the performance evaluation under the project. Dr. Evdorides will serve as senior advisor and work closely with Dr. Anthony Harris to provide quality assurance of all HDM-4 related inputs and deliverables. Dr. Sarah Hughes will oversee all data collection activities. Other Mathematica staff, including those with experience in analyzing quantitative and qualitative data, will assist key staff in carrying out the evaluation. Finally, we are in the process of recruiting a locally based evaluation coordinator who will assist us in identifying and obtaining secondary and administrative data, serve as a liaison between the U.S.-based evaluation team and local stakeholders, and oversee primary data collection activities.

E. Evaluation timeline and reporting schedule

We expect to pilot and conduct traffic counts during the rainy season in March of 2019. We will then pilot the road condition data, axle load surveys, and expanded vehicle intercept survey in May and the conduct data collection and focus group discussions in June 2019 which is at the end of the rainy season in most parts of Tanzania. We aim to begin collecting primary data that are not seasonal in nature, including administrative data on maintenance and the airport runway, KIIs, and the vehicle operator survey prior to the end of the rainy season, and to finish all data collection by August 2019.

We will produce the draft analysis report, which will integrate findings from all evaluation components by September 2019. We will then present the draft analysis report to stakeholders and obtain their feedback before finalizing it in October 2019.

Table V.1. Projected timeline for the evaluation

Month	J	F	M	A	M	J	J	A	S	O	N	D	J	F	M	A	M	J	J	A	S	O	N	D	J	F	M	A	M	J	J	A	S	O
Task	2017												2018												2019									
2. Design report																																		
Design report trip		□	▲																															
HDM-4 calibration																																		
Design report											▲					▲								▲										
3. Develop evaluation materials																																		
Draft TORs and hire data collection firm											▲													▲										
Prepare and pilot data collection instruments																													□	(▲)				
Data collection work plan																												▲						
IRB approval																										▲								
4. Supervise and prepare data collection																																		
Data collection trip																														□	○			
Data collection oversight																															▲			
Qualitative data collection (focus groups)																															□	○		
5. Analysis report and files																																		
Data collection inventory trip																												□						
Data cleaning and anonymization																																		▲
Analysis Report																																		▲
6. Impact Evaluation																																		
Acquire existing road roughness/materials data																																		
Develop "market access" measures																																		
Incorporate additional geospatial data for theory-informed heterogeneity																																		
Merge nighttime lights and analyze																																		
Daytime satellite imagery acquisition																																		
Develop machine learning algorithm to generate annual poverty estimates																																		
Analyze impacts																																		
7. Disseminate Results																																		▲

□ Trip to Tanzania

○ Trip Report

Δ Draft report/deliverable

▲ Report/deliverable

This page has been left blank for double-sided copying.

REFERENCES

Africon Limited. “Detailed Engineering Design for the Upgrading of Tunduma-Sumbawanga Road to Bituminous Standard; Package 2: Laela-Sumbawanga Section (96.5km): Design Report.” Pretoria, South Africa: Africon Limited, 2007.

Akee, Randall. “The Babeldaob Road: The Impact of Road Construction on Rural Labor Force Outcomes in the Republic of Palau.” IZA Discussion Paper Series no. 2452. Bonn, Germany: Institute of Labor Economics, November 2006.

Alevy, Jonathan E. “Impacts of the MCC Transportation Project in Nicaragua.” Final report submitted to Millennium Challenge Corporation. September 2014.

Asher, Sam, and Paul Novosad. “Market Access and Structural Transformation: Evidence from Rural Roads in India.” Unpublished job market paper. Oxford, UK: Department of Economics, University of Oxford, January 2016.

Atkin, David, and Dave Donaldson. “Who’s Getting Globalized? The Size and Implications of Intra-national Trade Costs.” NBER working paper no. 21439, National Bureau of Economic Research, Inc., July 2015.

Banerjee, Abhijit, Esther Duflo, and Nancy Qian. “On the Road: Access to Transportation Infrastructure and Economic Growth in China.” NBER working paper no. 17897, National Bureau of Economic Research, Inc., March 2012.

BCEOM and Inter-Consult, Ltd. “Consultancy Services for Detailed Engineering Design and Preparation of Tender Documents for Upgrading of Songea-Namtumbo-Tunduru Road (264 km) to Bitumen Standard: Draft Final Engineering Design Report.” Dar es Salaam, Tanzania: BCEOM and Inter-Consult, Ltd., 2008.

Bennett, C.R., and W.D.O. Paterson. “Guide to Calibration and Adaption of HDM-4.” Highway Development and Management Series, vol. 5, International Study of Highway Development and Management. Paris: World Roads Association, 2000.

Burningham, Sally, and Natalya Stankevich. “Why Road Maintenance Is Important and How to Get It Done.” Transport note no. TRN-4. Washington, DC: World Bank, 2005. Available at http://siteresources.worldbank.org/INTTRANSPORT/Resources/336291-1227561426235/5611053-1231943010251/TRN4_Road_Maintenance.pdf. Accessed March 2017.

Casaburi, Lorenzo, and Tristan Reed. “Incomplete Pass-Through and Interlinked Transactions.” Working paper F-6014-SLE-1. London, United Kingdom: International Growth Center, 2013. Available at http://www.theigc.org/wp-content/uploads/2012/04/Casaburi-Reed-2013-Working-Paper.pdf. Accessed November 2016.

Damsere-Derry, James, Gavan Palk, and Mark King. “Prevalence of Alcohol-Impaired Driving and Riding in Northern Ghana.” *Traffic Injury Prevention*, vol. 17, no. 3, 2016, pp. 226–232.

Dercon, Stefan, Daniel Gilligan, John Hoddinott, and Tassew Woldehanna. "The Impact of Roads and Agricultural Extension on Consumption Growth and Poverty in Fifteen Ethiopian Villages." CSAE working paper no. WPS 2007-01, Oxford, UK: Department of Economics, University of Oxford, May 2006.

Donaldson, Dave, and Richard Hornbeck "Railroads and American Economic Growth: A 'Market Access' Approach." *Quarterly Journal of Economics*, vol. 131, no. 2, pp. 799–858.

Donaldson, Dave. "Railroads of the Raj: Estimating the Impact of Transportation Infrastructure." *American Economic Review*, forthcoming, 2017.

Dr. Ahmed Abdel Warith Consulting Engineers and Dataconsult Ltd. (Warith and Dataconsult). "Detailed Engineering Design for the Upgrading of Tunduma – Sumbawanga Road Package 1: Tunduma – Laela Section (128km) to Bituminous Standard: Final Design Report." Dar es Salaam, Tanzania: Dr. Ahmed Abdel Warith Consulting Engineers and Dataconsult Ltd., 2007.

Economic Development Initiatives (EDI). "Consultancy Services for the Design and Implementation of Household Survey and Community Profile for Transport Sector." Final report submitted to Millennium Challenge Corporation. Kagera, Tanzania: EDI, 2009.

Escobal, Javier, and Carmen Ponce. "The Benefits of Rural Roads: Enhancing Income Opportunities for the Rural Poor." GRADE working paper no. 40, Lima, Peru: Grupo de Análisis para el Desarrollo (GRADE), November 2004

Fogel, Robert W. *Railroads and American Economic Growth: Essays in Econometric History*. Baltimore, MD: Johns Hopkins University Press, 1964.

Fortson, Kenneth, Randall Blair, and Kathryn Gonzalez. **"**Evaluation of a Rural Road Rehabilitation Project in Armenia." Washington, DC: Mathematica Policy Research, March 2015.

Foster, Vivien, and Cecilia Briceño-Garmendia. "Africa's Infrastructure: A Time for Transformation." Washington, DC: World Bank Publications, 2010.

Gollin, Douglas, and Richard Rogerson. "Productivity, Transport Costs and Subsistence Agriculture." *Journal of Development Economics*, vol. 107, 2014, pp. 38–48.

Gonzalez-Navarro, Marco, and Climent Quintana-Domeque. "Paving Streets for the Poor: Experimental Analysis of Infrastructure Effects." *The Review of Economics and Statistics*, vol. 98, no. 2, May 2015, pp. 254–267.

Government of Tanzania (GoT). "Agricultural Sector Development Strategy – II: 2015/2016 – 2024/2025." Dar es Salaam: GoT, 2016. Available at http://www.tzdpg.or.tz/fileadmin/documents/external/national_development_frameworks/ASDP2_Final_Document_20_May._2016__after_edit__1_.pdf. Accessed June 2017.

Guy, B.P., and J.D. Fricker. "*Guidelines for Data Collection Techniques and Methods for Roadside Station Origin-Destination Studies*." Publication FHWA/IN/JTRP-2005/27. Joint Transportation Research Program, Indiana Department of Transportation and Purdue University, West Lafayette, Indiana, 2005. Available at https://doi.org/10.5703/1288284313368. Accessed March 2017.

Gwilliam, Ken, Vivien Foster, Rodrigo Archondo-Callao, Cecilia Briceño-Garmendia, Alberto Nogales, and Kavita Sethi. "The Burden of Maintenance: Roads in Sub-Saharan Africa." AICD background paper 14. Washington, DC: The World Bank, 2009.

Hajek, Jaroslav J. "Optimal Sample Size of Roadside-Interview Origin-Destination Surveys." No. RR 208. Ontario, Canada: Ontario Ministry of Transportation & Communication, 1977.

Harral, Clell, Graham Smith, and William Paterson. "Maintaining Road Assets: A Fresh Look at The World Bank's 1988 Policy Paper 'Road Deterioration in Developing Countries.'" Washington, DC: World Bank, 2011. Available at http://documents.worldbank.org/curated/en/675111468152723420/Maintaining-road-assets-a-fresh-look-at-the-World-Banks-1988-policy-paper-Road-Deterioration-in-Developing-Countries. Accessed March 2017.

Heggie, Ian. "Management and Financing of Roads: An Agenda for Reform." World Bank technical paper no. 275. Washington, DC: The World Bank, 1994. Available at http://www.ssatp.org/sites/ssatp/files/publications/WorldBank-TechnicalPapers/TP275/TP275.pdf. Accessed March 2017.

Iimi, Atsushi, Eric Lancelot, Isabela Manelici, and Satoshi Ogita. "Social and Economic Impacts of Rural Road Improvements in the State of Tocantins, Brazil." Policy Research Working Paper no. WPS 7249. Impact Evaluation Series. Washington, DC: World Bank Group, April, 2015.

Jalan, Jyotsna, and Martin Ravallion. "Geographic Poverty Traps? A Micro Model of Consumption Growth in Rural China." *Journal of Applied Econometrics*, vol.17, no. 4, 2002, pp. 329–346.

Khandker, Shahidur R., Zaid Bakht, and Gayatri B. Koolwal. "The Poverty Impact of Rural Roads: Evidence from Bangladesh," *Economic Development and Cultural Change*, vol. 57, no. 4, 2009, pp. 685–722.

