DOCUMENT DE TRAVAIL DT/2020-11

Take the Highway? Paved Roads and Well-Being in Africa

Elodie DJEMAÏ

Andrew E. CLARK

Conchita D'AMBROSIO

UMR LEDa
Place du Maréchal de Lattre de Tassigny 75775 • Paris •Tél. (33) 01 44 05 45 42 • Fax (33) 01 44 05 45 45
DIAL • 4, rue d'Enghien • 75010 Paris • Tél. (33) 01 53 24 14 50 • Fax (33) 01 53 24 14 51
E-mail : dial@dial.prd.fr • Site : dial.ird.fr

Take the Highway? Paved Roads and Well-Being in Africa[1]

Elodie Djemaï
Université Paris-Dauphine, PSL Research University, IRD, LEDa, UMR 225, DIAL
elodie.djemai@dauphine.psl.eu

Andrew E. Clark
Paris School of Economics - CNRS
Andrew.Clark@ens.fr

Conchita D'Ambrosio
Department of Cognitive and Behavioral Sciences, University of Luxembourg
conchita.dambrosio@uni.lu

This version: July 28, 2020

Abstract. Public Goods aim to improve individual welfare. We investigate the causal consequences of roads on well-being for 24 African countries, instrumenting paved roads by 19th Century hypothetical lines between major ports and cities. We have data on over 32000 individuals, and consider both their objective and subjective well-being. Roads reduce material deprivation, in terms of access to basic needs. But at the same time those closer to roads evaluate their living conditions as being worse. This suggests that roads are a double-edged sword in Africa, either being associated with worse outcomes in non basic-needs domains, or increasing individuals' aspirations.

Résumé. La provision de biens publics vise à améliorer le bien-être individuel. Nous étudions les conséquences causales des routes sur le bien-être dans 24 pays africains, en

[1]We are grateful to Matthias Kroenke, David Margolis, Cecilia Poggi and participants at the DIAL conference (Paris, 2019), the London School of Economics, the ITEA annual congress (Paris, 2019) and the Workshop on The Distributional Impact of Social Protection (Addis Ababa, 2020) for helpful comments. Clément Baticle provided excellent research assistance in coding the road network. Data Availability Statement: The data used in this article can be obtained upon request to Afrobarometer (https://afrobarometer.org/fr/données/politique-dutilisation-des-données) and the Demographic and Health Surveys (www.dhsprogram.com). Additional replication materials will be provided in the Online Appendix if the manuscript is accepted for publication. Declarations of interest: none. Financial support from CEPREMAP and the Fonds National de la Recherche Luxembourg (Grant C18/SC/12677653) is gratefully acknowledged.

instrumentant les routes bitumées par les lignes hypothétiques reliant les ports et les villes principales au 19eme siècle. Nous disposons des données d'enquêtes collectées auprès de 32000 individus, et considérons le bien-être objectif et subjectif. Les routes réduisent la privation matérielle, en termes d'accès aux besoins de base. Mais les individus vivant à proximité des routes évaluent leurs conditions de vie comme étant plus mauvaises. Ceci suggère que les routes sont à double tranchant en Afrique, soit parce qu'elles détériorent certains aspects que les individus valorisent lorsqu'ils évaluent leurs propres conditions de vie, soit parce qu'elles augmentent les aspirations individuelles.

Journal of Economic Literature Classification Nos.: D63, I32, O18.

Keywords: Roads, Subjective Well-being, Basic Needs, Material Deprivation, Africa.

1 Introduction

We often think of social protection at the individual level, for example via transfers and social safety nets. We here consider the role of public goods in protecting individuals, not via transfers but by potentially improving the environment in which they live (De Janvry and Sadoulet 2015). Our public good here is paved roads.

There is by now a large body of literature in economics questioning the effect of transportation infrastructure on economic outcomes, both in developed and developing countries. The traditional outcomes investigated in this literature are pecuniary: poverty, consumption, income and investments (see, for example, Aggarwal 2018; Asher and Novosad 2020; Banerjee *et al.* 2020; Buys *et al.* 2010; Dercon *et al.* 2009; Dillon *et al.* 2011; Donaldson 2018; Gibbons *et al.* 2019; Gibson and Rozelle 2003; Jacoby 2000; Jacoby and Minten 2009; Khandker *et al.* 2009; Straub *et al.* 2008; Wang and Wu 2015). Here roads have been shown to be beneficial. However, only few contributions have focussed on the effects of roads on access to basic needs, confirming their positive effects (see Bucheli *et al.* 2018).

The overall effect of roads on well-being is nonetheless ambiguous, and has remained largely unexplored in the literature. Roads also bring problems, for example via disease and accidents leading to worse health, and environmental deterioration through reduced forest cover and biodiversity. In addition, transport infrastructure may affect individuals' evaluations of their lives via its effect on individual expectations, aspirations and the salience of different reference groups. Although economists have become increasingly open to the use of subjective data, the robust analysis of life evaluations remains only rare in some areas of the discipline and, in particular, in the analysis of developing countries.

We here estimate the effect of road infrastructure on well-being using individual data from Wave 5 of the Afrobarometer survey (2011-2013), covering most African countries. In particular we appeal to two distinct measures of well-being for the same individuals: one objective and one subjective. The comparison of these two kinds of well-being measures has been suggested as a useful tool in the analysis of poverty (for instance, Clark and D'Ambrosio 2019; Pradhan and Ravallion 2000; Ravallion 2014). Our first measure is an index of material deprivation in terms of the individual's access to basic needs (food, water, fuel and medical care). Material deprivation is a special case of multidimensional poverty when the dimensions considered are only related to material aspects of life. The second well-being measure we look at is the respondent's evaluation of her current living conditions. We will link both the objective and subjective indicators to the distance to the closest paved road. Our contribution to the literature is hence twofold: we estimate the effects of roads on multiple observable dimensions of poverty, adding to the very few

findings in this area, and we are the first, to the best of our knowledge, to analyse the causal relationship between roads and subjective well-being.

The same two individual well-being measures to which we appeal here were analysed in Clark and D'Ambrosio (2019), who use five rounds of Afrobarometer data over the 2004-2016 period to explore the association between the two. They find that, as expected, the more deprived in basic needs the individual is, the lower their evaluation of their current life, controlling for standard socio-economic and demographic variables, wave and region fixed effects. The estimated relationship is large in size: a one standard-deviation rise in the material-deprivation index is associated with a lower evaluation of current living conditions by around one quarter of a standard deviation. Our goal in the current paper is to investigate the causal determinants of these two variables.

Wave 5 of the Afrobarometer survey is geo-coded and, when matched to the road network from Bing Maps using ArcGIS, allows us to measure the distance between the residential location of the respondent and the nearest paved road. We then evaluate the correlation between this distance and respondent outcomes. These correlations do not allow us to estimate the causal effect of roads on well-being, as road placement is partly endogenous. The current literature on the effect of roads or railways instruments their location via the location of historical routes, including those followed by explorers (Agrawal *et al.* 2017; Baum-Snow *et al.* 2017; Duranton and Turner 2012; Martincus *et al.* 2017), and the straight lines connecting important locations in the past, such as cities in the 19th Century (Atack *et al.* 2010; Banerjee *et al.* 2020; Bird and Straub 2014; Donaldson 2018; Jedwab and Moradi 2016). Land characteristics such as slope have also been used to instrument the location of physical infrastructure in Batzilis *et al.* (2016), Dinkelman (2011), Djemaï (2018), Duflo and Pande (2007) and Lipscomb *et al.* (2013).

Existing work has suggested that access to roads reduces poverty and increases consumption; we hence might expect that individuals living further from a paved road are more likely to be deprived. However, the relationship with the subjective evaluation of the current living conditions remains an open empirical question. Living conditions may be better closer to roads, driven by better access to basic needs; on the contrary, we may also find offsetting effects via the consequences of roads on pollution, the environment (Damania *et al.* 2018), health (see Djemaï 2018 for HIV, and Riley-Powel *et al.* 2018 for qualitative evidence on dengue fever in Peru), crime, social capital, and exposure to richer individuals that may increase individual expectations and aspirations.

Our main results are that distance to roads increases material deprivation, while the results for self-assessed living conditions differ by specification and sample: in the naïve

regressions with uninstrumented distance, those living further from a paved road are less satisfied with their current living conditions. The IV results tell a different story: there is no effect of road distance on current living conditions. The difference between the OLS and IV estimates is consistent with both the sorting of individuals and the endogenous location of roads: those who move to live closer to the current road network are more likely to report good subjective living conditions, and roads are built where people report being better off. When we distinguish between urban and rural residents, the estimated coefficients for the first are not significantly different from their counterparts in the whole sample. Strikingly, while the uninstrumented coefficient for the rural sample is similar to that for the whole sample, the instrumented coefficient is significantly positive and large: rural residents living further from roads now report better living conditions. The protective effect of the instrumented distance to roads in rural areas, as opposed to no effect in the naïve regression, again reflects either the movement of individuals towards roads or the endogenous location of the latter.

We then drop sequentially locations that are "too close" to historical settlements, as these are unlikely to have been treated by the instrument. The estimated road coefficient for subjective well-being in the whole sample turns from insignificant to positive and significant for locations that are at least six kilometers away: living further from roads has a positive effect on individuals' evaluated living conditions. We find no change in the effect of roads on material deprivation, where the coefficient remains negative and significant.

The analysis of a variety of mediators reveals that public-good provision and urbanisation are more pronounced closer to roads, and are positively correlated with objective and subjective well-being. The mediating role of education, local labour markets, crime, health and social capital appears much more limited.

We conclude by considering the potential role of migration, in that certain types of individuals may move closer to roads. As the Afrobarometer does not include migration information, we here turn to the Demographic and Health Surveys, in which only objective well-being appears. Our results show a consistent negative effect of road distance on the latter. When we split our sample based on migration status we find that migration does not lie behind our core results: the effects of road distance on objective well-being is negative and significant (and of similar size) for both migrants and non-migrants.

The remainder of the paper is organised as follows. Section 2 describes the estimation strategy and the data. The estimation results are then presented and discussed in Sections 3 and 4. The potential role of mediators is investigated in Section 5, and Section 6 uses the Demographic and Health Surveys to explore the role of migration. Last, Section 7

concludes.

2 Empirical Strategy and Data

2.1 Estimation Equations

To quantify the effect of distance to the nearest road on subjective and objective well-being, we first estimate these well-being outcomes, WB, for individual i living in community j in country c as:

$$WB_{ijc} = \alpha + \beta distance_{jc} + \gamma X_{ijc} + \delta_c + \varepsilon_{ijc} \tag{1}$$

where WB is alternatively self-assessed current living conditions and the index of access to basic needs, and *distance* is the log of 1 + distance to the nearest paved road in kilometres. We use the log specification to account for our *a priori* expectation of a decreasing marginal effect of distance (so that an extra kilometre matters less at 100km from a road than at 10km from a road). As it turns out, the data prefer this specification (in terms of fit) to a regression where distance enters linearly. In Equation (1), X_{ijc} contains age, age-squared and sex. We control for country fixed effects, denoted by δ_c. The errors in Equation (1) above, ε_{ijc}, are clustered at the community level as our measure of distance to the nearest paved road is defined at this level. If access to roads improves well-being, then the estimated value of β will be negative (distance to roads is bad for well-being).

Equation (1) does not control for any of the other potentially endogenous explanatory variables such as education, labour-market status and urban/rural location, as these are all arguably partly determined by the distance to a paved road, and are thus bad controls. We will below in Section 4 consider a variety of candidate variables, which may be endogenous to road distance, as potential mediators of our main results regarding the effect of road access on living conditions and deprivation. For example, considering education as a control variable, we will estimate the following equation:

$$WB_{ijc} = \alpha^* + \beta^* distance_{jc} + \gamma^* X_{ijc} + \psi Educ_{ijc} + \delta_c^* + \varepsilon_{ijc}. \tag{2}$$

If, as we suspect, educational opportunities are better closer to roads, and education improves well-being outcomes, then the estimated value of ψ will be positive, and the estimate of β^* in Equation (2) will be less negative than the estimate of β in Equation (1): holding education constant turns off part of the well-being benefit of roads. We will carry

out analogous analyses for the mediating effect of labour-market status, health, living in an urban area, public goods, crime and social capital (trust and participation in social or religious groups). We estimate linear models to be able to compare the size of the road effect when we add the possible mediator and when we modify the sample. The core results are qualitatively unchanged when ordered probit models are estimated in Table O2.

Road placement is however likely endogenous in the above equations, and as such we carry out two-stage least squares estimation including a set of potential instrumental variables. This is described in the sub-section below.

2.2 The Non-random Location of Roads: IV Strategy

Our first-stage regressions is as follows, where Z_{jc} is the instrumental variable:

$$distance_{ijc} = \phi + \theta Z_{jc} + \pi X_{ijc} + \mu_c + \nu_{ijc}. \tag{3}$$

Building on previous research on the endogenous placement of transport infrastructure, we will consider the standard instrumental variables in this literature (see Redding and Turner 2015 for a review): routes that were used a long time ago, hypothetical lines between historical settlements, and land characteristics that may affect the cost of road-building. Section 2.4 below will describe how all three of these are measured in the African data that we use. These three are thought to be good instruments for the following reasons.

First, the location of explorer routes or routes that were used a long time ago could affect the placement of current roads without directly affecting our outcomes. As stated in the review by Redding and Turner (2015), the historical-route instrumental-variable approach relies on the location of old transportation routes as a source of quasi-random variation in the location of current transportation infrastructure. This approach has been used, for example, in Agrawal *et al.* (2017), Baum-Snow *et al.* (2017), Duranton and Turner (2012) and Martincus *et al.* (2017). The value of this instrument in our African case will be the distance between the respondent's location and the closest pre-colonial explorer route.

Second, the hypothetical lines connecting current or historical major cities are candidates for the instrumentation of the location of current roads and railways (Atack *et al.* 2010; Banerjee *et al.* 2020; Bird and Straub 2014; Donaldson 2018; Ghani *et al.* 2016; Jedwab and Moradi 2016; Michaels 2008). The straight line connecting historical settlements is a proxy for the most cost-effective way of linking major cities abstracting from natural constraints (e.g. lakes or mountains), and thus provides relevant exogenous variation. Along these lines, Michaels (2008) uses the straight-line distance of the centroid of the county to

the nearest city to instrument for the location of highways in the US. Here we will use the network of hypothetical lines from Jedwab and Moradi (2015 and 2016) that relates the urban network as of 1900, consisting of the capital city, the largest and second-largest cities, other cities with above 10,000 inhabitants, and ports:[2] there are on average 3.5 historical settlements per country, with a median figure of 2. These hypothetical lines could have joined major cities and ports at the end of the 19th Century, but were not necessarily built (given natural constraints and the absence of cooperation among colonial empires).

