Data for Children Collaborative
WITH UNICEF

Child Poverty and Access to Services Analysis Summary

Combining household surveys and geospatial data to examine if physical access to basic services is a determinant of multi-dimensional childhood poverty?

THE UNIVERSITY of EDINBURGH

unicef for every child

Table of Contents

Introduction

Child Poverty and Access to Services (CPAS) Analysis Summary: Combining household surveys and geospatial data to examine if physical access to basic services is a determinant of multidimensional childhood poverty?

This document describes the work conducted for the **Child Poverty and Access to Services (CPAS)** project. The ultimate aim of the project was to determine if **geographic access to health centres and schools correlated with multidimensional child poverty.** Any child who is deprived of the right to education, health, housing, nutrition, sanitation, or water is living in poverty. Research into pregnancy and delivery service shows that maternal age, education and household wealth as well as distance to health facilities can all be determinants of service use but the impact of these can vary. There is currently insufficient empirical evidence at a country level to understand the reasons why children are deprived of access to basic services. This has consequences for developing quantifiable information about where and how much to invest in particular services. There are complex multidimensional reasons for children not accessing services such as lack of affordability, poor education of household members, lack of time and physical barriers to travel such as distance, lack of transport options, poor weather. There have been studies in the past examining the correlation between spatial access to surgical care[1] and child birth delivery services[2]. This project focused on the physical accessibility of key child services (health services/clinics and schools) in selected countries. **Problem Question posed by UNICEF:** is physical access (measured in distance/time) to basic services (health, education, water supply) a determinant of childhood poverty in rural regions?

The overall goal of the project was to explore if access (measured in terms of distance and travel time) to certain services determined using geospatial data sets such as remotely sensed satellite data and open street map were a determinant of childhood poverty. A secondary objective was to develop a method for estimating access to services that could be replicated in multiple countries and thus overcoming a gap in current understanding about how physical access to basic services impacts childhood poverty. The Challenge for UNICEF in provisioning child services is that they rely on household surveys such as the Multiple Indicator Cluster Surveys (MICS) and the Demographic and Health Survey (DHS). MICS provides internationally comparable estimates of about 130 indicators to assess the situation of children, women and men in the areas of health, education and child protection and the DHS provides supplementary information on health and wellbeing. Included in the surveys are indicators on (1) time to travel to main water

[1] Tansley et al. (2017) *World J Surg*. 41(3);639-643.
[2] Nesbitt et al. (2014) *Int. J of health Geog*. 13(25)

source; (2) if clinics were visited during specific medical problems for children. However, the surveys often do not ask clarifying questions such as possible reasons that families do not seek treatment for poorly children. So this project was framed around the following research question: *"We know that children lacking access to services are poor, but why do they lack access?"* **The objective** was to identify if distance or travel time explained the lack of access by generating **national level** estimations of distance and travel time from water sources and health facilities.

Approach

The project was split into three phases:

1. Phase 1: An upfront mini project collated different sources of information to select a single country in Sub-Saharan Africa (SSA) on which we focused for the remainder of the project. The country was selected based on the following criteria:
    a. DHS household survey data included GPS locations for download;
    b. DHS household survey included the module on distance to travel to water source;
    c. The country had clinics recorded in the World Health Organisation (WHO) spatial database of health facilities[3] and included latitude of longitude locations for each health facility.
    d. Satellite data was available covering the whole country and had been acquired within 12 months of the DHS household survey for that country.

2. Phase 2: for the shortlisted country (Uganda), we developed fine spatial resolution estimations of access measured in travel time in minutes from communities to health clinics and schools. This phase was split into three sub-goals:
    a. Goal 1. Identified rural communities using the high-resolution settlement layer (HRSL)[4] within the selected country(s) and linked these to the survey locations using GIS techniques[5].
    b. Goal 2. Using the least cost allocation approach to combine Earth Observation land cover data, elevation and roads we estimated travel time from health clinic locations across rural areas of Uganda.
    c. Goal 3. Identify if there is an optimal way of measuring access to services? Are displaced DHS clusters adequate for calculating access or do we need un-displaced data points for clusters? How does the fine spatial resolution approach compare to the current 1km global version that is often used.

[3] Maina et al. (2019) *Scientific Data, 6(134)*
[4] Facebook Connectivity Lab and Center for International Earth Science Information Network - CIESIN - Columbia University. 2016.
[5] Grace et al. (2019) *Pop & Dev Review* 45(1);197-218

3. Phase 3: Using the travel time estimations along with other socioeconomic and environmental factors known to impact wellbeing and deprivation we examined statistical relationships between travel time to services and child multidimensional deprivation.
    a. Given the definition of child deprivation used by UNICEF, what we were interested to look at the specific (quantifiable) contribution of distance/geographic accessibility to health and school services to child poverty? Specific research questions that we answered:
        i. What is the Multidimensional poverty index score for each village?
        ii. What is the median distance travelled to school and the median travel time to clinics for each village? We will also aggregate these access estimates to the district level.
        iii. What is the variation between villages and districts?
        iv. What is the relationship between poverty and the physical access to water and clinics?