Lokshin, Michael, and Ruslan Yemtsov. "Has Rural Infrastructure Rehabilitation in Georgia Helped the Poor?" *The World Bank Economic Review*, vol. 19, no. 2, 2005, pp. 311–333.

Millennium Challenge Account-Tanzania (MCA-T). "Monitoring and Evaluation Plan (Final)." Dar es Salaam, Tanzania: MCA-T, 2013.

Millennium Challenge Account-Tanzania. "Tanzania COMPACT/Government Completion Report 2008–2015." Dar es Salaam, Tanzania: MCA-T, 2015.

Millennium Challenge Corporation (MCC). "Tanzania: Mainland Trunk Roads ERR Workbook." Washington, DC: The Millennium Challenge Corporation, 2007. Available at http://microdata.worldbank.org/index.php/catalog/2304#metadata-version. Accessed November 2016.

Millennium Challenge Corporation. "Tanzania Compact." Washington, DC: Millennium Challenge Corporation, n.d. Available at https://www.mcc.gov/where-we-work/program/tanzania-compact. Accessed June 6, 2017.

Millennium Challenge Corporation. "Principles into Practice on MCC Road Investments in the First 10 Years" Washington, DC: MCC, forthcoming.

Morosiuk, Greg, Mike Riley, and Tyrone Toole. "Applications Guide." *Highway Development and Management Series*, vol. 2, International Study of Highway Development and Management. Paris: World Roads Association, 2006.

Mu, Ren, and Dominique van de Walle. "Rural Roads and Market Development in Vietnam," *Journal of Development Studies*, vol. 47, no. 5, 2011, pp. 709–734.

Muzira, Stephen, and Damaris Hernandez de Diaz. "Rethinking Infrastructure Delivery: Case Study of a Green, Inclusive, and Cost-effective Road Program in Nicaragua." Washington DC: World Bank Group, 2014.

Mwaipungu, R.R., and D. Allopi. "Review of Sub-Saharan African Gravel Roads Management System: Tanzanian Case Study." *WIT Transactions on The Built Environment*, vol. 128, 2012, pp. 629–640.

National Opinion Research Center (NORC). "Impact Evaluation of the Transportation Project in Honduras." Final report submitted to Millennium Challenge Corporation. Bethesda, MD. December 2013a.

National Opinion Research Center. "Samtskhe-Javakheti Roads Activity Impact Evaluation." Final report submitted to Millennium Challenge Corporation. Bethesda, MD, January 2013b.

Odoki, J.B., and H.R. Kerali. "Analytical Framework and Model Descriptions." *Highway Development and Management Series*, vol. 4, International Study of Highway Development and Management. Paris: World Road Association, 2000.

Paterson, W., and Scullion, T. "Information Systems for Road Management: Draft Guidelines on System Design and Data Issues." Report No. INU 77. Washington, DC: World Bank Infrastructure and Urban Development Department, September 1990. Available at http://documents.worldbank.org/curated/en/196321468762908116/pdf/multi-page.pdf. Accessed July 26, 2017.

Pinard, M.I. "Progress on Commercialized Road Management in Sub-Saharan Africa." The Sub-Saharan Africa Transportation Policy Program (SSATP) working paper no. 92. Washington, DC: World Bank, 2012.

Pinard, Michael Ian. "Road Management Policy: An Approach to the Evaluation of Road Agency Performance." SSATP working paper no. 95605. Washington, DC: World Bank, 2015.

Queiroz, Cesar. "Financing of Road Infrastructure." In *Proceedings of the 5th Symposium on Strait Crossings*, pp. 45–57. Trondheim, Norway: 2009.

Robinson, R., and B. Thagesen, eds. *Road Engineering for Development*. London: Spon Press, 2004. Available at http://sutlib2.sut.ac.th/sut_contents/H100769.pdf. Accessed March 2017.

Schroeder, Larry D. "Managing and Financing Rural Road Maintenance in Developing Countries." Burlington, VT: Associates in Rural Development, Incorporated, 1990. Available at http://pdf.usaid.gov/pdf_docs/PNACB393.pdf. Accessed May 2017.

Storeygard, Adam. "Farther on Down the Road: Transport Costs, Trade and Urban Growth in Sub-Saharan Africa." *The Review of Economic Studies,* vol. 83, no. 3, 2016. pp. 1263-1295.

Teravaninthorn, Supee, and Gael Raballand. "Transport Prices and Costs in Africa: A Review of the International Corridors." World Bank, 2009

TRL. *Overseas Road Note 31 (4th Edition): A Guide to the Structural Design of Bitumen-Surfaced Roads in Tropical and Sub-Tropical Countries.* Crowthrone, Berkshire UK: TRL, 1993.

TRL and DFID. *Overseas Road Note 18: A Guide to the Structural Design of Bitumen-Surfaced Roads in Tropical and Sub-Tropical Countries,* Crowthrone, Berkshire UK: TRL, 1999.The World Bank. "Poverty and Shared Prosperity 2016: Taking on Inequality." Washington, DC: World Bank, 2016.

U.K. Department of Transport (UK DOT). *Design Manual for Roads and Bridges, Volume 5: Assessment and Preparation of Road Schemes.* London, UK: U.K. DOT, 2009. Available at http://www.standardsforhighways.co.uk/ha/standards/dmrb/vol5/index.htm. Accessed May 2017.

UWP Consulting (Pty) Limited South Africa (UWP). "Environmental Impact Assessment, Detailed Engineering Design and Preparation of Tender Documents for Upgrading of Peramiho Junction – Mbinga – Mbamba Bay Road (144 km) to Bitumen Standard." Johannesburg, South Africa: UWP Consulting, 2008.

Walker, Donald, Lynn Entine and Susan Kummer. *Asphalt Airfield Pavements: PASER Manual.* Washington, DC: the FAA, 2014. Available at: https://www.faa.gov/documentLibrary/media/Advisory_Circular/150-5320-17a-appendix-a.pdf. Accessed April 26, 2018.

This page has been left blank for double-sided copying.

APPENDIX A:

SUMMARY OF STUDIES IN THE LITERATURE REVIEW

This page has been left blank for double-sided copying.

Table A.1. Studies examining impacts on beneficiaries close to improved roads

Study	Country	Study design [a]	Short description
MCC studies			
NORC (2013)	Honduras	Matching with continuous treatment effect	Evaluates the impact of upgrading 65 km of secondary roads and 495 km of rural roads one to two years after project completion. Results include reduced travel times to health centers and municipal capital; reduced travel costs to hospitals, health centers, and basic markets. No evidence of impact on utilization of social services and on employment or school attendance.
Alevy (2014)	Nicaragua	DiD with matched comparison group	Evaluates the impact of upgrading secondary roads totaling 68 km in length one year after project completion. Does not find evidence of impact on the cost and availability of a basket of basic consumer goods, but finds increases in household consumption of perishable goods.
NORC (2013)	Georgia	DiD with matched comparison group	Evaluates the impact of rehabilitating 220 km of a regional highway connecting the economically isolated Samtskhe–Javakheti region to the capital one year after road completion. Finds increased traffic and average speed of vehicles and decreased travel times to the capital and local markets. Also increases in the number of industrial facilities along the improved road segments. No evidence of impact on land use or cropping patterns, utilization of health care centers, household income, consumption, or asset ownership.
Fortson et al. (2015)	Armenia	DiD with comparison group	Evaluates the impact of 27 rural road segments originally intended to be upgraded under the MCC compact and later financed by the World Bank after one to two years of exposure. Finds that the project decreased travel times by households to sell agricultural products and increased use of roads for noncommercial purposes (e.g., shopping and visiting relatives). Does not find impacts on utilization of social services, agricultural production, household income, consumption, and poverty.
Non-MCC Studies			
Iimi et al. (2015)	Brazil	DiD with matched comparison group	Evaluates the impact of roads improvements and related investments to allow year-round passage in the four poorest regions in the state of Tocantins one year after completion. Finds reduced travel times to population centers and increased use of public transportation, car ownership, and perceived access to health centers and schools when available in close proximity. Impact on girls' enrollment in school varied by region. No impact on labor composition across different sectors or household income.
Khandker (2009)	Bangladesh	Household fixed-effects	Evaluates the impact of two projects rehabilitating rural roads and market improvement three to four years after completion. Finds decreased travel costs, increased school enrollment and employment, lower fertilizer prices, and increased aggregate crop prices, agricultural output, and per-capita household expenditure.

Study	Country	Study design [a]	Short description
Non-MCC Studies (*continued*)			
Mu and van de Walle (2011)	Vietnam	DiD with matched comparison group	Evaluates the impact of a rural road rehabilitation project in 18 provinces two years after completion. Finds increases in the presence and availability of local open-air temporary markets for goods and services, primary and secondary school enrollment, and off-farm activities (mostly in the service sector).
Lokshin and Yemtsov (2005)	Georgia	DiD with matched comparison group	Evaluate the impact of rural roads improvement projects in 41 villages one to two years after completion. Finds increases in the number of non-agricultural small and medium enterprises and reductions in barter trade. No evidence of impact on the sale of agricultural products.
Akee (2006)	Palau	DiD with comparison group	Evaluates the impact of a newly built road connecting rural areas to more urban areas one year after project completion. Finds increases in non-agricultural wage-sector employment, reduced self-employment in agriculture, and decreases in the number of immigrants sent abroad. Finds increased car ownership but no impacts on household income or wages.
Gonzalez-Navarro et al. (2015)	Mexico	RCT	Evaluates the impact of upgrading feeder roads connecting peri-urban residential neighborhoods to larger road networks one year after project completion. Finds increases in property values along the newly paved roads, vehicle and appliance ownership, and home improvements.
Escobal and Ponce (2004)	Peru	Matched comparison group	Evaluates the impact of rehabilitating rural roads in high-poverty districts four years after the start of the project. Finds increases in household income, mainly from increases in non-agricultural wage employment. Finds no evidence of impacts on annual consumption from increased income, but finds increased investments in durable goods, namely livestock.
Dercon et al. (2007) [b]	Ethiopia	Household fixed-effects	Evaluates the impact of road improvements on consumption and poverty in rural areas. Finds increases in consumption and reduction in the incidence of poverty.
Jalan and Ravallion (2002) [b]	China	Dynamic growth modelling	Evaluates the impacts of roads constriction on consumption, assuming that initial road placements are exogenous. Finds that higher density of roads (kilometers per 10,000 people) positively correlates with higher rates of consumption.