Last, land characteristics such as the slope may determine the construction of roads in a given area as they can influence the associated costs: roads are more likely to be built in flatter than steeper areas. Some recent work has used slope measures to instrument the location of physical infrastructure.[3]

The simultaneous-equation model we estimate is:

$$\begin{cases} distance_{ijc} = \phi + \theta Z_{jc} + \pi X_{ijc} + \mu_c + \nu_{ijc} \\ \\ WB_{ijc} = \alpha + \beta distance_{ijc} + \gamma X_{ijc} + \delta_c + \varepsilon_{ijc}. \end{cases} \tag{4}$$

For the instrumental variable to be valid, it has to be correlated with the current distance to the nearest paved road and not correlated with the error term of the second-stage equation ε. The first correlation is discussed when presenting the results from the first-stage estimations in Section 3. The second requirement is the exclusion restriction: the instrument should have no effect on the outcome other than through the first-stage channel.

We will below address two critical points regarding the exclusion restriction. First, the exclusion restriction is conditional on all covariates and does not assume the unconditional orthogonality of the instrument and the outcome. This distinction is emphasized in Agrawal *et al.* (2017) and Duranton and Turner (2012). It is therefore crucial to control for a large set of exogenous variables, in particular community-level variables. In our model, it is difficult to argue that education and labour-market outcomes in the community, for example, are exogenous. We will add latitude, longitude and altitude as exogenous community characteristics, as proxy variables for variations in climate and agricultural outcomes within countries. One concern with using land slope as an instrument for road location

[2]Some examples are Ouagadougou and Bobo-Dioulasso in Burkina Faso, Kumasi in Ghana, Luanda in Angola and Mombasa in Kenya.

[3]Duflo and Pande (2007) instrument dam construction using river gradient across Indian districts. Dinkelman (2011) uses land gradient as an instrument for the locations chosen to benefit from an electrification project in South Africa. Land gradient has also been identified as a determinant of cellular-phone coverage in Malawi by Batzilis *et al.* (2016).

is that, in a rural setting, it may affect agricultural outcomes, as suggested in Dinkelman (2011). In our case, the direct impact of gradient on farm productivity would appear in living conditions and material deprivation, our two outcome variables. With respect to the two other candidate instruments, distances to explorer routes and hypothetical lines, these will likely affect current town size and urbanisation that will in turn affect our outcome variables. This concern will be taken into account in the mediation analysis of Section 5.

The second critical point is that some of the locations that are close to a hypothetical line were very close to the capital city or a major port, for example, in 1900, while others were far away from both. We should not consider those who were already close to an urban area as being treated by the hypothetical-line instrument: communities that were in the suburbs of Dakar in 1900 are still in the suburbs of Dakar in 2000. The instrument will thus provide less exogenous variation the closer the community is to a historical city or port. We will deal with this issue in Section 3.3 by progressively dropping communities that are within a certain number of kilometres from a historical city or port, where the number of kilometres considered will vary between 1 and 20. This strategy is close to that used in Ghani *et al.* (2016), who instrument the location of the Golden Quadrilateral highway network by the straight line connecting nodal cities (i.e. Delhi, Mumbai, Chennai and Kolkata) and focus on the effect of the distance to the network only for non-nodal districts. As we will see below, this sample restriction will make a considerable difference to the econometric results. The remainder of this section presents the Afrobarometer survey data that we use, and the various types of geographical data that we match to it.

2.3 The Afrobarometer Survey

We use data from the 5th round of the Afrobarometer surveys collected in 2011-2013 in 30 countries in Sub-Saharan Africa.[4] The Afrobarometer data has been widely used in analyses of development and resources in Africa, such as Nunn and Wantchekon (2011) and Cagé and Rueda (2016). This Afrobarometer round is the closest in time to the data on the current paved-road network (which comes from 2013), is geo-referenced, and includes questions on both self-assessed living conditions and access to basic needs.[5] We exclude from our sample the islands (Cape Verde, Madagascar and Mauritius), and the countries for which we do not observe the location of hypothetical lines as they were not included in

[4]See http://afrobarometer.org/ for more details.

[5]We use data from the 5th wave only as access to geo-coded Afrobarometer data is restricted: it is possible to obtain either data on all countries for one wave in this respect, or for one country over all Afrobarometer waves.

Jedwab and Moradi (2016) (Lesotho, South Africa and Swaziland).

The sample size is around 38,000 individuals living in 24 different countries (see Appendix Table A1). The number of respondents per country is around 1,200 in 16 of these 24 countries, and twice this size in the remaining eight.

As described in Afrobarometer (2015), the sample is nationally-representative and randomly-defined following a clustered, stratified, multi-stage design. The stratification is based on the largest subnational administrative unit and urban-rural location. If the strata is rural, secondary sampling units are randomly selected within each strata, and two Primary Sampling Units, also called Enumeration Areas (EA), are drawn from each secondary sampling unit. If the strata is urban, the EA are selected in a direct way, that is without the preliminary selection of secondary sampling units. This is also the case for rural areas in some countries (e.g. Cameroon). Once the EAs have been drawn, eight households are randomly-selected by starting from a certain geographical point in the EA and then walking in turn in four directions that are 90 degrees from each other, and sampling the 5th and 10th dwelling in each of the four directions. The design of the survey is for eight individuals to be interviewed per enumeration area: in over 90% of EAs, this exact figure was reached. Most of the remaining EAs include either 16 or 24 interviews, reflecting an over-sampling of certain EAs. There are 4,105 EAs in our data.

One individual per sampled household is interviewed. Interviewers alternate interviews between men and women to reach a gender balance. All household members age 18 or older are eligible to be interviewed. If there is more than one household member of eligible sex and age, one is randomly-selected from the list of members.

The questionnaire for each individual includes an initial section filled out by the supervisor. Part of this section indicates whether the EA is urban or rural: we thus typically have eight urban/rural evaluations per enumeration area. In 97% of cases, all of the urban/rural location evaluations are consistent with each other. For the remaining 3% (114 enumeration areas out of 4,105), some questionnaires in the enumeration area report this latter as being rural and some as urban. In these cases of disagreement, we designate an enumeration area as urban if 50% or more of the completed questionnaires in that area refer to the area as being urban.[6] We end up with 36% of enumeration areas being urban, and the rest rural.

In the context of distance to roads, it might be thought that urban areas are not informative, as they will all be close to a paved road. As we will see below, this is far from

[6]Our robustness tests in Section 4.2 will show that the results do not change if we designate all of the disagreement areas as urban, or all of them as rural.

being the case. Urban areas are indeed closer to paved roads than are rural areas, but the distances are quite substantial in both cases (the mean distances are 5km and 15km respectively).

The Afrobarometer data was geocoded after the data collection, using a double-blind methodology where two coders assigned latitude and longitude co-ordinates to each EA (in case of disagreement, there was an arbitration round: see BenYishay *et al.* 2017). There are fewer GPS points than EAs: in a third of cases, GPS points covered households in more than one EA. There are 4,105 EAs in the Afrobarometer data, but 3,396 GPS points. All observations are assigned to a GPS point. We will call these points communities below, and it is at the level that we calculate the distance to the nearest paved road.[7] We have information on the presence of various public goods, such as health centres and schools, for each EA: this is provided by the supervisor. With more than one EA per GPS point, it is thus possible for households in the same GPS point to have different levels of public goods.

2.4 Descriptive Statistics

The data on self-assessed current living conditions come from the question "*In general, how would you describe your own present living conditions?*" The answers are on a five-point scale from very bad to very good.[8] 4.3% of respondents reported that their living conditions were very good, 26% fairly good, 18.6% neither good nor bad, 30.9% fairly bad and 20.2% very bad. This distribution of answers is plotted in Panel a) of Figure A2. The means of all of our analysis variables appear in Table 1. The mean value of current living conditions, on the 1-5 scale, is 2.6. We also create a dummy variable for replying "very good or fairly good" to this question, which covers 30% of respondents.

Our second key dependent variable is material deprivation, where respondents are asked about their difficulty in satisfying their basic needs in five dimensions. The five questions here are "*Over the past year, how often, if ever, have you or anyone in your family gone without enough food to eat, clean water for home use, medicines or medical treatment, fuel to cook your food, a cash income?*" The answers to each of these are given on a five-point scale: Never (0), Just once or twice (1), Several times (2), Many times (3) and Always (4). Due to concerns about collinearity between cash income and the other four elements of material deprivation, we drop the former and calculate a deprivation index as the sum

[7] As noted above, we will therefore cluster at the community level (as all observations within the community have the same distance to paved road). Our main results are robust to clustering at the far more-aggregated region level to address potential spatial autocorrelation in the communities. There are on average 11 GPS points per region.

[8] The "Don't Knows" are recoded as missing values. These represent only 0.3% of the sample.

of the answers to the first four questions. We then invert this figure in order to provide an index of access to basic needs, or lack of material deprivation, where higher values refer to better outcomes. This index thus ranges from zero (for respondents who are "always" deprived in these four dimensions) to 16 (for respondents who are never deprived in any dimension): this is our objective measure of well-being. In Table 1 the average value of this index is 11.4. In terms of the individual elements of the index, 60% of respondents declared that they have never or just once or twice gone without food, with analogous figures of 59%, 56% and 69% for water, medical care and cooking fuel respectively. The distributions of the answers to the four retained basic-needs questions appear in Panels b)-e) of Figure A2. These are fairly similar across the domains, but with somewhat more material deprivation regarding medical care and less deprivation with respect to cooking fuel.

In terms of the exogenous control variables, average age is 36, and 50% of respondents are women (reflecting the sampling design). The remainder of the table refers to the variables that potentially mediate the effect of roads on well-being. Apart from urban, described above, these are education, labour-market status, health, public goods, crime and social capital (trust and participation in social or religious groups). In our sample, 12% of respondents have higher education as their highest education level, 34% secondary education, 32% primary education and 22% no education. Labour-force status is measured as being a part- or full-time paid employee (which applies to just under one-third of our respondents). Health status is not collected as part of the survey, but we do know whether the respondent visited a Public Health Centre at least once over the past 12 months (85% did).

A number of the other mediating variables are aggregate. Access to public goods within the enumeration area or within walking distance is reported by the interviewers. First, they report the presence of an electricity grid, piped water, sewage system and cell-phone service within the EA that most houses could access. 54% of the respondents live in a EA where there is electricity, 47% with piped water, 21% with a sewage system and 92% in an EA with cell-phone coverage. Second, field workers and field supervisors report whether there is a Post Office, school, Police Station, Health Clinic, market stalls (selling groceries and/or clothing) in the EA or within walking distance to the EA. The lowest percentage here refers to Post Offices, where only 18% have access. 59% have some access to a clinic, 88% to a school and 66% to a market.

Social capital may well also act as a mediator, being affected by access to roads and also affecting both self-assessed living conditions and objective well-being (if social support helps satisfy basic needs if the individual lacks them). We measure social capital as involvement

in associations and trust. Respondents report whether they belong to a religious group, and other voluntary associations or community groups (0 = No, 1 = Inactive member, 2 = Active member and 3 = Official leader): 53% belong to a religious group, and 40% to another form of association. The general trust question is: "*Generally speaking, would you say that most people can be trusted or that you must be very careful in dealing with people?*": 20% declared that most people can be trusted. Last, respondents are asked "*How much do you trust your relatives/ your neighbours/ other people you know?*" on a scale of 0 (Not at all), 1 (Just a little), 2 (Somewhat) and 3 (A lot). Average trust is highest for relatives (2.4), followed by neighbours (1.8) and others (1.4).

The last mediator refers to feelings of security and crime. For the latter there are two questions: "*During the past year, have you or anyone in your family (No, Once, Twice, 3 or more) had something stolen from your house?*" and "*During the past year, have you or anyone in your family (No, Once, Twice, 3 or more) been physically attacked?*".[9] For the former, the questions are "*Over the past year, how often, if ever, have you or anyone in your family (Never, Just once or twice, Several times, Many times, Always) (i) felt unsafe walking in your neighborhood?, (ii) feared crime in your own home?*" and "*During election campaigns in this country, how much do you personally fear becoming a victim of political intimidation or violence?" (Not at all, A little bit, Somewhat, A lot).* The average score for feeling unsafe or fearing crime is between Never and Just once or twice, and few people on average have experienced crime.

2.5 Geographical Data

The geographical data from the Afrobarometer surveys are combined with data on the 2013 road network from BingMap. The exact location of the individual respondent is not recorded, but we do know the coordinates of the enumeration area in which they live.[10] The latitude and longitude coordinates of the communities enable us to place each community on a country map and match in geographical data, especially regarding the road network.

Satellite-data from BingMap, available through ArcGIS and providing a satellite-based

[9] This last question was not asked in Tanzania.

[10] This produces measurement error in the measure of individual distance to the nearest paved road. However, the households interviewed in each enumeration area are quite close to each other, so that the gap between the actual (unobserved) distance and the measured distance will be only small, and small positive and negative gaps will likely compensate each other in the empirical analysis. This type of measurement error is common in the spatial literature. For instance, in Ghani *et al.* (2016) plant-level data are aggregated at the district level and the distance measure is the shortest straight-line from the district's edge to the Golden Quadrilateral highway network. In our case, the same measurement error is found for the current road network and our instrumental variables (hypothetical lines or explorer routes).

representation of the road network as of 2013 was used to construct shapefiles that can be used to calculate the straight-line distance in kilometres between the community and the nearest paved road. Figure 1 shows roads in West Africa in 2013 as replicated from the BingMap in ArcGIS on the left panel and the location of communities on the right panel. The same figure for the rest of our sample is in Figure A1. Using the road network and the geographic coordinates of each community, we can compute the straight-line distance between the community and the nearest paved road via ArcGIS. Taking the straight-line distance (also called the "as the crow flies" or the great circle distance) has a number of advantages over the use of more sophisticated measures such as the use of real distance or time distance, as no assumptions about the means of transportation owned and used by households need to be made. In addition, Combes and Lafourcade (2005) note that the straight-line distance and alternative measures of distance are highly correlated (with a correlation coefficient of over 0.97). Our respondents here live an average of 12km away from the nearest paved road (see Table 2).