Phase 1: Shortlisting

We developed a multi-criteria analysis (MCA) process to examine empirically the available data for countries across Africa. There were 51 countries contained in the WHO/Maina et al. (2019) health location database[6]. These 51 countries were cross referenced with available DHS data since 2009 that also had GPS data available. Any countries either with surveys pre-dating 2009 or not including GPS were removed. We also considered the high-resolution settlement layer availability (HRSL). This resulted in 31 countries on the long list *[note that originally we intended to calculate the location of every individual settlement in the selected country and then link the displaced DHS data to the nearest point to overcome the GPS displacement issues. However, subsequently during the project we managed to partner with DHS and they ran calculations for us to link the travel time estimates to the un-displaced locations which rendered some of the work that we did on HRSL void – it is included in this report to show what we did however].* We also considered the different DHS survey types and the appropriateness of each for the study. We defined appropriateness as surveys that contained questions and answers that could be used to calculate the child multidimensional poverty index as well as spatial data that could provide information on service use.

The Child multidimensional poverty index (MPI) contains three dimensions (health, education and living standards) and 11 indicator variables (Table 1)[7]. The following surveys available through the DHS database could not be used (1) AIDS Indicator surveys as there were not enough questions on assets[8] preventing us from calculating living standards; (2) special surveys as they are done by DHS for specific government policy making and can contain different variables in each so are often not standardised, and: (3) Continuous DHS surveys for example in Senegal are not replicated elsewhere. The focus on standardisation was important as this study was seen as a proof-of-concept for UNICEF and thus we needed to be able to apply the methodology to other countries in the future.

[6] Maina et al. (2019) *Scientific Data* 6(1)

[7] Alkire et al. 2017 OPHI briefing 46 - https://www.ophi.org.uk/wp-content/uploads/Brief_46_Child_MPI_2017-1.pdf

[8] DHS (2020) https://dhsprogram.com/What-We-Do/Survey-Types/AIS.cfm

Table 1: Child multidimensional poverty indicator variables

Health	Education	Living Standards
Nutrition	Years of schooling	Cooking fuel
Child mortality	School attendance	Sanitation
		Clean Drinking water
		Improved sanitation
		Electricity
		Floor type
		Household productive assets

Surveys that could be considered

The standard DHS survey[9] and the malaria indicator survey (MIS)[10] were the two most useful surveys. See table 2 for the breakdown of DHS survey questions we may need to use.

Round 2 shortlisting:

Round 2 shortlisting took the 31 countries on the long list and looked at the year of standard DHS Surveys and only considered countries that had completed a Standard DHS survey since 2009.

Round 3 shortlisting:

The final round of shortlisting involved the creation of the MCA. We looked at the dates of the DHS surveys and cross referenced them with land cover products available from various locations. We discounted any product over 30 m as this would be too coarse and not provide enough detail for building estimated travel times (the travel times would likely not vary between all DHS clusters). We found two different land cover products for Africa at three time periods:

1. Sentinel-2 CCI land cover classification at 20 m spatial resolution from 2016 with nine land cover classes.
2. Landsat IPCC land cover classification at 30 m spatial resolution from 2010 and 2014 Scheme 1 with six land cover classes
3. Landsat IPCC land cover classification at 30 m spatial resolution from 2010 and 2014 Scheme 2 with 17 land cover classes

[9] DHS (2020) https://dhsprogram.com/What-We-Do/Survey-Types/DHS.cfm

[10] DHS (2020) https://dhsprogram.com/What-We-Do/Survey-Types/MIS.cfm

We considered countries that had a:

1. landcover product and Standard DHS survey from the same year, or;
2. landcover and standard DHS survey with a 12-month gap between them, and;
3. second period where the land cover product and standard survey were either acquired in the same year or within 12 months.

This approach left the following countries:

- Malawi – DHS in 2015 & Sentinel-2 2016 ---- DHS 2010 & Landsat IPCC 2010
- Tanzania – DHS in 2015 & Sentinel-2 2016 ---- DHS 2010 & Landsat IPCC 2010
- Burundi – DHS in 2016 & Sentinel-2 2016 ---- DHS 2010 & Landsat IPCC 2010
- Ethiopia – DHS in 2016 & Sentinel-2 2016 ---- DHS 2011 & Landsat IPCC 2010
- Uganda - DHS in 2016 & Sentinel-2 2016 ---- DHS 2011 & Landsat IPCC 2010

We did a deep dive into the data available for these five countries and developed an MCA.

Table 2 Breakdown of questions available in Standard Survey and MIS survey

	Health		Education		Living Standard						Services	
	Nutrit	Mort.	Yr School	School att.	Cook Fuel	Sanitation	Drink H20	Elect	Floor Type	Asset	H20 Source	Treatment seeking
Std DHS[11,12]	W637; 638; 639	W201; 206; 207; 212	?	?	HH117; 118; 120	HH110: 111; 112; 114; 115; 149	yes	HH132(A)	HH152	HH132 B-G	HH101; 102; 103; 104; 106	W Section 6
MIS[13,14]		W201; 206; 207; 208; 210	No	No	HH108	HH106; 105; 107	HH101; 102; 104	HH114(A)	HH131	HH114 B-F; HH115	HH101; 102; 104	W406; 407; 409

Nutrit = child nutrition; Mort. – child mortality; Yr = years; Att. = attendance; Elect = electricity; Asset = DHS often includes a pre-built asset index with household download. But we may not want to use it depending on the variables used.

Treatment seeking – treatment seeking behaviour of women for their children; MIS = Malaria Indicator Survey; W = Women's survey; HH = Household survey

Std DHS = standard DHS survey. ? = the schooling data was often incomplete as for many countries the children either never attended or didn't attend in the last 12 months. So this wasn't considered from the DHS and MIS surveys.