[a] DiD = Difference-in-Difference; RCT = Randomized-controlled trial

[b] Study did not report exposure period

Table A.2. Studies examining broader economic effects

Study	Country	Study design	Short description
Donaldson (2017)	India	General equilibrium trade model	Examines the impact on income of railroad network construction in India in the 19th and early 20th century. Guided by the predictions from a general equilibrium trade model and using rich archival data, the study finds that access to the network reduced trade costs and interregional price gaps, increased interregional and international trade, and improved income.
Atkin and Donaldson (2015)	Ethiopia and Nigeria	Theoretical model	Investigates the cost of intranational trade for consumers in remote locations resulting from poor transportation networks and imperfect transport market. Finds that the effects of long distance on the cost of intranational trade are approximately four to five times larger in these two countries than in the U.S. Also finds that intermediaries are able to capture a greater fraction of the benefits of price reductions, leaving little to pass on to the consumers in communities in remote locations.
Storeygard (2016)	Fifteen Sub-Saharan countries	Natural experiment (exogenous changes in oil prices)	Examines the role of transport costs in determining income of cities in 15 Sub-Saharan countries where the largest city is a port. Using satellite-based nighttime lights data to measure economic activity, the study finds that exogenous changes in transport costs, caused by world oil price fluctuations, led to increases in the income of cities near the port compared to similar cities 500 km away.
Donaldson and Hornbeck (2015)	U.S.	General equilibrium trade model	Examines the impacts of the expansion of the U.S. railroad network on the agricultural sector in 1890. Uses historical county-level data on agricultural land value combined with satellite data on railroad expansion to identify a counterfactual scenario where no railroad network existed. Finds that railroad access led to moderate increases in U.S. gross national product.
Banerjee et al. (2012)	China	Natural experiment (exogenous variation in distance from straight lines connecting historical cities)	Examines the long-term impacts of access to transportation network on income and growth across different sectors in China during a period of rapid economic growth. Finds that regions closer to transportation networks have higher levels of GDP per capita, income inequality, number of firms, and average firm profits, but the level differences are small in magnitude. Also finds no difference in income growth between regions closer and farther from transportation network during the two decades of rapid economic growth between 1986 and 2006.
Asher and Novosad (2016)	India	Regression discontinuity	Examines the impacts of a national rural road construction program that connected isolated villages to the regional transportation network. Finds that the program led to large movements of labor across sectors—decreasing the share of households and workers participating in agriculture and increasing the share participating in wage labor—and also decreased agricultural production and increased income from wage labor.
Gollin and Rogerson (2014)	Uganda	General equilibrium model	Examines the reasons a large fraction of the workforce in poor economies live in rural areas and engage in subsistence agriculture, using a multisector multiregion general equilibrium model. Finds that a closed economy with lower agricultural productivity needs more people to engage in subsistence farming to meet food requirements through domestic supply, and high transportation costs are an important contributor for households to locate near the source of production. Concludes that productivity increases have little impact on resource reallocation in remote areas because high transportation costs prevent households from selling increased output.

Study	Country	Study design	Short description
Casaburi et al. (2013)	Sierra Leone	Regression discontinuity	Examines the impact of a rural road rehabilitation program on crop prices in Sierra Leone. Finds that the program lowered the price of rice and cassava, more so in areas further away from large urban centers and with lower productivity, likely because improved road quality decreases the cost of reaching local markets where farmers and traders conduct business. Also finds that the effects on prices were weaker in areas with better cell phone coverage because reductions in travel times were less relevant for farmers and traders who could negotiate prices over phones.

APPENDIX B:

LIST OF VARIABLES FOR HDM-4 ANALYSIS

This page has been left blank for double-sided copying.

Table B.1. HDM-4 variables and data sources

Variable description	Likelihood of calibration	Data source
Climate data		
Average rainfall	Calibrate	Secondary
Duration of dry season	Calibrate	Secondary
Mean temperature	Calibrate	Secondary
Moisture index	Calibrate	Secondary
Number of days temperature exceeds 32 C	Calibrate	Secondary
Temperature range	Calibrate	Secondary
Pavement characteristics, road inventory and road works		
Altitude	Calibrate	As-built drawings, roads agency
Carriageway width	Calibrate	As-built drawings, roads agency
Construction age	Calibrate	As-built drawings, roads agency
Effective number of lanes	Calibrate	As-built drawings, roads agency
Horizontal curvature	Calibrate	As-built drawings, roads agency
Number of surface layers	Calibrate	As-built drawings, roads agency
Posted speed limit	Calibrate	As-built drawings, roads agency
Preventative treatment age	Calibrate	As-built drawings, roads agency
Rise plus fall	Calibrate	As-built drawings, roads agency
Shoulder width	Calibrate	As-built drawings, roads agency
Superelevation	Calibrate	As-built drawings, roads agency
Surface type	Calibrate	As-built drawings, roads agency
Surfacing age	Calibrate	As-built drawings, roads agency
Area potholed	Calibrate	Engineering survey
Base type	Calibrate	Engineering survey
Drainage condition	Calibrate	Engineering survey
Mean rut depth	Calibrate	Engineering survey
Environmental Coefficient of Roughness	Calibrate	Secondary
Subgrade California Bearing Ratio (CBR)	Potentially calibrate	As-built drawings completion reports
Structural number	Potentially calibrate	As-built drawings, completion reports
Drainage factor	Potentially calibrate	Secondary
Area with all cracking	Potentially calibrate	Engineering survey
Area with wide cracking	Potentially calibrate	Engineering survey
Crack initiation factor	Use default values	Secondary
Crack progression factor	Use default values	Secondary
Pothole progression	Use default values	Secondary
Ravelling initiation	Use default values	Secondary
Ravelling progression	Use default values	Secondary
Roughness age term	Use default values	Secondary
Roughness progression	Use default values	Secondary
Rut depth progression	Use default values	Secondary
Sand patch texture depth	Use default values	Secondary
Seasonal effects on Structural No	Use default values	Secondary
Skid resistance	Use default values	Secondary
Roughness	Calibrate	Engineering survey
Time lapses to patching	Potentially calibrate	Secondary

Variable description	Likelihood of calibration	Data source
Project finance data		
Analysis period	Calibrate	Model assumption
Discount rate	Calibrate	Model assumption
Unit cost for construction	Calibrate	Project documents
Interest rate	Calibrate	Secondary
Road user data		
Cost of fuel	Calibrate	Vehicle operator survey
Cost of maintenance labor	Calibrate	Vehicle operator survey
Cost of oil	Calibrate	Vehicle operator survey
Cost of overhead	Calibrate	Vehicle operator survey
Cost of retreaded tire	Calibrate	Vehicle operator survey
Cost of tire	Calibrate	Vehicle operator survey
Cost of vehicle/price	Calibrate	Vehicle operator survey
NMT Variables	Calibrate	Vehicle operator survey, traffic counts
Equivalent standard axles	Potentially calibrate	Axle load survey
Annual loading	Potentially calibrate	Transport sector reports, road agency
Cost of travel time/cost of passenger work time	Potentially calibrate	Transport sector reports, road agency
Cost of cargo	Potentially calibrate	Vehicle operator survey
Cost of crew	Potentially calibrate	Vehicle operator survey
Average service life	Use default values	Transport sector reports, road agency
Base number of retreads - NR0	Use default values	Transport sector reports, road agency
Desired speed	Use default values	Transport sector reports, road agency
Engine speed - a0	Use default values	Transport sector reports, road agency
Engine speed - a1	Use default values	Transport sector reports, road agency
Engine speed - a2	Use default values	Transport sector reports, road agency
Engine speed - a3	Use default values	Transport sector reports, road agency
Engine speed - Idle	Use default values	Transport sector reports, road agency
Number of axles	Use default values	Transport sector reports, road agency
Number of wheels	Use default values	Transport sector reports, road agency
Operating weight	Use default values	Transport sector reports, road agency
Optimal life depreciation parameters	Use default values	Transport sector reports, road agency
Percentage of private use	Use default values	Transport sector reports, road agency
Power - braking	Use default values	Transport sector reports, road agency
Power - driving	Use default values	Transport sector reports, road agency
Power - rated	Use default values	Transport sector reports, road agency
Projected frontal area	Use default values	Transport sector reports, road agency
Time driven on wet roads	Use default values	Transport sector reports, road agency
Travel on snow covered roads	Use default values	Transport sector reports, road agency
Travel on wet roads	Use default values	Transport sector reports, road agency
Tyre type	Use default values	Transport sector reports, road agency
Wheel diameter	Use default values	Transport sector reports, road agency
Aerodynamic drag coefficient	Use default values	Vehicle operator survey
Aerodynamic drag coefficient multiplier	Use default values	Vehicle operator survey
Average annual utilisation	Use default values	Vehicle operator survey
Hours driven	Use default values	Vehicle operator survey

Variable description	Likelihood of calibration	Data source
Traffic data		
Annual average daily traffic	Calibrate	Traffic counts
Hourly distribution of traffic	Calibrate	Traffic counts
Traffic growth rate	Calibrate	Traffic counts/secondary

This page has been left blank for double-sided copying.

APPENDIX C:

SAMPLE SIZE ESTIMATION METHODOLOGY FOR THE EXPANDED VEHICLE INTERCEPT SURVEY

This page has been left blank for double-sided copying.

A. Formula

The U.K. Department of Transport (U.K. DOT) Design Manual for Roads and Bridges includes the following formula for the estimation of appropriate sample sizes (S) for origin-destination surveys (U.K. DOT 2009; Guy and Fricker 2005).

$$S = \frac{P(1-P)Q^3}{\left(\frac{E}{1.96}\right)^2 (Q-1) + P(1-P)Q^2}, \text{ where}$$

Q = total traffic flow, in number of vehicles

E = the acceptable error or accuracy (expressed as the number of vehicles); and

P = the proportion above which it is unlikely that the proportion of flow that is of interest falls (that is, if we wished to estimate the number of vehicles traveling to destinations outside Tanzania, and believed that no more than 10 percent of traffic was traveling to international destinations, P = 0.10).

Solving the equation for E, in order to calculate the sampling error for a given sample size yields:

$$E = 1.96 * \sqrt{\frac{\left(\frac{P(1-P)Q^3}{S}\right) - P(1-P)Q^2}{Q-1}}$$

B. Parameter estimates

To generate pre-implementation economic rates of return (ERRs) for the mainland roads activity, MCC estimated average daily traffic (AADT) for each trunk road through 2027 (the time horizon of the ERR estimates) by applying a one-time increase and annual growth rates to the estimates of AADT from the feasibility stage. The growth rates were based on data from other, similar roads in Tanzania that had been upgraded (MCC 2007). We used the resulting AADTs for 2017 as an estimate of Q for each trunk road.

For S, we used 20 percent of these AADT estimates, which is a proportion of traffic that falls between the 10 percent bidirectional interviewing that is considered adequate for descriptive studies and the 50 percent that is recommended for surveys that are used to predict future travel patterns (Hajek 1977). Moreover, this is a proportion of traffic flow that has been successfully implemented in Sub-Saharan Africa (Damsere-Derry et al. 2016).

To identify plausible values for P for a range of outcomes, we reviewed pre-implementation feasibility studies and design reports (BCEOM and Inter-Consult 2008; Africon Ltd. 2007; Warith and Dataconsult. 2007; UWP 2008). This review yielded the results of baseline origin destination surveys for only one of the trunk road segments: Songea-Namtumbo. O-D surveys were conducted at two sites along the roads that were ultimately selected for upgrading and at Gumbiro, which is 50 km northwest of Songea and located along a trunk road that was paved prior to project implementation.