We calculate all three possible instrumental variables at the community level in ArcGIS. We construct the measure of land gradient using the SRTM digital-elevation map[11] and we calculate the straight-line distances between the communities and historical routes or hypothetical lines. Figure 1c shows the location of explorer routes (in red) and hypothetical lines (in blue). We use the GIS shapefile from Nunn and Wantchekon (2011) that provides the location of the explorer routes used during the pre-colonial and early colonial periods (between 1768 and 1894); this information comes from the Century Company (1911).

Jedwab and Moradi (2016) provide the hypothetical lines or "straight-lines" connecting as of 1900 major cities (those with over 10,000 inhabitants), capital cities and ports. More precisely Jedwab and Moradi create an Euclidean Minimum Spanning Tree (EMST) network based on the initial urban network in 1900 (including the capital, largest and second-largest cities, and the other cities in each country) and define this network, EMST, as *"the network that the colonial powers would have built if they had collaborated to optimally connect the initial cities while minimizing construction costs (using the Euclidean distance between them)"* (page 275).

As shown in the first column of Table 2, the percentage slope ranges from 0 to 28, with an average figure of 1.6; respondents live on average 97km away from the nearest hypothetical line, and 157km from the nearest explorer route. If we compare the samples living in urban and rural areas (columns 2 and 3 of Table 2 respectively), the distance to the nearest paved road is 5.8km on average in urban areas and 14.9km in rural areas.

[11] Using Shuttle Radar Topography Mission data, at a resolution or cell size of approximately 90 meters.

The standard deviations of this distance are substantial in both areas, at 16.6 and 20.1. The distance to hypothetical lines is about twice as large in rural than urban areas, while the distance to explorer routes is lower (140km vs. 187km). The distance between the communities and the nearest historical settlement is calculated in ArcGIS: this is 127km in the whole sample, and 83km and 153km for the urban and rural samples respectively.

3 Results

3.1 The Non-random Location of Roads: First-stage Results

The first-stage results appear in Table 3. This table has three columns and six panels, referring to two specifications for each of our three potential instruments: distance to hypothetical lines, slope, and distance to explorer routes. The instruments appear first as levels and then in logs. The first column refers to the entire sample while the second and the third separate the urban and the rural samples. All equations include the instrument, country fixed effects (so that we estimate the effect of the instrument within a country), and the exogenous demographic and community characteristics (sex, a quadratic in age, altitude, longitude and latitude). Appendix Table A3 shows the full results, including the estimated coefficients on the control variables.

The results in Table 3 show that distance to the hypothetical line is the only instrument with a convincing first-stage F-statistic (usually considered to be greater than 10: see Staiger and Stock, 1997). In column (1) of Table 3, the log of 1 + the distance to the nearest hypothetical line in Panel B attracts an F-statistic of around 240, and the alternative linear specification (with distance divided by 100 to make the coefficients easier to read) an F-statistic of about 70. The estimated coefficient in the log specification implies an elasticity of current road distance to the distance to historical hypothetical lines that is around 0.3. The other potential instruments, slope and distance to explorer routes, attract F-statistics of under 10.

We now separately show the first-stage results for the urban and rural samples in columns (2) and (3) of Table 3. These again explore the six different potential instrument specifications. The resulting F-statistics are entirely in line with those in column 1, with only the log of 1 + distance to the nearest hypothetical line and the same distance in levels being consistently satisfactory. There turns out to be relatively little difference between the urban and rural samples with respect to the size of the first-stage estimated coefficient on the log of the distance to the hypothetical line, and the F-statistics remain

satisfactory. As discussed in Section 2, there remains substantial variation in the distance to paved roads in the areas that enumerators describe as urban.

3.2 The Effects of Road Distance on Well-being: Second-stage Results

We first discuss the second-stage results for the whole sample, before progressively dropping communities that are close to historical cities or ports, as noted above in Section 2.2.

The first of our outcome variables is self-assessed current living conditions (taking values from 1 to 5). The results for the whole sample are shown in the left part of Panel A in Table 4. These are linear regressions. As we might think that the effect of roads differs by location, we also estimate the model separately for urban and rural residents in Panels B and C. All of the specifications in this table include country fixed effects and the exogenous controls (age, age-squared, female, altitude, longitude and latitude).

The results for self-assessed living conditions depend on the specification. In the naïve regressions with uninstrumented distance in column (1), those living further from a paved road are less satisfied with their current living conditions. The effect size is quite small here: a one standard-deviation rise in the log of distance (1.29, from Table 1) is estimated to lead to a fall in subjective living conditions of around 3% of a standard deviation. The IV results in column (2) tell a different story: here there is no effect of road distance on current living conditions. The difference between the OLS and IV estimates is consistent with both the sorting of individuals and the endogenous location of roads. In the first case, those who move to live closer to the current road network are more likely to report good subjective living conditions, and this sorting is naturally stronger for the observed road network (in the uninstrumented results) than for the predicted road network using the instrumental variable. In the latter case, the unexplained part of road location is correlated with the subjective living conditions of local residents: roads are built where people report being better off.[12]

Panels B and C of Table 4 distinguish between the living conditions of urban and rural residents. The general pattern of results here is similar to that for the whole sample in Panel A, in that the instrumented effect of distance is more positive (so that distance is less harmful) than in the naïve regressions. In the urban sample, the estimated coefficients in columns (1) and (2) are not significantly different from their counterparts for the whole sample in Panel A. Strikingly, while the uninstrumented coefficient for the rural sample in

[12] Asher and Novosad (2020) note that, in India, "new roads are disproportionately built in villages that are growing for other reasons" (page 799).

Panel C is similar to that in the other panels, the instrumented coefficient is now significantly positive (and large, with a marginal effect of one standard deviation of log distance (1.26) being about one fourth of a standard deviation in living conditions (1.19)): rural residents living further from roads report better living conditions. The protective effect of the instrumented distance to roads in rural areas, as opposed to no effect in the naïve regression, again reflects either the movement of individuals towards roads or the endogenous location of the latter.

Appendix Table A4 shows the full results for the whole sample, including the estimated coefficients on all of the control variables. There is a notable U-shape in age in both living conditions and lack of deprivation (as is very often found in the literature on subjective well-being and life satisfaction), and women report slightly less-good outcomes for both measures. Regarding the within-country estimates for the geographical variables (as we include country fixed effects), these are significant only for latitude (with living conditions being better in African countries that are further North in our sample); the estimated coefficients on altitude and longitude are insignificant.

We now switch from looking at subjective living conditions to objective measures of deprivation or functioning failure. The results appear in columns (3) and (4) of Table 4. The dependent variable here is the lack of deprivation in four dimensions, each of which is measured from 0 to 4. Our resulting lack of deprivation index ranges from 0 to 16, with higher numbers reflecting better outcomes: 0 refers to someone who is always lacking in all four dimensions, and 16 to someone who never lacks in any dimension. We will also carry out robustness checks on each dimension separately below in Table 6.

The results show that distance to roads increases deprivation (i.e. it reduces the lack of deprivation). This result holds in both the uninstrumented and instrumented regressions. The size of the estimated coefficient in the IV regression in column (4) of Panel A implies that a one standard-deviation rise in the log of distance reduces the lack of deprivation (i.e. increases material deprivation) by just over one-fifth of a standard deviation.

In Panel A, the instrumented coefficient for lack of deprivation in column (4) is more negative than the naïve version in column (3). This is the opposite result to that for subjective living conditions. Were roads to be randomly allocated, those living far from them would be more deprived than those who live far from the observed road network. In terms of endogeneity, roads are either located closer to individuals who are more deprived, or it is the more deprived who move closer to the existing road network.

The endogeneity of roads does not then appear to work in the same way for living conditions and material deprivation: they are located in areas of better subjective living

conditions but more objective material deprivation. Roads have then served to improve the objective living conditions of the deprived in Africa. One way of interpreting this opposition is to consider that living conditions cover all of the aspects that individuals find important (and much more than just the four basic needs we have considered here), and that the non-basic needs element of living conditions is negatively correlated with the four aspects of material deprivation. This is certainly possible as basic needs and living conditions are only relatively weakly correlated (with a correlation coefficient of 0.22).

3.3 Instrument Validity and Distance from Historical Settlements

Section 2.2 raised a major concern about the validity of distance to the hypothetical lines (connecting historical settlements) as an instrument. The instrument will not be valid for locations that were already very close to a city or major port in 1900 and will still be so today, and the location of the city or port in 1900 cannot be considered as exogenous (with settlements being more likely where living conditions are more favourable).

We deal with this potential problem by re-estimating the regressions in Tables 3 and 4, progressively dropping locations that are within a certain number of kilometres from historical settlements. The value of this exclusion distance varies from 1 to 20 kilometres. Figure 2 depicts the F-statistic from the first stage in each of these 20 new regressions: the instrument remains sufficiently strong in all cases (the horizontal line shows the standard value of 10 below which the instrument is considered to be not strong). The F-statistic falls as we progressively exclude locations that are further from the 1900 historical settlements. This fall is consistent with our first-stage results in Table 3, where the F-statistic was larger for urban than rural locations (with the former being more likely to be close to historical settlements). The high F-statistic in Table 3 then partly reflects the mechanical correlation for communities that were already close to 1900 historical settlements.

The results for the 20 corresponding second-stage regressions are plotted in Figure 3. Those on the right refer to the effect of the distance to a road on the lack of deprivation. These turn out to be remarkably stable across all of the regressions (which is consistent with the similarity of the urban and rural figures in column 4 of Table 4). The effect of the distance to a current road on material deprivation is the same for individuals who are far from a historical settlement and those who are much nearer.

The left-hand side of Figure 3 shows the corresponding results for the effect of road distance on living conditions. In sharp contrast to those for material deprivation, the estimated coefficients here start off insignificant (as in Panel A of Table 4), but turn significantly positive once we drop locations that are within six kilometres of a historical

settlement. Individuals report better living conditions when further from a road (as in our rural sample in Table 4). The fact that the estimated coefficient becomes increasingly positive as we exclude those who are close to a historical settlement suggests that the correlation between road distance and living conditions for those close to a historical settlement is in fact negative. This can be understood in terms of the endogeneity of the location of historical settlements: these were built where living conditions were good.

In what follows, and in particular for all of the robustness tests, we will restrict the sample to locations that are situated over 10kms from a historical settlement. The descriptive statistics for this restricted sample appear in Appendix Table A5 (which shows that the means and standard deviations of the outcome variables are very similar to those in the whole sample). The benchmark first- and second-stage results (as well as the OLS counterparts) appear in Table 5. The F-statistic in column (1) is satisfactory. As suggested in Figure 3, the results for deprivation are similar to those for the whole sample in Table 4. On the contrary, the instrumented coefficient for road distance in the living-conditions regression is now positive and significant as compared to the previous insignificant effect for the whole sample in Table 4. The effect size (which is similar to that estimated for the rural sample in Table 4) is that a one standard-deviation rise in distance to a road improves living conditions by 20% of a standard deviation. Alternatively, moving from 10kms to 20kms from a paved road is estimated to increase subjective living conditions by 0.12 points or 10% of a standard deviation, while the same movement is estimated to reduce access to basic needs by one half of a point (13% of a standard deviation).

4 Robustness Checks

We now turn to a number of robustness tests. These are all carried out on the sample of locations over 10kms from a historical settlement, the baseline results for which appeared above in Table 5. Our tests are split up into two groups: the first refers to the measures of well-being and road distance, and the second to the estimation sample.

4.1 Robustness Checks 1: Measures

Measures of outcome variables We above took a counting approach to material deprivation, adding up the unweighted sum of the scores in each of the four dimensions. This relies on two assumptions: that each dimension is equally important, and that the scores for each dimension are cardinal (in the sense that moving from "Never" (0) to "Just once

or twice" (1) represents the same rise in deprivation as moving from "Many times" (3) to "Always" (4)). We relax both of these assumptions in Table 6, first by considering each of our four dimensions separately, and then by considering a dummy variable for low deprivation (and thus high well-being), defined as the individual reporting having had difficulty in satisfying their basic needs in the dimension under consideration over the past year never or just once or twice.

We first consider the effect of road distance on the four separate dimensions of the deprivation index, considered cardinally (i.e. from 0 to 4), in the left-hand part of Table 6. The results show that distance is harmful for three of the four dimensions: the exception is cooking fuel. The difference between the naïve and instrumented regressions for these three dimensions is analogous to that for the overall index in Table 5.

The right-hand side of Table 6 then drops the cardinality assumption, and considers dummy variables for low deprivation, to see whether the effect of roads is concentrated at one or other of the extremes of the scale. The estimation here is linear (the same qualitative results are found in probit regressions). The estimated coefficients for distance to roads in these binary regressions are all similar to those using all of the values of the dependent variable in the left-hand side of Table 6.[13] As such, roads shift some people from being more deprived (2 to 4) to less deprived (0 or 1), as well as shifting some individuals within these categories.[14]

For completeness, Panel E of Table 6 presents analogous results for subjective living conditions. The left-hand side shows the baseline results from the cardinal regression in Table 5. We then create a dummy variable for good or very good living conditions, the results of which appear on the right-hand side of Panel E. As was the case for the dimensions of deprivation, the binary results are entirely consistent with the cardinal results.

Measures of distance to the nearest paved road Table 7 checks whether the results from the instrumented regressions are robust to changes in the distance measure. Panel A here reproduces the benchmark estimates from Table 5. We then create three dummy variables for road distance as enumeration areas that are located more than 5km, 10km and 20km from a paved road: the results appear in Panels B to D. Distance is treated

[13]The estimated coefficient on distance to roads in the IV cooking-fuel regressions is positive, and significant in the binary variable specification, perhaps reflecting that the observed road network was built further from forests.

[14]The Afrobarometer also includes an observable binary measure of deprivation: whether the dwelling has a solid roof. The 2SLS analysis of this variable, using the same approach as in the right-hand side of Table 6, reveals a negative significant relationship between road distance and solid roofs. These results appear in Online Appendix Table O3.

as a level (rather than a log) in row E and as a quadratic in row F. Last, the measure in row G is one of road proximity: the inverse of one plus the distance in km. The results in Table 7 show that our conclusions are not sensitive to the measure of distance: roads reduce material deprivation but lead to worse reported living conditions.[15]

4.2 Robustness Checks 2: Samples

We now consider changing the estimation sample in the four following ways.

