ROADS AND CHILD HEALTH IN SUB-SAHARAN AFRICA

Luisito Bertinelli, Evie Graus,
Jean-François Maystadt,
Silvia Peracchi

LIDAM Discussion Paper IRES
2025 / 17

UCLouvain

LIDAM
IRES | Louvain Institute of Data Analysis and Modeling in economics and statistics

Roads and child health in Sub-Saharan Africa*

Luisito Bertinelli† Evie Graus‡ Jean-François Maystadt§ Silvia Peracchi¶

October 24, 2025

Abstract

This paper examines the causal impact of road access on child health in Sub-Saharan Africa between 1980 and 2012 by combining geolocated data on child anthropometric outcomes with spatial data on road networks. To address endogeneity, we employ an instrumental variable approach based on the inconsequential units framework, constructing hypothetical road networks that connect historical cities and active mines. Our results show that closer proximity to paved roads significantly improves child health. The main mechanisms operate through improved healthcare access and utilization, higher household wealth, early signs of structural transformation, and cropland expansion. We find no evidence that these gains are offset by adverse environmental or epidemiological effects of improved road access. Overall, the findings underscore the role of road infrastructure in fostering development across Sub-Saharan Africa.

Keywords: roads, Sub-Saharan Africa, child health, causal analysis.

JEL Classification: O15, I15, O18, O55.

*This work was supported by the Fonds de la Recherche Scientifique - FNRS under Grant T.0150.23 and by the Luxembourg National Research Fund (INTER/FNRS/22/17545821).

†Department of Economics and Management, University of Luxembourg, L-1359 Luxembourg.

‡Department of Economics and Management, University of Luxembourg, L-1359 Luxembourg.

§IRES/LIDAM, UCLouvain; FNRS, Belgium, and Department of Economics, Lancaster University.

¶IRES/LIDAM, UCLouvain; FNRS, Belgium.

1 Introduction

Road deficiencies are widely recognized as a major impediment to economic development (Ali et al., 2015). Road infrastructure enhances market access, reduces production costs by lowering input and transportation expenses, and ultimately decreases the prices of consumer goods and services (Jayachandran, 2022). Yet, with only 204 kilometers of road per 1,000 square kilometers of land area, Sub-Saharan Africa has a road density five times lower than the world average, and a network less than 30% compared to the next least-developed region in terms of roads - South Asia (Ali et al., 2015, p. 6). Beyond insufficient road coverage, spatial inefficiencies in Africa's current road network are argued to further contribute to significant welfare losses: it has been estimated that Africa could gain 1.3% of total welfare from reorganizing its current network and 0.8% from optimally expanding it (Graff, 2024). Given the region's abundance of labor, land, and water, road infrastructure could be a major driver in accessing Sub-Saharan Africa's substantial agricultural and demographic potential. Such prospect justified a World Bank investment of $5.9 billion (USD) in the transport sector in 2013, an amount exceeding allocations to education, health, and social services combined, with roads accounting for 60% of the total lending (Ali et al., 2015). In parallel, since its launch in 2013, the Chinese Belt and Road Initiative has maintained a strong presence. For instance, the $480 million loan for the Yaoundé–Douala highway project in Cameroon remains under implementation today (Goodman et al., 2024), reflecting the continued scale of road infrastructure investment across the region.

In this paper, we analyze the impact of road access on child health in Sub-Saharan Africa. We focus on child health for three main reasons. First, the relationship between road infrastructure and health outcomes has received limited attention in the empirical literature. Although recent evidence have begun to explore this link, such as Dasgupta et al. (2024) for India and Li (2025) for selected Sub-Saharan African countries, comprehensive, region-wide evidence on road access and health outcomes remains scarce. Second, the potential mechanisms linking road access to health are diverse and context specific, warranting further investigation. Roads may improve health through enhancing market-access, which can increase the availability of essential goods and facilitate access to healthcare facilities (Faber, 2014; Aggarwal, 2021; Jedwab and Storeygard, 2022), by promoting structural transformation and income growth through shifts toward non-farm employment (Asher and Novosad, 2020; Fiorini and Sanfilippo, 2022; Dumas and Játiva, 2025), and by fostering agricultural development and food security, including cropland expansion, crop diversification, and improved production techniques (Stifel and Minten, 2017; Berg et al., 2018; Shamdasani, 2021; Gebresilasse, 2023; Adamopoulos, 2025). At the same time, roads may generate adverse health externalities through several channels, including deforestation, which can affect local environments and food systems (Damania et al., 2018; Kleinschroth et al., 2019; Acharya et al., 2020; Asher et al., 2020; Baehr et al., 2023), air pollution from roadside activities and traffic (Rosales-Rueda and Triyana, 2019; Heft-Neal et al., 2025), waste mismanagement (Gennaioli and Narciso, 2025), and increased exposure to communicable diseases through higher mobility (Djemaï, 2018). Third, child health is a fundamental component of human capital formation and a strong predictor of long-term nutritional and economic outcomes (Moradi, 2010; Perkins et al., 2016). In Sub-Saharan Africa, an estimated 31% of children under five are stunted, well above the global average of 22% (United Nations Children's Fund (UNICEF) et al., 2023), making the study of its determinants a pressing development priority. As such, understanding how road infrastructure affects child health is essential for assessing the broader welfare implications of infrastructure investments in a region that continues to face intertwined challenges of limited health outcomes and low transport connectivity.

To address our research question, we use geolocated data on roads from Jedwab and Storeygard (2022) and information on health for children aged 0 to 5 years, specifically height-for-age and weight-for-age Z-scores, from the Demographic and Health Surveys (DHS). The latter serve as proxies for child health, as lower height-for-age and weight-for-age values typically indicate poorer physical health, reflecting chronic and acute health deficiencies, respectively. Our analysis covers 31 Sub-Saharan African countries over the period 1980 to 2012.

However, establishing a causal relationship between road infrastructure and child health is challenging due to potential endogeneity concerns. Road building and upgrading are likely not random, as unobserved factors such as economic development or agricultural potential may simultaneously influence both road expansion and health outcomes. Additionally, it remains uncertain whether roads are built in areas that already have better health conditions, or whether improved health is a consequence of road development. To address these sources of endogeneity, we employ an instrumental variable (IV) approach using the "inconsequential units framework" from Chandra and Thompson (2000). This framework posits that roads are constructed between two economically consequential places, while smaller intermediate areas, deemed "inconsequential", do not influence the route choice but are nonetheless affected by the road's presence. As a result, for these inconsequential units, road placement can be considered random. We apply this method by constructing hypothetical road networks for each year in our sample. More specifically, we connect historical cities (Jedwab and Moradi, 2016) and active mines using a Euclidean Minimum Spanning Tree (EMST) algorithm. These hypothetical road networks represent the road network that should have been created if cities and mines were to be connected in the most efficient way possible, thereby minimizing total network length. We argue that in Sub-Saharan Africa, both cities and mines are important determinants in shaping road development, as roads not only facilitate connectivity between cities but are also essential to expedite the extraction of mineral resources from mines for international export via national or local ports (Bonfatti and Poelhekke, 2017; Bonfatti et al., 2022).

Our reduced-form estimates suggest that proximity to paved roads significantly improves child health. In particular, a 10% increase in proximity to a paved road improves height-for-age and weight-for-age Z-scores by approximately 0.0248 and 0.0097 standard deviations, respectively. To put this into perspective, the mean height-for-age and weight-for-age Z-scores in our full sample are -1.279 and -1.157 standard deviations, respectively. Since our favorite specification includes road distance in logarithms, the resulting non-linear relationships show that the effect of distance on child health declines sharply, with the steepest gradient within the first 100 kilometers from paved roads. Importantly, this is also where most DHS respondents are located, suggesting that for most households in our sample, the effects of road access on child health are meaningful. Our findings remain robust across multiple specifications and sensitivity analyses, and appear to be mainly driven by improved access and utilization of healthcare services, higher household wealth accompanied by shifts in labor away from agriculture, and cropland expansion. By contrast, we find no evidence that these health gains are offset by adverse effects such as deforestation, air pollution, or increased malaria exposure.

The literature documenting the contribution of transport infrastructure to economic growth and development is extensive (Duranton and Turner, 2012; Redding and Turner, 2015; Donaldson, 2018; Banerjee et al., 2020; Allen and Atkin, 2022; Jedwab and Storeygard, 2022). Closest to our work, Dasgupta et al. (2024) examine the effects of a large-scale rural road construction program in India on fertility and investments in child health, showing that improved road access reduced fertility and enhanced child health through higher immunization rates and lower infant mortality. More recently, Li (2025) studies the effects of Chinese-aided transport infrastructure on child height in eleven Sub-Saharan African countries, finding that an additional year of exposure increases height-for-age Z-scores and reduces stunting. While both studies provide valuable evidence, they focus on specific interventions or subsets of countries. In particular, Chinese-financed infrastructure may differ from road investments in their placement and purpose, thereby leaving open the question of how general road access affects child health across Sub-Saharan Africa.

Related work also highlights the broader welfare implications of remoteness and road infrastructure in Africa. Stifel and Minten (2017) show that households in more remote areas of Ethiopia experience lower consumption, higher food insecurity, and reduced school enrollment rates. Similarly, Djemaï et al. (2024) examined the effects of road paving on well-being across Sub-Saharan Africa, employing an instrumental variable approach to account for the endogeneity of road paving, and find that greater distance from paved roads is negatively related with objective but not subjective well-being.

Our paper contributes to this growing literature on the developmental consequences of road investments in Sub-Saharan Africa in two main ways. First, we provide a comprehensive regional analysis of the causal impact of road access on child health, covering 31 countries over three decades. We extend the range of development outcomes by focusing on child anthropometric indicators, which offer objective measures of health less prone to recall bias or misreporting than income or consumption data, thereby reducing the risk of "hiding the poor" (Deaton, 2007). These indicators capture not only nutritional intake but also exposure to disease and access to healthcare, making them particularly relevant in rural African contexts where income measures may fail to reflect deprivation (Strauss and Thomas, 1998, 2008). Second, we contribute to the empirical identification strategy by extending the IV framework proposed by Jedwab and Moradi (2016). Specifically, we incorporate mines as additional nodes in constructing hypothetical road networks, generating a time-varying instrument that better captures the historical and economic determinants of road placement. This allows us to identify the causal effects of road access on child health at a continental scale, complementing existing country- or project-specific analyses.

The remainder of this paper is structured as follows. Section 2 describes the data, while Section 3 outlines the empirical methodology and IV specification. Section 4 presents our main OLS and IV results, including multiple robustness checks. Section 5 examines the potential mechanisms at play. Finally, Section 6 concludes.

2 Data

We begin by describing the primary data sources used to construct our dependent and independent variables: child health and road access, along with the additional covariates used in the analyses. Summary statistics for the full sample are presented in Table 1.

2.1 Child health

We use anthropometric child health indicators sourced from the Demographic and Health Surveys (DHS). Specifically, we focus on height-for-age and weight-for-age Z-scores for children aged 0 to 5 years. DHS is a program funded by the U.S. Agency for International Development (USAID) and other donors. It represents a collection of representative cross-sectional surveys recording several health characteristics of the population for a rich set of countries, available for several years. The main units of interest in DHS are women of reproductive age (15–49). Eligible women are identified from selected households, and all household members (including children) and several of their characteristics are recorded. The primary sampling unit of DHS surveys is called a cluster. It consists of a number of adjacent households in a geographical area and generally corresponds either to a census enumeration area or a subdivision thereof if the enumeration area is large. We use information on the location of DHS clusters to merge DHS with other data. Figure A.1 illustrates the countries with geolocated DHS data included in our analysis.[1]

Height-for-age and weight-for-age Z-scores provide standardized measures of comparing children's variations from normal height and weight, accounting for differences by age. The scores are benchmarked against the World Health Organization's (WHO) Child Growth Standards, where a Z-score indicates the number of standard deviations (SD) a child's measurement deviates from the median of the reference population.[2] The values of height-for-age and weight-for-age Z-scores range from -5 to 5 SD, with a Z-score below -2 SD indicating stunting or being underweight, respectively. These anthropometric indicators serve as proxies for child health, as children who are significantly shorter or lighter than expected for their age typically exhibit signs of poorer physical health at early age. While weight is considered a short-term indicator of child health, height reflects cumulative health conditions from fetal development onward and has been shown to be a strong predictor of long-term nutritional deficiencies and economic well-being (Moradi, 2010; Perkins et al., 2016).

2.2 Road access

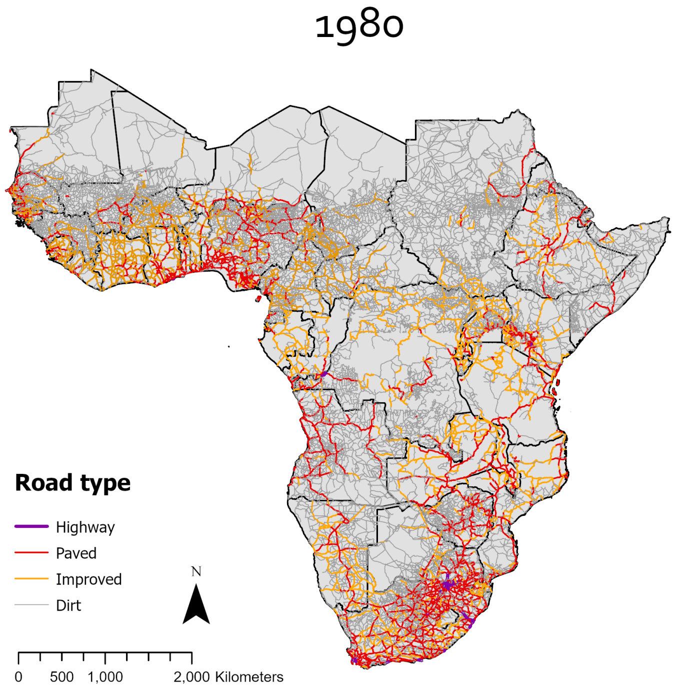
Sub-Saharan African road data is sourced from Jedwab and Storeygard (2022), who compiled a comprehensive dataset of road segments spanning various time periods and regions. The dataset originates from 64 Michelin maps covering the period 1961 to 2014, and digitized by Jedwab and Storeygard (2022). The data provides information on the total road length (in kilometers) and road quality, classifying road segments into four main categories: highways, paved roads, improved roads, and dirt roads.[3] Hence, this dataset allows us to track changes in road quality over time, distinguishing

[1] The full sample, disaggregated by country and year of birth, is presented in Table A.1.

[2] The WHO Child Growth Standards are based on a multi-country reference population of children from the United States, Oman, Norway, Brazil, Ghana, and India, raised in socioeconomic conditions favorable to growth and characterized by low mobility.

[3] Highways are paved roads with at least three lanes per side. Paved roads consist of asphalt with one or two lanes per side. Improved roads are unpaved but made of laterite or gravel, while dirt roads are unpaved and composed of native material, often becoming unusable during rainfall. A visualization of these road types is provided in Figure A.2.

Figure 1. Roads in Sub-Saharan Africa in 1980 vs. 2010



Notes: Authors' own calculations based on data from Jedwab and Storeygard (2022).

between upgrading, where unpaved roads transition to a paved status, and deteriorating, where paved roads lose quality and turn into unpaved roads. Figure 1 visualizes how road quality changed between 1980 and 2010, illustrating that while the majority of roads in Sub-Saharan Africa remain unpaved, some upgrading to paved roads has occurred over time, particularly, in West African countries such as Côte d'Ivoire and Nigeria, and in Central Africa in countries like Burundi and Rwanda.