We relied on the estimates from Gumbiro for two main reasons:

1. We anticipate that data from a previously paved road is more representative of traffic on the MCC-financed roads after they were upgraded than the unpaved roads later upgraded by MCC.
2. Data from Gumbiro covers a wider range of outcomes of interest than the surveys conducted along the MCC roads, including more disaggregated O-D data and more specific information on the contents of trucks.

Using the detailed origin-destination matrix collected from Gumbiro, we estimated mutually exclusive proportions of intraregional traffic, interregional traffic excluding traffic to/from Dar es Salaam, to/from Dar es Salaam, and to/from Zambia. In addition, we estimated the proportion of vehicles that were trucks, carrying agricultural inputs/outputs, and traveling for noncommercial purposes (for example, school, shopping, or medical purposes). This analysis yielded the following estimates (Table C.1).

Table C.1. Proportion of vehicles by O-D and purpose of travel

Outcome	Proportion of AADT (%)
Origin-destination outcomes	
Intra-regional traffic	16
Traffic to/from Dar es Salaam	33
Traffic to/from other regions	51
International traffic	<1
Other outcomes	
Vehicles traveling for school, shopping, or medical purposes	11
Vehicles that are trucks	36
Trucks carrying agricultural inputs/outputs	25

Source: Authors’ calculations based on BCEOM and Inter-Consult Ltd. (2008).

No baseline data were available to carry out the same calculations on the Zanzibar rural roads.

C. Sampling error estimation

We used these estimates of P and the AADT of the different trunk roads to estimate E for each outcome and trunk road. This results in estimated sampling errors across all trunk roads of between 2 and 4 percent of AADT for O-D outcomes and between 1 and 6 percent of AADT for other outcomes when sampling 20 percent of all vehicles (Table C.2).

Table C.2. Estimated sampling error by trunk road and outcome

Outcome	Type of vehicle	Projected AADT (Q)	Sampling rate (%)	Average surveys per day	Number of days	Estimated sample (S)	Number of vehicles	Estimated sampling error (E) (% of AADT)
Mtwara corridor								
Origin-destination								
Intraregional traffic	All	385	20	77	3	231	11.5	3.0
Traffic to/from Dar es Salaam	All	385	20	77	3	231	14.8	3.8
Traffic to/from other regions	All	385	20	77	3	231	15.7	4.1
Other outcomes								
Vehicles traveling for school, shopping, or medical purposes	All	385	20	77	3	231	9.9	2.6
Traffic carrying agricultural inputs/outputs	Trucks only	138	20	28	3	83	8.2	5.9
Tanga-Horohoro								
Origin-destination outcomes								
Intraregional traffic	All	1265	20	253	3	759	20.9	1.6
Traffic to/from Dar es Salaam	All	1265	20	253	3	759	26.8	2.1
Traffic to/from other regions	All	1265	20	253	3	759	28.5	2.3
Other outcomes								
Vehicles traveling for school, shopping, or medical purposes	All	1265	20	253	3	759	17.8	1.4
Traffic carrying agricultural inputs/outputs	Trucks only	453	20	91	3	272	14.7	3.3
Tunduma Sumbawanga								
Origin-destination outcomes								
Intraregional traffic	All	1124	20	225	3	675	19.7	1.8
Traffic to/from Dar es Salaam	All	1124	20	225	3	675	25.3	2.2
Traffic to/from Dar es Salaam	All	1124	20	225	3	675	26.8	2.4
Other outcomes								
Vehicles traveling for school, shopping, or medical purposes	All	1124	20	225	3	675	16.8	1.5
Traffic carrying agricultural inputs/outputs	Trucks only	402	20	80	3	241	13.9	3.5

Source: Authors' calculations based on BCEOM and Inter-Consult Ltd. 2008; MCC 2007.

Notes: Traffic flows to/from Dar es Salaam, share of traffic that are trucks, and share of trucks carrying agricultural inputs/outputs are based on a seven-day O-D survey conducted as part of a pre-implementation design report by BCEOM and Inter-Consult Ltd (2008).

This page has been left blank for double-sided copying.

APPENDIX D:

DATA COLLECTION COMPARISON TABLE

This page has been left blank for double-sided copying.

The tables below summarizes the data collection efforts proposed for the Tanzania Transport project evaluation and how these differ from the specifications included in annex 4- Data Collection Quality Protocols of the contract.

Table D.1. Road Data Collection Comparison

Measurement	Standards included in MCC data collection protocol (Annex 4)	Deviations between MPR data collection plans and annex 4, and rationale for deviation
IRI	• Outer wheel path of each lane per relevant ASTM standards using a Class 3 or better IRI measuring device.	• No deviation
Deflection	• Measurements on all selected roads during or at the end of the rainy season so as to obtain the modulus of every pavement layer and subgrade. • Measurements at intervals of no more than 100 meters for the entire chainage during both rainy and dry seasons. • Deflection measurement shall be taken by the FWD for all bitumen surfaced roads. • Based on the deflection results and traffic counts (see below), the Contractor shall determine the remaining structural life of the roads investment using an appropriate fatigue curve to be identified by the Contractor and approved by MCC before use.	• After conducting a sensitivity analysis that assessed the impact of various structural numbers on the ERR, we recommend not collecting deflection data for this evaluation.
Geotechnical	• Determine the Adjusted Structural Number (SNC) based on the As-Built drawings and geotechnical results and compare to the thicknesses checked with a Ground Penetrating Radar (GPR) with GPS capability for a thorough verification of the thickness of the built road vis a vis the design. • Based on the geotechnical results provided in the As-built drawings, the Contractor shall determine the subgrade modulus and its resulting California Bearing Ratio (CBR), the modulus of every layer and Adjusted Structural Number for use in HDM-4.	• The Adjusted Structural Number (SNC) will be based on the As-Built drawings.
Road condition	• Evaluate the road condition, namely any distress encountered (cracking, bleeding, raveling, rutting, potholing, etc.) in accordance with the LTPP Distress Identification Manual (June 2003) or other relevant and approved methodology and determine the cause(s) of deterioration. • Any maintenance operation(s) performed since construction should also be identified. Note all performed measurements will also need to be shown in accordance with HDM4 data input requirements (this will require a calibration between the LTPP and HDM4 cracking values for example.) • The Contractor shall graphically illustrate the distress(s) found on the aerial imagery collected for the entire chainage, with appropriate LTPP distress identification and severity shown on the aerial image(s). For example, fatigue cracking of moderate severity should be shown on the image in the location on the aerial image with coloring to show moderate severity. MCC suggests using Red for High, Yellow for Moderate and Green for Low.	• We will pursue a sampling approach following World Bank International Quality Level standards based on Bennett and Paterson (2000) and TRL Road Note 18 (TRL and DFID 1999) for inspection of road condition. We will adjust the frequency based on visual observations of road conditions. In roads that show more distress, we will oversample. • Data will be collected during dry season. We will graphically illustrate road condition on the itinerary diagram. We will be collecting GPS linked video capture.

Measurement	Standards included in MCC data collection protocol (Annex 4)	Deviations between MPR data collection plans and annex 4, and rationale for deviation
Axle load	• The traffic evaluation shall distinguish between domestic and international traffic and shall consider the degree of vehicle overloading for the purpose of estimating the truck factor (average ESALs perheavy vehicle) to be used. Each road shall be measured over a period of one week from 6am to 8pm. • The stations will be integrated into the aerial imagery and itinerary diagrams.	• We will collect axle load along each road on mainland Tanzania. Because feeder roads of Pemba have limited truck traffic, we will include 1 axle load station. • We will measure axle load for a duration of three days, but for a 12-hour period in order to avoid working during night hours
Expanded vehicle intercept- O/D module &road user module	• The surveys will be conducted for two consecutive market and non-market days at each site from 6am to 8pm. The stations should not be located near urban areas so as to avoid inclusion of local traffic in the survey. The primary elements to be collected are the 1) Origin and Destination, journey purpose, travel time, vehicle classification, passengers per vehicle, number of passengers in employment, number of crew, type and approximate weight of merchandise or goods transported. • A minimal interview sample rate of 20% of each vehicle type will have to be achieved at each site. • The stations will be integrated into the aerial imagery and itinerary diagrams.	• The surveys will be conducted during daylight hours for three days, one of which will be a market day. • Interviews will be conducted with a representative sample of traffic over the course of the day and we will not oversample to ensure a 20% sampling rate by vehicle type.
Vehicle operating cost	• The survey shall be carried out at major transport operators, travel companies, garages, and vehicle concessionaires and on a sample of private road users from the O-D work and in accordance with HDM4 Volume 5 requirements.	• No deviation
Traffic counts	• Develop a traffic counting procedure in accordance with the Traffic Monitoring Guide issued by the US Federal Highway Administration or equivalent. • The stations will be integrated into the aerial imagery and itinerary diagrams.	• Traffic counting procedures will be developed in accordance with TANROADS' Traffic Data Collection and Analysis standards[a] based on recommendations from local experts. TANROADS standards follow international standards. These standards were used at baseline and will allow for better comparability of results. • We will collect traffic count data during the wet and the dry season. • Local and international experts have recommended the following traffic count length: seven-day traffic count to include three-day, 24-hour counts and four-day, 12-hour counts, with one of the 24-hour count days being a regional market day.

Measurement	Standards included in MCC data collection protocol (Annex 4)	Deviations between MPR data collection plans and annex 4, and rationale for deviation
Satellite imagery	• The Contractor shall collect recent satellite imagery for pre-construction work at a resolution of 50cm or better and aerial imagery at a resolution of 5cm or better for the constructed works of each road and overlay the collected data.	• We are not proposing to collect aerial imagery because we will collect GPS-linked video, mitigating the need for 5 cm or better aerial imagery. • We will use satellite imagery in the impact evaluation (30m Landsat) and are also collecting video imagery of the roads.

[a] https://www.vegvesen.no/_attachment/336339/binary/585485 .

Table D.2. Airport Runway Data Collection Comparison

Measurement	Standards included in MCC data collection protocol (Annex 4)	Mathematica proposed data collection plans and rationale for deviation
IRI	The Contractor shall measure the IRI using a Class 3 or better IRI measuring device per relevant ASTM standards including equipment calibration on reference sections to provide appropriate precision and bias	In design discussions with MCC, we decided to focus data collection on the key components of the theory of change for the runway. These are increased traffic via the landing of larger aircraft. We will conduct a visual inspection of the runway condition and gather data on air traffic.
Deflection	The Contractor shall perform deflection measurements with a Falling Weight Deflectometer (FWD) during or at the end of the rainy season so as to obtain the modulus of every pavement layer. The deflection measurements shall be performed at 50 meter increments on each runway section in accordance with the relevant ASTM standards.	See above
Traffic	The Contractor shall provide updated air traffic counts by plane classification for each runway project including the verification of the assumptions made in the design of the runway and perform a structural design analysris based on the estimated layer moduli in order to determine the residual structural life.	Will obtain traffic counts from Tanzanian Airport Authority records.
Runway condition	The Contractor shall evaluate the runway condition namely any distress encountered and determine the cause(s) and associate to the pavement performance curve used in in order to update the maintenance and rehabilitation assumptions used in the lifecycle cost analysis	We will conduct a visual inspection of the runway using the PASER Asphalt Airfield Pavements Manual published by the FAA (Walker, Entine, and Kummer 2014).