1. Dropping observations that are "too close" to the nearest paved road.
2. Removing individuals who are aged 70 or older (this applies to 2.7% of the sample).
3. Removing one country at a time from the sample in a Jackknife analysis.
4. Reconsidering the definition of urban locations.

The results for Checks 1 and 2 appear in Table 8. The first row of this table includes the benchmark results, for ease of comparison. We first drop observations that are less than 5 and then 10 kilometers from the current paved road. We do so to make sure that our results are not only driven by local observations that are close to roads, rather than being more global. This is a particular concern in our case as we use log distance, where the variation is higher for smaller distance values. In rows 2 and 3, the results for more-distant observations are of the same sign and significance as the benchmark results: our findings then seem to be global rather than local.[16] Equally, dropping older respondents in row 4 makes very little difference to the results.

The results from the Jackknife analysis appear in Appendix Table A6. Again the first line refers to the benchmark results. The remaining 24 lines list the estimated coefficients on the distance to roads in the well-being regressions when the country in the first column is dropped. No one country drives our benchmark coefficients.

Last, we reconsider the way in which we determine whether a location is urban or rural. As we noted in Section 2.3, in each location there are multiple evaluations of whether the location is urban or rural. In 114 out of 4,105 enumeration areas (5%) there is disagreement in these evaluations, in that the location is described as rural in some questionnaires and

[15] The results in Panel G are of the right sign, but are not significant. This may reflect that the inverse of distance approaches its asymptotic value faster than does log, so that identification is more concentrated on small distance values.

[16] We may also worry about distance values that are "too large": top trimming the 1% (over 98km) or 5% (over 51km) of the observations with the highest values for distance to the nearest paved road makes very little difference to the results.

urban in others. We above designated these areas as urban if 50% or more of the completed questionnaires described the area as urban. We now check that our results are not sensitive to this choice. In Table O4 we consider all areas where there is disagreement as urban (in Panel B) and then as rural (in Panel C); Panel A shows the results using our initial classification method. Note that the results in Panel A here refer to the restricted sample that is found in locations over 10km from historical settlements, and thus differ from those in Table 4 which came from the entire sample.[17] The estimated living-conditions coefficients in Table O4 are consistent across the three panels. This is also the case for lack of deprivation, except for the rural sample when disagreement areas are defined as urban, where we lose significance in Panel B.

5 Potential Mediators

Roads may affect many outcomes, which themselves help determine well-being: these include health, social-capital accumulation, criminality, public goods, urbanisation, education and labour-market opportunities.

We evaluate the role of these different mediators in the effect of road distance on well-being by adding a sequence of additional control variables to the instrumented benchmark Equation (4): these are urban/rural location, education, labour-market variables, health, public goods, social capital and crime. The estimated road-distance coefficients once we control for each of these different mediators in turn appear in Table 9 (where the first row shows the benchmark values) and are illustrated in Figure 4 (along with their respective 95-percent confidence intervals). Table 9 has two columns, the first for living conditions and the second for lack of deprivation.

Every specification in column (1) produces a significant positive effect of road distance on living conditions. In terms of mediation, the road-distance coefficient is notably larger when we control for urbanisation and the provision of public goods in rows B and G (although only that controlling for public goods is significantly different from the coefficient in the benchmark specification in row A). This tells us that public-good provision is better closer to roads and is positively correlated with subjective living conditions. None of our other potential mediators changes the road-distance coefficient substantially.

We now switch to the role of mediators in the objective measure of lack of deprivation (see the left-panel of Figure 4 and Column 2 of Table 9). The mediation results for basic

[17] As might be suspected, the major difference between Tables 4 and O4 is found for the urban sample, which is much smaller in the latter table once we exclude those who are close to historical settlements

needs are very similar to those for living conditions. Only holding public goods constant in row G renders the estimated road coefficient notably less negative (the difference between the coefficient here and that in the baseline in row A produces a t-statistic of 1.58, for a significance level of 11%): were people closer to and further from paved roads to have the same level of public goods, there would be no difference in their deprivation.

The last row in Table 9 introduces all of the mediators at the same time (except for experience of crime, for which one question is missing for Tanzania) continues to produce a positive effect of road distance on living conditions, but with an estimated coefficient for lack of deprivation that is insignificant (as when we controlled for urban location or public goods in lines B and G).

The fact that there almost always remains a significant road coefficient in this mediation analysis means that there must be an omitted variable here making those closer to roads feel worse off (and which we do not measure in the Afrobarometer data): this could reflect, for example, pollution, health, family break-up, worse community life, unwanted pregnancies or observing richer people (including tourists and traders) closer to roads.

6 The Role of Migration

Our analysis above has corrected for the endogenous placement of roads via an instrumental variable. However, we have not yet addressed the question of the endogenous placement of individuals. If individuals of a certain type systematically move towards (or away from) exogenously-placed roads, then our estimated coefficients will be biased.

Unfortunately, the Afrobarometer data does not include information about migration status. We thus turn to Demographic and Health Survey (DHS) data. Of the 24 countries that appeared in our Afrobarometer analysis above, there are 15 that appear in the DHS data and for which we have geo-coded location information and migration status.[18] We take the latest available geo-coded wave for each country, which gives us observations spread out between 2003 and 2017.

We try to replicate as far as possible our Afrobarometer analysis on DHS data. We calculate distance to observed roads in the same way as above, and continue to use distance to hypothetical lines as an instrument. We do not have the same outcome measures, and here consider the wealth index constructed by the DHS on the basis of objective information on housing characteristics and durable-good ownership. We standardise this index to have

[18]Regressions on the Afrobarometer data dropping the countries that do not appear in the DHS produce estimated coefficients with larger standard errors but which continue to be significant.

a mean of zero and a standard deviation of one. The other controls are as in our analysis above, except that we now have information on migration status. The exact question in the DHS is as follows: *"How long have you been living continuously in (name of current city, town or village of residence)?"* The answers were either always, a number of years, or a special code for those who do not live there but are visitors from outside.[19]. Those who report a number of years are migrants, picking up both international and internal migration, and at all respondent ages (including migration as a child).

The initial sample size is $276,346$ for these 15 countries. Appendix Table A7 lists the countries, sample size and year in the DHS surveys that we use. Dropping the $5,624$ visitors reduces this figure to $270,722$. We will apply the same method as in Table 5 to the DHS data, dropping locations that are 10km or closer to historical settlements (referring to 33,968 observations). This produces a DHS analysis sample of $236,754$. The descriptive statistics for this DHS sample appear in Appendix Table A8 and the location of the respondents in Appendix Figure A3. Almost half of our DHS sample have changed location at least once in their life. Figure 5 plots the percentage of migrants by distance to the nearest current paved road: as can be seen, the distribution is fairly uniform, with a slightly higher percentage who are either close to a road or very far from one (Jacoby and Minten, 2009, note that migrants may decide to live in remoter areas in order to profit from lower land prices and greater agricultural-land availability).

The first-stage results are shown in Appendix Table A9. As was the case for the Afrobarometer, the distance to the hypothetical line is the only instrument that has an F-statistic of over 10. The ensuing second-stage results from the benchmark regressions on DHS data appear in Panel A of Table 10. These show a consistent negative effect of road distance on material well-being, that is larger in the IV results in column (2) (analogous to our findings for the lack of deprivation in Afrobarometer data in column (4) of Table 5).

The DHS Panel A results did not make any use of the migration information. We now do so, and estimate the following system of equations:

$$\begin{cases} distance_{ijc} = \delta + \theta Z_{jc} + \pi X_{ijc} + \psi_c + \nu_{ijc} \\ \\ WB_{ijc} = \alpha + \beta distance_{ijc} + \lambda distance_{ijc} \times Migrant_{ijc} + \zeta Migrant_{ijc} + \gamma X_{ijc} + \delta_c + \varepsilon_{ijc}. \end{cases} \tag{5}$$

[19]We do not include visitors in our analysis, as we observe neither their permanent place of residence (and its distance to roads) nor their wealth index.

We here introduce both a dummy for migration in the well-being equation and an interaction of this dummy with distance to the paved road. The estimated coefficient on the distance to roads for non-migrants is given by β; that for migrants is given by the sum of β and λ. We wish to establish whether our main results continue to hold in the sample of non-migrants (in the same spirit as the work on the returns to education, program effectiveness and migration reviewed in Strauss and Thomas 1995).

The results appear in Panel B of Table 10. These first reveal that the material well-being of migrants is higher. Second, and central to our analysis, the estimated road coefficient for non-migrants (β) is negative and not significantly different from the distance estimate in Panel A: distance to roads has a sizeable negative effect on the material well-being of non-migrants. The estimated value of λ is also negative and significant, so that road distance reduces well-being somewhat more for migrants than for non-migrants, but is only 20% of the size of β in our preferred IV regression. This provides suggestive evidence that migration is not behind our core Afrobarometer results, as the effect of roads on well-being is negative and significant (and of broadly similar size) for both migrants and non-migrants.

7 Conclusion

This paper has matched road-network information to nationally-representative household survey data from 24 African countries in the early 2010s to analyze the relationship between the distance to paved roads and both living conditions and the lack of material deprivation (regarding food, water, cooking fuel and medical care). The naïve correlations between distance to roads and both of the well-being outcomes are negative: those who live further from roads are less well-off. However, roads are not located exogenously (and neither are people). We instrument road location by the lines drawn between major cities and ports at the end of the 19th Century. The instrumented effect of road distance on the lack of deprivation continues to be negative; however, that on self-assessed living conditions is now zero. As such, we may conclude that roads have improved access to basic needs (as picked up by the lack of deprivation measure), but have not led to better evaluations of lives, very likely because these latter include many other aspects of life that may be negatively affected by roads (such as health, family break-down, crime and insecurity).

We do not believe that this is the end of the story. In the context of spatial analyses, lines drawn between major cities and ports in 1900 are undoubtedly correlated with the current location of roads, but may not be good instruments for areas that were already close to an urban area in 1900: communities that were in the suburbs of Dakar in 1900

are still in the suburbs of Dakar in 2000. The areas that are more likely to be treated by the instrument are those that were further from historical settlements in 1900. We ensure the validity of our instrument by excluding areas that are "too close" to historical settlements. This makes little difference to the results for lack of deprivation, but does affect the estimated coefficient on distance to roads in the subjective living conditions regression, which changes from essentially zero to being positive and significant. This is our preferred specification. We also provide some supporting evidence that these estimated road coefficients do not entirely reflect the migration of certain kinds of individuals towards areas with roads.

Roads are therefore a double-edged sword in Africa. While they provide access to basic needs, and likely reduce inequality in this dimension, they produce worse self-assessed living conditions. The positive effects on access to water, food and medical care seem to be outweighed by other (unmeasured) aspects of life that matter to individuals, and overall lead them to report worse living conditions. As we know that subjective evaluations predict individual outcomes, we may well then expect (ceteris paribus) proximity to roads to have consequences in terms of future health and family structure, for example. While our mediation analysis here was not able to identify the life dimensions that render roads bad for subjective living conditions, we hope that future data will be able to match road location to both subjective measures of well-being and variables such as the environment, health risks, child mortality and family life. These will help our understanding of the role of roads in determining the quality of life.

References

Afrobarometer (2015), Data codebook for a Round 5 Afrobarometer survey in 34 African countries. Prepared by: Chunho Park Michigan State University July 2015. Report (107 pages).

Afrobarometer Data, Burundi, Benin, Burkina Faso, Botswana, Cameroon, Cote d'Ivoire, Ghana, Guinea, Kenya, Liberia, Mali, Malawi, Mozambique, Namibia, Niger, Nigeria, Senegal, Sierra Leone, Sudan, Tanzania, Togo, Uganda, Zambia and Zimbabwe, Round 5, 2011-2013, available at http://www.afrobarometer.org.

Aggarwal, S. (2018), "Do rural roads create pathways out of poverty? Evidence from India", *Journal of Development Economics*, 133, 375–395.

Agrawal, A., Galasso, E., and Oettl, A. (2017), "Roads and Innovation", *Review of Economics and Statistics*, 99, 417–434.

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

Atack, J., Bateman, F., Haines, M. and Margo, R.A. (2010), "Did Railroads Induce or Follow Economic Growth? Urbanization and Population Growth in the American Midwest, 1850-60", *Social Science History*, 34, 171–197.

Banerjee, A., Duflo, E. and Qian, N. (2020), "On the Road: Access to Transportation Infrastructure and Economic Growth in China", *Journal of Development Economics*, 145, 102442.

Batzilis, D., Dinkelman, T., Oster, E., Thornton, R. and Zanera, D. (2016), "New Cellular Networks in Malawi: Correlates of Service Rollout and Network Performance", in Edwards S., Johnson S. and Weil D., (Eds.), *African Successes: Modernization and Development*, University of Chicago Press, 215–246.

Baum-Snow, N., Brandt, L., Henderson, JV., Turner, MA. and Zhang, Q. (2017), "Roads, Railroads and Decentralization of Chinese Cities", *Review of Economics and Statistics*, 99, 435–448.

Bird, J. and Straub, S. (2014), "The Brasilia Experiment: Road Access and the Spatial Pattern of Long-term Local Development in Brazil", TSE Working Paper No. 14-495.

Buys, P., Deichmann, U. and Wheeler, D. (2010), "Road Network Upgrading and Overland Trade Expansion in Sub-Saharan Africa," *Journal of African Economies*, 19, 399–432.

BenYishay, A., Rotberg, R., Wells, J., Lv, Z., Goodman, S., Kovacevic, L. and Runfola, D. (2017), "Geocoding Afrobarometer Rounds 1-6: Methodology & Data Quality," AidData. Available online at *http://geo.aiddata.org*.

Bucheli, J.R., Bohara, A.K. and Villa, K. (2018), "Paths To Development? Rural Roads and Multidimensional Poverty in the Hills And Plains Of Nepal," *Journal of International Development*, 30, 430–456.

Cagé, J. and Rueda, V. (2016), "The Long-Term Effects of the Printing Press in Sub-Saharan Africa", *American Economic Journal: Applied Economics*, 8, 69–99.

Clark, A.E. and D'Ambrosio, C. (2019), "Living Conditions and Well-Being: Evidence from African Countries", *South African Journal of Economics*, 87, 91–109.