A key advantage of using the DHS data is its geolocated cluster structure, where each DHS cluster consists of multiple households. This allows us to combine the DHS and road data by calculating road access at the cluster level. More specifically, for each child, we compute the Euclidean distance to the nearest paved road at birth, which serves as our main independent variable. A paved road is defined as either a highway or a standard paved road.

We rely on distance to the closest paved road as a more direct and interpretable measure of access to goods and services, essential to child health.[4] Compared to the broad class of market

[4]Several alternative measures have been employed in the literature to proxy accessibility. For instance, Jedwab and Storeygard (2022), in their analysis of the determinants of urban population growth, follow Donaldson and Hornbeck (2016) and construct a market access measure, a gravity-style index that combines population size and distance to urban centers to capture potential economic interactions. Other studies use least-cost market access, which identifies

access measures, distance to the nearest paved road offers a more exogenous measure of accessibility, with fewer embedded assumptions and a clearer causal interpretation.[5] It also facilitates a more straightforward application of instrumental variable strategies to address potential endogeneity.

To complement our continuous distance measure, we also construct a binary indicator for road proximity and a local road density measure, in the spirit of Ibisch et al. (2016). These alternative specifications are discussed in Section 4.3.4.

2.3 Other covariates

We include additional covariates to account for climatic conditions, conflict intensity, local economic development, and geographic characteristics.

Climatic covariates for temperature and precipitation are sourced from PRIO-GRID (Tollefsen et al., 2012). This dataset compiles annual grid-level information from multiple sources across a range of dimensions, spanning the period from 1980 to 2014. Grids are defined as 0.5 by 0.5 decimal degrees squares (corresponding to roughly 55km x 55km at the equator). We match DHS cluster coordinates to PRIO-GRID information by assigning the DHS point to the grid cell with the closest centroid.[6] As PRIO-GRID data are generally available at a yearly frequency, we match covariate values to the child's year of birth in the main analysis. For PRIO-GRID variables available at lower frequencies, each child is assigned the most recent available value from preceding years.

We control for local conflict intensity, sourcing variables from Uppsala Conflict Data Program's (UCDP) Georeferenced Event Dataset (GED, Sundberg and Melander, 2013). The data coverage

the lowest-cost destination based on travel time or expense, and has been applied to explain outcomes such as local GDP and agricultural productivity (Faber, 2014; Damania et al., 2018). In line with Djemaï et al. (2024), who examine the relationship between accessibility and well-being, we use Euclidean distance to the nearest paved road.

[5] By contrast, market access measures, while designed to reflect both geographic proximity and the economic weight of potential destinations, introduce several complications. First, they tend to obscure the interpretation of estimated coefficients, given their composite nature. More importantly, they are subject to greater endogeneity concerns: population and economic activity at destination points, core components of these indices, are likely to be influenced by the very infrastructure whose effects we seek to identify. This two-way relationship complicates causal inference, particularly in the case of health, where both population density and health outcomes can co-evolve with infrastructure investments. Furthermore, market access measures rest on behavioral assumptions about spatial interaction, typically that individuals interact more with larger and closer markets. These assumptions may not always hold in practice, for example, access to healthcare or food assistance may not follow a gravitational logic.

[6] We impose the constraint that the closest centroid must fall within the same country.

starts in 1989 and collects a systematic list of events of fatal violence occurring at a specific place and time, providing detailed information on the type of violence recorded (e.g., if state-induced, the actors are involved). We measure conflict intensity by assigning conflict events to a DHS cluster if an event occurred within a 25 kilometer wide buffer around the DHS point.[7] Our preferred measures consist in a dummy for the occurrence of any death (of any actor in the conflict) and a dummy for the occurrence of casualties of at least 10 civilians, in the child's year of birth.

To account for local economic development, we exploit nighttime lights data sourced from PRIO-GRID (Tollefsen et al., 2012), which reports annual mean nightlight intensity at the grid-cell level. The underlying data originate from the United States Air Force Defense Meteorological Satellite Program (DMSP) Operational Linescan System (OLS), which has recorded global nighttime light emissions since 1992. Nighttime lights have been widely employed as a proxy for local economic activity, particularly in contexts where disaggregated economic data are unavailable or official statistics are limited (Henderson et al., 2012). Finally, we retrieve elevation data from NASA's Shuttle Radar Topography Mission (STRM) Global 3 arc-second V003 dataset and compute slope estimates using ArcGIS to approximate for geographic characteristics.

2.4 Descriptive statistics

Table 1 presents summary statistics for the final sample, split by the median distance to the nearest paved road, which is close to 9 kilometers. Households located beyond this threshold seem to have lower height-for-age and weight-for-age Z-scores, indicating poorer child health. Additionally, in these households, mothers seem less likely to be mid- or highly educated. These households are also less frequently located in urban areas but, on average, have a similar distance to unpaved roads.

[7]GED reports the level of location precision at which the level could be recorded. We only keep conflict data if either: i) the exact location is available; ii) the event location is known to be within a 25 kilometer radius from a precise point; iii) the second-order administrative division where an event happened was known.

Table 1. Descriptive statistics, by road distance

	Road distance (paved)									
	Below median					***Above median***				
	Mean	Standard Dev.	N	Min	Max	Mean	Standard Dev.	N	Min	Max
Height for age, Z-score	-1.131	1.495	205591	-5.000	5.000	-1.427	1.558	205592	-5.000	5.000
Weight for age, Z-score	-1.032	1.240	205591	-5.000	5.000	-1.282	1.262	205592	-5.000	5.000
Gender: Male	0.506	0.500	205591	0.000	1.000	0.504	0.500	205592	0.000	1.000
Multiple birth	0.031	0.173	205591	0.000	1.000	0.028	0.164	205592	0.000	1.000
First child	0.220	0.414	205591	0.000	1.000	0.174	0.379	205592	0.000	1.000
Nr. sisters alive	1.087	1.254	205591	0.000	11.000	1.246	1.325	205592	0.000	11.000
Nr. brothers alive	1.087	1.251	205591	0.000	10.000	1.229	1.308	205592	0.000	9.000
Child age	29.954	17.220	205591	0.000	60.000	29.720	17.308	205592	0.000	60.000
Mother's age	29.645	6.819	205591	15.000	49.917	29.680	7.071	205592	15.000	49.917
Mother's age, squared	925.334	428.549	205591	225.000	2491.674	930.895	445.357	205592	225.000	2491.674
Mother's education: Low	0.335	0.472	205591	0.000	1.000	0.298	0.457	205592	0.000	1.000
Mother's education: Mid	0.231	0.421	205591	0.000	1.000	0.099	0.298	205592	0.000	1.000
Mother's education: High	0.033	0.179	205591	0.000	1.000	0.006	0.077	205592	0.000	1.000
Mother's literacy	0.072	0.258	205591	0.000	1.000	0.059	0.236	205592	0.000	1.000
Female headed household	0.192	0.393	205591	0.000	1.000	0.158	0.364	205592	0.000	1.000
Nightlights, PRIO-GRID	0.066	0.047	174507	0.000	0.569	0.046	0.027	186577	0.000	0.535
Precipitations, dev.	0.014	0.165	205591	-0.763	1.944	0.020	0.169	205592	-0.763	2.997
Precipitations, dev. squared	0.027	0.084	205591	0.000	3.779	0.029	0.151	205592	0.000	8.980
Temperature, dev.	0.006	0.032	205455	-0.324	0.429	0.007	0.031	205311	-0.311	0.463
Temperature, dev. squared	0.001	0.004	205455	0.000	0.184	0.001	0.003	205311	0.000	0.215
Slope	2.449	3.546	204877	0.000	37.414	2.972	4.631	204586	0.000	58.101
Conflict dummy, any death	0.070	0.255	199696	0.000	1.000	0.027	0.161	203833	0.000	1.000
Conflict dummy, civilian deaths >=10	0.015	0.122	199696	0.000	1.000	0.006	0.075	203833	0.000	1.000
Child's year of birth	2004.960	7.086	205591	1985.000	2014.000	2005.145	6.305	205592	1985.000	2014.000
Original road-data year	2003.240	6.613	205591	1985.000	2014.000	2003.376	5.975	205592	1985.000	2014.000
Distance DHS - MA centroid	4389.581	1907.446	205591	0.000	27904.204	4493.493	2114.251	205592	0.000	26994.751
Distance DHS - PRIO centroid	22356.668	8926.757	205591	0.000	64109.844	22041.320	9266.184	205592	292.372	98277.807
Urban area	0.464	0.499	205591	0.000	1.000	0.107	0.310	205592	0.000	1.000
Road distance, km, paved	2.775	2.471	205591	0.000	8.982	56.757	70.777	205592	8.985	955.588
Road distance, km, unpaved	10.954	13.098	205591	0.000	89.691	10.337	14.091	205592	0.000	138.032
IV, MST cities and mines	70955.746	86269.018	188193	0.621	673219.125	134174.981	117754.260	191919	3.553	950744.438
Far from consequential point	0.569	0.495	188193	0.000	1.000	0.854	0.353	191919	0.000	1.000
Distance to city or mine	85942.050	88095.991	188193	3.318	672829.250	154257.958	114028.339	191919	259.604	954654.938

Notes: Summary statistics based on the complete DHS sample. The sample is split by whether clusters are located below or above the median distance to the nearest paved road (8.985 kilometers). Overall, the mean height-for-age and weight-for-age Z-scores are –1.279 and –1.157 standard deviations, respectively.

3 Empirical methodology

3.1 Basic specification

We seek to estimate the relationship between road access and child health in Sub-Saharan Africa. To quantify this, we start with using the following ordinary-least-squared (OLS) equation:

$$\text{Health}_{ijcy} = \beta_0 + \beta_1 \text{DistanceRoad}_{jcy} + X'_{ijcy}\beta_2 + Z'_{jcy}\beta_3 + \delta_{my} + \mu_r + \epsilon_{ijcy} \tag{1}$$

where $Health_{ijcy}$ refers to either the height-for-age or weight-for-age Z-score of child i in DHS cluster j from country c born in year y, and $DistanceRoad_{jcy}$ refers to the distance in logarithms of DHS cluster j to the nearest paved road.[8] We include distance to the nearest paved road in logarithms to account for diminishing marginal effects, i.e. distance matters most at small values, matching thereby other behavioral processes such as travel time sensitivity. The main parameter of interest, β_1, captures how changes in road proximity (in %) impact the height-for-age and weight-for-age Z-scores of children aged 0 to 5 years. On average, if road proximity is expected to contribute to improved health, we would expect $\beta_1 < 0$.

In specification (1), we control for several confounders. Child-specific controls, X'_{ijcy}, include additional information obtained from the DHS about the child, the mother, and the household, i.e. the gender of the child, whether the child is the first born and/or part of a multiple birth, the number of sisters and brothers alive, the age of the child, the age, literacy, and education of the mother, and whether the household is female-headed. We address time gaps in our road data by interpolating between gaps of years, taking the previous observation as constant. We match the road data as closely as possible to the child's birth year and adjust for discrepancies by including the time difference between the birth year and the road data year, which we call the *Time distance to road data* variable. DHS-cluster-specific controls, Z'_{jcy}, include climatic variables for precipitation and temperature, local conflict intensity, slope, and nighttime lights.

Additionally, we include fixed effects (FEs) for birth month-year, δ_{my}, and DHS-region, μ_r. By

[8] In all regressions, we account for denormalized DHS weights to give more weight to countries with larger populations.

employing these FEs, we control for any regional time-invariant or seasonal and time-specific aggregate characteristics that stay unobserved. For example, for birth-date FEs, this means that we control for any seasonality effects related to birth timing that may influence child health. We include DHS-region FEs to control for regional differences that could affect child health, such as agricultural productivity.[9] By using these FEs, we compare individuals within the same geographic context, net of any aggregate time effects. Furthermore, we cluster standard errors, ϵ_{ijcy}, at the DHS-region level because households within the same region are likely not independent: they may share similar road access and be exposed to similar conditions, leading to potential correlation of errors.

3.2 Identification strategy

The relationship between road building and child health is most likely not exogenous as the placement of paved roads is non-random. Several factors contribute to this endogeneity. First, there might be reversed causality: proximity to roads improves health, but areas with better health may also be more likely to be connected to a road network. Second, there might be omitted variable bias (OVB), as unobserved factors such as economic development or agricultural productivity could influence both road paving and child health. Wealthier or more agriculturally productive areas, for instance, might have better health outcomes and may also be more likely to attract road investment. Third, measurement error might arise: there could be errors in the road data, either due to imprecise road digitization or displacement of DHS household locations.[10] If any of these sources of endogeneity hold, our OLS results will provide biased estimates.

To address these endogeneity concerns and identify causal relationships, we use an instrumental variable (IV) strategy. Equation (2) depicts the two-stage least squares (2SLS) regression:

[9] In our sample, there are 642 DHS regions. We could actually use more granular levels of fixed effects (e.g. DHS clusters) without altering much our main conclusions. However, as demonstrated in Section 3.3, the use of DHS-region FEs seems to be the right geographical level to be left with a meaningful variation in the road data.

[10] In the DHS, urban households are randomly displaced by 0 to 2 kilometers, and rural households by 0 to 5 kilometers. 1% of rural households are randomly displaced by 0 to 10 kilometers.

$$\begin{cases} \text{DistanceRoad}_{jcy} = \alpha_0 + \alpha_1 \text{DistanceIV}_{jcy} + X'_{ijcy}\alpha_2 + Z'_{jcy}\alpha_3 + \delta_{my} + \mu_r + \eta_{ijcy} \\ \text{Health}_{ijcy} = \beta_0 + \beta_1 \widehat{\text{DistanceRoad}}_{jcy} + X'_{ijcy}\beta_2 + Z'_{jcy}\beta_3 + \delta_{my} + \mu_r + \epsilon_{ijcy} \end{cases} \tag{2}$$

where $DistanceIV_{jcy}$ represents our IV: the distance in logarithms of DHS cluster j to the nearest hypothetical road.

For our IV, we construct hypothetical road networks using the "inconsequential units framework" (Chandra and Thompson, 2000; Faber, 2014; Redding and Turner, 2015). This approach states that roads are constructed between two economically important places. Economically smaller units located in between these two places do not influence the route choice but are still affected by the road's presence, making the road treatment random for these "inconsequential" places.