APPENDIX E:

QUANTITATIVE DATA COLLECTION STANDARDS TABLE

This page has been left blank for double-sided copying.

The table below describes the required data standards for the quantitative data collection in the evaluation. This table will be part of the contract between Mathematica and the relevant data collection subcontractor.

Table E.1. Required data standards

Data Type	Sample unit	Sample size	Relevant instruments/ equipment	Data Standard	Data Processing	Anticipated Quality Control Protocols
Roughness	All roads	470 km of mainland trunk roads 35 km of Pemba rural roads	Laser Profilometer with GPS manufactured by Data Collection Limited (DCL) as part of the ROMDAS system If the Laser Profilometer is not available on Pemba, the subcontractor will use a Bump Integrator	Devices compliant with Class 3 device or better as per Word Bank Technical paper 46 ASTM Standards: Laser Profiler is a a Class I Inertial Profiler and conforms to ASTM-E950. Bump Integrator uses ASTM E1448 (calibration) and ASTM E1082 (test method)	Data will be processed using Road Measurement Data Acquisition System (ROMDAS).	• Roughness data linked to GPS readings. • Photographic evidence of field application. • Video data will be collected from the vehicle carrying the instrument to ensure that it is driven in the wheel tracks and at the designated speed (where safe to do so). **Calibration:** • Calibration of the bump integrator will be conducted following ASTM E1448. • Verification tests of the Laser Profilometer will be conducted before work begins at permanent calibration sites in Morogoro region or on a test section to establish the repeatability and bias of the IRI measurements produced by the instrument. • Calibration of the Laser Profiler will be conducted following ASTM E950 and the manufacturer’s instructions.

Data Type	Sample unit	Sample size	Relevant instruments/ equipment	Data Standard	Data Processing	Anticipated Quality Control Protocols
Surface condition	Road segment	Full length using analysis using video and photographs of road surface Sampled segments (10% of length) along all trunk and feeder roads for manual assessment	Manual inspection and video	Surface condition data will be collected in accordance with TRL Road Note 18 (TRL and DFID 1999) and complemented by the Tanzania Pavement and Materials Design Manual (1999). On sampled segments: the data collector shall walk along the road in order to identify and record the type, location, extent and severity of the various distresses present. Information Quality Level for surface condition data, will meet the IQL-2 requirements for HDM-4. The video data collection process meet IQL-3. Manual surface condition data will meet IQL-2.	Pavement segments will be classified into coarse condition measures using an analysis of video captured along the complete road length. Pavement distress identified during the manual surface condition assessment on a sample of road segments will be translated into indicators required for HDM-4	• GPS linked video capture to check random sections of road against manual classification and inspection of surface condition. • Spot checks by local engineering consultant. • Collective training of pavement raters to ensure that pavement condition is interpreted in a consistent and uniform way. • Periodic rater reviews to verify raters are collecting uniform and consistent data.

Data Type	Sample unit	Sample size	Relevant instruments/ equipment	Data Standard	Data Processing	Anticipated Quality Control Protocols
Axle load survey	Road segment	One per road segment on the mainland trunk roads and one axle load survey on Pemba island, 3 days for 12-hour counts. Traffic counts will be conducted alongside axle load surveys	Mobile weigh stations use portable axle weight pad or equivalent. Existing permanent TANROADS weigh stations will be used, where available.	Classification of vehicles used on paper forms will be based on standard TANROADs vehicle classification.	Individual Axle load measurements will be translated into 8-tonne Equivalent Standard Axle measurements	• Spot checks by local engineering consultant **Calibration:** • Documents certifying periodic calibration of mobile weigh stations and permanent weigh station will be requested from TANROADS • Before data collection, a truck with known weight will be used to calibrate the weigh station
Traffic Counts	Select census stations/road segment	A total of 13 stations in dry season and 7 stations in wet season; see Table IV.2; 7-day counts	Tablets or paper based forms	Classification of vehicles used on paper forms will be based on standard TANROADs vehicle classification.	Classified traffic counts will be converted to Annual Average Daily Traffic (AADT) using seasonal adjustment factors provided by TANROADs	• Training of enumerators. • Photographic evidence of traffic counts being conducted. • Spot checks by local engineering consultant.

Data Type	Sample unit	Sample size	Relevant instruments/ equipment	Data Standard	Data Processing	Anticipated Quality Control Protocols
Expanded vehicle intercept survey	Points along road segment, coinciding with location of traffic count stations. Sample of vehicles will be representative of traffic patterns during survey	Survey will be conducted for 3 consecutive days, including 1 market day. More days may be required for rural roads on Pemba.Surveys will be conducted from 6am to 8pm	Tablet-based data collection	Purpose-designed survey	N/A	• Pre-testing and piloting of data collection instruments. • Training and testing of enumerators. • Data quality checks based on uploaded data. • Data quality checks for 10% of the sample based on uploaded data.
Vehicle Operating Cost Survey	Vehicle operators, vendors and maintenance providers	5 major transport operators 5 garage /mechanics 5 freight companies 5 vehicle sellers – second hand and new. Coverage by vehicle type based on common vehicles sold and/or used.	Paper based survey	VOC forms based on HDM-4 input requirements	N/A	• Back checks by local consultant to verify contacts and information collected • Pilot surveys to test questions and responses

APPENDIX F:

SUMMARY OF HDM-4 SENSITIVITY TESTING

This page has been left blank for double-sided copying.

In order to assess the added value of collecting deflection data, we conduct sensitivity tests using an HDM-4 model that is developed using input data from the ex-ante HDM-4 models[22] as well as actual construction costs and, where available, traffic count data for 2013. The initial condition of the roads is based on reports in the feasibility studies, and we developed the base case (without project case) by estimating how the road agency would likely respond to the normal traffic observed on the roads.

Our sensitivity analysis runs the HDM-4 model for three road segments – one from each of the three mainland trunk roads – using different scenarios, each defined by a structural number. All other HDM-4 inputs are unchanged across scenarios. The scenarios are:

1. Minimum case (weakest pavement): Structural number = 2
2. Original design: Structural number varies by road segment depending on layer thickness and materials
3. Maximum case (strongest pavement): Structural number = 5

The rest of the memo provides background information on the roads and compiles key HDM-4 outputs for each run. We then summarize the findings and key lessons. The appendix includes more information about the input data extracted from the ex-ante HDM-4 model for the Mtwara corridor.

A. Background:

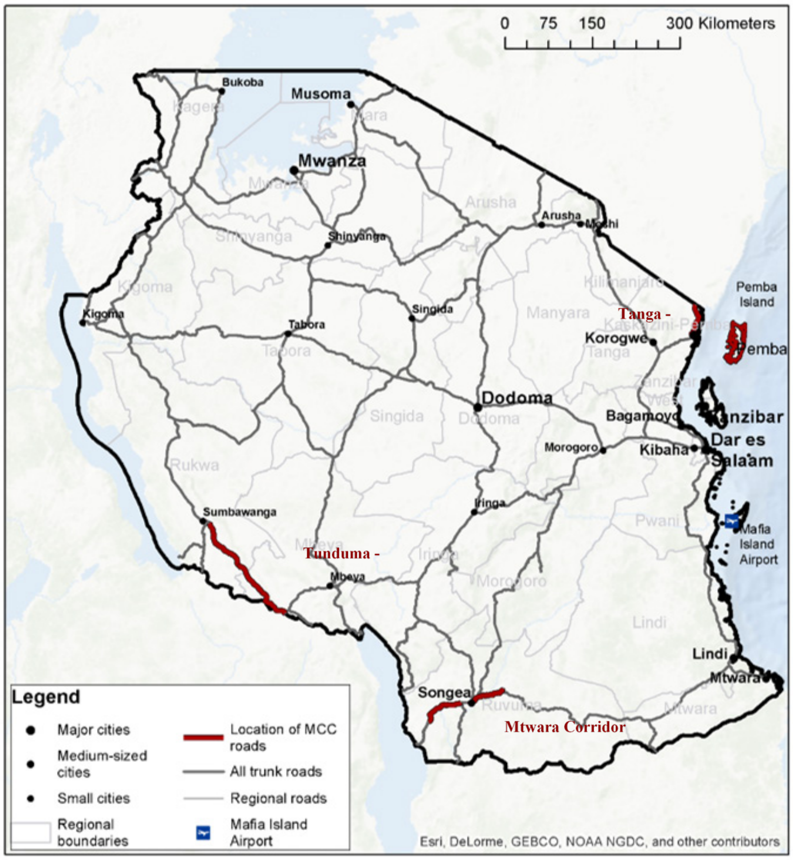
We model ERRs using HDM-4 for the following road segments:

- Tanga – Horohoro (1 of 1 segment)
- Songea – Namtumbo (1 of 2 segments in the Mtwara Corridor)
- Tunduma – Laela (2 of 3 segments in the Tunduma – Sumbawanga road)

Each segment represents part or all of the full length of road in each of the three mainland trunk roads. The location of the roads are shown in Figure F.1 below. We have not modelled the Peramiho Junction – Mbinga road, which is the second segment of the Mtwara Corridor or the Laela – Sumbawanga segment, which is the final segment of the Tunduma – Sumbawanga road. In both cases, the road design for the missing segments is the same as rest of the road. Construction costs for the Peramiho Junction - Mbinga segment are almost identical to the Songea – Namtumbo segment while traffic volumes are lower. Construction costs for the Laela – Sumbawanga segment are 30% higher than the other segments and traffic volumes in the initial year are similar. It is therefore unlikely that conducting a sensitivity analysis on the remaining segments will lead to different conclusions regarding the importance of collecting deflection data.

[22] We use input data from the Mtwara Corridor ex ante HDM-4 model because this is the only HDM-4 object file we have access to at this time. This model must be run using an older version of HDM-4.

Figure F.1. Location of trunk roads in Mainland Tanzania



The Jacobs Due Diligence report provides basic information on the assumptions behind the ex-ante ERRs. These are reported in Table F.1 from the report and the roads we study are outlined in red.