Century Company. 1911, "The Century Atlas: Africa [map]" Buffalo, NY: Matthews-Northrup (470 English statute miles to 1 inch).

Combes, P.-P. and Lafourcade, M. (2005), "Transport costs: measures, determinants, and regional policy implications for France", *Journal of Economic Geography*, 5, 319–349.

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

De Janvry, A. and Sadoulet, E. (2015), Development Economics: Theory and Practice. Routledge, 2015.

Dercon, S., Gilligan, D.O., Hoddinott, J. and Woldehanna, T. (2009), "The Impact of Agricultural Extension and Roads on Poverty and Consumption Growth in Fifteen Ethiopian Villages", *American Journal of Agricultural Economics*, 91, 1007–1021.

Dillon, A., Sharma, M. and Zhang, X. (2011), "Estimating the Impact of Rural Investments in Nepal", *Food Policy*, 36, 250–258.

Dinkelman, T. (2011), "The Effects of Rural Electrification on Employment: New Evidence from South Africa", *American Economic Review,* 101, 3078–3108.

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

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

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

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

Ghani, E., Goswami, A and Kerr, W. (2016), "Highway to Success: The Impact of the Golden Quadrilateral Project for the Location and Performance of Indian Manufacturing", *Economic Journal*, 126, 317–357.

Gibbons, S., Overman, H., Lyytikainen, T. and Sanchis-Guarner, R. (2019), "New Road Infrastructure: The Effects on Firms", *Journal of Urban Economics*, 110, 35–50.

Gibson, J. and Rozelle, S. (2003), "Poverty and Access to Roads in Papua New Guinea", *Economic Development and Cultural Change*, 52, 159–185.

Jacoby, H.G. (2000), "Access to Markets and the Benefits of Rural Roads", *Economic Journal*, 110, 713–737.

Jacoby, H.G. and Minten, B. (2009), "On Measuring the Benefits of Lower Transport Costs", *Journal of Development Economics*, 89, 28–38.

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

Jedwab, R. and Moradi, A. (2015), "Replication data for: The Permanent Effects of Transportation Revolutions in Poor Countries: Evidence from Africa", doi:10.7910/DVN/29908, Harvard Dataverse, V1.

Khandker, S.K., Bakht, Z. and Koolwal, G.B. (2009), "The Poverty Impact of Rural Roads: Evidence from Bangladesh", *Economic Development and Cultural Change*, 57, 685–722.

Lipscomb, M., Mobarak, A.M. and Barham, T. (2013), "Development Effects of Electrification: Evidence from the Topographic Placement of Hydropower Plants in Brazil", *American Economic Journal: Applied Economics* 5, 200–231.

Martincus, C.V., Carballo, J. and Cusolito, A. (2017), "Roads, exports and employment: Evidence from a developing country", *Journal of Development Economics* 125, 21-39.

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

Nunn, N. and Wantchekon, L. (2011), "The Slave Trade and the Origins of Mistrust in Africa", *American Economic Review*, 101, 3221–3252.

Pradhan, M. and Ravallion, M. (2000), "Measuring Poverty Using Qualitative Perceptions of Consumption Adequacy", *Review of Economics and Statistics*, 82, 462–471.

Ravallion, M. (2014), "Poor, or Just Feeling Poor? On Using Subjective Data in Measuring Poverty", In A.E. Clark and C. Senik (Eds.), *Happiness and Economic Growth: Lessons from Developing Countries*. Oxford: Oxford University Press.

Redding, S.J. and Turner, M.A. (2015), "Transportation Costs and the Spatial Organization of Economic Activity", in Cheshire P. and Mills E.S. (Eds.), *Handbook of Regional and Urban Economics*, Volume 5B, 1339–1398.

Riley-Powell A.R., Lee G.O., Naik N.S., Kelly E. Jensen, K.E., O'Neal, C., Salmon-Mulanovich, G., Hartinger, S.M., Bausch, D.G. and Paz-Soldan, V.A. (2018), "The Impact of Road Construction on Subjective Well-Being in Communities in Madre de Dios, Peru", *International Journal of Environmental Research and Public Health*, 15, 1271.

Staiger D. and Stock J.H. (1997), "Instrumental Variables Regression with Weak Instruments", *Econometrica*, 65(3): 557–586.

Straub, S., Vellutini, C. and Warlters, M. (2008), "Infrastructures and Economic Growth in East Asia", Policy Research Working Paper No. 4589. World Bank.

Strauss, J and Thomas, D. (1995), "Human Resources: Empirical Modeling of Household and Family Decisions", in J. Behrman and T.N. Srinivasan (Eds.), *Handbook of Development Economics*, Volume III, 1883–2023.

Wang, Y. and Wu, B. (2015), "Railways and the Local Economy: Evidence from Qingzang Railway", *Economic Development and Cultural Change*, 63, 551–588.

Tables and Figures

Table 1: Sample Descriptive Statistics

VARIABLES	(1) N	(2) Mean	(3) SD	(4) Min	(5) Max
Current living conditions (LC)	38,249	2.63	1.19	1	5
Current LC, good or very good	38,249	0.30	0.46	0	1
Lack of depriv., over 4 items	37,985	11.39	3.74	0	16
Age	38,027	36.38	14.12	18	105
Female	38,395	0.50	0.50	0	1
Urban	38,395	0.36	0.48	0	1
Primary education	38,290	0.32	0.47	0	1
Secondary education	38,290	0.34	0.47	0	1
Higher education	38,290	0.12	0.32	0	1
Paid work	38,216	0.31	0.46	0	1
Regional paid work rate	38,395	0.31	0.17	0	0.79
Visited Public Health Centre	38,124	0.85	0.36	0	1
Electricity	38,395	0.54	0.50	0	1
Piped water	38,331	0.47	0.50	0	1
Sewage	38,108	0.21	0.41	0	1
Mobile	38,371	0.92	0.27	0	1
Post Office	38,283	0.18	0.38	0	1
School	38,267	0.88	0.32	0	1
Police Station	38,203	0.34	0.47	0	1
Clinic	38,196	0.59	0.49	0	1
Market	38,291	0.66	0.47	0	1
Religious group, inactive	38,157	0.17	0.37	0	1
Religious group, active	38,157	0.29	0.45	0	1
Religious group, leader	38,157	0.07	0.25	0	1
Association, inactive	38,049	0.14	0.34	0	1
Association, active	38,049	0.20	0.40	0	1
Association, leader	38,049	0.06	0.24	0	1
Trust general	37,618	0.20	0.40	0	1
Trust relatives	38,240	2.44	0.87	0	3
Trust neighbours	38,260	1.82	1.01	0	3
Trust others	38,148	1.35	1.02	0	3
Regional general trust rate	38,395	0.20	0.14	0	0.93
Feeling unsafe	38,238	0.79	1.18	0	4
Fearing crime	38,242	0.67	1.14	0	4
Experience stolen	38,358	0.53	0.93	0	3
Experience attacked	35,929	0.13	0.48	0	3
Fearing election	37,726	1.04	1.18	0	3

Notes: Unweighted statistics. Sample: Whole.

Table 2: Geographical Descriptive Statistics

	(1) Whole sample N = 38,395		(2) Urban sample N = 13,968		(3) Rural sample N = 24,427	
	Mean	SD	Mean	SD	Mean	SD
Distance to paved road (km)	11.58	19.35	5.84	16.57	14.87	20.05
Log distance to paved road	1.69	1.29	1.04	1.07	2.07	1.25
Distance to hypothetical line (km)	97.14	110.38	63.06	101.68	116.63	110.43
Log distance to hypothetical line	3.73	1.58	2.90	1.73	4.20	1.27
Distance to explorer route (km)	156.81	168.85	186.68	193.02	139.73	150.70
Log distance to explorer route	4.44	1.28	4.62	1.29	4.34	1.26
Slope (%)	1.63	2.60	1.41	2.34	1.75	2.72
Log slope	0.72	0.62	0.66	0.58	0.76	0.63
Distance to historical settlement (km)	127.37	122.81	82.74	114.47	152.90	120.10

Note: Unweighted statistics. Sample: Whole.

Table 3: First-Stage OLS Regressions

Dependent variable: Log distance to the nearest paved road

	Whole (1)	Urban (2)	Rural (3)
A. Dist to hypothetical lines	0.258***	0.230***	0.128***
(divided by 100)	(0.031)	(0.053)	(0.034)
N	37902	13822	24080
Adjusted R^2	0.202	0.185	0.166
F (excluded instruments)	67.97***	19.03***	14.21***
B. Log Dist to hypothetical lines	0.260***	0.177***	0.180***
	(0.017)	(0.021)	(0.028)
N	37902	13822	24080
Adjusted R^2	0.253	0.220	0.182
F (excluded instruments)	236.25***	73.98***	42.52***
C. Slope (%)	0.028***	0.021*	0.026**
	(0.010)	(0.012)	(0.011)
N	37397	13403	23994
Adjusted R^2	0.167	0.160	0.156
F (excluded instruments)	7.212***	2.95*	4.99**
D. Log slope	0.082*	0.030	0.077
	(0.050)	(0.059)	(0.058)
N	37397	13403	23994
Adjusted R^2	0.166	0.159	0.155
F (excluded instruments)	2.71	0.25	1.74
E. Dist to explorer routes	0.045*	0.123***	0.014
(divided by 100)	(0.027)	(0.034)	(0.034)
N	37902	13822	24080
Adjusted R^2	0.168	0.165	0.157
F (excluded instruments)	2.93*	12.83***	0.16
F. Log Dist to explorer routes	0.037	0.078*	0.014
	(0.028)	(0.042)	(0.029)
N	37902	13822	24080
Adjusted R^2	0.167	0.156	0.157
F (excluded instruments)	1.73	3.36*	0.23

Notes: Standard errors clustered at the community level appear in brackets. Controls: Age, age-squared, female, altitude, longitude, latitude and country fixed effects. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Sample: Whole.

Table 4: The Effect of Roads on Self-assessed Current Living Conditions and Lack of Deprivation

Outcomes	Living Conditions		Lack of Deprivation	
	OLS	2SLS	OLS	2SLS
	(1)	(2)	(3)	(4)
A. Whole sample				
Log distance to paved road	-0.029***	-0.005	-0.274***	-0.631***
	(0.007)	(0.023)	(0.031)	(0.099)
N	37764	37764	37498	37498
Adjusted R^2	0.084	0.083	0.097	0.085
B. Urban sample				
Log distance to paved road	-0.026**	-0.016	-0.138***	-0.441**
	(0.013)	(0.047)	(0.050)	(0.176)
N	13754	13754	13671	13671
Adjusted R^2	0.086	0.086	0.100	0.093
C. Rural sample				
Log distance to paved road	-0.012	0.231***	-0.170***	-0.270
	(0.009)	(0.069)	(0.037)	(0.226)
N	24010	24010	23827	23827
Adjusted R^2	0.085	0.031	0.093	0.092

Notes: Standard errors clustered at the community level appear in brackets. Controls: Age, age-squared, female, altitude, longitude, latitude and country fixed effects. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Whole Sample.

Table 5: The Effect of Roads for Communities over 10km from Historical Settlements

Dependent variable	Dist. to Road	Living Conditions		Lack of Deprivation	
Model	OLS	OLS	2SLS	OLS	2SLS
	(1)	(2)	(3)	(4)	(5)
Log distance to hyp. line	0.146***				
	(0.022)				
Log distance to paved road		-0.021***	0.181***	-0.233***	-0.741***
		(0.008)	(0.064)	(0.032)	(0.237)
N	32882	32769	32769	32545	32545
Adjusted R^2	0.198	0.085	0.046	0.102	0.077
F^*	43.29***				

Notes: Standard errors clustered at the community level appear in brackets. Controls: Age, age-squared, female, altitude, longitude, latitude and country fixed effects. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Sample: Communities over 10kms from historical settlements.

Table 6: Robustness Checks: The Dependent Variables

Model	Categorical variable		Binary variable	
	OLS	2SLS	OLS	2SLS
	(1)	(2)	(3)	(4)
Dimensions of basic needs				
A. Having enough food				
Log distance to paved road	-0.059***	-0.360***	-0.021***	-0.124***
	(0.009)	(0.073)	(0.003)	(0.027)
N	32825	32825	32825	32825
Adjusted R^2	0.074	.	0.068	0.008
B. Having enough water				
Log distance to paved road	-0.082***	-0.208**	-0.027***	-0.062**
	(0.012)	(0.090)	(0.004)	(0.031)
N	32826	32826	32826	32826
Adjusted R^2	0.069	0.058	0.059	0.052
C. Having enough medical care				
Log distance to paved road	-0.084***	-0.264***	-0.027***	-0.086***
	(0.010)	(0.073)	(0.004)	(0.026)
N	32732	32732	32732	32732
Adjusted R^2	0.096	0.071	0.082	0.063
D. Having enough cooking fuel				
Log distance to paved road	-0.010	0.098	-0.001	0.043*
	(0.010)	(0.073)	(0.003)	(0.026)
N	32708	32708	32708	32708
Adjusted R^2	0.064	0.053	0.060	0.047
Dummy variable for Living Conditions				
E. Having good or very good living conditions				
Log distance to paved road	-0.021***	0.181***	-0.006**	0.088***
	(0.008)	(0.064)	(0.003)	(0.026)
N	32769	32769	32769	32769
Adjusted R^2	0.085	0.046	0.068	0.011

Notes: In cols. 2 and 4, the log of 1 plus the distance to the nearest paved road is instrumented by the log of 1 plus the distance to the nearest hypothetical lines. Standard errors clustered at the community level appear in parentheses. Controls: Age, age-squared, female, altitude, longitude, latitude and country fixed-effects are included in each model. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Cols. 1 and 2 use the ordinal measures of whether each basic need is satisfied (values from 0 to 4) and living conditions (1 to 5); cols. 3 and 4 use the dummy variables for reporting having had difficulty in satisfying their basic needs in each dimension over the past year never or just once or twice, and for having good or very good living conditions. Sample: Communities over 10kms from historical settlements.