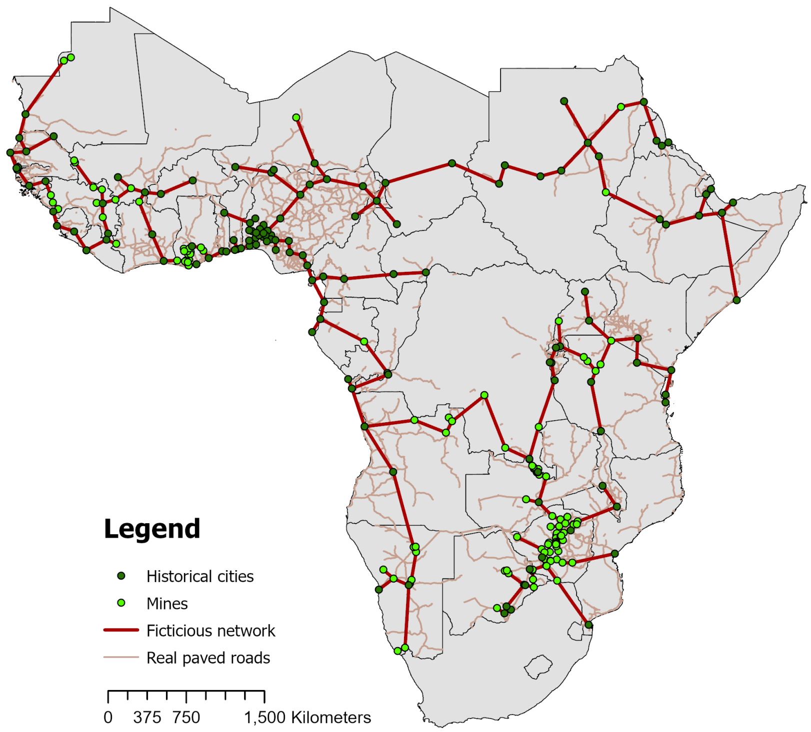
In our analyses we consider mines and historical cities to be consequential places. Hence, for each year in our analysis, we construct hypothetical road networks based on connecting historical cities[11] and active mines[12] to mimic the real road networks from 1980 to 2012. In this process, we apply a Euclidean Minimum Spanning Tree (EMST) algorithm which connects the nodes of our network in the most optimal way, thereby minimizing the total length of the network.[13] In other words, our hypothetical road networks represent a thought experiment envisioning what the networks would have looked like if political leaders across Sub-Saharan Africa had cooperated to optimally connect mines

[11]Historical city data are obtained from Jedwab and Moradi (2016). A historical city is considered as any city with more than 10,000 people in 1890 or 1900, the capital city, largest city, and second largest city at independence. By using historical cities we avoid reverse causality issues that could arise from population growth attracting roads or vice versa.

[12]Coordinates of mines are sourced from the S&P Capital IQ dataset, which capture mining projects and facilities worldwide. We select mines based on two criteria: (i) located in Sub-Saharan Africa, and (ii) having production data available at any point in time between 1980 and 2014. Due to the unavailability of historical city data for South Africa, Lesotho, and Eswatini, mines in these countries are also excluded. Potential concerns about reverse causality (where road paving might lead to mining discovery) are limited in our study for two reasons. First, mine locations are determined primarily by geological characteristics. Second, our analysis focuses on large-scale industrial mines, which are planned and developed over long time horizons. Small artisanal mines, which might be more responsive to short-term changes in infrastructure, are excluded. As a robustness check, we also construct a counterfactual network based on solely historical cities (Section 4.3.4).

[13]Rather than constructing networks based on Euclidean distances, one could also define linkages between consequential points using geographic constraints, such as elevation, to generate least-cost paths. We do not pursue this approach because geographic features such as slope and topography may influence child health through agricultural potential and food security, which could introduce endogeneity into an IV based on least-cost path networks.

Figure 2. Hypothetical road network and actual paved roads in 2000



Notes: The map shows the hypothetical Euclidean Minimum Spanning Tree (EMST) network, linking historical cities and active mines, alongside the actual paved roads in 2000. The EMST network is used to construct the instrumental variable.

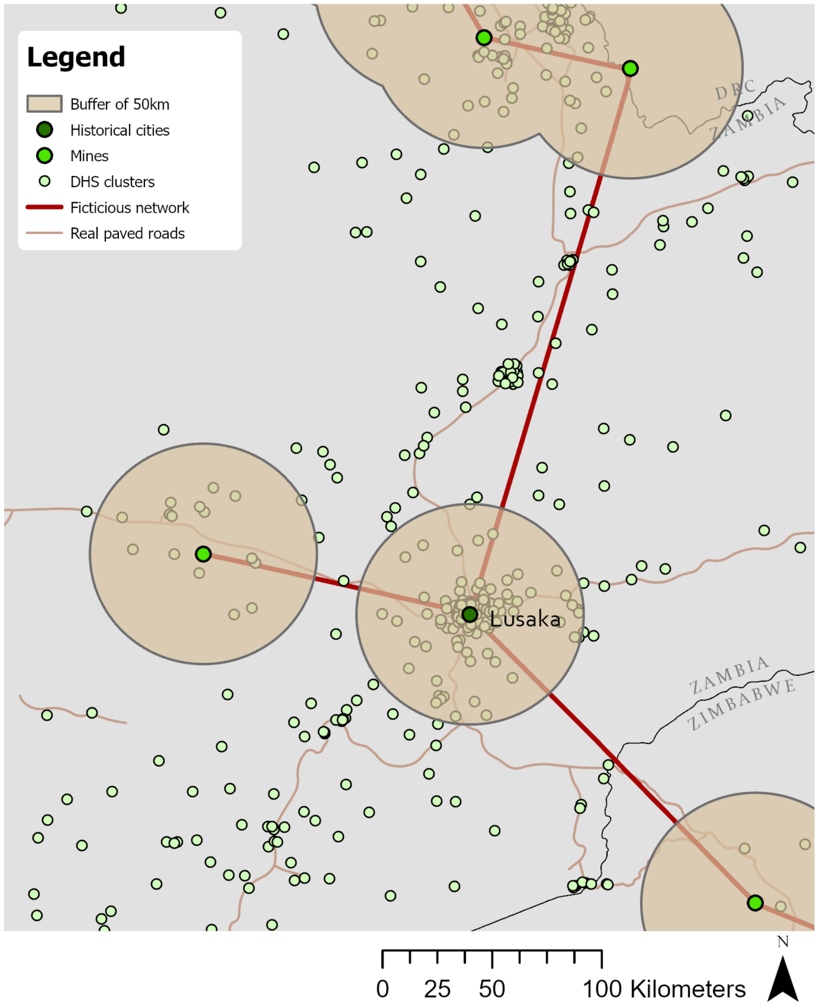
and cities without topographical constraints. Figure 2 depicts the correspondence between the EMST network generated for 2000 and the actual paved road network observed that same year.

While Jedwab and Moradi (2016) use only historical cities to create a hypothetical network, we extend their approach by including mines as additional nodes, recognizing that besides creating urban connectivity, roads in Sub-Saharan Africa are built to facilitate the extraction of mineral resources from mines for international export overseas (Bonfatti and Poelhekke, 2017; Bonfatti et al., 2022). In this way, we also make the IV network time-variant, accounting for the opening and closure of mines over time. As roads depreciate and anticipation effects may occur with new mine discoveries, we include mines up to two years before their opening and two years after their closure in constructing the EMST networks.[14]

Our IV is valid if it satisfies both the relevance condition and exclusion restriction. For the

[14]This assumption appears reasonable given that Lebrand et al. (2025) find that in the Democratic Republic of Congo, the quality of rehabilitated roads deteriorates by 60% within three years. As such, the EMST networks we construct likely underestimate the actual extent of usable road infrastructure, making our measure a conservative estimate.

Figure 3. Close-up of the hypothetical road network and actual paved roads in 2000



Notes: The map provides a close-up of the hypothetical Euclidean Minimum Spanning Tree (EMST) network, linking historical cities and active mines, alongside the actual paved roads in 2000. The EMST network is used to construct the instrumental variable. DHS clusters within 50 km of the nodes are excluded from the analyses.

relevance condition to hold, the IV must be strongly correlated with the actual road network distance. We expect this because we create road networks based on important points that political leaders or investors would likely prioritize for road construction. For the exclusion restriction to be respected, the IV should affect child health only through its impact on the distance to the real road network and not through any other channels. This requires the IV to be uncorrelated with any (un)observed determinants of health. The exclusion restriction would fail if there were factors that influence both the distance to the hypothetical road network and child health. To make this assumption more plausible and account for possible spillover effects, we exclude households belonging to DHS clusters within a 50-kilometer radius of the consequential points, i.e. the historical cities and active mines. The excluded clusters are illustrated in the example provided in Figure 3. We offer further identification checks in Section 4.3.2.

3.3 Residual variation in the road data and identification

Our main explanatory variable of interest is the distance to the nearest paved road at the moment of the child's birth, in the vicinity of the location where the mothers reside. Temporal variation in our measure depends on: i) record changes in road quality; ii) the timing of updates in the road data. Because the source data are not updated annually, we assign the most recently available road values to fill the intervening years.[15] Both these conditions imply a relatively high level of persistency in road variables. As discussed above, we include DHS-region fixed effects in our preferred specifications, and cluster standard errors at this level. This choice is motivated by a trade-off between obtaining highly granular fixed effects that capture most unobserved heterogeneity, while still leaving out meaningful variation in the road data to explore.

We summarize this trade-off in Figure 4, where we plot the residual variance in the road distance, after absorbing four alternative fixed-effects specifications. At the top-left, we only regress the log of road distance on the time distance between the birth year and the original road-data year and a constant. We cluster standard errors at the DHS-cluster ID level, which is survey-specific and thus varies with time. Most of the variation for the road distance is left unexplained, the residuals' distribution (in blue) mimics the variation of the regressand (log road distance, in green). At the top-right, we introduce DHS cluster FEs. This means that we rely on variation within the same DHS cluster in the year of births, combined with variation in road density given conditions i) and ii) mentioned above. This specification turns out to be too restrictive and misleading, as almost the entire variation in log-distance to paved roads is explained by this set of FEs (the regression's R^2 is above 99%). In grey, we add variation in the instrumental variable, given the same fixed effects specification. For this specification, the FE also absorb most of the IV's variation (the R^2 is above

[15]We control for a variable representing the time distance between the original road data year, and the matched year of birth. This variable captures the degree of precision in recording the timing of road updates.

98%). At the bottom-left, we replace DHS cluster FEs with the combination of DHS-region FEs and birth year-month FEs. We cluster standard errors at the regional level. In this case, the R^2 is around 33% (and around 48% for the IV), meaning that the amount of absorbed variation still leaves space to exploit the remaining variability in paved-road exposure and its instrumental variable. Finally, at the bottom-right, we use a combination of PRIO-GRID FEs and birth year-month FEs. We cluster standard errors at grid level. The grids are more granular than the DHS-regions, so this specification absorbs more variance (the R^2 is 64.3% for road distance, the R^2 is 80% for road distance). While some variation still stays unexplained with this last specification, the first-stage estimates would require a strong instrument. The relation between fictitious and actual roads that we exploit for our IV is in part static. We therefore choose DHS-regional FEs to avoid absorbing too much of the static but meaningful first stage variation. Within these regional FEs, we make a plausible assumption that any time-constant, omitted variable would then be accounted for thanks to the use of our instrument.

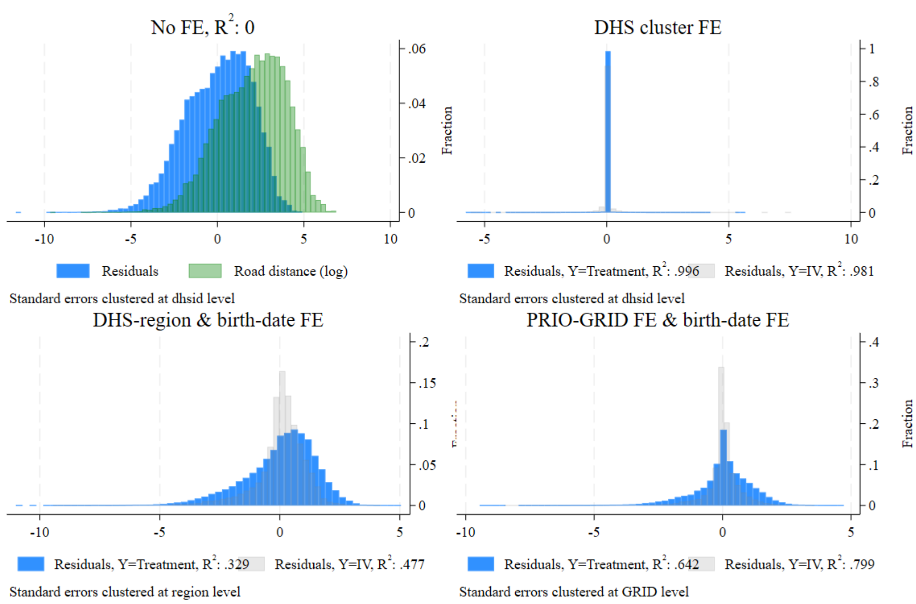


Figure 4. Fixed effects selection

4 Baseline results

4.1 OLS results

Table 2 presents the OLS regression results, progressively incorporating additional control variables across specifications. Column 1 estimates the relationship between the log distance to the nearest paved road on child health (height-for-age or weight-for-age Z-scores), while accounting for the potential temporal gap between the original year of recorded road changes and the child's birth year. Column 2 introduces climate and geographical controls, whereas column 3 further incorporates individual, household and maternal controls. Column 4 extends the regression by including conflict-related controls. Column 5 accounts for country-specific trends. In contrast, column 6 replaces the country trends with a variable capturing the total length of paved roads (in kilometers) within each country. Finally, column 7 introduces a dummy variable indicating whether a households is located within 100 kilometers of a mine, port, or historical city, along with an interaction term between this variable and road distance. The latter aims to assess potential spillover effects from mines, ports, or historical cities on child health outcomes.

Across all OLS specifications, we observe a statistically significant negative relationship between road distance and child health. In other words, children living closer to paved roads tend to exhibit better anthropometric outcomes, as reflected in higher height-for-age and weight-for-age Z-scores. The estimated coefficients range from -0.024 to -0.050 for height-for-age, and from -0.021 to -0.040 for weight-for-age. Using our preferred specification (column 6), a 10% shorter distance to the nearest paved road is, on average, correlated with higher Z-scores of 0.0031 SD and 0.0022 SD for height-for-age and weight-for-age, respectively. Results from column 7 further indicate that the correlation with height-for-age is more pronounced in areas located within 100 kilometers of a city, mine, or port. A more detailed discussion, particularly of non-linear effects within a causal framework, is provided in Section 4.2.

In general, the OLS coefficients suggest that proximity to paved roads is more strongly related to indicators of long-term health than to short-term conditions. The correlations are systematically larger for height-for-age than for weight-for-age, consistent with the idea that weight responds more readily to short-term changes in nutrition and living conditions, whereas height reflects cumulative health and nutrition over several years, particularly during early childhood and infancy. As a result, height may be more affected by the presence of paved roads, as better accessibility can have a lasting impact on long-term health.

Table 2. OLS results

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Dependent variable: Height-for-age						
Log road distance, paved	-0.050***	-0.049***	-0.032***	-0.033***	-0.032***	-0.031***	-0.024***
	(0.007)	(0.008)	(0.007)	(0.007)	(0.007)	(0.007)	(0.009)
Within 100km from city/port/mine=1							0.096***
							(0.027)
Within 100km from city/port/mine=1 × Log road distance, paved							-0.022**
							(0.010)
Observations	270854	269450	263061	263061	263061	263061	263061
Adjusted R^2	0.09	0.09	0.13	0.13	0.13	0.13	0.13
	Dependent variable: Weight-for-age						
Log road distance, paved	-0.040***	-0.039***	-0.024***	-0.024***	-0.022***	-0.022***	-0.021***
	(0.005)	(0.006)	(0.005)	(0.005)	(0.005)	(0.005)	(0.007)
Within 100km from city/port/mine=1							0.016
							(0.031)
Within 100km from city/port/mine=1 × Log road distance, paved							-0.004
							(0.008)
Time distance to road data	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Climate	No	Yes	Yes	Yes	Yes	Yes	Yes
Individual Controls	No	No	Yes	Yes	Yes	Yes	Yes
HH Controls & Nightlights	No	No	Yes	Yes	Yes	Yes	Yes
Mother Controls	No	No	Yes	Yes	Yes	Yes	Yes
Conflict	No	No	No	Yes	Yes	Yes	Yes
Geo	No	Yes	Yes	Yes	Yes	Yes	Yes
Observations	270854	269450	263061	263061	263061	263061	263061
Adjusted R^2	0.09	0.10	0.13	0.13	0.13	0.13	0.13
Country trends					✓		
GEO FE	✓	✓	✓	✓	✓	✓	✓
TIME FE	✓	✓	✓	✓	✓	✓	✓
Level of GEO-FE and S.E. cluster	Region	Region	Region	Region	Region	Region	Region

Notes: OLS results, weighted regressions with denormalized weights, DHS final sample. Observations within a 50km distance from consequential points, i.e., the nodes constituting the IV road network, are excluded. Stars correspond to the following p-values: * p < .10, ** p < .05, *** p < .01.