[11] DHS (2020) DHS Model Household Questionnaire
[12] DHS (2020) DHS Model Women's Questionnaire
[13] DHS (2020) MIS Model Household Questionnaire
[14] DHS (2020) MIS Model Women's Questionnaire

Developing factors for a Multi-Criteria Analysis

To identify a country to work on for this project we followed an MCA[15] approach. This involved developing a simple scoring system which is described in more detail below and in Table 3.

- Clinics Types: all countries in the list had clinics available so we considered if each country had different types of facilities listed which could be used to grade clinics later in the project.
- School completeness: was based on the availability of publicly available school lists, often from the country's Ministry of Education website, and school data through Open Street Map (OSM). We compared the number of school location points we were able to extract from the Open street map 'points of interest' datasets and with how many were reported as being within a country from available reports.
- DHS Year 1 and DHS year 2 indicated if the country had a Standard DHS survey since 2009 and if so, if it had a second one that could be used for change analysis in the future.
- OSM office:
    - OSM (collaborator) had offices located in Tanzania, Uganda and DRC and this could help us to validate the eventual cost surfaces [note this validation wasn't possible due to covid].
- Land cover type number 1 and number 2 were graded 0-3 (points awarded)
    - 0 = not available
    - 1 = MODIS 500 m data product
    - 2 = Landsat 30 m data product
    - 3 = Sentinel-2 20 m data product
- Land cover within survey – indicates how close in time the land cover product was to the Standard DHS surveys and was graded as 0-4:
    - 0 = over 2 years difference
    - 1 = >12 months but less than 24 months
    - 3 = within 12 months
    - 4 = same year (the dates match)
    - We gave a score of 3 to within 12 months to reflect that it is quite important for the two data products to be close in time.
- OSM speeds – indicated if in the OSM road data there were estimated speeds for the major roads. These are often only for a very small number of roads but could be applied to other roads in the country.
- OSM No. rd type – indicated the number of road types listed in the dataset, most had over 20 different types of roads, tracks, paths.

---

[15] Communities and Local Govt (2009) MC: A manual http://eprints.lse.ac.uk/12761/1/Multi-criteria_Analysis.pdf

- OSM barriers to travel – the OSM land cover product contained some land cover types which may indicate barriers to travel: military sites, quarries and cemeteries however they are often patchy.

Electricity grid data – Ethiopia and Uganda have data from the humanitarian data exchange for the electricity grid which was initially thought might be useful when looking at poverty but was not in the end used in subsequent analyses for this project.

Table 3 Factors to consider for MCA with scores the points awarded are indicated in the columns.

	0	1	2	3	4
Clinics types	none available	multiple types listed.			
HRSL	not available	available			
Schools	not available	available			
Sch completeness?	low completeness	unknown		good completeness.	
DHS Yr 1	not available since 2009	available since 2009			
DHS Yr 2	not available since 2009	available since 2009			
Land Cover type # 1	not available	modis 500 m	landsat 30 m	sentinel-2 20 m	
land cover within survey	over 24 months difference	>1 yr <2 years		within 12 months	same year
land cover type # 2	not available	modis 500 m	landsat 30 m	sentinel-2 20 m	
land cover 2 within survey	over 24 months difference	>1 yr <2 years		within 12 months	same year
OSM RD Available	not available	available			
OSM Speeds	not available	available			
OSM No rd types	no road types listed	multiple road types listed			
OSM barrier to travel	not available	available			
electricity grid data	not available	available			
OSM office	not available	available			

MCA method:

- The scores in Table 3 were allocated to each variable for each country in the 31-country long list (even though we had 5 in the shortlist we ran this for the longer list so UNICEF can see which countries could be analysed next)
- The scores were summed together.

MCA results:

The full results are supplied at the end of the document (Appendix 1) and the country with the highest scores were listed in Table 4. The scores were driven by the land cover product availability which was key to being able to estimate accurate travel times. The top three countries were Uganda (1st), Malawi (2nd) and Ethiopia (3rd).

- Uganda was top because it had multiple clinic types available in the Maina et al. (2019) database, appeared to have more accurate school numbers in the Open Street Map (OSM) data when compared to government lists of schools, had 20 m spatial resolution Sentinel-2 land cover data and DHS data from the same year (2016) as well as having Landsat 30 m spatial resolution IPCC land cover data form 2010 and DHS data from 2011 providing an opportunity for looking at a change in access if needed.

- Malawi was second because it had multiple clinic types, some schools corresponded between OSM school locations and government lists although there appeared to be a large number of missing schools in the OSM compared to government data, Sentinel-2 data from 2016 and DHS data from 2015, Landsat IPCC data from 2010 and DHS data from 2010.
- Ethiopia was third because it had multiple clinic types in the database, some schools in the OSM database although appeared to be a large underestimate compared to publicly available list of government schools and it had Sentinel-2 data and DHS from 2016.

Since we were not planning on running time series analysis or change analysis we also examined the MCA results when the criteria around second surveys were removed. In this case, Uganda again came out top, with Ethiopia second, Burundi third and Malawi fourth.

Table 4: MCA scores for each country on the shortlist.

Country	SUM
Uganda	26
Malawi	22
Ethiopia	18
Burundi	17
Tanzania	15
Zambia	13
Rwanda	12
Lesotho	11
South Africa	8
Zimbabwe	8
Benin	8
Angola	7

Conclusions from Phase 1

Uganda was the highest scoring country in the MCA. Uganda was also relatively small and had landcover data and DHS survey data from the same year (2016). So this country was selected for Phase 2 and Phase 3 of the work.