Table F.1. Ex-ante ERR estimates and key model assumptions extracted from Due Dilligence report

Table 8-18. EIRR reported in feasibility studies and initial rankings

Measure	Tunduma - Sumbawanga				Tanga - Horohoro					Tunduru - Songea			
	Tunduma - Laela	Laela - Sumbawanga			Main road			Bypass and port access road		Songea - Tunduru		Songea - Mbambabay	
Sections	Tunduma - Laela	Laela - Mkina	Mkina - Kiswite	Kiswite - Sumbawanga	Chumbageni RB - Mabokweni	Mabokweni - Km 46.00	Km 46.00 - Horohoro	New bypass	Port access road	Tunduru - Namtumbo	Namtumbo - Songea (start of Bypass)	Songea (End of Bitumen) - Mbinga	Mbinga - Mbambabay
Km by section	128.0	38.0	50.5	6[illegible]	11.8	34.5	19.4	2.3	0.8	1[illegible]0.7	70.0	77.6	66.3
Km total	128.0	95.3			65.7			3.1		263.2		143.9	
Existing standard (base year)	Earth / gravel IRI- (3.5-10)	Earth / gravel IRI- (3.10-10)			Gravel/paved - IRI 6.0 (3.5 - 11)	Gravel - IRI 14.0	Gravel - IRI 18.0	Gravel	Paved - IRI 3.4	Earth IRI-16.7	Gravel IRI-10.9	Earth IRI-[illegible]0.6	Earth IRI-12.3
Proposed standard (opening year)	Paved	Paved			Paved - IRI 2.5			Paved - IRI 2.5		Paved - IRI-2.2		Paved - IRI-2.2	
Traffic base year	2002	2002			2005				2005	2005		2003	2005
Daily average including bicycles	NA	NA	NA	NA	3,099	1,032	1,032	0	1,644	NA	NA	NA	NA
Motorised AADT	94	75	121	94[illegible]	838	262	262		438	40	187	78	63
Motorised excluding motorcycles	80	72	110	73[illegible]	640	215	215		280	40	187	78	63
Opening year	2007	2007			2010			2010		2008		2008	
Without project													
Daily average including bicycles	121	97	156	60[illegible]	3,110	1,036	1,036		1,761	NA		NA	
Motorised AADT	102	93	141	54[illegible]	962	319	319	37	415	48	226	107	76
With project													
Diverted traffic	yes	yes	yes	yes				yes		NA	NA	NA	NA
Generated traffic	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes
Motorised AADT	1,903	367	625	3,17[illegible]	1216	453	473		455	[illegible]13	448	191	175
Traffic growth per year (earlier years)	6.5%	6.5%			8.7%			8.2%		7.7%		7.7%	
Traffic growth per year (later years)	3.0%	3.0%			6.6%			6.7%		6.5%		6.5%	
Price Base Year	2002	2002	2002	200[illegible]	2005	2005	2005	2005	2005	[illegible]003	2003	2003	2003
Construction cost by section (Financial in $m)	31.93	[illegible]0.66	15.07	3.[illegible]	4.29	13.07	6.77	1.38	0.27	2[illegible].87	9.80	8.99	12.90
Construction cost by section (economic in $m)	24.27	8.11	11.45	2.[illegible]	3.35	10.13	5.25	1.07	0.21	2[illegible].99	8.33	7.65	10.97
Total econ cost ($m)	24.27	21.95			18.73			1.28		30.32		18.62	
Cost per km ($m)	0.19	0.21	0.23	0.[illegible]	0.28	0.29	0.27	0.47	0.26	[illegible].12	0.12	0.10	0.17
Benefits ($m)													
Benefit/km (undiscounted)					0.85	1.19	1.26		0.13	0.46	1.41	[illegible].78	0.53
VOC/km (% of total benefit/km)	85%	86%	89%	88[illegible]	93%	91%	91%	95%	89%	[illegible]9%	71%	69%	70%
Time/km (% of total benefit/km)	15%	14%	11%	12[illegible]	7%	9%	9%	5%	11%	[illegible]1%	29%	31%	30%
Other Benefit/km (% of total benefit/km)										[illegible]0%	0%	0%	0%
EIRR by section (%)	13.4%	[illegible]2.0%	15.2%	8.9[illegible]	7.4%	14.2%	14.4%	-1.0%	-2.5%	1[illegible]2%	32.8%	22.0%	9.5%
Total EIRR (%)	13.4%	13.5%			13.0%			-1.4%		17.0%		16.2%	
Ranking	4	3			5			6		1		2	

Table F.2 summarizes key outcomes and assumptions for the each of the roads. Ex-ante ERRs all exceed the MCC hurdle rate. Table F.2 also shows the initial level of normal traffic, the initial road condition on each road segment and the ex-ante and actual cost per kilometer.

Table F.2. Ex-ante ERRs, initial traffic and key outcomes

Road segment	Ex-ante ERR	Initial normal traffic (motorized AADT)	Initial roughness (IRI)	Ex-ante cost per kilometer ($)	Actual cost per kilometer ($)
Tanga – Horohoro	13.0%	291[a]	13.7	280,000	771,000
Songea – Namtumbo	32.8%	187	10.9	120,000	815,000
Tunduma – Laela	13.4%	80	7	190,000	1,040,000

[a] AADT for segment is calculated as the weighted average of multiple AADT measures along homogenous sections of the Tanga – Horohoro road segment.

B. Updates to the ex-ante model based on actual observations

For each of the segments modelled we use the cost per kilometer based on the actual costs and the traffic assumptions. Across all three cases, actual costs of construction are multiples higher than the costs used in the ex-ante model and contribute to much lower post-project ERRs. We also update traffic assumptions based on: i) a close review of assumptions in the model and ii) an analysis of traffic count data collected in 2013. MCA-T conducted traffic counts in 2013 for the Tanga-Horohoro road and the Mtwara Corridor roads. Construction of the Tunduma – Sumbawanga was not complete in 2013 and so no traffic counts were done. In each case, we made the following changes to traffic assumptions:

- **Tanga - Horohoro road:** We calibrated generated traffic and traffic growth rates using traffic volumes observed in 2013. The growth rates we use are similar to those in the ex-ante model, but our estimates of the incremental increase in generated traffic in the opening year of the project is lower for the post-project ERR. The with-project AADT in the opening year (including normal and generated traffic) of the road is 502 vehicles per day, which is smaller than the opening year AADT of 596 vehicles per day in the ex-ante model.
- S**ongea – Namutumbo road:** We calibrated growth rates for traffic using traffic volumes observed in 2013. The previous model had unrealistically high growth rates, and possible errors in the way that generated traffic was modelled. We use the ex-ante assumptions for the increase in generated traffic observed in the opening year of the upgraded road. However, we adjust down the growth rates for generated traffic, which are too high in the ex-ante model. The with-project AADT in the opening year of the road (including normal and generated traffic) is 448 vehicles per day, which is the same as the opening year traffic in the ex-ante model.
- **Tunduma – Laela road:** We do not have post-project traffic counts for this road. However, we revise down the estimates of generated and diverted traffic volumes. Base traffic in the ex-ante model was 80 vehicles per day, increasing to 1,903 vehicles per day in the year the road opens. Instead we assume that generated traffic is 200% of normal traffic in the year the road opens. Opening year traffic in the post-project model is 331 vehicles per day.

C Summary of model outputs and sensitivity analysis

The following tables report the key economic outcomes for each road segment that we analyzed, varying the structural number as we describe above.

Table F.3. Sensitivity of economic outcomes to different SN for Tanga – Horohoro

Parameter	SN: 2.0 Low	SN: 3.19 Constructed	SN: 5.0 Maximum	Comments
Roughness average IRI				
Initial roughness IRI before construction	13.7	13.7	13.7	Same Roughness for all cases with or without project case
Initial roughness IRI after construction	2.5	2.5	2.5	Roughness in 2010 after 3 years of construction 2007-2009
Final roughness IRI over year 20	8.0	6.32	3.0	For With Project case in year 2024; lower IRI due to hypothetical increase of SN from 2.0 to 5.0

Parameter	SN: 2.0 Low	SN: 3.19 Constructed	SN: 5.0 Maximum	RAC: Present Value of Road Agency Cost
Present value total agency costs - RAC				
Base Case US$ m	2.329	2.329	2.329	Same case
With project case US$ m	30.617	30.617	30.617	Capital Investment Cost was the same for all cases

Parameter	SN: 2.0 Low	SN: 3.19 Constructed	SN: 5.0 Maximum	C: Present Value of Increase of Road User Costs RUC
Increase in agency costs - C				
With project case US$ m	28.288	28.288	28.288	Difference between RAC with Base Case and with Project Case

Parameter	SN: 2.0 Low	SN: 3.19 Constructed	SN: 5.0 Maximum	B: Present Value of Benefits of Road User Costs RUC
Decrease in User Costs - B				
With project case US$ m	22.785	23.853	24.549	Stronger SN imply better average IRI and lower RUC (VOC + Travel Time) and bigger decrease

Parameter	SN: 2.0 Low	SN: 3.19 Constructed	SN: 5.0 Maximum	Comments
Net Present Value - NPV US$ m	**-5.503**	**-4.435**	**-3.739**	at 12% discount rate Reduced RUC implies savings in RUC and increase in NPV. Costs are bigger than benefits.
Internal Rate of Return - IRR %	9.00%	9.00%	10.00%	Reduced RUC implies savings in RUC and increase in IRR

Note: Tanga – HoroHoro Roads, Route A-14

Length: 65.7

Structural Number (SN) Sensitivity Analysis Comparison and Use Calibrated Traffic to Real values

Table F.4. Sensitivity of economic outcomes to different SN for Songea – Namtumbo road

Parameter	SN: 2.0 Low	SN: 3.19 Constructed	SN: 5.0 Maximum	Comments
Roughness average IRI				
Initial roughness IRI before construction	10.9	10.9	10.9	Same Roughness for all cases with or without project case
Initial roughness IRI after construction	2.2	2.2	2.2	Roughness in 2008 after 3 years of construction 2005-2007
Final roughness IRI over year 20	3.88	3.44	3.31	For With Project case in year 2024; lower IRI due to hypothetical increase of SN from 1 to 5

Parameter	SN: 2.0 Low	SN: 3.19 Constructed	SN: 5.0 Maximum	RAC: Present Value of Road Agency Cost
Present value total agency costs - RAC				
Base Case US$ m	8.257	8.257	8.257	Same case for each scenario
With project case US$ m	40.355	40.355	40.355	Capital Investment Cost was the same for all cases

Parameter	SN: 2.0 Low	SN: 3.19 Constructed	SN: 5.0 Maximum	C: Present Value of Increase of Road User Costs RUC
Increase in agency costs - C				
With project case US$ m	32.098	32.098	32.098	Difference between RAC with project case and RAC base case

Parameter	SN: 2.0 Low	SN: 3.19 Constructed	SN: 5.0 Maximum	B: Present Value of Benefits of Road User Costs RUC
Decrease in User Costs - B				
With project case US$ m	17.145	17.275	17.304	Stronger SN imply better average IRI and lower RUC (VOC + Travel Time) and bigger decrease

Parameter	SN: 2.0 Low	SN: 3.19 Constructed	SN: 5.0 Maximum	Comments
Net Present Value - NPV US$ m	**-14.953**	**-14.823**	**-14.794**	at 12% discount rate Negative value due to User Benefits B are lower than increase in Road Agency Costs C
Internal Rate of Return - IRR %	6.00%	6.10%	6.10%	In case of very low level of traffic, implies reduced VOC and savings in RUC and small variations in IRR due to strongest SN

Note: Namtumbo – Songea Road, Mtwara Corridor

Length: 71.4 km

Structural Number (SN) Sensitivity Analysis Comparison and Use of Traffic Real Values