Table 7: Robustness Checks: Measures of Distance to the Nearest Paved Road

Outcome	Living Conditions (1)	Lack of Deprivation (2)
A. Log distance to road (Benchmark)	0.181***	-0.741***
	(0.064)	(0.237)
B. Dist to road > 5km	0.614**	-2.527***
	(0.246)	(0.934)
C. Dist to road > 10km	0.416***	-1.707***
	(0.141)	(0.566)
D. Dist to road > 20km	0.416***	-1.704***
	(0.139)	(0.534)
E. Distance to road (divided by 100)	0.893***	-2.738**
	(0.341)	(1.198)
F. Distance to road (divided by 100)	1.657***	-10.066***
	(0.592)	(3.159)
Squared distance to road (divided by 100)	-0.007	0.066**
	(0.005)	(0.032)
G. Inverse of (1+ Dist. to road)	-1.585	6.579
	(1.665)	(5.619)

Notes: Instrument: Distance to the nearest hypothetical lines (the log of 1 plus, or levels, or in quadratic, or 1 over 1+distance to HL). Standard errors clustered at the community level appear in parentheses. Controls: Age, age-squared, female, altitude, longitude, latitude and country fixed effects. Sample: Communities over 10kms from historical settlements. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 8: Robustness Checks regarding the Sample

Outcome	Living Conditions (1)	Lack of Deprivation (2)
Benchmark	0.181***	-0.741***
	(0.064)	(0.237)
N	32769	32545
Adjusted R^2	0.046	0.077
Removing obs. less than 5 kms from road	0.194***	-0.704**
	(0.068)	(0.279)
N	16588	16473
Adjusted R^2	0.073	0.074
Removing obs. less than 10 kms from road	0.273**	-0.928**
	(0.116)	(0.434)
N	12405	12319
Adjusted R^2	0.070	0.063
Removing people aged over 70	0.171***	-0.768***
	(0.064)	(0.239)
N	31942	31719
Adjusted R^2	0.049	0.074

Notes: These are the instrumented estimates for the log of 1 plus distance to the nearest paved road in linear regressions. Standard errors clustered at the community level appear in brackets. Controls: Age, age-squared, female, altitude, longitude, latitude and country fixed effects. Sample: Communities over 10km from historical settlements. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 9: Mediation Analysis

Outcome	Living Conditions (1)	Lack of Deprivation (2)
A. Benchmark	0.181***	-0.741***
	(0.064)	(0.237)
B. A + urban location	0.334***	-0.470
	(0.110)	(0.327)
C. A + education	0.259***	-0.504**
	(0.074)	(0.236)
D. A + paid work	0.208***	-0.666***
	(0.068)	(0.236)
E. A + paid work and regional rate	0.252***	-0.613**
	(0.073)	(0.239)
F. A + visited Public Health Centre	0.181***	-0.699***
	(0.065)	(0.235)
G. A + public goods	0.455***	-0.105
	(0.136)	(0.325)
H. A + association	0.185***	-0.702***
	(0.066)	(0.240)
I. A + trust general	0.183***	-0.757***
	(0.066)	(0.242)
J. A + trust general and regional rate	0.163**	-0.801***
	(0.065)	(0.242)
K. A + trust w.r.t. relatives, neig. and other	0.158**	-0.775***
	(0.064)	(0.246)
L. A + feelings of security and crime	0.188***	-0.763***
	(0.065)	(0.238)
M. A + experiences of crime	0.180***	-0.592**
	(0.067)	(0.235)
N. A + H, I, K, L, M	0.157**	-0.613**
	(0.070)	(0.249)
O. A + All Mediators (except M)	0.664***	0.279
	(0.217)	(0.403)

Notes: These are the instrumented estimates for the log of 1 plus distance to the nearest paved road in linear regressions. Standard errors clustered at the community level appear in parentheses. Controls: Age, age-squared, female, altitude, longitude, latitude and country fixed-effects. The models in rows M and N exclude Tanzania. Sample: Communities over 10kms from historical settlements. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

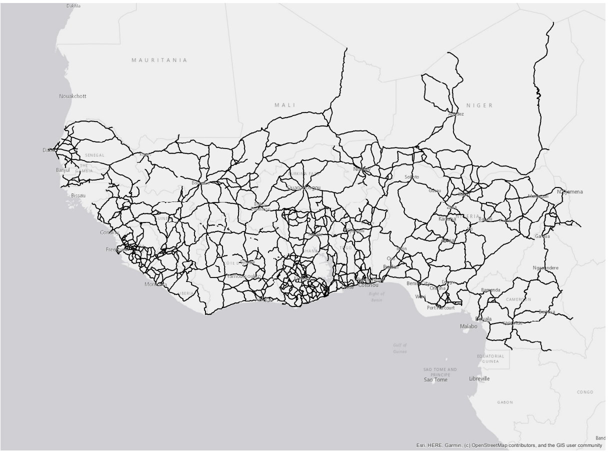
Table 10: The Effect of Roads on the Wealth Index: DHS Data

Dependent Variable: Standardized Wealth Index

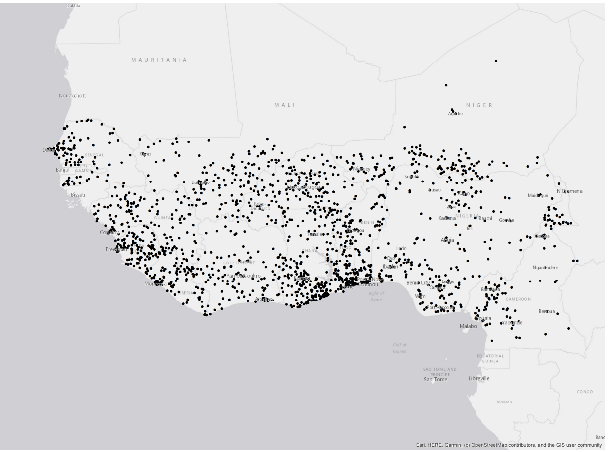
	(1) OLS	(2) 2SLS
A. Average effects		
Log distance to paved road	-0.240***	-0.740***
	(0.011)	(0.086)
N	234962	234962
Adjusted R^2	0.089	.
B. Heterogeneous effects		
Log distance to paved road	-0.181***	-0.648***
	(0.009)	(0.082)
Log distance to paved road × migrant	-0.106***	-0.140**
	(0.010)	(0.063)
Migrant	0.364***	0.327***
	(0.022)	(0.119)
N	234601	234601
Adjusted R^2	0.099	.

Notes: Standard errors clustered at the community level appear in parentheses. Controls: Age, age-squared, female, altitude, longitude, latitude and country fixed effects. Sample: Communities over 10kms from historical settlements. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

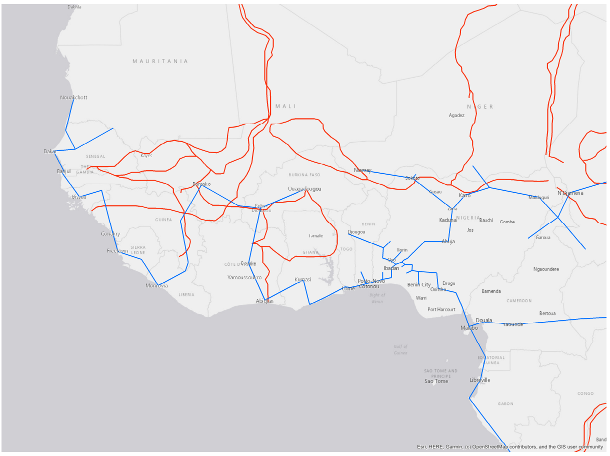
Figure 1: Geographic data, West Africa



(a) Location of current paved roads



(b) Location of the enumeration areas



(c) Instruments - Location of the hypothetical lines (blue) and explorer routes (red)

Figure 2: F-statistic from the First-stage Regressions, by Minimum Distance to the Nearest Historical Settlement

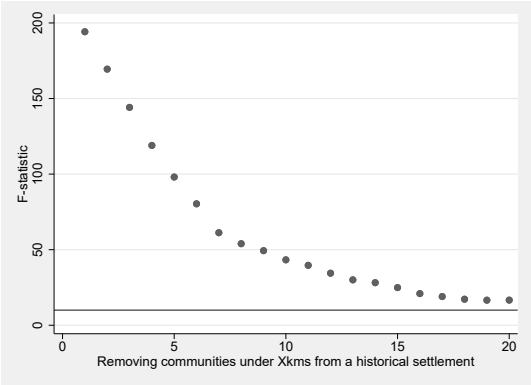
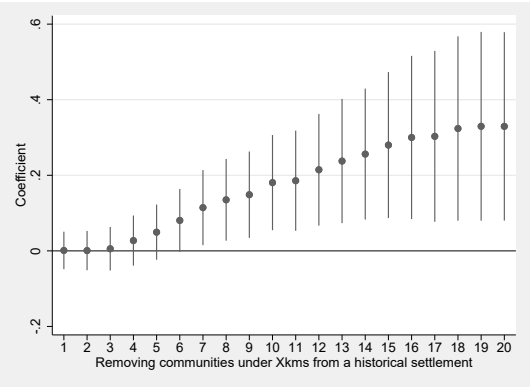
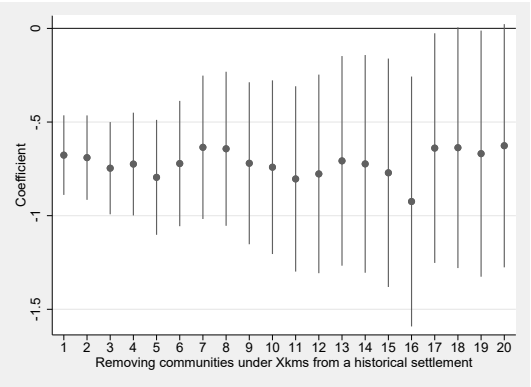


Figure 3: Effects of Roads, by Minimum Distance to the Nearest Historical Settlement

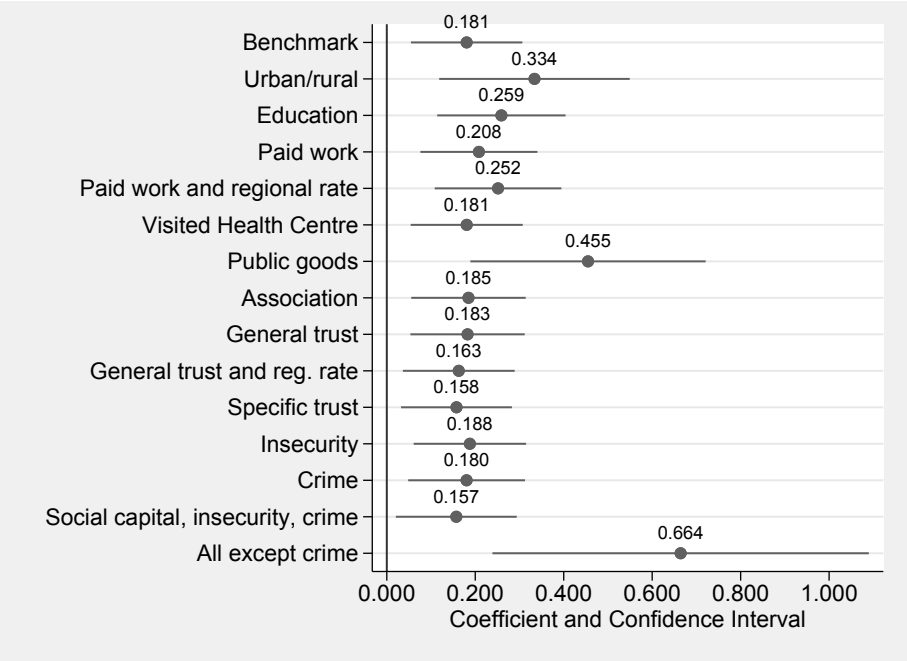


(a) Effects on subjective living conditions

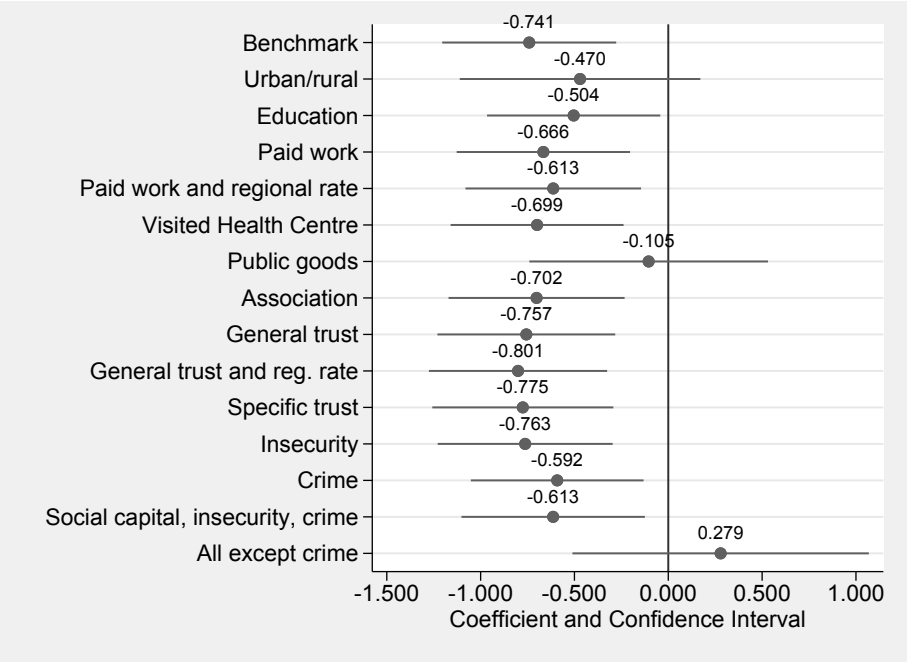


(b) Effects on the lack of deprivation

Figure 4: Mediation in the effects of roads on well-being

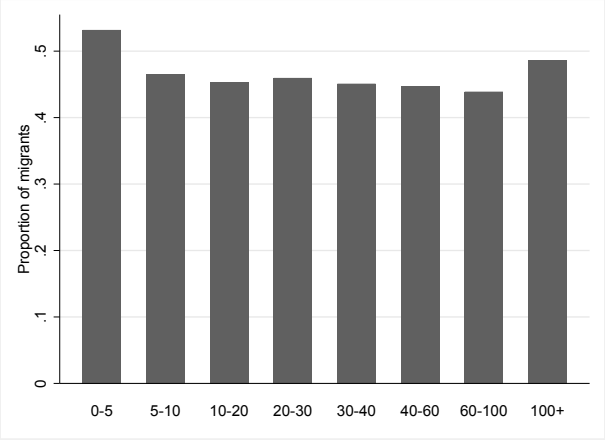


(a) Effects on subjective living conditions



(b) Effects on lack of deprivation

Figure 5: Proportions of ever migrants, by distance to the current paved road



Appendix

Table A1: List of Countries, Sample Size and Means of Key Variables

Countries	N	Living Conditions	Lack of Deprivation	Distance to Road	Log Distance to Road
Benin	1,200	2.58	11.83	3.52	1.05
Burkina Faso	1,200	2.90	10.31	14.12	1.84
Burundi	1,200	2.59	9.81	8.37	1.67
Botswana	1,200	2.39	12.53	14.48	1.65
Cameroon	1,200	2.93	10.04	11.57	1.52
Cote d'Ivoire	1,200	2.62	10.23	6.80	1.33
Ghana	2,400	2.52	14.01	2.51	0.83
Guinea	1,200	2.62	10.02	7.65	1.58
Kenya	2,399	1.96	12.09	6.79	1.24
Liberia	1,199	3.07	10.77	5.73	1.22
Mali	1,200	2.70	11.82	11.45	1.69
Malawi	2,408	2.51	11.51	9.89	1.88
Mozambique	2,400	2.86	11.98	24.14	2.32
Namibia	1,200	3.04	12.89	14.25	1.62
Niger	1,199	3.09	9.67	11.65	1.66
Nigeria	2,400	2.91	11.49	11.72	1.74
Senegal	1,200	2.73	10.12	4.83	1.11
Sierra Leone	1,190	2.99	11.09	5.80	1.30
Sudan	1,199	2.60	11.46	19.55	1.63
Tanzania	2,400	2.13	11.49	17.30	2.22
Togo	1,201	2.30	9.66	4.05	1.12
Uganda	2,400	2.33	11.15	12.34	2.08
Zambia	1,200	3.21	11.94	24.38	2.52
Zimbabwe	2,400	2.76	11.38	16.53	2.54

Notes: Unweighted statistics. Whole Sample.