Phase 2: Developing fine-spatial resolution estimates for travel time to key services

The highest resolution global travel time maps have a spatial resolution of 1km which we considered too course for analysis at the DHS cluster or village level. This is because the DHS cluster points are displaced from their original locations for confidentiality reasons and the range in travel time values in a 1km grid cell could be substantial in some regions which would give inaccurate estimates of access and therefore result in unreliable statistical results. We developed an open source and transferable method for estimating travel time at 20 m spatial resolution using a cost allocation method. The method was created using Python and required the following inputs: (1) Land cover data; (2) roads data shapefile; (3) digital elevation model and (4) destination GPS points of schools or health centres. The data that we used were:

1. 20 m spatial resolution Sentinel-2 CCI land cover data from the European Space Agency
2. A combination of Open Street Map (OSM) roads and MapwithAi roads data
3. The SRTM 30 m DEM
4. Maina et al. (2019) health facility GPS locations or the OSM school locations.

The method was subsequently converted into an out-of-the-box python software and released as open source through Zenodo[16] (10.5281/zenodo.4638563). The work has been described in detail in a Nature Scientific Data[17] article that was accepted for publication in April 2022. For the publication we estimated travel time to health facilities in Uganda, Zimbabwe, Tanzania, and Mozambique (Figure 1). Travel times were calculated for different facility types in each country such as Dispensary, Health Centre, Hospital all of the output datasets are available free to download from the Data for Children Edinburgh Data Share (https://datashare.ed.ac.uk/handle/10283/3898). The cost allocation surfaces and travel times are provided for child walking speeds but can be altered easily to account for adult walking speeds and motorised transport.

[16] Hagdorn (2021) Child Povety Access to Services/cpas: initial release https://zenodo.org/record/4638563#.YjxRxZPP2u4

[17] Watmough et al. (in press) Using open-source data to construct 20 metre resolution maps of children's travel time to the nearest health facility, **Scientific Data**

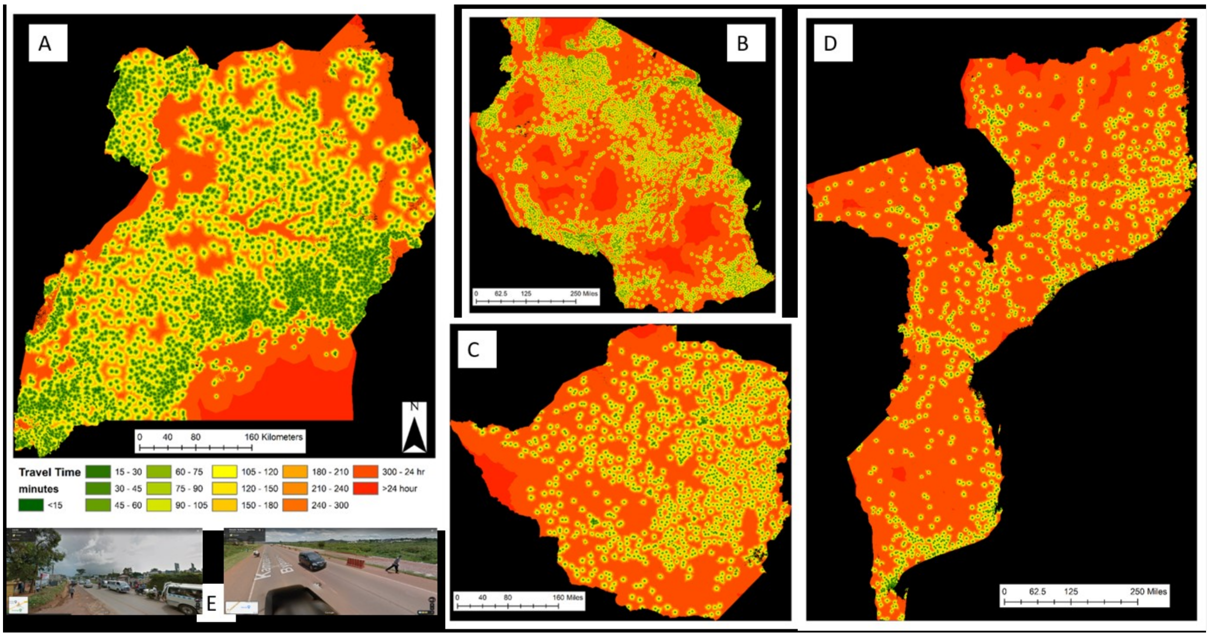

Figure 1 Travel time to any health centre in (A) Uganda, (B) Tanzania, (C) Zimbabwe (D) Mozambique calculated using the CPAS 20-m resolution grids.

Summary of Method

A least cost path method was used to calculate travel time (minutes) from each pixel in a country to the nearest health facility location. The method is comprised of two main phases: (1) creation of a 'cost' allocation surface (also known as an effort or friction surface)[18]. This represents the effort to travel across a particular pixel by considering land cover type and slope angle as well barriers to travel such as open water, and; (2) the cost allocation (effort surface) is used in a least cost path analysis to estimate travel time from every pixel to the nearest destination location (in this case health centres or schools).