Table F.5. Sensitivity of economic outcomes to different SN for Tanga – Horohoro

Normal traffic Parameter	SN: 2.0 Low	SN: 3.56 Constructed	SN: 5.0 Maximum	Comments
Parameter/generate traffic AADT	200%	200%	200%	Generated traffic obtained as percentage of 200% of Normal Traffic
Roughness average IRI				
Initial roughness IRI before construction	7.0	7.0	7.0	Same Roughness for all cases with or without project case
Initial roughness IRI after construction	2.5	2.5	2.5	Roughness in 2007 after 3 years of construction 2005-2007
Final roughness IRI over year 20	8.14	5.54	4.95	With project case in year 2026

Normal traffic Parameter	SN: 2.0 Low	SN: 3.56 Constructed	SN: 5.0 Maximum	RAC: Present Value of Road Agency Cost
Present value total agency costs - RAC				
Base Case US$ m	2.375	2.375	2.375	Same case for each scenario
With project case US$ m	63.542	63.504	63.504	Capital Investment Cost was the same for all cases US$m 63.333

Normal traffic Parameter	SN: 2.0 Low	SN: 3.56 Constructed	SN: 5.0 Maximum	C: Agency costs
Increase in agency costs - C				
With project case US$ m	61.167	61.129	61.129	Difference between RAC with project case and RAC base case

Normal traffic Parameter	SN: 2.0 Low	SN: 3.56 Constructed	SN: 5.0 Maximum	B: Present Value of Benefits of Road User Costs RUC
Decrease in User Costs - B				
With project case US$ m	15.893	16.844	17.021	Strong SN imply better road conditions and more benefits for users (VOC + Travel Time)

Normal traffic Parameter	SN: 2.0 Low	SN: 3.56 Constructed	SN: 5.0 Maximum	Comments
Net Present Value - NPV US$ m **With project case**	**-45.274**	**-44.285**	**-44.108**	at 12% discount rate Negative value due to User Benefits B are lower than increase in Road Agency Costs C
Internal Rate of Return - IRR %	6.00%	6.10%	6.10%	In case of very low level of traffic, implies reduced VOC and savings in RUC and small variations in IRR due to strongest SN

Note: Tunduma – Laela Road

Length: 128 km

Structural Number (SN) Sensitivity Analysis Comparison of several Structural Number from SN: 1.0 to SN: 5.0

D. Sensitivity analysis results:

The ERRs are lower across all HDM-4 models. This is driven by the big change in construction costs and by revisions to the traffic assumptions. The key findings from this analysis are:

1. Increased agency costs associated with road repairs do not differ greatly, if at all, across all three cases. Small costs associated with resealing and crack sealing arise over the lifetime of the road, but traffic is not sufficiently high to cause major damage over the 20 year period of analysis.
2. Lower SN leads to the emergence of more defects and increased IRI, which affects road users. This is most pronounced in the minimum SN case for each of the roads. However, the effect on the ERR is within 1 percentage point for each of the modelled roads.

For each road, across the range of values for the structural number, most of the difference in benefits arises from differences in the Road User cost savings due to road condition. However, these differences are small, and the costs imposed on road users are greatest at the end of the project's life and are therefore heavily discounted. There is little to no difference in the timing of defects and maintenance costs across all three SN cases and therefore little to no difference in the present value of total agency costs in the with-project case.

These findings reflect the sensitivity analysis reported in the HDM-4 Manual Volume 5 (calibration and adaptation), which states on p. 32 that "variations in structural variables…are sensitive to [a] high degree only when the pavement structural adequacy is low or moderate compared with the loading." In these cases structural variables have impact elasticities >0.5. When the pavement structural adequacy is high to moderate, the impact elasticities range from 0.2 to 0.5. The relevant paragraph is:

1. High impacts, Class S-I (> 0.5)

The most sensitive inputs include pavement structural variables, traffic and roughness, which are all measured data items. Variations in the structural variables, Modified Structural Number or deflection, and annual traffic loading (in million ESA), affect most major results including periodic maintenance and rehabilitation alternatives and the economic returns. They are sensitive to this high degree only when the pavement structural adequacy (PSA – see Section 4.2) is low or moderate compared with the loading (under other conditions when the PSA is high to moderate the impacts are moderate, that is, Class S-II). Variation of the traffic volume, which determines the number of users deriving benefits, has strong influence on the economic returns but low influences (S-II or S-III) on physical impacts. Variation of the pavement roughness, which affects the unit savings gained by each user, has high impacts on all economic results and on rehabilitation needs, but little on surface distress (S-III).

2. Annex: Input variables from ex-ante HDM-4 model for Songea – Namtumbo road

We extracted the basic input data by installing an old version of HDM-4. The following tables show the characteristics of the base option. Input data that is not specific to the Songea-Namtumbo road is used in the other models because we do not have the original HDM-4 models used in the ex-ante ERR analysis for those roads.

Road characteristics:

Section 4.2 / Namtumbo to Start of Bypass

Definition

Section name:	Namtumbo to Start of Bypass	Climate zone:	Humid/Tropical	Shoulder width:	0,00 m
Section ID:	Section 4.2	Road class:	Primary or Trunk	Number of lanes:	2
Link name:		Surface class:	Unsealed	Motorised AADT:	187
Link ID:		Pavement type:	Gravel	NM AADT:	0
Speed flow type:	Two Lane Road	Length:	60,80 m	AADT year:	2005
Traffic flow pattern:	Free-Flow	Cway width:	9,00 m	Flow direction:	Two-way

Geometry

Rise + fall:	28 m/km	Speed limit:	30 km/h
Avg horiz curvature:	76 deg/km	Altitude:	923 m

Pavement

Surface material:	Lateritic gravel	Compaction method:	Mechanical
Subgrade material:	Well-graded gravel-sands with small clay content (GC)	Last gravel year:	1993

Condition

Condition year:	2003	Gravel thickness:	10 mm	IRI:	9,50 m/km

Speed related

Num rises + falls:	4 no./km	XNMT:	1,00	XMT:	1,00
Superelevation:	3,00 %	XFRI:	1,00	Speed limit enforcement:	1,10
Sigma adral:	0,10 m/s²				

Surface Material Gradation

Max particle size:	4,75 mm	% passing 2.00mm sieve:	99,00 %	% passing 0.075mm sieve:	46,00 %
Plasticity index:	24,00 %	% passing 0.425mm sieve:	75,00 %		

Subgrade Material Gradation

Max particle size:	4,75 mm	% passing 2.00mm sieve:	99,00 %	% passing 0.075mm sieve:	46,00 %
Plasticity index:	24,00 %	% passing 0.425mm sieve:	75,00 %		

Shoulders and NMT Lanes

Num shoulders:	2	Num NMT lanes:	0	NMT lane surface type :	Bituminous
Separate NMT lanes:	No				

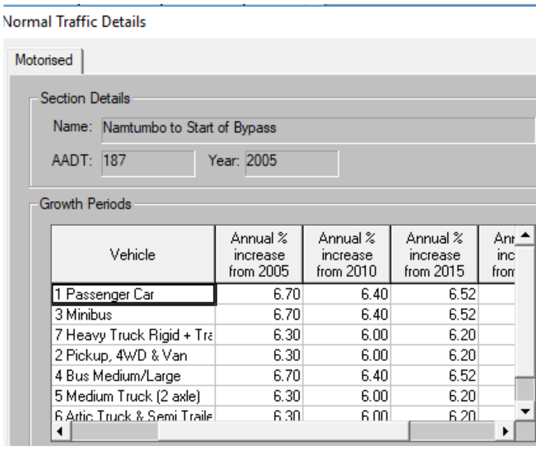
Roughness Model Calibration

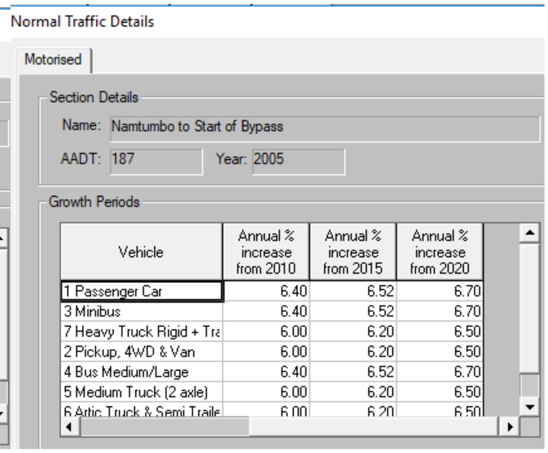
Derivation:	Computed/derived	Surface minimum:	2,77 m/km	Subgrade minimum:	1,12 m/km
Surface maximum:	19,97 m/km	Subgrade maximum:	20,30 m/km		

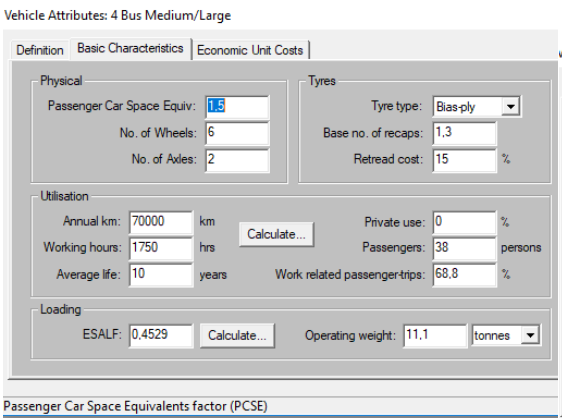
Material Loss Calibration

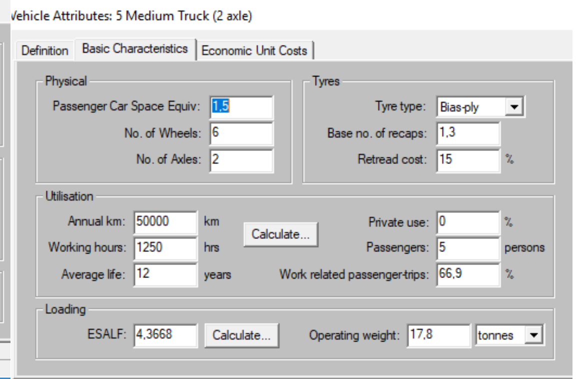
Surface loss factor:	1,00	Subgrade loss factor:	1,00	Subgrade traffic induced:	1,00
Surface traffic induced:	1,00				

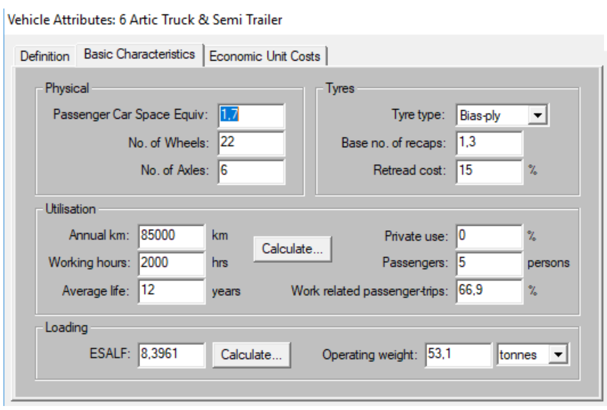
Traffic characteristics:

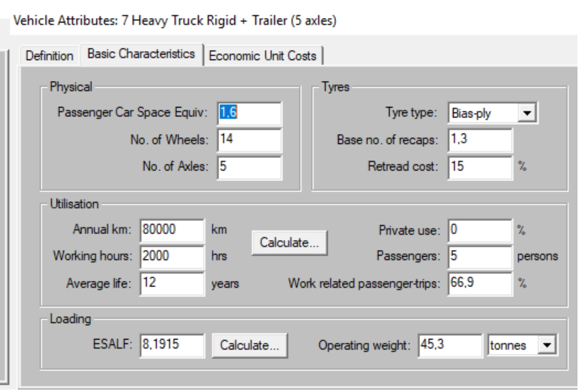












APPENDIX G:

RESPONSES TO MCC COMMENTS ON THE DRAFT EVALUATION DESIGN REPORT

This page has been left blank for double-sided copying.