Table A2: Sample Descriptive Statistics

	N	Mean	SD	Min	Max
Having enough food [0;4]	38,323	2.84	1.23	0	4
Having enough water [0;4]	38,325	2.76	1.39	0	4
Having enough medical care [0;4]	38,207	2.69	1.31	0	4
Having enough cooking fuel [0;4]	38,184	3.08	1.21	0	4
Having enough food [0;1]	38,323	0.60	0.49	0	1
Having enough water [0;1]	38,325	0.59	0.49	0	1
Having enough medical care [0;1]	38,207	0.56	0.50	0	1
Having enough cooking fuel [0;1]	38,184	0.69	0.46	0	1

Notes: Unweighted statistics. Sample: Whole.

Table A3: First-Stage OLS Regression - All Controls

Dependent variable: Log distance to paved road

	(1)	(2)	(3)	(4)	(5)	(6)
Log distance to hypothetical line	0.260***					
	(0.017)					
Distance to hypothetical line/100		0.258***				
		(0.031)				
Log distance to explorer route			0.037			
			(0.028)			
Distance to explorer route/100				0.045*		
				(0.027)		
Log slope					0.082*	
					(0.050)	
Slope (%)						0.028***
						(0.010)
Age	0.006***	0.007***	0.008***	0.008***	0.008***	0.008***
	(0.002)	(0.002)	(0.003)	(0.003)	(0.003)	(0.003)
Age-squared/100	-0.005*	-0.005*	-0.006*	-0.006**	-0.006**	-0.006**
	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
Female	0.005	0.008**	0.008**	0.008**	0.008**	0.008**
	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
Latitude	0.040***	0.051***	0.087***	0.093***	0.083***	0.084***
	(0.014)	(0.015)	(0.015)	(0.016)	(0.015)	(0.015)
Longitude	0.025*	0.030*	0.049***	0.048***	0.046***	0.045***
	(0.014)	(0.016)	(0.016)	(0.016)	(0.016)	(0.016)
Altitude	-0.000***	-0.000*	-0.000**	-0.000**	-0.000***	-0.000***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)

Notes: Standard errors clustered at the community level appear in parentheses. Country fixed-effects are included but the coefficients are not shown. Sample: Whole. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A4: Distance to Roads and Well-being - All Controls

	Living Conditions		Lack of Deprivation	
	OLS	2SLS	OLS	2SLS
	(1)	(2)	(3)	(4)
Log distance to paved road	-0.029***	-0.005	-0.274***	-0.631***
	(0.007)	(0.023)	(0.031)	(0.099)
Age	-0.028***	-0.028***	-0.052***	-0.049***
	(0.002)	(0.002)	(0.007)	(0.007)
Age-squared/100	0.024***	0.024***	0.038***	0.036***
	(0.002)	(0.002)	(0.008)	(0.008)
Female	-0.032***	-0.032***	-0.063**	-0.060**
	(0.011)	(0.011)	(0.030)	(0.030)
Latitude	0.024***	0.022***	-0.018	0.012
	(0.005)	(0.005)	(0.018)	(0.021)
Longitude	-0.001	-0.002	-0.024	-0.006
	(0.005)	(0.005)	(0.021)	(0.022)
Altitude	-0.000	-0.000	0.000**	0.000
	(0.000)	(0.000)	(0.000)	(0.000)
Constant	3.486***	3.472***	11.838***	12.062***
	(0.175)	(0.175)	(0.775)	(0.810)
N	37764	37764	37498	37498
Adjusted R^2	0.084	0.083	0.097	0.085

Notes: Standard errors clustered at the community level appear in parentheses. Country fixed-effects are included but the coefficients are not reported. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A5: Restricted Sample Descriptive Statistics

Variables	(1) N	(2) Mean	(3) SD	(4) Min	(5) Max
Current living conditions (LC)	33,220	2.62	1.19	1	5
Lack of depriv., over 4 items	32,998	11.29	3.74	0	16
Distance to road, km	33,341	13.16	20.29	0	248.05
Log distance to road	33,341	1.86	1.29	0	5.52
Distance to hypothetical line, km	33,341	111.46	111.68	0.02	733.56
Log distance to hypothetical line	33,341	4.13	1.26	0.02	6.60
Age	32,999	36.67	14.23	18	105
Female	33,341	0.50	0.50	0	1
Urban	33,341	0.29	0.45	0	1
Primary education	33,261	0.34	0.47	0	1
Secondary education	33,261	0.33	0.47	0	1
Higher education	33,261	0.09	0.29	0	1
Paid work	33,207	0.30	0.46	0	1
Regional paid work	33,341	0.30	0.17	0	0.79
Visited Public Health Centre	33,097	0.86	0.35	0	1
Electricity	33,341	0.50	0.50	0	1
Piped water	33,285	0.43	0.49	0	1
Sewage	33,125	0.18	0.38	0	1
Mobile	33,317	0.91	0.29	0	1
Post Office	33,245	0.17	0.37	0	1
School	33,221	0.88	0.33	0	1
Police Station	33,197	0.31	0.46	0	1
Clinic	33,165	0.57	0.50	0	1
Market	33,237	0.64	0.48	0	1
Religious group, inactive	33,159	0.17	0.37	0	1
Religious group, active	33,159	0.29	0.45	0	1
Religious group, leader	33,159	0.07	0.25	0	1
Association, inactive	33,069	0.14	0.35	0	1
Association, active	33,069	0.21	0.40	0	1
Association, leader	33,069	0.06	0.25	0	1
Trust general	32,673	0.21	0.40	0	1
Trust relatives	33,203	2.45	0.87	0	3
Trust neighbours	33,234	1.86	1.01	0	3
Trust others	33,143	1.38	1.02	0	3
Regional general trust rate	33,341	0.21	0.15	0	0.93
Feeling unsafe	33,216	0.75	1.16	0	4
Fearing crime	33,223	0.64	1.11	0	4
Experience stolen	33,311	0.52	0.92	0	3
Experience attacked	31,144	0.13	0.47	0	3
Fearing election	32,777	1.04	1.18	0	3

Notes: Unweighted statistics. Sample: Communities over 10kms from historical settlements.

Table A6: Jackknife Analysis

	(1) Living Conditions	(2) Lack of Deprivation
All	0.181***	-0.741***
	(0.064)	(0.237)
Burundi	0.164**	-0.817***
	(0.066)	(0.252)
Benin	0.181***	-0.767***
	(0.064)	(0.238)
Burkina Faso	0.179***	-0.726***
	(0.066)	(0.243)
Botswana	0.213***	-0.740***
	(0.068)	(0.241)
Cameroon	0.167**	-0.875***
	(0.066)	(0.255)
Cote d'Ivoire	0.178***	-0.785***
	(0.065)	(0.243)
Ghana	0.151***	-0.576**
	(0.058)	(0.224)
Guinea	0.182***	-0.760***
	(0.064)	(0.237)
Kenya	0.210***	-0.679***
	(0.066)	(0.227)
Liberia	0.181***	-0.926***
	(0.068)	(0.261)
Mali	0.174***	-0.732***
	(0.065)	(0.240)
Malawi	0.178***	-0.793***
	(0.063)	(0.232)
Mozambique	0.225***	-0.857***
	(0.071)	(0.258)
Namibia	0.173***	-0.776***
	(0.065)	(0.247)
Niger	0.184***	-0.692***
	(0.063)	(0.229)
Nigeria	0.158**	-0.485**
	(0.063)	(0.228)
Senegal	0.196***	-0.743***
	(0.068)	(0.244)
Sierra Leone	0.188***	-0.741***
	(0.066)	(0.242)
Sudan	0.214***	-0.703***
	(0.069)	(0.246)
Tanzania	0.176***	-0.636***
	(0.067)	(0.242)
Togo	0.174***	-0.654***
	(0.067)	(0.238)
Uganda	0.137**	-0.641***
	(0.063)	(0.247)
Zambia	0.187***	-0.780***
	(0.068)	(0.252)
Zimbabwe	0.168**	-0.924***
	(0.068)	(0.251)

Notes: Standard errors clustered at the community level appear in parentheses. Controls: Age, age-squared, female, altitude, longitude, latitude and country fixed effects. Sample: EA over 10km from historical settlements. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table A7: Sample size, DHS

Country	Year	Obs.	Restricted Obs.
Benin	2017	23320	19138
Burkina Faso	2003	15894	14098
Cameroon	2004	15188	12115
Ghana	2008	9396	8201
Guinea	2005	10959	9953
Kenya	2009	11649	10335
Liberia	2007	12889	10674
Mali	2006	18550	15780
Malawi	2010	29756	27516
Namibia	2007	13128	11492
Nigeria	2008	48153	43065
Senegal	2005	17846	15834
Tanzania	2015	16280	14555
Uganda	2006	9725	8020
Zimbabwe	2015	17989	15978
Total		270722	236754

Table A8: Descriptive Statistics, DHS

Variable	*N*	Mean	Std. Dev.	Min	Max
Migrant	236,388	.49	.49	0	1
Std. wealth index	236,754	0	1	-6.25	11.56
Wealth index (raw, continuous)	236,754	-11340.24	260297.2	-1638620	2996530
Log distance to paved road	236,754	1.95	1.22	.00003	5.31
Log distance to hyp. line	236,754	4.08	1.21	.002	6.69
Dist. to hyp. line (/100)	236,754	1.03	1.02	.00002	8.07
Log distance to explorer route	236,754	4.57	1.38	.0121201	6.829656
Dist. to expl. route (/100)	236,754	1.92	2.07	.0001	9.24
Log slope	232,831	.63	.598	0	4.14
Slope (%)	232,831	1.48	2.54	0	62.02
Age	236,754	29.21	10.30	15	64
Age-squared (over 100)	236,754	959.38	663.61	225	4096
Female	236,754	.70	.46	0	1
Latitude	236,754	2.00	11.52	-28.57	20.18
Longitude	236,754	11.48	17.45	-17.47	41.83
Altitude	234,962	573.53	498.15	1	2951

Notes: Unweighted statistics. Sample: EA over 10kms from historical settlements (permanent residents).

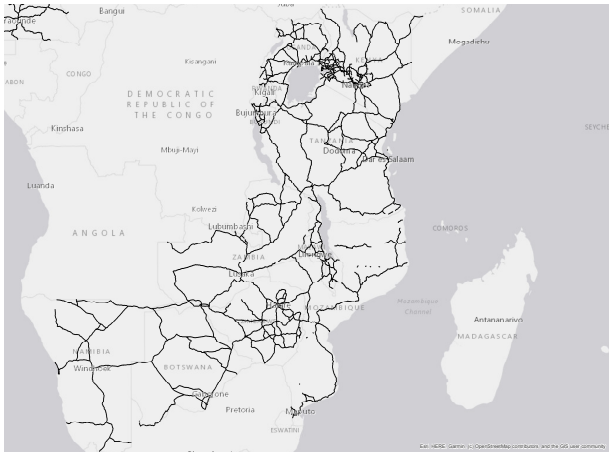
Table A9: First-Stage OLS Regressions (DHS)

Dependent variable: Log distance to the nearest paved road

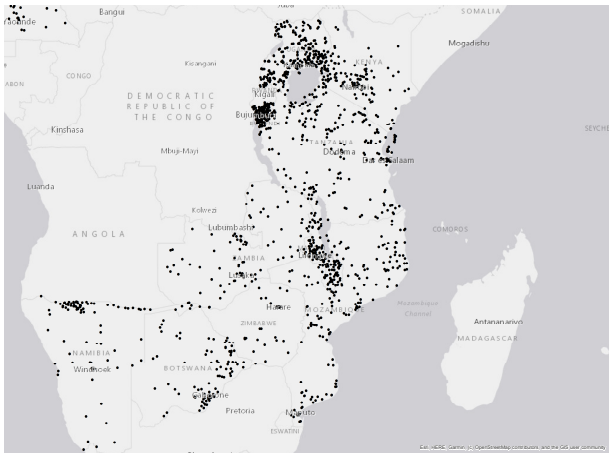
	(1)
A. Dist to hypothetical lines	0.182***
	(0.020)
N	234962
Adjusted R^2	0.123
F (excluded instruments)	84.74***
B. Log Dist to hypothetical lines	0.176***
	(0.014)
N	234962
Adjusted R^2	0.131
F (excluded instruments)	160.30***
C. Slope (%)	0.014**
	(0.006)
N	232831
Adjusted R^2	0.103
F (excluded instruments)	6.24**
D. Log slope	0.028
	(0.031)
N	232831
Adjusted R^2	0.102
F (excluded instruments)	0.803
E. Dist to explorer routes	-0.021
	(0.016)
N	234962
Adjusted R^2	0.107
F (excluded instruments)	1.82
F. Log Dist to explorer routes	-0.010
	(0.016)
N	234962
Adjusted R^2	0.107
F (excluded instruments)	0.42