Transport mode choice

There was little published information on how patients in Uganda access health services. However, in neighbouring Kenya over 80% of patients were found to access health services by walking[19]. Uganda Transport Policy documentation revealed that non-motorised transport was used for 50% of journeys[20] and the national census of Uganda showed at least 50% of households in rural Uganda did

[18] Etherington, T.R. (2016) *Current Landscape Ecology Reports* https://doi.org/10.1007/s40823-016-0006-9
[19] Noor, A. M. *et al.* (2006) *Tropical Medicine and International Health* doi: 10.1111/j.1365-3156.2005.01555.x
[20] Ministry of Works and Transport (2012) Draft Non-Motorised Transport Policy, Ministry of Works and Transport, Republic of Uganda, https://wedocs.unep.org/bitstream/handle/20.500.11822/25414/uganda_nmtpolicy.pdf?sequence=1&isAllowed=y

not own any form of transport asset[21]. Public transport mostly exists in Uganda through boda bodas (motorcycle taxis) or matutus (shared minibuses). It is likely that when accessing health services further away or when in need of particular medical attention that motorised transport of some form would be used. However, we did not at the time of analysis have access to transport speed information. Thus, we only considered walking/pedestrian travel in this data.

Child walking speed definitions

We used published walking speeds for specific road surface and land cover types taken from studies in Niger[22] and the Horn of Africa[23] which had a maximum walking speed of 5 km/hr (1.39 m/s) that is commonly used to be representative of adult walking speeds. However, we reduced this by 22% as Bouterse and Wall-Scheffler[24] observed that adults travelling with children had an average speed of 0.773 m/s compared with 1.001 m/s when walking without a child.

Extracting travel times to DHS clusters

Travel time surfaces for Uganda were shared with the DHS GIS team who extracted the travel times to health centres to un-displaced Cluster GPS points for Uganda in 2016. We also extracted the travel times to the displaced GPS Cluster points so that we could compare the two. The health facility types available in Uganda[6] were: clinics, health centres (Level II, III and IV), hospitals and regional referral hospitals. The Uganda Hospital and Health Centre Survey Census[25] was used to identify services available at each facility type and we focused on services that would be used by children and women. Level III and Level IV facilities provide maternity care and inpatient care whilst Level II health centres provide only basic medical care. Level IV facilities were the only facility type to have qualified doctors[25]. Hospitals provide access to surgical care and specialist health services. Therefore, in Uganda we calculated travel time to any type of facility, level III health centre, level IV health centres, and level IV health centres and hospitals. Regional and national referral hospitals were excluded, as to access these facilities a patient must be referred from another health facility (Level IV). Clinics were excluded because they are undefined in the census document. Although we did not calculate travel time to Level II facilities these are the most frequent and therefore are covered in the map of travel time to any facility. DHS ran extractions for all four types of health facilities

---

[21] Uganda Bureau of Statistics (2016) The National Population and Housing Census 2014 – Main Report, Kampala, Uganda, https://unstats.un.org/unsd/demographic/sources/census/wphc/Uganda/UGA-2016-05-23.pdf

[22] Blanford, J. I. et al. (2012) *International Journal of Health Geographics* doi: 10.1186/1476-072X-11-24

[23] Pozzi, F. and Robinson, T. (2008) *IGAD LPI Working Paper* https://core.ac.uk/download/pdf/132642458.pdf

[24] Bouterse, L. and Wall-Scheffler, C. (2018) *PeerJ* doi: 10.7717/peerj.5547

[25] Uganda Ministry of Health (2014) Uganda Hospital and Health Centre IV Survey Census, World Health Organization, African Development Bank and Republic of Uganda Ministry of Health https://www.who.int/healthinfo/systems/SARA_H_UGA_Results_2014.pdf

in Uganda as well as the OSM schools. For a full breakdown and description of the methods see Watmough et al. (in press) [17].

Schools analysis

The OSM school data was linked with a database of government schools that was available online. Attempts were made to link the two datasets on the names of the schools so that an assessment could be made of the completeness of the OSM dataset. However, the list from the government was not a complete list of schools as it only contained government registered schools and no private schools were included. The OSM Schools are also known not to be complete. Linking the datasets proved difficult as we were unable to identify the completeness of either dataset. Thus, school locations were taken from the OSM data and used to calculate travel times to the nearest school. However, we are aware that this is not a complete dataset so results for access to schools should be interpreted carefully with this caveat in mind.

Results

The overall output from this phase of the work was a database of travel times to different types of clinics and OSM schools that could be linked to DHS cluster GPS points and be used in Phase 3 of the project. Figure 2 shows the travel time to nearest health centre (any type) and figure 3 shows the travel time to level IV health centres. Travel times to the centres with more facilities are much larger.

Outputs from Phase 2:

- Open source software to estimate travel time
- Friction surfaces and travel time maps for Uganda, Tanzania, Mozambique, Zimbabwe
- Accepted manuscript in Nature Scientific Data describing the 20 m travel time data.
- Invite from UNICEF to use the above outputs on the GenVax project (negotiations ongoing)