4.2 2SLS results

As previously discussed, the OLS results are likely subject to endogeneity due to the non-random placement of paved roads. To address this, we run 2SLS regressions using the distance to EMST networks connecting historical cities and mines as an instrument for actual paved roads. Results are presented in Table 3. The specifications of the columns follow the same structure as the OLS specifications in Table 2, with the exception of column 7, which is omitted here since mines and cities are inherently incorporated into the IV.

The first section of Table 3 presents the first-stage results with the Kleibergen-Paap F-statistic to assess the relevance of the IV. The second section of the table displays the main coefficients for height-for-age and weight-for-age, providing insights into the causal impact of road accessibility on child health.

The first-stage estimates and F-statistic indicate that the IV is strong across all specifications.[16] Furthermore, in each specification, the IV exhibits a positive and statistically significant relationship with real paved road distance. Specifically, a 10% reduction in the distance to the IV network is associated, on average, with a 1.56% to 1.69% decrease in the distance to the actual paved road network.

Since the IV satisfies the relevance condition and is assumed to meet the exclusion restriction, we interpret the estimates to capture causal effects. In line with the OLS results, Table 3 shows a negative relationship between paved road distance and child anthropometric outcomes, with robust coefficients across different specifications. The estimated coefficients range from -0.247 to -0.256 for height-for-age and from -0.093 to -0.104 for weight-for-age. Based on the preferred specification (rightmost column), a 10% reduction in the distance to the nearest paved road leads, on average, to a 0.0248 SD increase in height-for-age and a 0.0097 SD increase in weight-for-age.

To put these magnitudes in context, the average child in the full sample has height-for-age and weight-for-age Z-scores of -1.279 SD and -1.157 SD, respectively, and lives 29.8 kilometers from the

[16] For visualization, Figure A.3 presents overlaid histograms of distances (in kilometers) to the IV network and to actual paved roads.

nearest paved road. A 10% reduction in this distance (from 29.8 km to about 26.8 km) would therefore raise the average height-for-age to -1.254 SD and weight-for-age to -1.147 SD. Overall, these results suggest that improved road access has a meaningful positive impact on child health, particularly for height, which reflects longer-term health conditions.

Table 3. 2SLS results

First stage:

			Dependent variable: Road distance			
	(1)	(2)	(3)	(4)	(5)	(6)
Log IV, MST cities and mines	0.156**	0.169***	0.166***	0.165***	0.166***	0.167***
	(0.049)	(0.046)	(0.044)	(0.043)	(0.042)	(0.042)
Observations	270854	269450	263061	263061	263061	263061
Kleibergen-Paap F	10.14	13.60	14.32	14.64	15.43	15.64

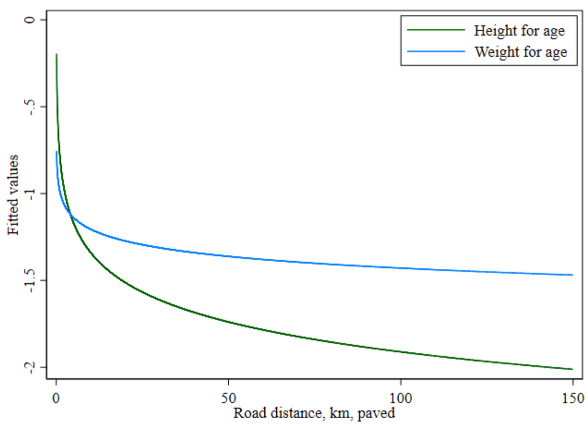
Second stage:

			Dependent variable: Height-for-age			
	(1)	(2)	(3)	(4)	(5)	(6)
Log road distance, paved	-0.256**	-0.254**	-0.247**	-0.247**	-0.253**	-0.248**
	(0.104)	(0.096)	(0.097)	(0.096)	(0.096)	(0.095)
Observations	270854	269450	263061	263061	263061	263061
			Dependent variable: Weight-for-age			
Log road distance, paved	-0.095	-0.104*	-0.095*	-0.094*	-0.093*	-0.097*
	(0.069)	(0.058)	(0.057)	(0.057)	(0.055)	(0.057)
Time distance to road data	Yes	Yes	Yes	Yes	Yes	Yes
Climate	No	Yes	Yes	Yes	Yes	Yes
Individual Controls	No	No	Yes	Yes	Yes	Yes
HH Controls & Nightlights	No	No	Yes	Yes	Yes	Yes
Mother Controls	No	No	Yes	Yes	Yes	Yes
Conflict	No	No	No	Yes	Yes	Yes
Geo	No	Yes	Yes	Yes	Yes	Yes
Observations	270854	269450	263061	263061	263061	263061
Country trends					✓	
GEO FE	✓	✓	✓	✓	✓	✓
TIME FE	✓	✓	✓	✓	✓	✓
Level of GEO-FE and S.E. cluster	Region	Region	Region	Region	Region	Region

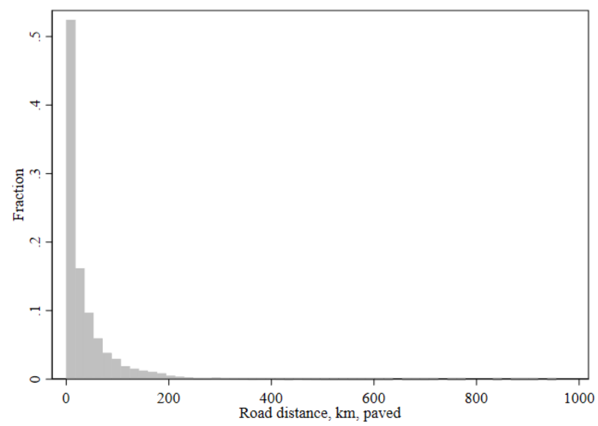
Notes: 2SLS results, weighted regressions with denormalized weights, DHS final sample. Observations within a 50km distance from consequential points, i.e., the nodes constituting the IV road network, are excluded. Stars correspond to the following p-values: * $p < .10$, ** $p < .05$, *** $p < .01$.

As our favorite specification includes road distance in logs, we study the non-linearity of the estimated effects in Figure 5. We observe a sharply decaying effect of road distance on both health outcome in subfigure (a), with a steeper decrease for weight-for-age than for height-for-age. After the first 100 kilometers from paved roads, the road impact fades away completely for weight-for-age, while it is still marginally present for height-for-age. As shown in the histogram of subfigure (b), most DHS respondents in the regression sample are located within the first 100 kilometers from roads. Accordingly, the health benefits of improved road access are most substantial for those children situated near paved roads.

Figure 5. Result visualization



(a) Effect visualization



(b) Road distance histogram, regression sample

Notes: Visualization of the effects of interest from the 2SLS. Subfigure a) shows the predicted values of height-for-age (green) and weight-for-age (blue) at different kilometers from the paved roads (for sake of visualization, the graph breaks at 150km). Subfigure b) shows the distribution of road distance in the regression sample.

When comparing OLS (Table 2) and IV (Table 3) estimates, the OLS results likely underestimate the true impact of road distance on child health. Three potential sources of endogeneity can explain this discrepancy. First, as previously discussed, reverse causality and omitted variable bias may influence the results. If an unobserved factor negatively or positively impacts both health and road infrastructure, it could cause the OLS coefficient to underestimate the true effect. For instance, poor local governance may lead to both inadequate road development and worse health outcomes, resulting in an underestimation of the true effect of road access. Second, measurement error in road network data and DHS cluster coordinates could introduce attenuation bias, causing OLS estimates to be biased toward zero. In particular, we can expect some measure discrepancies between real and recorded roads, which we recover from digitized Michelin maps as gathered by Jedwab and Storeygard (2022). These potential measurement errors would weaken the estimated relationship between road proximity and child health. Third, the IV and OLS estimates may capture different treatment effects. While OLS aims to estimate the Average Treatment Effect (ATE), the IV identifies the Local Average Treatment Effect (LATE) for households classified as compliers. These are those households whose actual road proximity was influenced by their proximity to the hypothetical EMST network. While LATE provides valuable insights into the treatment effect for compliers, it may not be fully generalizable to the entire population.

4.3 Robustness checks

To ensure the robustness of our findings, we have conducted several additional analyses, including: (i) using alternative dependent variables related to child health, (ii) controlling for potential confounding variables and excluding subsamples, (iii) performing a leave-one-country-out analysis, (iv) testing the IV as a binary variable for different distance thresholds, exploring alternative IV constructions, and using a paved road density variable. Additionally, we conduct two placebo analyses, one that randomly reassigns road proximity across observations within DHS regions and another that uses future road developments occurring after the child's birth Finally, we investigate sample selection issues, and in particular the role of selective migration on potential bias in our results.

4.3.1 Alternative dependent variables

Table 4 presents results using alternative child anthropometric indicators, specifically stunting, wasting, and underweight, along with their severe classifications. A child is classified as stunted when their height-for-age Z-score falls below -2 SD from the reference median and severely stunted when it is below -3 SD. Similarly, wasting is defined based on the weight-for-height Z-score, while underweight is determined by the weight-for-age Z-score, both following the same -2 SD and -3 SD thresholds.

The results in Table 4 indicate that increased proximity to paved roads reduces the likelihood of stunting, severe stunting, and underweight status. Specifically, an increase in distance to the nearest paved road from 1 km to 5 km (a change of 1.609 log units) is associated with a 0.113 increase in the probability of being stunted (11.3 percentage points), according to the IV estimates. This represents an increase by about 33% compared to the average stunting rate in the regression sample (0.34). These results are not generalizable to wasting-based indicators but are confirmed for underweight, which is more likely to capture short-term variations in food security. Overall, our findings suggest that access to road infrastructure continues to play an important role in particularly long-term health indicators such as stunting.

Table 4. Other outcomes

	Stunted	*Severely stunted*	*Wasted*	*Severely wasted*	*Underweight*	*Severely underweight*
				OLS		
Log road distance, paved	0.008***	0.005**	0.001**	0.000	0.007***	0.004***
	(0.002)	(0.002)	(0.001)	(0.000)	(0.002)	(0.001)
Observations	263061	263061	263061	263061	263061	263061
				2SLS		
Log road distance, paved	0.070**	0.048**	-0.017	-0.015	0.042**	0.010
	(0.024)	(0.015)	(0.019)	(0.011)	(0.018)	(0.008)
Time distance to road data	Yes	Yes	Yes	Yes	Yes	Yes
Climate	Yes	Yes	Yes	Yes	Yes	Yes
Individual Controls	Yes	Yes	Yes	Yes	Yes	Yes
HH Controls & Nightlights	Yes	Yes	Yes	Yes	Yes	Yes
Mother Controls	Yes	Yes	Yes	Yes	Yes	Yes
Conflict	Yes	Yes	Yes	Yes	Yes	Yes
Geo	Yes	Yes	Yes	Yes	Yes	Yes
Observations	263061	263061	263061	263061	263061	263061
Kleibergen-Paap F	15.64	15.64	15.64	15.64	15.64	15.64
GEO FE	✓	✓	✓	✓	✓	✓
TIME FE	✓	✓	✓	✓	✓	✓
S.e. cluster	Region	Region	Region	Region	Region	Region
Mean Outcome	0.34	0.14	0.09	0.02	0.27	0.06

Notes: OLS and 2SLS results, weighted regressions with denormalized weights, DHS final sample. Observations within a 50km distance from consequential points, i.e., the nodes constituting the IV road network, are excluded. A (severely) stunted child is defined as a child with height-for-age Z-score below (-3.0) -2 SD from the reference median. A (severely) wasted child has weight-for-height Z-score below (-3.0) -2 SD from the reference median. A (severely) underweight child has a weight-for-age Z-score below (-3.0) -2 SD from the reference median. Stars correspond to the following p-values: * p < .10, ** p < .05, *** p < .01.

4.3.2 IV robustness checks

In order to make our identification more credible, additional robustness checks for our main IV specification are reported in Appendix Table A.2. In columns 1 and 2, we control for the distance to mines and ports as well as the mother's height and weight. In columns 3 to 5, we exclude specific subsamples based on geographic and mobility constraints: (i) households within 100 kilometers of coastlines, (ii) households within 100 kilometers of consequential points, and (iii) relocated households in the past 10 years. The results for height-for-age remain largely robust under these additional specifications. However, the coefficient for weight-for-age loses statistical significance when controlling for maternal anthropometrics or when excluding coastal areas. Additionally, when we exclude all households within 100 kilometers of consequential points, the results become statistically insignificant. However, given that 100 kilometers is a considerable distance and most effects occur in the initial distances within this range (as shown in Figure 5), this exclusion removes a substantial portion of the sample and the strength of our IV.

4.3.3 Country sensitivity analysis

To further validate the IV approach, Appendix Figure A.4 reports the results of a leave-one-country-out analysis, where the IV specification is re-estimated after sequentially omitting each country. The first- and second-stage coefficients remain stable across specifications, indicating that the IV remains robust and is not driven by particular countries.

4.3.4 Binary IV and alternative specifications

We also test alternative constructions of the treatment and instrument. In Appendix Figure A.5, we modify the independent variable to a binary indicator, equal to 1 if the DHS cluster is located within a specified distance threshold (x-axis) from a paved road. The IV is then adjusted accordingly, taking a value of 1 if the DHS cluster falls within the same distance threshold from the EMST-generated network. The IV remains valid at thresholds of 40 kilometers and above, while the second-stage coefficients confirm the earlier findings that smaller distance thresholds are associated with greater improvements in children's height-for-age Z-scores compared to larger thresholds.

In addition to modifying the IV into a binary form, we explore alternative IV constructions as illustrated in Appendix Table A.3. These alternatives include constructing other EMST networks based on using different consequential points: (i) historical cities, mines, and ports; (ii) historical cities only; (iii) the minimum distance between the baseline IV and a historical city-only EMST. Additionally, we consider a four-nearest-neighbor network as an alternative to the minimum spanning tree. Although, this approach is less commonly used in the literature, it can potentially generate denser connections than the baseline EMST by forcing each city to connect to its 4 closest consequential points. Across these alternative constructions, the IV remains valid, as indicated by F-statistics exceeding the conventional threshold of 10. The corresponding second-stage estimates are slightly smaller in magnitude, and insignificant for the four-nearest-neighbor network.

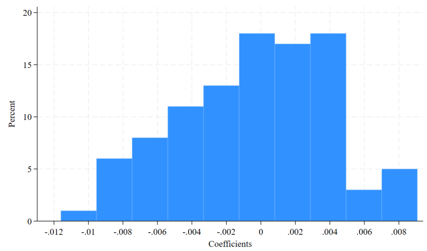
Finally, we propose a different measurement of the main independent variable to distance. Precisely, we compute a measure of paved road density, which represents the total paved road length within a 25km buffer around a DHS cluster. We maintain the same instrument and present the results in Appendix Table A.4. While the first stage slightly loses precision, the coefficients of road density are positive and significant for both our outcomes of interest, which is in line with our baseline results.