Notes: Standard errors clustered at the community level appear in parentheses. Controls: Age, age-squared, female, altitude, longitude, latitude and country fixed effects. Sample: Communities over 10kms from historical settlements. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

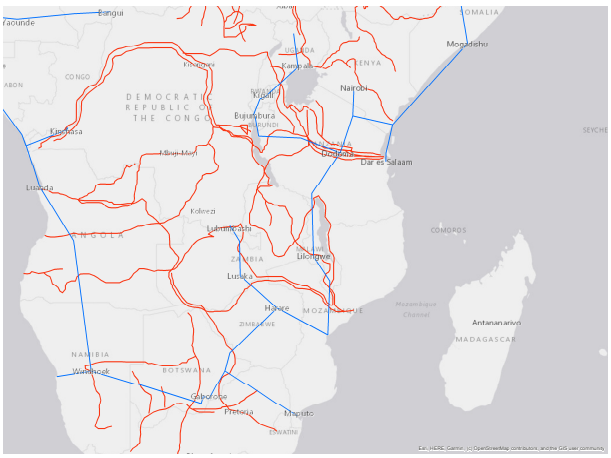
Figure A1: Geographic data, Africa (second part of the sample)



(a) Location of the current paved roads

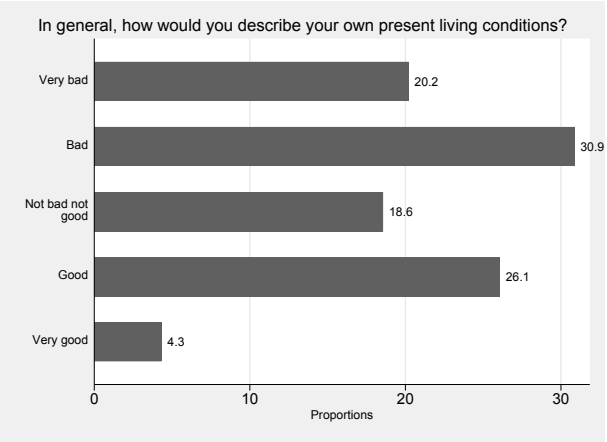


(b) Location of the enumeration areas

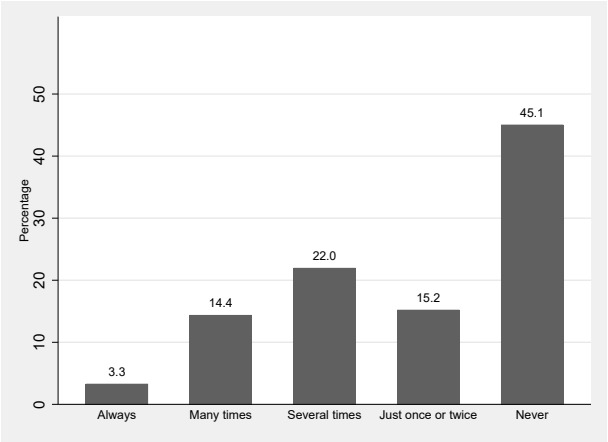


(c) Instruments - Location of the hypothetical lines (blue) and explorer routes (red)

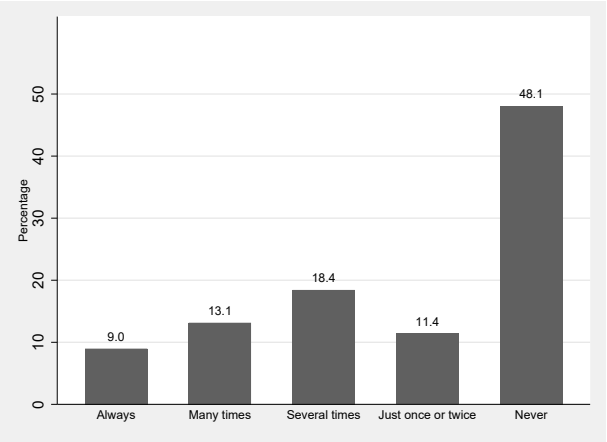
Figure A2: Distribution of the Categorical Variables



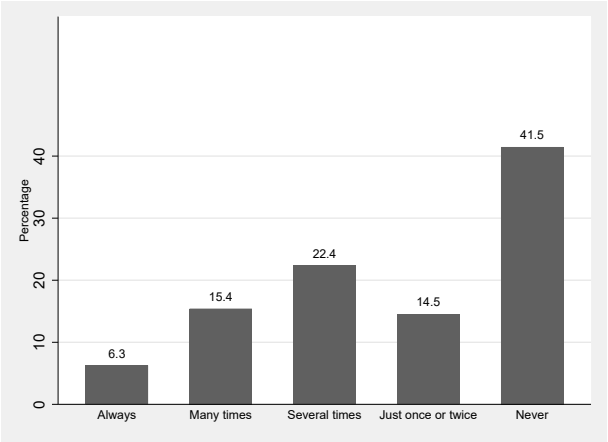
(a) Subjective well-being



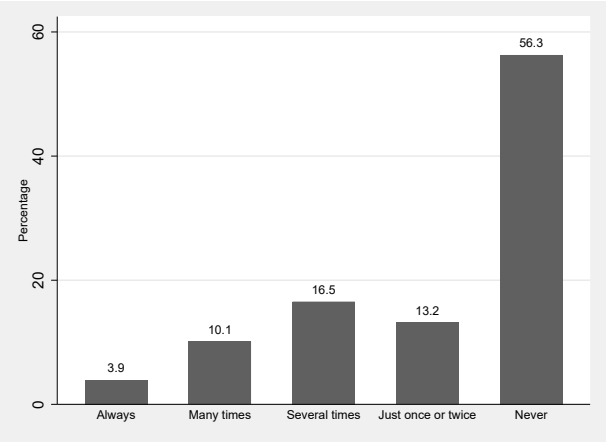
(b) Gone without enough food



(c) Gone without enough water

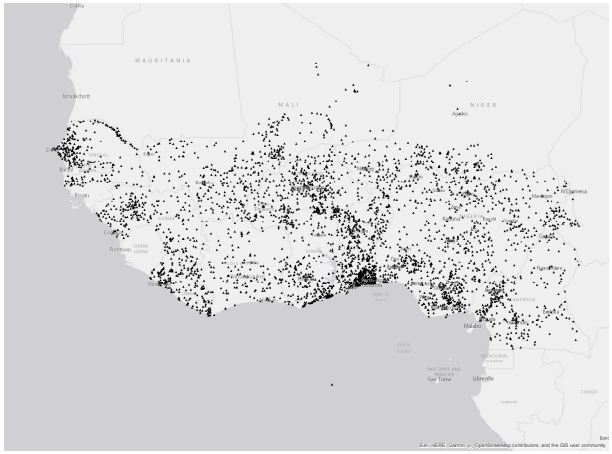


(d) Gone without enough medical care

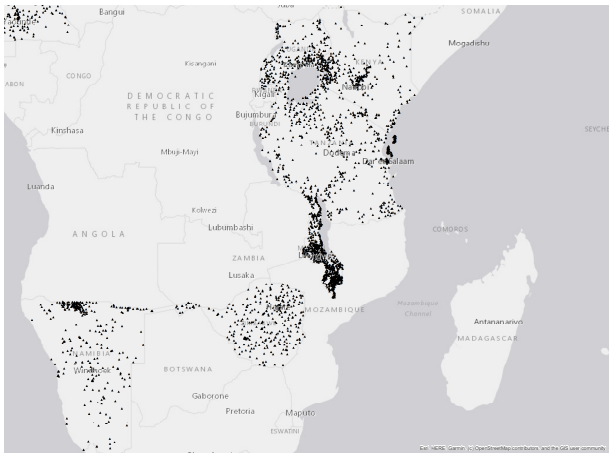


(e) Gone without enough cooking fuel

Figure A3: Location of the EA in the DHS



(a) West Africa



(b) South-East

Table O1: Robustness checks - Main results clustering the standard errors at the regional level

	Dist to road	Living Conditions		Lack of Deprivation	
	OLS	OLS	2SLS	OLS	2SLS
	(1)	(2)	(3)	(4)	(5)
Log distance to hyp. line	0.146***				
	(0.027)				
Log distance to paved road		-0.021**	0.181*	-0.233***	-0.741**
		(0.008)	(0.105)	(0.037)	(0.347)
N	32882	32769	32769	32545	32545
Adjusted R^2	0.198	0.085	0.046	0.102	0.077
F (excluded instruments)	29.44***				

Notes: Standard errors clustered at the regional level appear in parentheses. Controls: Age, age-squared, female, altitude, longitude, latitude and country fixed effects. Sample: Communities over 10kms from historical settlements. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table O2: The Effect of Roads on Self-assessed Current Living Conditions and Lack of Deprivation - Ordered Probit Estimations

Outcomes	Living Conditions		Lack of Deprivation	
Instrumented	No	Yes	No	Yes
	(1)	(2)	(3)	(4)
A. Whole sample				
Log distance to paved road	-0.025***	-0.002	-0.081***	-0.194***
	(0.007)	(0.021)	(0.009)	(0.028)
N	37764	37764	37498	37498
B. Urban sample				
Log distance to paved road	-0.023*	-0.011	-0.039***	-0.138***
	(0.012)	(0.044)	(0.015)	(0.052)
N	13754	13754	13671	13671
C. Rural sample				
Log distance to paved road	-0.010	0.208***	-0.051***	-0.096
	(0.009)	(0.058)	(0.011)	(0.065)
N	24010	24010	23827	23827
D. Communities over 10 kms from historical settlements				
Log distance to paved road	-0.018**	0.169***	-0.069***	-0.229***
	(0.007)	(0.056)	(0.009)	(0.066)
N	32769	32769	32545	32545

Notes: Standard errors clustered at the community level appear in parentheses. Controls: Age, age-squared, female, altitude, longitude, latitude and country fixed effects. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table O3: Robustness checks - using solid roof as the outcome variable

	(1)	(2)
	OLS	2SLS
Log distance to paved road	-0.033***	-0.206***
	(0.004)	(0.041)
N	32727	32727
Adjusted R^2	0.271	0.100

Notes: Standard errors clustered at the community level appear in parentheses. Controls: Age, age-squared, female, altitude, longitude, latitude and country fixed effects. Sample: Communities over 10km from historical settlements. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table O4: Robustness Checks for the Definition of the Urban and Rural Samples

Outcome	Living Conditions (1)	Lack of Deprivation (2)
A. Define as urban EAs where at least 50% are reported as urban		
Urban sample	0.186	-0.697
	(0.190)	(0.624)
N	9422	9369
Adjusted R^2	0.058	0.075
Rural sample	0.366***	-0.544*
	(0.112)	(0.325)
N	23347	23176
Adjusted R^2	.	0.083
B. Define as urban EAs where there is disagreement		
Urban sample	0.222*	-0.229
	(0.135)	(0.418)
N	12496	12415
Adjusted R^2	0.045	0.093
Rural sample	0.335***	-0.681*
	(0.113)	(0.350)
N	20273	20130
Adjusted R^2	.	0.072
C. Define as rural EAs where there is disagreement		
Urban sample	0.234	-0.831
	(0.203)	(0.689)
N	7949	7903
Adjusted R^2	0.034	0.040
Rural sample	0.332***	-0.514
	(0.104)	(0.318)
N	24820	24642
Adjusted R^2	.	0.102

Notes: These are the estimated coefficients on distance to road in the different well-being regressions. Standard errors clustered at the community level appear in parentheses. Controls: Age, age-squared, female, altitude, longitude, latitude and country fixed effects. Sample: Communities over 10km from historical settlements. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table O5: First-stage regressions controlling for the possible mediators

Dependent variable: Log distance to the nearest paved road

	(1) Coef.	(2) SE.	(3) F-stat
A. Benchmark	0.146***	(0.022)	43.29***
B. A adding urban location	0.103***	(0.022)	22.69***
C. A adding education	0.139***	(0.022)	39.09***
D. A adding paid work	0.144***	(0.022)	41.96***
E. A adding paid work and regional rate	0.143***	(0.023)	39.71***
F. A adding visited Public Health Centre	0.146***	(0.022)	42.71***
G. A adding public goods	0.095***	(0.021)	19.66***
H. A adding association	0.143***	(0.022)	41.93***
I. A adding trust general	0.145***	(0.022)	42.05***
J. A adding trust general and regional rate	0.146***	(0.023)	41.05***
K. A adding trust w.r.t. relatives, neig. and other	0.143***	(0.022)	40.89***
L. A adding feelings of security and crime	0.144***	(0.022)	42.65***
M. A adding experiences of crime	0.150***	(0.023)	41.18***
N. A adding all proxy var. for crime and social capital	0.142***	(0.023)	37.13***

Notes: Standard errors clustered at the community level appear in parentheses. Controls: Age, age-squared, female, altitude, longitude, latitude and country fixed effects. Sample: Communities over 10kms from historical settlements. Rows M and N do not include Tanzania, for which the experience of being attacked variable is missing. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table O6: Second-stage estimations from the Afrobarometer data, using only the 15 countries that appear in the DHS

Outcome	Living Conditions (1)	Lack of Deprivation (2)
Log distance to paved road	0.313***	-0.585*
	(0.111)	(0.341)
N	21685	21563
Adjusted R^2	.	0.109

Notes: Standard errors clustered at the community level appear in parentheses. Controls: Age, age-squared, female, altitude, longitude, latitude and country fixed effects. Sample: Communities over 10km from historical settlements. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table O7: The Number of Historical Settlements per Country

Country	Number
Angola	3
Benin	4
Botswana	5
Burkina Faso	2
Burundi	2
Cameroon	3
Cote d'Ivoire	2
Ghana	4
Guinea	3
Liberia	2
Malawi	2
Mali	2
Mozambique	2
Namibia	2
Niger	2
Nigeria	35
Senegal	3
Sierra Leone	2
Sudan	10
Tanzania, United Rep of	6
Togo	2
Uganda	2
Zambia	2
Zimbabwe	2