Table G.1. Evaluator comments on design report

Page number (please reference the number at the bottom of the page)	Comment	Evaluator responses
17, deflection	If we don't collect deflection data, what is the impact on the precision of the ex-post ERR estimate you will do? I understand from econ colleagues that if we have a completion measurement of deflection, and the roads appear to be in good condition, then they wouldn't recommend collecting it. If we don't have that completion measure, but we have CBR and structural number, then they also don't recommend collecting it. What do you think?	After conducting a sensitivity analysis that assessed the impact of various structural numbers on the ERR, we recommend not collecting deflection data or conducting DCP tests for this evaluation.
17, modified structural number	This should usually be stated in our construction contracts. Do you have those? I can look if not.	We do not have the construction contracts. It would be helpful to know if the information on SN is available in the contracts.
19, "collect a core sample, and conduct DCP testing."	Why is this necessary to add?	We have removed this text because we are no longer planning to collect DCP tests for this evaluation.
19, Traffic	When will it be possible to propose the locations of the traffic, OD and axle load surveys? And the focus group sampling appraoch? Since traffic counts are planned to start in March, it should be soon.	Whenever possible, we will use the same locations used in the feasiblity reports for traffic and OD surveys. We plan to identify the locations as part of the work planning process and confirm them after the pilot in the form of a memo or the final workplan.
19, Traffic	This sounds like local traffic will be prohibited, which isn't right. Is the intent to *avoid* local traffic?	This text comes from Annex 4 of our contract. The traffic counts are intended to capture the main types of vehicles using the road and avoid overcounting certain types that are making a short trips within villages/towns/urban areas along the road.
19, Axel Load	Is the number the same as the number of traffic counting stations? Will the axel loading be done during the wet or dry season? Or both?	Axle load surveys will only be conducted during the dry season. We will conduct 4 axle load surveys (one for each road on the mainland, and one for Pemba). We will conduct have 7 traffic count stations in the rainy season and 14 in the dry.

Page number (please reference the number at the bottom of the page)	Comment	Evaluator responses
20, Deflection	Why was the decision taken to measure deflection everywhere, instead of the selected sample that we had originally discussed (just wondering the rationale)? Why are there no implications for the budget? I believe the cost to measure it everywhere was significantly more (around $100K)	We are no longer planning on measuring deflection for this evaluation.
20, Deflection	Do we have any deflection measurements on the completed road? These are usually in the Supervisory Engineer's final report or the Independent Engineer Report. If you don't have those I can provide. Or maybe the government has some?	We do not have these reports.
31, Table III.2	Can you explain how the HDM-4 estimates of time savings are related to the evaluation question 2d?	This bullet was in the wrong place. We have moved it to address subquestion 2c.
31, Table III.2, evaluation question 2d	Why delete value?	Based on our experience in Ghana, drivers were not able to answer this question, but we pilot the question again in Tanzania and keep it in the text of the EDR.
32, Traffic count data	This should include a discussion of TANROADS full-time traffic monitoring and weightbridge systems, how well they align to inform seasonal patterns for our locations of interest, and access to these detailed full-time data.	We have added to this section. According to meetings with TANROADS officials in October 2017, no traffic counting machines were in use (for full-time automated counts), but they were in the process of procuring the equipment. If these have been added prior to data collection, we will attempt to use this data as well.
OD module	Must these be mutually exclusive, or will you allow multi-purpose trips?	The OD survey captures multiple trips.
32, Road Users Module	Why delete "weight and value of cargo, and vehicle capacity in tons."? They seem important for the economic model, and for evaluation question 3.	See response above on line 11.
33, "we will use the EDI baseline data to guide…"	Can the OD survey help to improve representation of road users?	The EDI data is the only data available for comparison roads. We can use the VIS data to help select communities on the MCC roads.

Page number (please reference the number at the bottom of the page)	Comment	Evaluator responses
33, " we will sample at least 20 percent (approximately every fifth vehicle) of all traffic moving in both directions along the roads"	How will periods of unusually heavy traffic (that exceed the interviewers' capacity) be handled?	Traffic is stopped at regular intervals, so the sampling rates are not constant throughout the day. During periods of heavy traffic, the sampling intervals will be lower.
36, "In addition, we will track transportation firm characteristics such as fleet size, number of employees, and the share of the firm's routes that utilize MCC project roads, to assess share of the market."	What was the rationale for deleting this?	Based on our experience in working with private sector firms, they would likely be reluctant to share this information and would not have data on trips taken on MCC project roads.
46, "For each urban area, we will calculate the change in TravelTime_it that is attributable to MCC's trunk road investments and use, β, to estimate the resulting change in economic activity."	A necessary intermediate outcome in this model is that things or people must be moving between locations in order to facilitate economic growth. How might the OD survey evidence be used to either test or improve the predictive power of this model?	This model is looking at trade integration across the entire network of urban areas in Tanzania, so not all urban areas need to trade directly with each other in order to achieve economic growth. We can use evidence from the OD survey to assess whether the frequency of trips between origins and destinations reflects the relative travel costs of moving between these places, as predicted by the model.

Page number (please reference the number at the bottom of the page)	Comment	Evaluator responses
51, We propose to complete the initial EDI design on Pemba only because the household surveys conducted on Pemba are more complete than the surveys conducted on the mainland, and the roads that were upgraded were designed largely to benefit these rural communities (unlike the trunk roads, which were also intended to connect other parts of the country).	How strong is this instrument? 1. Were the travel-times and costs prior to the road improvements large enough to inhibit behaviorally significant income generating opportunities (and were the reductions sufficient to overcome these barriers)? 2. Can we determine whether there were either compensating, or adverse changes in the travel-times and costs experienced along the control roads over the same period? 3. Can we determine how homogeneous income growth was between the treatment and control locations prior to changes in transport times and costs?	We propose removing the validation exercise from the evaluation design report given the time constraints for completing the evaluation. If there is a strong interest in conducting it, we can add it back in and use information from the fieldwork to verify the quality of the instrument.
55, Vehicle operator survey	How can we avoid a convenient sample bias here? An ideal sample frame for this would be the OD or traffic count survey populations. Can the transport operators, freight companies and vehicle sellers be chosen to approximately represent the survey population?	Because of the small number of operators and the challenges associated with collecting VOCs, we can't really avoid the convenience sample. The OD survey is unlikely to provide sufficient data on VOCs because it covers a very small number of vehicles.
56, "The final interval length will be determined by available resources and the condition of the road. "	Is this still relevant? Delete?	We have deleted this sentence.
59, ..as well as trip purpose..	Multi-purpose allowed?	Yes, see response on line 14.

Page number (please reference the number at the bottom of the page)	Comment	Evaluator responses
59, …we will ask questions about cargo type..	Is this ad-hoc list preferable to 1 digit SITC groupings?	We used a modified version of the SITC groupings to develop the instrument in Ghana. These categories were also used in the pre-project feasibility studies and will allow pre-post comparisons. The modifications were necessary to capture the type of freight that is typically traveling on the roads, and make it easier for respondents to provide accurate answers. We will make sure to recode the data into SITC groupings in the final dataset.
65, "conduct additional data quality reviews"	Can these also be used better estimate medium term (e.g., monthly) traffic volumes?	Yes, but it depends on cost (for automated methods) and availablty in country.
107, Traffic counts	How did you reach these numbers? What is the rationale for 7 days? I don't see it in the EDR but I might be missing it (apologies in advance if so)	The Traffic Monitoring Guide issued by the US Federal Highway Administration recommends 7-day counts, espeically where there is important day-of-week variation. The total number of stations was determined by looking at the road map and identifying the number of significant intersections joining the roads. Each segment where we may see different traffic volume because vehicles are joining the road has its own traffic count station.
Budget	I think option B is now irrelevant and can be removed, as can the note at the bottom	We have removed Option B.
5, description of activities	Include a summary of the total project costs and approx. costs per km of the roads (distinguishing any that are unusually high or low)	We have added this.

Page number (please reference the number at the bottom of the page)	Comment	Evaluator responses
19, "Traffic will be collected along multiple points on the road network to ensure that local traffic is excluded from the count and that any segments defined by important exits or entries onto the road are covered"	This sounds like local traffic will be prohibited, which isn't right. Is the intent to avoid local traffic?	See response above on line 6.
21, "Although HDM-4 is widely used…"	Ugh. This is consistent with Jacobs' 2007 Due Diligence Report (p.8-28), and the fact that this was a known issue during due diligence might be important – especially if the observed road condition is significantly different from HDM4 predictions.	This is true, and something to highlight for the lessons learned.
22	Do you have the Jacobs Due Diligence Report?	Yes
24, "Access to initial ERR data"	Again, look at the Jacobs report. While it does not include the detailed model parameters, it does critically assess an original set of ERRs and provides parameters used to calculate alternatives	We have reviewed this. We will use some of this information in our analysis.
4, Table III	Need to document how we reached this	We used this estimate for budgeting purposes and we will finalize the number of FGDs and sampling in a memo prior to doing the FGDs.
4, "and a graphic representation of all traffic counting stations"	Including GIS coordinates	Agree.

Page number (please reference the number at the bottom of the page)	Comment	Evaluator responses
5, OD module, "cargo type"	Rather than the ad hoc list provided, it is preferable to base this on a comparable but established categories of traded goods to facilitate comparison with other data sources. If comparability to baseline data is not an issue, then Version 4 of SITC single-digit codes is probably the most commonly used coding: https://unstats.un.org/unsd/publication/SeriesM/SeriesM_34rev4E.pdf	See response on Line 23.
9, "()"	link missing?	We have added the link.
Table III	*Are there power calculations underlying* • *the number of 1km detailed study segments for surface condition for road condition sampling, and* • *the number of core samples for geotechnical measurements?*	The total number of 1km segments is based on the 10% estimate. We are no longer planning on conducting DCP tests for this evaluation.