4.3.5 Placebos

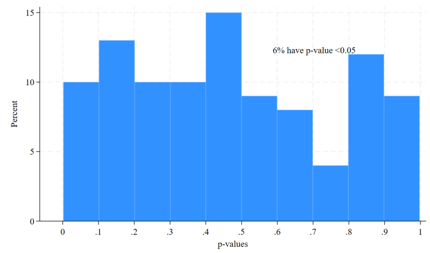
Results from our placebo exercises appear in Figure 6 and Table 5. Figure 6 shows the result from a randomized inference estimation that reshuffles road distances 100 times across geographic units within DHS regions. The specification used is equivalent of the one appearing in column 6 in the baseline results. In the figure, we report histograms of the coefficients (panel (a) and (c)) and their associated p-values (panel (b) and (d)), which represent the empirical distribution of these values. Top and bottom parts of the graph represent results for height-for-age and weight-for-age, respectively. We expect to find coefficients from randomized road distance to have an expected value of 0. We indeed observe the histogram in panels (a) and (c) to be centered around this value. Additionally, we can observe that our baseline coefficients (-0.031 and -0.022) constitute extremely rare cases for height-for-age and weight-for-age, respectively, if compared to the plotted distribution. This strengthens the validity of our findings.

Figure 6. Randomization inference placebo

Height-for-age

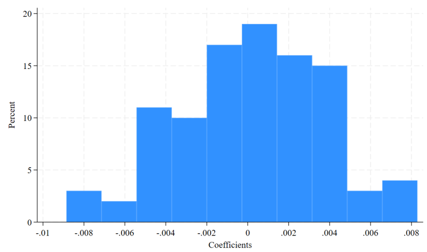


(a) Placebo β

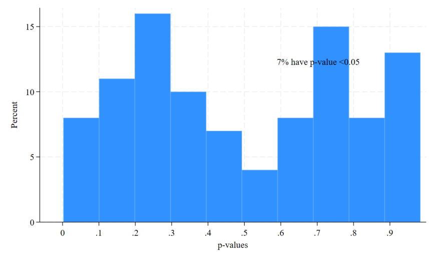


(b) Placebo p-values

Weight-for-age



(c) Placebo β



(d) Placebo p-values

Notes: Results from 100 regressions, with a placebo distance to paved roads. Placebo distances are obtained by permuting the original distances across DHS clusters within DHS regions. Results pertain to the preferred specification from the baseline table (OLS).

Table 5 presents results from a second placebo exercise, in which we add distances to paved roads recorded one year after the DHS survey interview took place (t+1), as well as road distances at the time of the interview. We first add these covariates individually (columns 1 and 2) and then jointly (column 3). We do not expect future (t+1) road updates to affect current malnutrition indices, except through their correlations with current and past road developments. We might expect some effect of paved road distances at interview year and children nutrition, in the case that recent road paving after their birth had immediate beneficial impacts on children. Note that depending on the child's age, road distance at birth and road distance at interview year could coincide.

Our findings in Table 5 reveal that height-for-age is only impacted by birth-year road paving, while no significant effects appear for future (t+1) road paving, nor for those linked to the interview year.[17] Regarding weight-for-age, we observe that interview year results absorb part of the significant

[17]In our analysis, we do not implement a Difference-in-Differences (DiD) framework, which we consider suboptimal in our setting due to data limitations. In particular, the exact year of road upgrading is often unknown, which could blur event-study estimates. Moreover, part of the effect of road access is static, therefore including fixed effects or time interactions in a DiD framework could absorb long-term effects, potentially underestimating the true impact. In

effect found at baseline. Note that while future road distances appear significant in column 2, this correlation seems to be completely derived by the correlation between future road updates with those observed in the interview year, as the coefficient drops in significance and switches sign in column 3. All in all, we conclude that the effects on height-for-age display more persistence, while weight-for-age appear to fluctuate more with variations on road changes. This is in line with height-for-age being a more stable index of long-term health, as anticipated in Section 2.1.

this context, the placebo using future road updates (t+1) helps assess whether pre-existing trends might influence the results. Specifically, if children living near roads that are upgraded in the future already exhibit different nutritional outcomes before the upgrade, this would indicate pre-existing trends unrelated to road improvements. However, as shown in Table 5 and elaborated in the main text, we do not find evidence of such pre-trends.

Table 5. Placebo: Future road paving

	Dependent variable: Height-for-age		
	(1)	(2)	(3)
Log road distance, paved	-0.034**	-0.032**	-0.032**
	(0.014)	(0.014)	(0.015)
Log road distance (interview t), paved	0.002		0.026
	(0.016)		(0.050)
Log road distance (t+1), paved		-0.001	-0.027
		(0.017)	(0.048)
Adjusted R^2	0.14	0.14	0.14
	Dependent variable: Weight-for-age		
Log road distance, paved	-0.015	-0.001	-0.000
	(0.010)	(0.010)	(0.011)
Log road distance (interview t), paved	-0.004		-0.046
	(0.010)		(0.031)
Log road distance (t+1), paved		-0.018*	0.027
		(0.011)	(0.030)
Time distance to road data	Yes	Yes	Yes
Climate	Yes	Yes	Yes
Individual Controls	Yes	Yes	Yes
HH Controls & Nightlights	Yes	Yes	Yes
Mother Controls	Yes	Yes	Yes
Conflict	Yes	Yes	Yes
Geo	Yes	Yes	Yes
Observations	219660	185192	185125
Country trends			
GEO FE	✓	✓	✓
TIME FE	✓	✓	✓
Level of GEO-FE and S.E. cluster	Region	Region	Region

Notes: OLS results, weighted regressions with denormalized weights, DHS final sample. Observations within a 50km distance from consequential points, i.e., the nodes constituting the IV road network, are excluded. Results pertain to the preferred specification from the baseline table (OLS). Stars correspond to the following p-values: * $p < .10$, ** $p < .05$, *** $p < .01$.

4.3.6 Sample selection

In our main specifications, we compare children within DHS regions, based on the distance of their residential location to the nearest paved roads. In the baseline estimates, we cannot exclude that mothers migrated to the interview areas after the paving of roads. If that was the case, one concern could be that migrants might be self-selected: those who have migrated from no-road to road areas might be systematically in better or worse health conditions than those that stayed in the no-road places. For example, if only healthier individuals migrated, this positive selection would lead us to overestimate the negative impact of road distance on health outcomes. Equivalently, another possibility is that mothers with pre-existing better (or worse) characteristics migrated from the road-areas to consequential points, therefore disappearing from our sample. This would lead to a negative (or positive) sample selection in the stayers at the treated areas, unless the equivalent selective migration had occurred across all road distances.

As a solution to this, we look into the years since arrival in the place of interview and explore selective migration in Table 6. First, we look at a short-term (ST) migration response after road paving (ST-named columns in the table). We build a dummy that equals 1 if the mother immigrated to the place of interview within 5 years from the road updating. Additionally, we investigate a long-term (LT) migration response after road paving (LT-named columns in the table), with a dummy that takes the value 1 if the mother immigrated to the place of interview after 5 years or more since the road paving.

We study whether ST or LT migration response depends on the area being a consequential point and whether this dependence is stronger based on mothers' height and weight (left-hand side of the table). Furthermore, we analyse whether the likelihood of migrating depends on road distance and whether this is based on mothers' height and weight (right-hand side of the table).

We find no evidence of selective migratory response in the short run towards consequential points, as shown in the insignificant coefficients in column 1 and 3. However, we observe that immigration is more likely if an area is a consequential point (column 2 and 4), and migrants in consequential points are negatively selected with respect to the incumbent individuals (as the interaction coefficients are negative). In addition, we find no evidence that mothers' height influences how paved roads affect the likelihood of migrating in the short run (column 5) nor in the long run (column 6). However, we observe that mothers of lower weight are more likely to be immigrants, but this difference decreases at higher road distances. Immigration is less likely in areas of higher road distance, for mothers with lowest weight, but increases for mothers of greater weight. This selective pattern on mothers' weight is not found in the long run.

Finally, we complement these tables with Figure A.6, in which we estimate the heterogeneous effects of road paving on height-for-age (left panels) and weight-for-age (right panels), depending on the paving being recent of not (meaning that it occurred at most 3 years before the child birth), as presented in the top panels, and depending on the mothers living in the area before the road update, as shown in the bottom panels. Focusing on the IV results, which are plotted in red, we find no significant differences in the results across recent and past road updates and between immigrants and incumbents.

We conclude that while fluctuations in weight-for-age could be partially resulting from short-term migration responses after road updates, height-for-age results appear stable and robust to any selective migration patterns analyzed. The negative and significant coefficient is maintained if we focus on individuals that did not move after road paving at destination. This reassures us against selectivity concerns. We should nonetheless acknowledge that we cannot observe exactly the origin and destination trajectories that migrants undertook with the data at hand.

Table 6. Selection

	Immigrant selection to consequential points (OLS)				Immigrant selection to road areas (IV)			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	ST	***LT***	***ST***	***LT***	***ST***	***LT***	***ST***	***LT***
Consequential point=1	0.027	0.268***	-0.005	0.089***				
	(0.039)	(0.064)	(0.008)	(0.021)				
Mother's height	0.002	0.019			-0.046	-0.098		
	(0.013)	(0.022)			(0.070)	(0.164)		
Consequential point=1 × Mother's height	-0.015	-0.150***						
	(0.024)	(0.039)						
Mother's Weight			-0.000	0.001***			-0.002**	0.001
			(0.000)	(0.000)			(0.001)	(0.001)
Consequential point=1 × Mother's Weight			0.000	-0.001***				
			(0.000)	(0.000)				
Log road distance, paved					-0.051	-0.053	-0.068*	0.021
					(0.050)	(0.108)	(0.036)	(0.057)
Log road distance, paved × Mother's height					0.019	0.050		
					(0.027)	(0.065)		
Log road distance, paved × Mother's Weight							0.001**	0.000
							(0.000)	(0.001)
Climate	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Individual Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
HH Controls & Nightlights	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Mother Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Conflict	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Geo	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	250243	250243	250517	250517	178287	178287	178468	178468
Kleibergen-Paap F					7.22	7.22	6.89	6.89
GEO FE	✓	✓	✓	✓	✓	✓	✓	✓
TIME FE	✓	✓	✓	✓	✓	✓	✓	✓
S.e. cluster	Region	Region	Region	Region	Region	Region	Region	Region

Notes: OLS and 2SLS results, weighted regressions with denormalized weights, DHS final sample, mothers only. Consequential points are the areas around the nodes constituting the IV road network. We consider consequential points those areas within 20km from the centroid of the node. The outcome variable in the first four columns represents a short-term migration response after road paving. It is a dummy that equals 1 if the mother immigrated to the place of interview within 5 years from the road updating. The outcome variable in columns 5 to 8 represents a long-term migration response after road paving. It is a dummy that takes the value 1 if the mother immigrated to the place of interview after 5 years or more since the road paving. To study migration selectivity towards consequential points, we interact the consequential point dummy with the mother's height (columns 1 to 2) or weight (columns 3 to 4). To study migration selectivity towards road areas, we interact the road distance with the mother's anthropometrics (height in columns 5 and 6, weight in columns 7 and 8). Stars correspond to the following p-values: * $p < .10$, ** $p < .05$, *** $p < .01$.

5 Mechanisms

Our analysis has shown that proximity to paved roads improves child anthropometric outcomes. To better understand the pathways underlying this relationship, we re-estimate the baseline model using the IV framework with alternative outcomes capturing potential mechanisms. Table 7 summarizes the results. Three mechanisms emerge as particularly relevant: (i) improved healthcare access and utilization, (ii) higher household wealth and structural transformation; and (iii) agricultural expansion. We also test for potential adverse health effects, including deforestation, air pollution, and malaria exposure, but find no evidence of such negative spillovers.[18]

Healthcare accessibility. Proximity to paved roads is associated with greater utilization of maternal and child health services. Specifically, mothers living closer to paved roads are more likely to receive antenatal visits (panel A) and postnatal check-ups (panel B). Likewise, children in these areas are more likely to receive all appropriate vaccinations (panel C). In contrast, distance to health facilities when being sick is not commonly reported as a barrier to seeking care (Panel D). These findings indicate that improved road connectivity facilitates access and utilization to healthcare for both mothers and children, consistent with prior evidence linking transport infrastructure to better healthcare accessibility (Aggarwal, 2021; Dasgupta et al., 2024).

Household wealth and labor. Improved road access is also linked to higher household wealth and changing labor patterns. Households located closer to paved roads exhibit higher wealth, as measured by the DHS wealth index (panel E). Moreover, road proximity increases the likelihood of women engaging in paid work (panel G) and reduces reliance on agricultural employment for both mothers and partners (panels H–I). These results are consistent with structural transformation dynamics, in which road infrastructure facilitates labor reallocation toward non-farm activities (Fiorini and Sanfilippo, 2022; Gertler et al., 2024). Higher incomes, in turn, can enable households to invest more

[18]Tables A.5 to A.10 present the detailed results, including outcomes not discussed here.

in nutrition, sanitation, and medical care, contributing to better child health.[19]

Agricultural expansion. A third mechanism operates through agriculture. Improved connectivity can promote cropland expansion, crop diversification, and greater agricultural productivity (Stifel and Minten, 2017; Berg et al., 2018; Shamdasani, 2021; Gebresilasse, 2023; Adamopoulos, 2025), thereby strengthening local food security and resilience. In our data, proximity to paved roads is associated with larger cropland areas (panel J), suggesting potentially improved local food availability. However, we find no evidence that road access is linked to changes in dietary diversity among children (panel K), crop composition or production techniques used.[20]

Negative externalities. Finally, we assess whether the positive health effects of road access are partly mitigated by environmental or epidemiological externalities. Prior studies have linked road construction to deforestation (Damania et al., 2018; Kleinschroth et al., 2019; Baehr et al., 2023), which can disrupt local biodiversity and food systems, and to greater disease exposure (Djemaï, 2018). Other research documents the detrimental effects of air pollution on child health (Rosales-Rueda and Triyana, 2019; Heft-Neal et al., 2025), which may intensify with higher traffic volumes and roadside activities associated with improved road access. In contrast, our results show no systematic evidence of such adverse effects (panels L-N), suggesting that, within our study period and country sample, the positive health effects of road infrastructure were not offset by the potential environmental or epidemiological channels analyzed.[21,22]

[19]We also test whether road proximity is associated with improved access to electricity and reduced travel time to the main water source, but find no significant effects (Appendix Table A.8).

[20]Using the Spatial Production Allocation Model (SPAM) dataset (You et al., 2014), we test several indicators for whether households closer to paved roads cultivate different crops or adopt more advanced production methods, but find no significant differences (Appendix Table A.10).

[21]The lack of significant change in deforestation echoes the findings from Asher et al. (2020) who find no causal effect of new rural roads on local deforestation in India. We should nonetheless acknowledge that the deforestation data are rather coarse. Cloud cover, topography, and spectral confusion with crops or shrublands can produce classification errors, particularly in tropical Africa (Molinario et al., 2017). We should also recognize that the mitigating role of deforestation is complex. Although deforestation might be associated with agricultural expansion, it can also have a detrimental effect on food security.