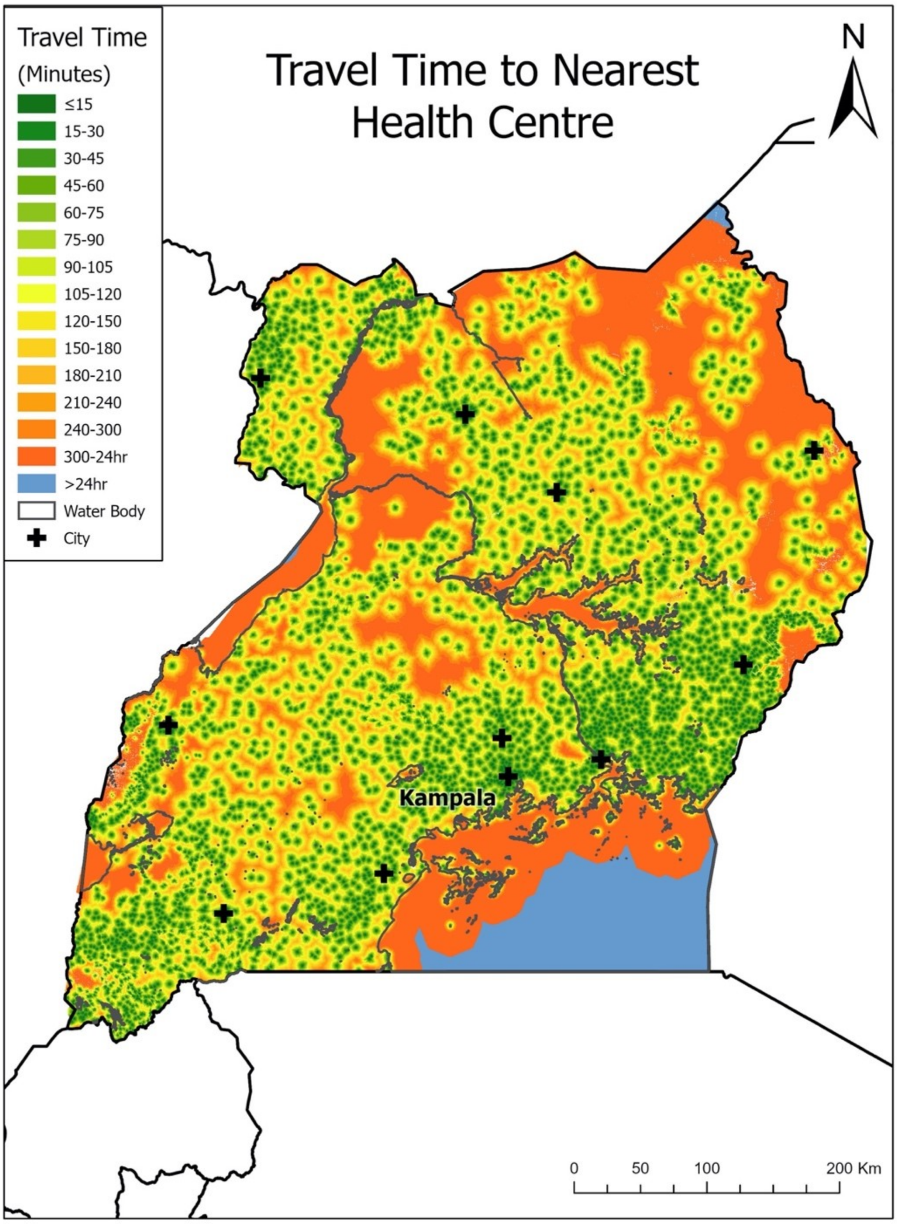

Figure 2 Travel Time (in minutes) to the nearest health facility - which was any type of clinic and centre in Uganda calculated using the CPAS software.