[22]We proxy air pollution using a DHS indicator equal to 1 if the child had a cough within the past 24 hours or two weeks, following Gennaioli and Narciso (2025), who employ a related measure based on whether the child experienced short, rapid breathing as an indicator of respiratory effects of air pollution.

Table 7. Summary table on mechanisms

	(1)	(2)	(3)
	Coef.	(SE)	[Obs.]
I: Healthcare access			
A. Antenatal visit[a]	-0.107*	(0.057)	[175,714]
B. Check post birth[b]	-0.114**	(0.042)	[111,082]
C. Child vaccine[c]	-0.053*	(0.030)	[144,481]
D. Distance to health facility[d]	0.023	(0.038)	[144,892]
II: Wealth and labor			
E. Wealth index[e]	-0.355**	(0.158)	[155,217]
F. Any Work[f]	0.057	(0.035)	[191,046]
G. Paid work[g]	-0.088**	(0.042)	[112,567]
H. Agriculture[h]	0.121*	(0.072)	[191,046]
I. Agriculture Partner | Work[i]	0.069**	(0.03)	[173,726]
III: Agricultural expansion			
J. Share Cropland[j]	-17.864***	(3.428)	[36,356]
K. Child diet diversity[k]	0.042	(0.039)	[102,082]
IV: Negative externalities			
L. Forest loss[l]	-0.263	(0.623)	[32,592]
M. Malaria[m]	0.013	(0.014)	[33,134]
N. Child cough[n]	0.024	(0.025)	[262,184]

Notes: Only the coefficients for the main variable of interest are reported, with side-lined associated standard errors in parentheses and number of observations in squared brackets. Birth month-year and DHS-region fixed effects, as well as climatic, conflict controls are included. Ind. and HH controls are also included, unless the analysis is implemented at the cell level (see "III. Land use"): [a] Col. (1) in Table A.5; [b] Col. (2) in Table A.5; [c] Col. (3) in Table A.5; [d] Col. (3) in Table A.6; [e] Col. (1) in Table A.7; [f] Col. (2) in Table A.7; [g] Col. (3) in Table A.7; [h] Col. (4) in Table A.7; [i] Col. (6) in Table A.7; [j] Col. (3) in Table A.9; [k] Col. (4) in Table A.5; [l] Col. (1) in Table A.9; [m] Col. (2) in Table A.9; [n] Col. (5) in Table A.5;. Stars correspond to the following p-values: * p < .10, ** p < .05, *** p < .01.

6 Conclusions

This study examines the impact of road infrastructure on child health in Sub-Saharan Africa, combining geolocated data on child anthropometric measures with information on road networks between 1980 and 2012. By employing an IV approach based on the inconsequential units framework and constructing hypothetical road networks between historical cities and active mines, we mitigate endogeneity concerns and identify the causal effects of road proximity on child health outcomes. The results show that proximity to paved roads significantly improves child health, as measured by height-for-age and weight-for-age Z-scores, with the effect being stronger for height, a long-term indicator of nutritional and health status. Moreover, the estimated relationship is non-linear, decaying sharply with distance and concentrated within the first 100 kilometers of paved roads, where most surveyed households are located.

Our mechanism analysis suggests that these improvements are primarily driven by enhanced access to and utilization of healthcare services, higher household wealth accompanied by early signs of structural transformation, and cropland expansion. Households living closer to paved roads are more likely to receive antenatal and postnatal care, experience higher wealth levels, and engage less in agricultural labor and more in paid work, while road-adjacent areas exhibit greater cropland shares, suggesting improved food availability. We find no systematic evidence that road access exacerbates adverse health externalities, such as deforestation, air pollution, or malaria exposure. Together, these results reinforce the role of transport infrastructure as a catalyst for both economic and human development. A tangible example of how our findings translate into daily life come from a road upgrading project in the Democratic Republic of Congo between Kinshasa, Kwango and Kwilu, where a project implementation coordinator explains: "*Today, people living along the road can get more value from their daily produce. They can sell more easily because vehicles now have direct access to their villages. One mother, for example, no longer needs to travel to Kinshasa or the market to sell a bag of cassava or charcoal: she can sell it in front of her house. It's a real change in their daily lives*" (African Development Bank Group, 2025).

While our findings underscore the positive role of road access in improving child health, important questions remain. In particular, potential negative externalities of road construction, affecting people through environmental degradation and unequal distribution of benefits, remain largely understudied. Given the substantial investments in road infrastructure across Sub-Saharan Africa by, for instance, the World Bank, the African Development Bank, and China's Belt and Road Initiative, developing a more comprehensive understanding of the multifaceted impacts of roads represents an important avenue for future research.

Bibliography

Acharya, Y., S. Naz, L. P. Galway, and A. D. Jones (2020). Deforestation and Household- and Individual-Level Double Burden of Malnutrition in Sub-saharan Africa. *Frontiers in Sustainable Food Systems 4*.

Adamopoulos, T. (2025). Spatial Integration and Agricultural Productivity: Quantifying the Impact of New Roads. *American Economic Journal: Macroeconomics 17*(1), 343–378.

African Development Bank Group (2025, August). Renovated National Road 1 between Kinshasa, Kwango and Kwilu boosting economic activity in south-west Democratic Republic of Congo. Accessed: 2025-08-22.

Aggarwal, S. (2021). The long road to health: Healthcare utilization impacts of a road pavement policy in rural India. *Journal of Development Economics 151*, 102667.

Ali, R., A. F. Barra, C. Berg, R. Damania, J. Nash, and J. Russ (2015). *Highways to Success or Byways to Waste: Estimating the Economic Benefits of Roads in Africa.* Washington, DC: World Bank; and Agence Française de Développement.

Allen, T. and D. Atkin (2022). Volatility and the Gains from Trade. *Econometrica 90*(5), 2053–2092.

Asher, S., T. Garg, and P. Novosad (2020). The Ecological Impact of Transportation Infrastructure. *The Economic Journal 130*(629), 1173–1199.

Asher, S. and P. Novosad (2020). Rural Roads and Local Economic Development. *American Economic Review 110*(3), 797–823.

Baehr, C., A. BenYishay, and B. Parks (2023). Highway to the forest? Land governance and the siting and environmental impacts of Chinese government-funded road building in Cambodia. *Journal of Environmental Economics and Management 122*, 102898.

Banerjee, A., E. Duflo, and N. Qian (2020). On the road: Access to transportation infrastructure and economic growth in China. *Journal of Development Economics 145*, 102442.

Berg, C. N., B. Blankespoor, and H. Selod (2018). Roads and Rural Development in Sub-Saharan Africa. *The Journal of Development Studies 54*(5), 856–874.

Bonfatti, R., Y. Gu, and S. Poelhekke (2022). Priority Roads: the Political Economy of Africa's Interior-to-Coast Roads. *STEG Working Paper WP031*.

Bonfatti, R. and S. Poelhekke (2017). From mine to coast: Transport infrastructure and the direction of trade in developing countries. *Journal of Development Economics 127*, 91–108.

Chandra, A. and E. Thompson (2000). Does public infrastructure affect economic activity?: Evidence from the rural interstate highway system. *Regional Science and Urban Economics 30*(4), 457–490.

Damania, R., J. Russ, D. Wheeler, and A. F. Barra (2018). The Road to Growth: Measuring the Tradeoffs between Economic Growth and Ecological Destruction. *World Development 101*, 351–376.

Dasgupta, A., A. Karandikar, and D. Raghav (2024). Road access, fertility, and child health in rural india. *Population and Development Review 50*(1), 117–147.

Deaton, A. (2007). Height, health, and development. *Proceedings of the National Academy of Sciences 104*(33), 13232–13237.

Djemaï, E. (2018). Roads and the spread of hiv in africa. *Journal of Health Economics 60*, 118–141.

Djemaï, E., A. E. Clark, and C. D'Ambrosio (2024). Take the highway? Paved roads and well-being in Africa. *World Development 183*, 106691.

Donaldson, D. (2018). Railroads of the Raj: Estimating the Impact of Transportation Infrastructure. *American Economic Review 108*(4-5), 899–934.

Donaldson, D. and R. Hornbeck (2016). Railroads and american economic growth: A "market access" approach. *The Quarterly Journal of Economics 131*(2), 799–858.

Dumas, C. and X. Játiva (2025). Better roads, better off? Evidence on improving roads in Tanzania. *The World Bank Economic Review 39*(1), 104–123.

Duranton, G. and M. A. Turner (2012). Urban Growth and Transportation. *Review of Economic Studies 79*(4), 1407–1440.

Faber, B. (2014). Trade Integration, Market Size, and Industrialization: Evidence from China's National Trunk Highway System. *Review of Economic Studies 81*, 1046–1070.

Fiorini, M. and M. Sanfilippo (2022). Roads and Jobs in Ethiopia. *The World Bank Economic Review 36*(4), 999–1020.

Gebresilasse, M. (2023). Rural roads, agricultural extension, and productivity. *Journal of Development Economics 162*, 103048.

Gennaioli, C. and G. Narciso (2025). Toxic roads. *Environmental and Resource Economics 88*(6), 1773–1803.

Gertler, P. J., M. Gonzalez-Navarro, T. Gracner, and A. Rothenberg (2024). Road Maintenance and Local Economic Development. *Journal of Urban Economics 143*, 103687.

Goodman, S., S. Zhang, A. Malik, B. Parks, and J. Hall (2024). AidData's Geospatial Global Chinese Development Finance Dataset. *Scientific Data 11*, 529.

Graff, T. (2024). Spatial Inefficiencies in Africa's Trade Network. *Journal of Development Economics 171*, 103319.

Hansen, H., P. Potapov, R. Moore, M. Hancer, S. Turubanova, A. Tyukavina, D. Thau, S. Stehman, S. Goetz, T. Loveland, A. Kommareddy, A. Egorov, L. Chini, C. Justice, and J. Townshed (2013). High-Resolution Global Maps of 21st-Century Forest Cover Change. *Science 342*(6160), 850–853.

Hay, S. I. and R. W. Snow (2006). The Malaria Atlas Project: Developing Global Maps of Malaria Risk. *PLoS Med 3*(12), e473.

Heft-Neal, S., M. Philipp Heger, V. Rathi, and M. Burke (2025). Identifying child growth effects of elevated pollution levels during pregnancy. *Environmental Research Letters 20*(1), 014019.

Henderson, J., A. Storeygard, and D. Weil (2012). Measuring Economic Growth from Outer Space. *American Economic Review 102*(2), 994–1028.

Ibisch, P. L., M. T. Hoffmann, S. Kreft, G. Pe'er, V. Kati, L. Biber-Freudenberger, D. A. DellaSala, M. M. Vale, P. R. Hobson, and N. Selva (2016). A global map of roadless areas and their conservation status. *Science 354*(6318), 1423–1427.

Jayachandran, S. (2022). How Economic Development Influences the Environment. *Annual Review of Economics 14*(1), 229–252.

Jedwab, R. and A. Moradi (2016). The Permanent Effects of Transportation Revolutions in Poor Countries: Evidence from Africa. *The Review of Economics and Statistics 98*(2), 268–284.

Jedwab, R. and A. Storeygard (2022). The Average and Heterogeneous Effects of Transportation Investments: Evidence from Sub-Saharan Africa 1960–2010. *Journal of the European Economic Association 20*(1), 1–38.

Kleinschroth, F., N. Laporte, W. F. Laurance, S. J. Goetz, and J. Ghazoul (2019). Road expansion and persistence in forests of the Congo Basin. *Nature Sustainability 2*(7), 628–634.

Lebrand, M., H. Mueller, P. Schouten, and J. Steinbuks (2025). Road Investment and Violence in DRC: Perishable Peace Dividends. *World Bank Policy Research Working Paper 11071*.

Li, J. (2025). Chinese Aid for Transportation Infrastructure and Child Health in Africa. *Health Economics 00*, 1–15.

Molinario, G., M. C. Hansen, P. V. Potapov, A. Tyukavina, S. Stehman, B. Barker, and M. Humber (2017). Quantification of land cover and land use within the rural complex of the democratic republic of congo. *Environmental Research Letters 12*(10), 104001.

Moradi, A. (2010). Nutritional status and economic development in sub-saharan africa, 1950–1980. *Economics Human Biology 8*(1), 16–29.

Perkins, J. M., S. Subramanian, G. Davey Smith, and E. Özaltin (2016). Adult height, nutrition, and population health. *Nutrition Reviews 74*(3), 149–165.

Redding, S. J. and M. A. Turner (2015). Chapter 20 - Transportation Costs and the Spatial Organization of Economic Activity. In G. Duranton, J. V. Henderson, and W. C. Strange (Eds.), *Handbook of Regional and Urban Economics*, Volume 5 of *Handbook of Regional and Urban Economics*, pp. 1339–1398. Elsevier.

Rosales-Rueda, M. and M. Triyana (2019). The Persistent Effects of Early-Life Exposure to Air Pollution: Evidence from the Indonesian Forest Fires. *Journal of Human Resources 54*(4).

Shamdasani, Y. (2021). Rural road infrastructure & agricultural production: Evidence from india. *Journal of Development Economics 152*, 102686.

Stifel, D. and B. Minten (2017). Market Access, Well-being, and Nutrition: Evidence from Ethiopia. *World Development 90*, 229–241.

Strauss, J. and D. Thomas (1998). Health, Nutrition, and Economic Development. *Journal of Economic Literature 36*(2), 766–817.

Strauss, J. and D. Thomas (2008). Health over the Life Course. In P. Schultz and J. Strauss (Eds.), *Handbook of Development Economics*, Volume 4, pp. 3375–3484. Amsterdam: Elsevier.

Sundberg, R. and E. Melander (2013). Introducing the UCDP Georeferenced Event Dataset. *Journal of Peace Research 50*(4), 523–532. Publisher: SAGE Publications Ltd.

Tollefsen, A. F., H. Strand, and H. Buhaug (2012). PRIO-GRID: A unified spatial data structure. *Journal of Peace Research 49*(2), 363–374. Publisher: SAGE Publications Ltd.

United Nations Children's Fund (UNICEF), World Health Organization (WHO), and International Bank for Reconstruction and Development/The World Bank (2023). Levels and trends in child malnutrition: UNICEF / WHO / World Bank Group Joint Child Malnutrition Estimates: Key findings of the 2023 edition. Technical report, UNICEF and WHO, New York. CC BY-NC-SA 3.0 IGO.

You, L., S. Wood, U. Wood-Sichra, and W. Wu (2014). Generating global crop distribution maps: From census to grid. *Agricultural Systems 127*, 53–60.

Appendix Tables and Figures for:

Roads and child health in Sub-Saharan Africa

October 24, 2025

This appendix contains the supplementary tables and figures cited in the main text.

Figure A.1. Sub-Saharan African (SSA) countries included in our study

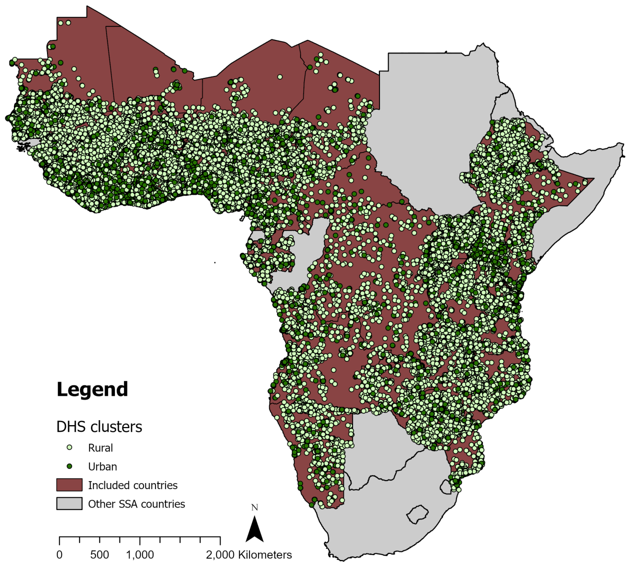


Figure A.2. Classification of road types

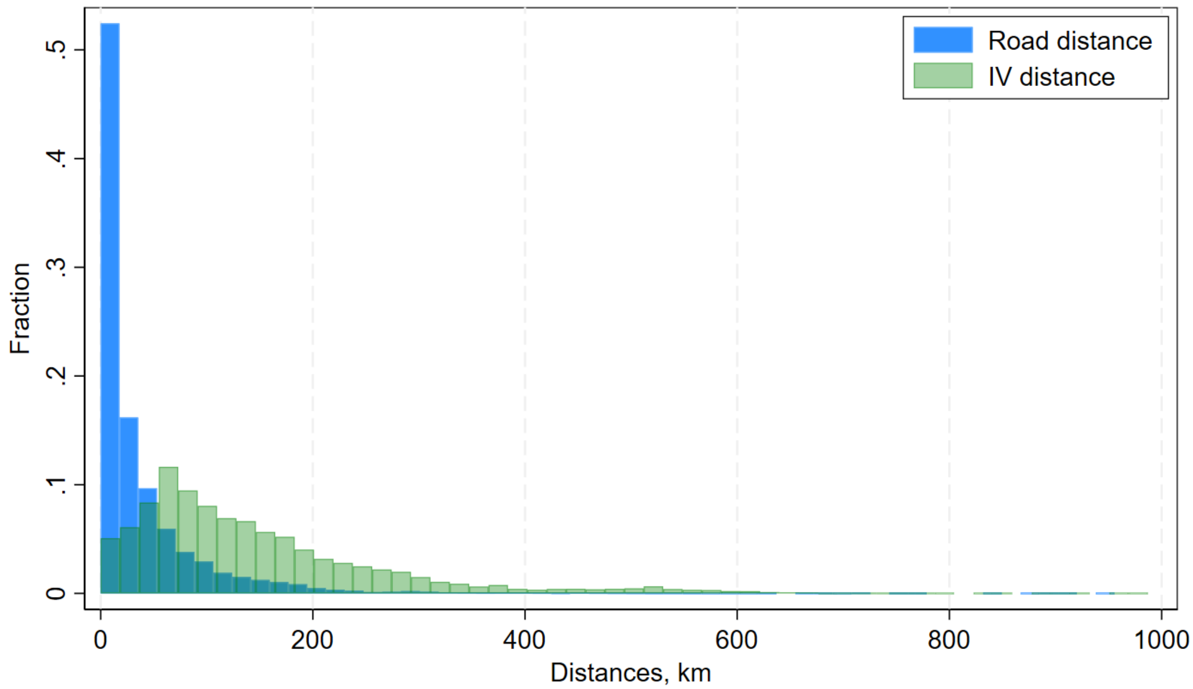
Notes: Obtained from teaching slides of Jedwab and Storeygard (2022).

Table A.1. DHS, available geolocalized BR data, final sample

	1985	1986	1987	1988	1989	1990	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	Total
AO	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	44	1004	1221	0	0	2269
BF	0	0	0	561	759	720	833	997	105	504	641	641	714	1325	1566	1622	1500	1613	1328	0	207	1123	1208	1282	1253	1018	0	0	0	0	21520
BJ	0	0	0	0	0	0	0	0	251	638	871	671	586	676	794	860	778	0	0	0	0	0	1106	1539	1400	1509	1495	120	1898	2139	17331
BU	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	46	644	662	703	749	603	79	1159	0	0	4645
CD	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	232	601	616	653	717	262	24	1274	1413	1465	1491	0	0	8748
CF	0	0	0	0	0	0	43	681	716	800	26	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2266
CI	0	0	0	0	0	0	190	1116	1104	1146	254	289	353	324	28	0	0	0	0	0	0	0	364	535	662	655	744	116	0	0	7880
CM	0	174	403	378	461	583	285	0	0	0	0	0	0	0	253	602	525	700	745	297	0	501	866	963	1021	1268	372	0	177	858	11432
ET	0	0	0	0	0	0	0	0	0	0	1088	1671	1673	1678	1809	747	797	727	684	820	313	1472	1947	1824	1693	1908	1496	1613	0	0	23960
GA	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	391	580	650	721	813	140	0	0	3295
GH	0	0	0	0	0	39	554	600	613	450	500	482	575	680	574	639	587	701	522	483	418	461	522	378	63	481	503	563	574	491	12453
GM	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	5	5
GN	0	0	0	0	0	0	0	0	0	130	438	516	557	780	505	266	456	482	479	621	197	0	99	584	622	583	666	506	263	652	9402
KE	0	0	0	0	0	0	0	0	0	0	0	0	0	383	893	922	927	1102	443	904	977	1045	974	1070	1258	3631	3801	3685	0	0	22015
LB	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	563	835	847	818	1025	90	336	590	583	619	803	221	29	7359
ML	0	0	0	0	0	0	0	16	1202	1340	1843	1347	1774	1629	1658	2399	982	1947	2070	2019	2472	1544	5	846	915	858	837	776	283	1484	30246
MR	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	29	29
MW	0	0	0	0	0	0	0	0	0	0	346	1494	1750	1840	2148	2905	1390	1434	1895	1618	327	854	909	961	950	502	935	1025	0	0	23283
MZ	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	420	1693	1810	1792	2189	1238	0	0	0	9142
NG	318	897	1020	1020	1297	796	0	0	0	0	0	0	0	294	756	838	811	1030	1193	3546	3792	3603	3943	5474	4437	4724	4633	5690	1767	2139	54018
NI	0	0	335	637	701	809	1105	456	0	0	643	1178	1464	547	0	0	0	0	0	0	0	0	418	936	1004	935	1034	362	0	0	12564
NM	0	0	0	0	0	0	0	0	0	0	48	490	487	521	707	642	1	478	644	660	759	967	78	88	294	325	355	440	0	0	7984
RW	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	340	555	806	736	806	349	777	811	873	785	1265	747	747	0	0	9597
SL	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	110	354	399	421	516	355	761	848	784	917	490	260	6215
SN	0	0	2	528	680	742	792	1002	98	0	0	0	0	0	0	300	490	504	604	641	207	493	675	1530	2535	3568	3987	5626	6256	6568	37828
TD	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	8	1810	1920	1851	1664	2132	9385
TG	148	441	465	331	0	0	0	0	0	0	728	1048	1297	340	0	0	0	0	0	0	0	0	0	4	511	644	612	683	668	44	7964
TZ	0	0	0	0	0	0	0	0	0	69	435	445	507	509	484	0	0	0	0	0	896	1262	1241	1399	1447	436	1532	1611	0	0	12273
UG	0	0	0	0	0	0	0	0	0	0	35	707	818	846	1032	977	164	391	423	442	502	348	371	388	416	432	562	789	0	0	9643
ZM	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	373	956	1024	1077	1129	485	184	2128	2205	2278	2297	0	0	14136
ZW	0	0	0	0	0	0	0	0	0	94	471	453	544	567	421	119	716	755	742	833	710	607	719	803	881	1316	963	952	0	0	12666
Total	466	1512	2225	3455	3898	3689	3802	4868	4089	5171	8367	11432	13099	12939	13657	14386	10944	14437	15645	17245	15866	20066	20647	25838	30454	36694	35773	35669	14261	16830	417424

Notes: Height-for-age and weight-for-age are available for children aged 0-5 years old.

Figure A.3. Distribution of paved road distances and IV road distances



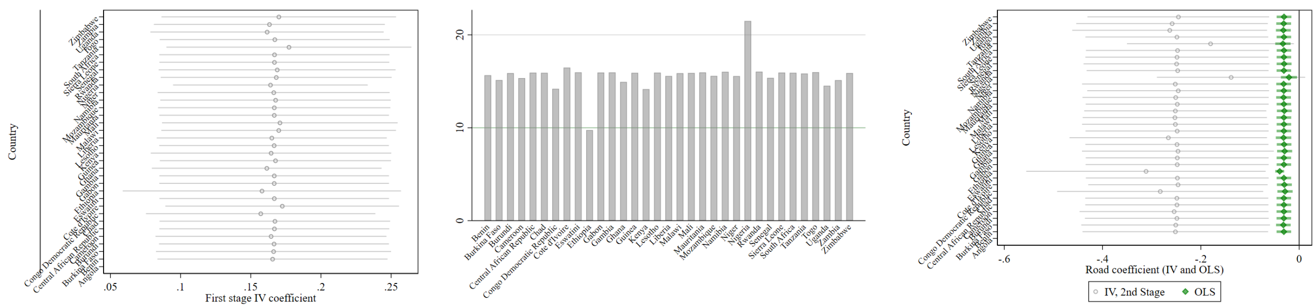
Notes: Histograms showing the distance to the nearest paved road (blue) and to the nearest hypothetical road from the Euclidean Minimum Spanning Tree (EMST) network (green), based on the regression sample.

Table A.2. IV Robustness

	Height-for-Age				
	Controling for:		Excluding:		
	(1)	(2)	(3)	(4)	(5)
	Distance to mines and ports	***Mum's height and weight***	***Coastlines***	***100 km from cons. points***	***New residents***
Log road distance, paved	-0.297**	-0.159*	-0.268**	-0.085	-0.598**
	(0.104)	(0.083)	(0.114)	(0.134)	(0.258)
	Weight-for-Age				
Log road distance, paved	-0.111*	-0.021	-0.105	-0.046	-0.284**
	(0.065)	(0.046)	(0.066)	(0.135)	(0.120)
Climate	Yes	Yes	Yes	Yes	Yes
Individual Controls	Yes	Yes	Yes	Yes	Yes
Household Controls	Yes	Yes	Yes	Yes	Yes
Mother Controls	Yes	Yes	Yes	Yes	Yes
Conflict	Yes	Yes	Yes	Yes	Yes
Geo	Yes	Yes	Yes	Yes	Yes
Observations	263061	240452	232459	176026	98751
Kleibergen-Paap F	11.38	15.49	13.58	4.98	5.89
GEO FE	✓	✓	✓	✓	✓
TIME FE	✓	✓	✓	✓	✓
S.e. cluster	Region	Region	Region	Region	Region

Notes: 2SLS results, weighted regressions with denormalized weights, DHS final sample. Observations within a 50km (100km in column 4) distance from consequential points, i.e., the nodes constituting the IV road network, are excluded. With respect to the baseline results, column 1 adds the log of distance from mines and the log of distance for ports to the controls set, measured in km. Column 2 controls for the mother's anthropometrics (height and weight). Column 3's specification excludes areas within 100km of coastlines. Column 5 only keeps residents of the area for at least 10 years. Stars correspond to the following p-values: * $p < .10$, ** $p < .05$, *** $p < .01$.

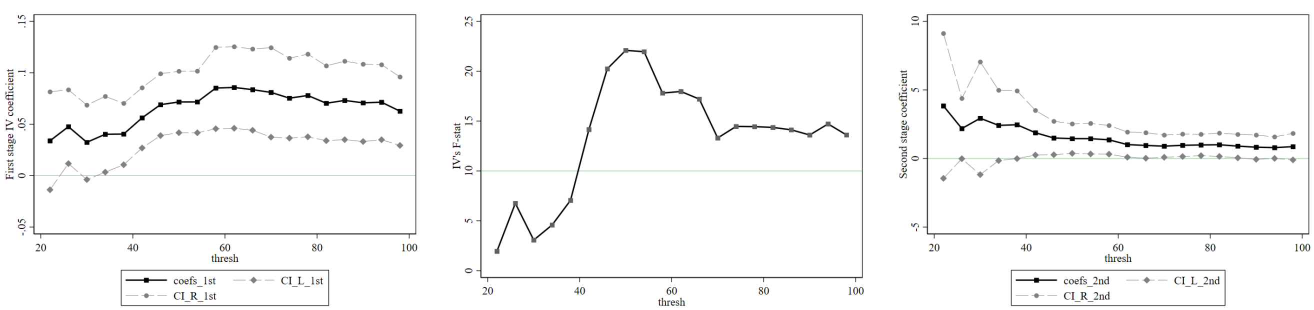
Figure A.4. Leaving-one-country-out sensitivity



(a) **First-stage β** (b) **IV's F-stat** (c) **Second-stage β**

Notes: 2SLS results, weighted regressions with denormalized weights, DHS final sample. Observations within a 50km distance from consequential points, i.e., the nodes constituting the IV road network, are excluded. Each dot represents either a point estimate or an F-statistic obtained from a regression of height-for-age in which one country is excluded at a time. The specifications used include all controls, date of birth and region FEs. Standard errors are clustered at the region level.

Figure A.5. Sensitivity to distance bins



(a) **First-stage β** (b) **IV's F-stat** (c) **Second-stage β**

Notes: 2SLS results, weighted regressions with denormalized weights, DHS final sample. Observations within a 50km distance from consequential points, i.e., the nodes constituting the IV road network, are excluded. The first stage's dependent variable is a binned distance dummy = 1 if the DHS cluster results within that km distance threshold from a paved road (in the x-axis). The IV is a binned distance dummy = 1 if the DHS cluster results within that distance threshold from the fictitious road. The specifications used refer to height-for-age as dependent variable and include all controls, date of birth and region FEs. Standard errors are clustered at the region level.

Table A.3. IV alternatives

	First Stage				
	Baseline	*4NN*	*EMST variants*		
	(1)	(2)	(3)	(4)	(5)
Log IV, EMST cities and mines	0.167***				
	(0.042)				
Log IV, 4NN cities		0.180***			
		(0.034)			
Log IV, EMST cities, ports, mines			0.166***		
			(0.041)		
Log IV, EMST cities				0.193***	
				(0.044)	
Log IV, min EMST					0.186***
					(0.039)
Kleibergen-Paap F	15.96	28.16	16.09	18.95	22.52
	Height-for-Age				
	(1)	(2)	(3)	(4)	(5)
Log road distance, paved	-0.248**	-0.099	-0.234**	-0.148**	-0.202**
	(0.095)	(0.073)	(0.092)	(0.067)	(0.065)
	Weight-for-Age				
Log road distance, paved	-0.100*	-0.072	-0.090*	-0.079	-0.112**
	(0.057)	(0.052)	(0.055)	(0.052)	(0.049)
Climate	Yes	Yes	Yes	Yes	Yes
Individual Controls	Yes	Yes	Yes	Yes	Yes
Household Controls	Yes	Yes	Yes	Yes	Yes
Mother Controls	Yes	Yes	Yes	Yes	Yes
Conflict	Yes	Yes	Yes	Yes	Yes
Geo	Yes	Yes	Yes	Yes	Yes
Observations	263061	273143	262147	273143	263061
GEO FE	✓	✓	✓	✓	✓
TIME FE	✓	✓	✓	✓	✓
S.e. cluster	Region	Region	Region	Region	Region

Notes: 2SLS results, weighted regressions with denormalized weights, DHS final sample. Observations within a 50km distance from consequential points, i.e., the nodes constituting the IV road network, are excluded. With respect to the baseline specification of column 1, column 2 uses an alternative construction of the fictitious roads using straight lines connecting the 4 nearest consequential points. Columns 3 to 5 vary the baseline instrument by including all ports (column 3), only considering cities and not mines (column 4), taking the minimum distance between the baseline IV and the IV with cities only (column 5). Stars correspond to the following p-values: * $p < .10$, ** $p < .05$, *** $p < .01$.