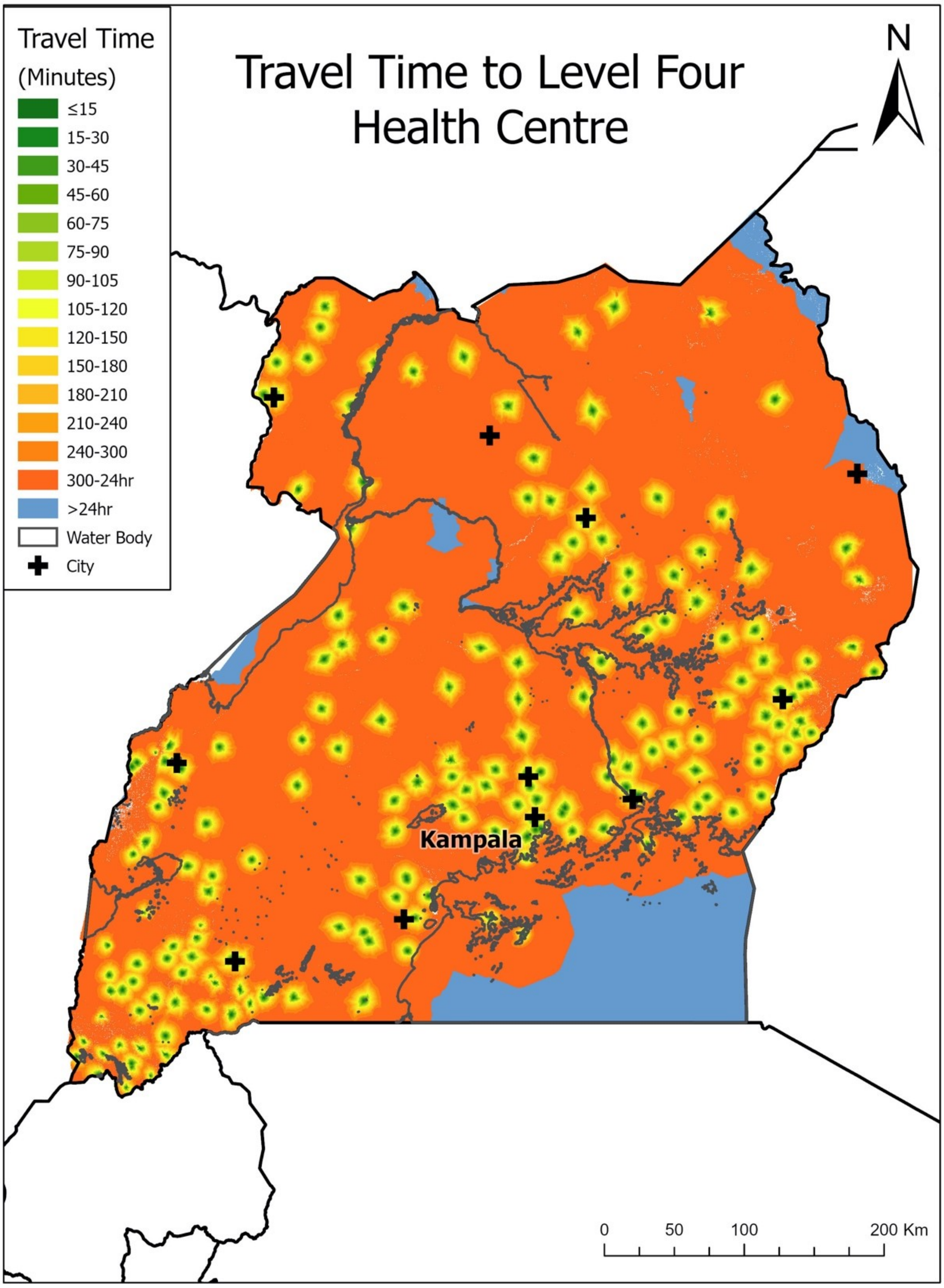

Figure 3 Travel Time (in minutes) to the nearest level IV health facility - calculated using the CPAS software

Phase 3: Examining the statistical evidence for travel time to services as a determinant of child deprivation

We know that access to facilities such as schools and health clinics are important for children. But is physical access (measured in travel time) associated with childhood deprivation? Increased travel times can result in decreased utilisation of health facilities[1, 26]. In the report from the 2016 Uganda DHS standard survey 44% of rural Ugandan women stated that distance to health centre contributed to decision not to access them[27]. Is this an issue for children? There is currently insufficient evidence at the country level to understand the reasons why children are deprived of access to basic services such as healthcare and schooling. Children that do not attend school or do not get the health care they need have much poorer outcomes throughout life. However, access can be determined by physical distances and travel times as well as financial, religious, cultural and educational issues. Therefore, phase 3 of the project used travel time estimations along with the DHS derived multidimensional childhood deprivation index to identify if access based on travel time is associated with deprivation. Furthermore, the socioeconomic and demographic data within the DHS was used in models along with the travel time to health centres and schools to examine the level of these relationships with multidimensional child deprivation.

Methods

The socioeconomic and demographic data used included data on (1) Individual Children [deprivation scores, age, gender, birth order]; (2) the Household [wealth index, mothers age, mothers literacy, household head education, household age, , ownership of livestock, hectares of agricultural land owned], and; (3) the Cluster [travel time to health centres and school, geospatial data on environment and climate]. There were approximately 50,000 children in the DHS data spread across 15,421 households and 603 clusters. We examined the relationships between travel time and deprivation at the individual child level and at the cluster level.

The deprivation index contained six dimensions some of which contained several sub-components and each had different thresholds used to determine severe deprivation and moderate deprivation:

- Housing/Shelter
    - Severe overcrowding = 5+ people per room
    - Moderate overcrowding = 3+ people per room

[26] Gabrysch & Campbell (2009) BMC Pregnancy and Childbirth, 9(34). https://doi.org/10.1186/1471-2393-9-34
[27] Uganda Bureau of Statistics and ICF (2018) Uganda Demographic and Health Survey 2016

- Water
    - Severe water deprivation = children who only have access to surface water
    - Moderate water deprivation = children with no access to improved water source
- Sanitation
    - Severe deprivation = lack of access to toilet facility of any kind
    - Moderate deprivation = lack of access to improved sanitation facility
- Nutrition
    - Severe deprivation = children under 5 who are severely stunted
    - Moderate deprivation = children under 5 who are moderately stunted.
- Health
    - Immunisation
        - Severely deprived = child (12-35 months) has not received a measles, DPT1 DPT2 or DPT3 vaccine
        - Moderately deprived: child has received 1-3 of the above vaccines.
    - Acute respiratory infection (ARI)
        - Severe deprivation = child had ARI symptoms and no treatment was sought.
        - Moderate deprivation = child had ARI symptoms and treatment was not sought at an appropriate medical facility.
    - Access to contraception
        - Severe deprivation = Girls (15-17) who do not want to become pregnant but are not using contraception
        - Moderate deprivation = Girls (15-17) who do not want to becomes pregnant but are reliant on traditional methods of contraception.
- Education
    - 6 to 14 years old
        - Severe deprivation = child has never attended school
        - Moderate deprivation = child attended school at some point but did not attend in the year of the survey
    - 15 to 17 years old
        - Severe deprivation = child is not attending school and did not finish primary school
        - Moderate deprivation = child is not attending school currently, but finished secondary school.

The variable thresholds are country specific and the syntax for Uganda was provided by UNICEF. The index is based on the international rights of the child and takes into account fundamental and material resources. The approach also does not assume poverty, if the information does not exist for that child or in that country survey then the child is considered non-deprived in that component. When the data exists in the survey, each child is recorded as being deprived or not deprived in each dimension. We created a deprivation index to give a count of the number of dimensions each child was deprived in for both severe and moderate thresholds. At the cluster level the deprivation scores were aggregated into categories showing the proportions of children within a cluster that were moderately or severely poor (Figure 4).