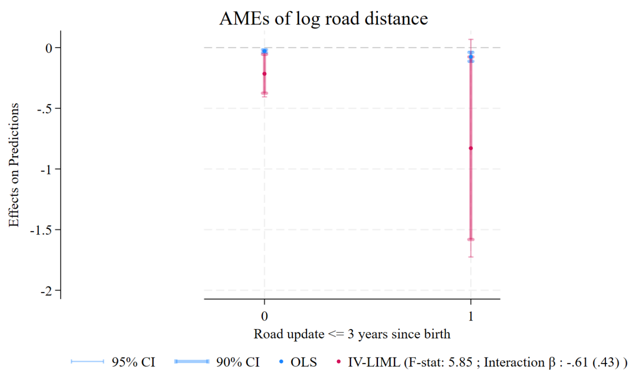
Table A.4. Road density robustness

	2SLS	
	Height-for-age	*Weight-for-age*
Road density, paved	0.262**	0.102*
	(0.116)	(0.062)
Time distance to road data	Yes	Yes
Climate	Yes	Yes
Individual Controls	Yes	Yes
HH Controls & Nightlights	Yes	Yes
Mother Controls	Yes	Yes
Conflict	Yes	Yes
Geo	Yes	Yes
Observations	263061	263061
Kleibergen-Paap F	6.05	6.05
GEO FE	✓	✓
TIME FE	✓	✓
S.e. cluster	Region	Region

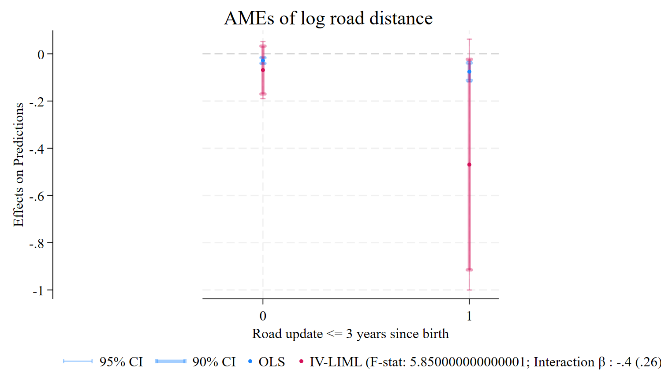
Notes: 2SLS results, weighted regressions with denormalized weights, DHS final sample. Observations within a 50km distance from consequential points, i.e., the nodes constituting the IV road network, are excluded. Stars correspond to the following p-values: * $p < .10$, ** $p < .05$, *** $p < .01$.

Figure A.6. Heterogeneity by road improvement

Recent paving

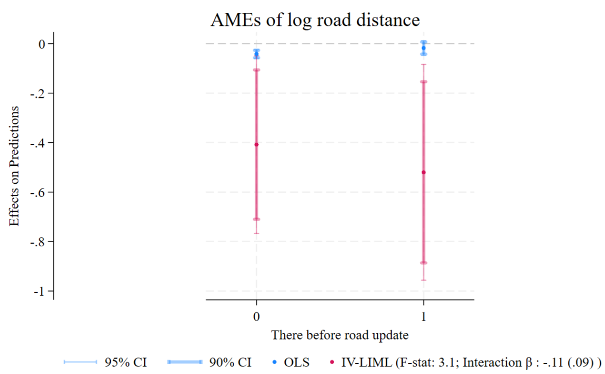


(a) Outcome: Height-for-age

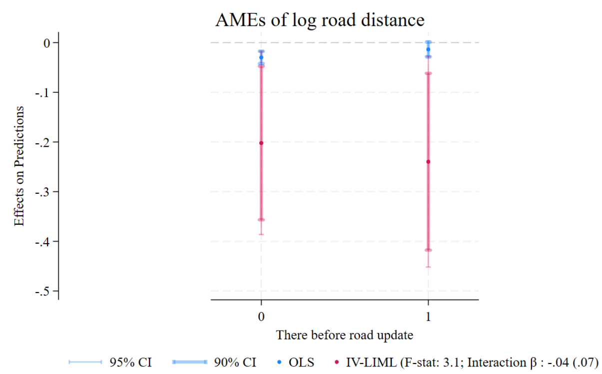


(b) Outcome: Weight-for-age

Living there before road update



(c) Outcome: Height-for-age



(d) Outcome: Weight-for-age

Notes: Average marginal effects from a specification interacting road distance with the years passed between the child birth and a recorded reduction in distance to the paved road. Results on the left and on the right relate respectively to the child's height-for-age and weight-for-age as the dependent variables.

Table A.5. Mechanisms - healthcare

	2SLS				
	Antenatal visit	***Check post birth***	***Child Vaccine***	***Child Diet***	***Child Cough***
Log road distance, paved	-0.107*	-0.114***	-0.053*	0.042	0.024
	(0.057)	(0.042)	(0.030)	(0.039)	(0.025)
Climate	Yes	Yes	Yes	Yes	Yes
Individual Controls	Yes	Yes	Yes	No	Yes
HH Controls & Nightlights	Yes	Yes	Yes	Yes	Yes
Mother Controls	Yes	Yes	Yes	Yes	Yes
Conflict	Yes	Yes	Yes	Yes	Yes
Geo	Yes	Yes	Yes	Yes	Yes
Observations	175714	111082	144481	102082	262184
Kleibergen-Paap F	13.50	13.22	19.65	7.69	15.65
GEO FE	✓	✓	✓	✓	✓
TIME FE	✓	✓	✓	✓	✓
S.e. cluster	Region	Region	Region	Region	Region
Outcome mean	0.79	0.40	0.46	0.30	0.25

Notes: 2SLS results, weighted regressions with denormalized weights, DHS final sample. Observations within a 50km distance from consequential points, i.e., the nodes constituting the IV road network, are excluded. Outcome variables differ by column. In column 1 and 2, the dependent variable equals 1 if the woman recorded at least one antenatal care visit during the woman's last pregnancy and if she received a professional health check after birth, respectively. In column 3, the outcome relates to children at least one-year-old and is 1 if the child received all appropriate vaccines (one dose against tuberculosis, 3 doses of the pentavalent (DPTHib, Hep-B) and oral polio vaccines, and one dose of Measles-Rubella vaccine). In column 4, the outcome variable is an indicator of diet diversity that equals one if in the last 24 hours the child ate meals containing at least 4 different food groups. In column 5, the outcome variable is a binary indicator equal to 1 if the child experienced a cough within the last 24 hours or the past two weeks. Stars correspond to the following p-values: * $p < .10$, ** $p < .05$, *** $p < .01$.

Table A.6. Mechanisms - mother's outcomes

	2SLS					
	Recent visit	***Wanted last child***	***Distance***	***Permission***	***Travel Alone***	***Money***
Log road distance, paved	-0.017 (0.041)	-0.001 (0.026)	0.023 (0.038)	-0.027 (0.030)	0.029 (0.026)	0.061 (0.056)
Climate	Yes	Yes	Yes	Yes	Yes	Yes
Individual Controls	No	Yes	No	No	No	No
HH Controls & Nightlights	Yes	Yes	Yes	Yes	Yes	Yes
Mother Controls	Yes	Yes	Yes	Yes	Yes	Yes
Conflict	Yes	Yes	Yes	Yes	Yes	Yes
Geo	Yes	Yes	Yes	Yes	Yes	Yes
Observations	183604	193678	144892	144867	141000	144886
Kleibergen-Paap F	14.44	15.26	13.79	13.88	13.56	13.85
GEO FE	✓	✓	✓	✓	✓	✓
TIME FE	✓	✓	✓	✓	✓	✓
S.e. cluster	Region	Region	Region	Region	Region	Region
Outcome mean	0.53	0.71	0.48	0.15	0.25	0.57

Notes: 2SLS results, weighted regressions with denormalized weights, DHS final sample. Observations within a 50km distance from consequential points, i.e., the nodes constituting the IV road network, are excluded. Outcome variables differ by column. In column 1, the dependent variable is 1 if the interviewed woman visited a health facility for any reason within 12 months before the interview. Column 2 looks at whether the last pregnancy was wanted. Columns 3 to 6 record problems preventing from getting medical care, in case the respondent is sick. The dependent variables take value of one if the woman signaled that the problem was relevant. The issues investigated are, respectively: distance to the health facility, having permission to go, not wanting to travel alone, money constraints. Stars correspond to the following p-values: * $p < .10$, ** $p < .05$, *** $p < .01$.

Table A.7. Mechanisms - mother's wealth and labor outcomes

	2SLS					
	Wealth Index	***Any Work***	***Paid Work***	***Agri***	***Agri / Works***	***Partner Agri/ Works***
Log road distance, paved	-0.355**	0.057	-0.088**	0.121*	0.155	0.069**
	(0.158)	(0.035)	(0.042)	(0.072)	(0.100)	(0.030)
Climate	Yes	Yes	Yes	Yes	Yes	Yes
HH Controls & Nightlights	Yes	Yes	Yes	Yes	Yes	Yes
Mother Controls	Yes	Yes	Yes	Yes	Yes	Yes
Conflict	Yes	Yes	Yes	Yes	Yes	Yes
Geo	Yes	Yes	Yes	Yes	Yes	Yes
Observations	155217	191046	112567	191046	128716	173726
Kleibergen-Paap F	15.38	15.08	11.53	15.08	11.55	16.20
FE	Region; Birth Date; Interview Yr	Region; Birth Date; Interview Yr	Region; Birth Date; Interview Yr	Region; Birth Date; Interview Yr	Region; Birth Date; Interview Yr	Region; Birth Date; Interview Yr
S.e. cluster	Region	Region	Region	Region	Region	Region
Outcome mean	2.57	0.67	0.67	0.37	0.55	0.58

Notes: 2SLS results, weighted regressions with denormalized weights, DHS final sample. Observations within a 50km distance from consequential points, i.e., the nodes constituting the IV road network, are excluded. Outcome variables differ by column. In column 1, we report results for a household wealth index (in quintiles, poorest to richest). In column 2 to 6, the outcome variable indicates labor market outcomes for the mother or her partner, as follows: whether she works at all, whether she performs paid work, whether her work is in agriculture, the same outcome but restricting the sample to workers, and whether the partner works in agriculture contingent on working. Stars correspond to the following p-values: * p < .10, ** p < .05, *** p < .01.

Table A.8. Mechanisms - complementary infrastructure

	2SLS	
	Has electricity	***Time to water source***
Log road distance, paved	-0.085 (0.056)	-0.175 (0.194)
Climate	Yes	Yes
Individual Controls	Yes	Yes
HH Controls & Nightlights	Yes	Yes
Mother Controls	Yes	Yes
Conflict	Yes	Yes
Geo	Yes	Yes
Observations	195184	155678
Kleibergen-Paap F	15.18	15.06
GEO FE	✓	✓
TIME FE	✓	✓
S.e. cluster	Region	Region
Outcome mean	0.16	32.12

Notes: 2SLS results, weighted regressions with denormalized weights, DHS final sample. Observations within a 50km distance from consequential points, i.e., the nodes constituting the IV road network, are excluded. Outcome variables differ by column. In column 1, the dependent variable equals 1 if the household has electricity. In column 2, the dependent variable is the logarithm of the time (in minutes) required for the household to reach its main water source; households with water on the premises (travel time = 0) are excluded. Stars correspond to the following p-values: * p < .10, ** p < .05, *** p < .01.

Table A.9. Mechanisms - land-use

	First-stage		
	Dependent variable: Road distance		
	(1)	(2)	(3)
Log IV, MST cities and mines	0.220***	0.224***	0.218***
	(0.042)	(0.041)	(0.038)
Observations	32592	33134	36356
	2SLS		
	Forest loss	***Malaria***	***Share Cropland***
Log road distance, paved	-0.263	0.013	-17.864***
	(0.623)	(0.014)	(3.428)
Climate	Yes	Yes	Yes
Conflict	Yes	Yes	Yes
Geo	Yes	Yes	Yes
Observations	32592	33134	36356
Kleibergen-Paap F	27.06	29.73	33.61
GEO FE	Region	Region	Region
TIME FE	✓	✓	✓
S.e. cluster	Region	Region	Region
Outcome mean	4.58	0.28	46.89

Notes: 2SLS results, unweighted regressions based on DHS geocodes as geographic units. Observations within a 50km distance from consequential points, i.e., the nodes constituting the IV road network, are excluded. Outcome variables differ by column group. In column 1, we report results for forest loss within a 31km radius of the DHS household, original data comes from Hansen et al. (2013). Column 2 looks at malaria incidence, from Hay and Snow (2006). Column 3 uses the share of cropland within a 31km radius of the DHS cluster, from ESA Climate Change Initiative (CCI) Land Cover dataset, 1992–2015. Stars correspond to the following p-values: * p < .10, ** p < .05, *** p < .01.

Table A.10. Mechanisms - Agricultural development

	2SLS										
	Number of crops			Land share under production technique				Total production by technique			
	Total	Subsistence	Cash	Irrigated HI	Rainfed HI	Rainfed LI	Rainfed subsistence	Irrigated HI	Rainfed HI	Rainfed LI	Rainfed subsistence
Log road distance, paved	-0.079	0.080	-0.001	-2.280	-0.010	2.203	-2.801	-2695.052	-1158.639**	-801.539	-2500.759***
	(0.366)	(0.198)	(0.196)	(2.284)	(0.874)	(1.839)	(2.888)	(2101.798)	(508.726)	(865.953)	(743.014)
Observations	28896	28896	28896	28896	28896	28896	28896	28896	28896	28896	28896
2SLS - agri share >10 percent											
Log road distance, paved	-1.386	1.059	0.326	-15.466	4.761	13.937	-3.164	-8003.434	-5032.799	-3550.951	-9010.321
	(1.883)	(1.219)	(0.802)	(15.785)	(5.853)	(12.292)	(13.250)	(9765.696)	(4921.785)	(5548.451)	(6839.861)
Climate	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Conflict	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Geo	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	12541	12541	12541	12541	12541	12541	12541	12541	12541	12541	12541
Kleibergen-Paap F	1.52	1.52	1.52	1.52	1.52	1.52	1.52	1.52	1.52	1.52	1.52
GEO FE	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
TIME FE	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
S.e. cluster	Region	Region	Region	Region	Region	Region	Region	Region	Region	Region	Region
Outcome mean	10.59	5.35	5.23	3.31	13.49	36.35	45.03	1843.80	3408.42	5373.76	6287.81

Notes: 2SLS results, Spatial Production Allocation Model (SPAM) data sample, from You et al. (2014). Regressions are based on pooled cross-sections and matching DHS clusters from 2000, 2005, 2010 with the other covariates. The first three columns report results for the total number of cultivated crops, by category. Columns 4 to 7 relate to land shares under each production technique (in the column names, HI means high-input, LI stands for low-input). Columns 8 to 11 report results on total production, by technique. Stars correspond to the following p-values: * p < .10, ** p < .05, *** p < .01.

INSTITUT DE RECHERCHE
ÉCONOMIQUES ET SOCIALES

Place Montesquieu 3
1348 Louvain-la-Neuve

ISSN 1379-244X D/2025/3082/17

IRES
Louvain Institute of Data Analysis and Modeling in economics and statistics