Analysis proceeded using both overall moderate deprivation and overall severe deprivation scores as well as only focused on the health dimension.

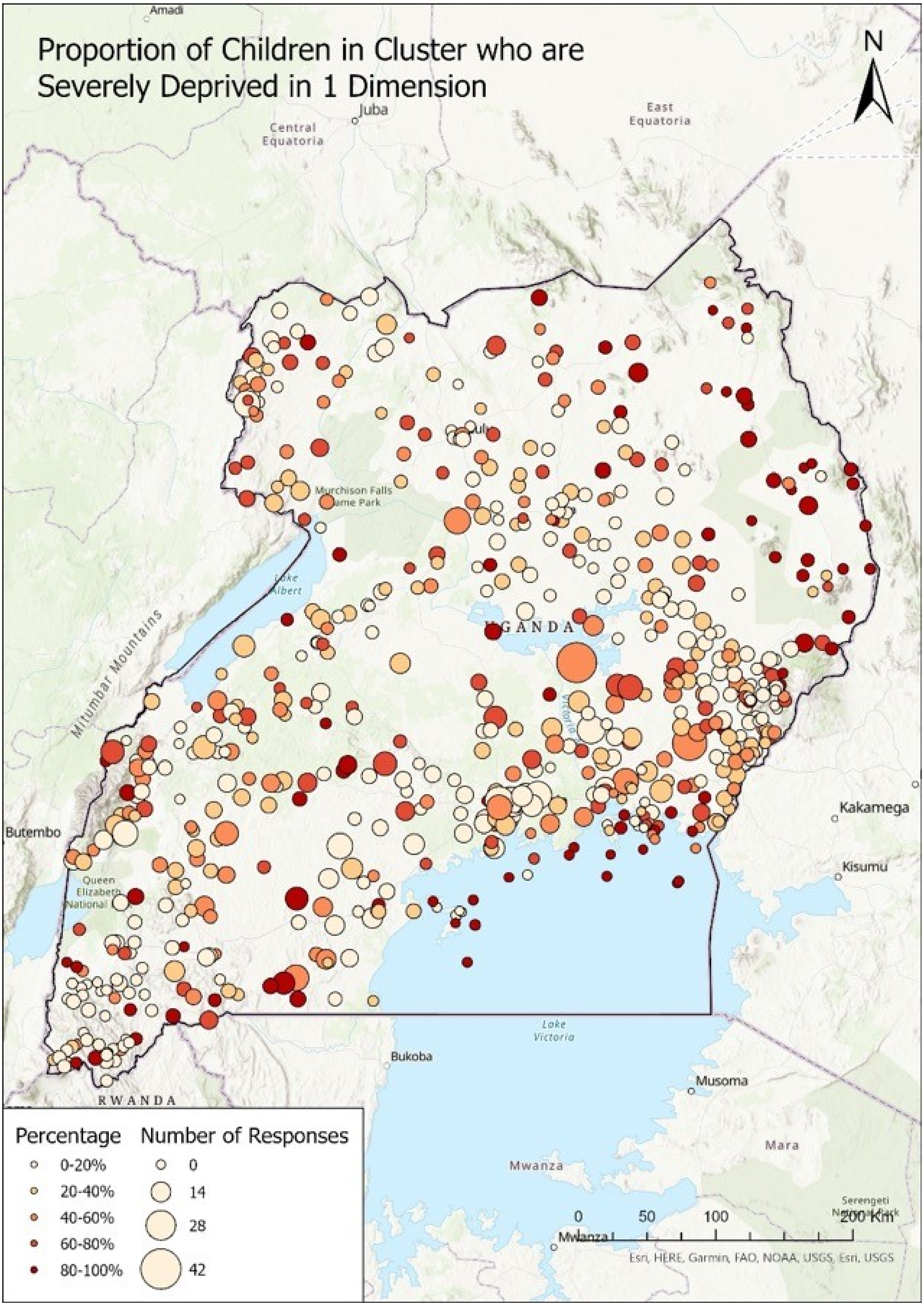

Figure 4 proportion of children in cluster who are severely deprived in at least 1 dimension

Moderately deprived analysis – methods and results:

We calculated the proportion of children in the cluster that were moderately poor in at least two dimensions and split into five categories:

- 0-20% - of children in the cluster were deprived defined here as being deprived in at least 2 dimensions (n=103) - these therefore are relatively speaking the wealthiest clusters. 20-40% - of children in the cluster were deprived (n=78)
- 40-60% - of children in the cluster were deprived (n=160)
- 60-80% - of children in the cluster were deprived (n=144)
- 80-100% - of children in the cluster were deprived (n=118), these are therefore relatively speaking the poorest clusters.

Spearman rank correlations were run on the cluster level deprivation and the following travel time estimations:

- Travel time to school
- Travel time to any health facility
- Travel time to Level III health facilities
- Travel time to Level IV health facilities.

Spearman Correlations showed that the travel time to schools were all insignificant ($p>0.05$) and so were dropped from subsequent analyses. The travel time to any health facility had the lowest rho for both moderate and severe deprivation and was dropped from subsequent analysis. This was somewhat expected given the large number of basic health facilities across Uganda meaning travel times were much lower to any facility than to other types of facility and the uncertainty in the school location data that we were able to collate. The Level III and Level IV health facility travel times had significant correlations and relatively large rho values (ranging from 0.31 to 0.51). Table 5 indicates the importance of using non-displaced values when analysing DHS data. It is often the case in the literature that displaced GPS points/data are spatially linked to geospatial data such as travel times. Our results indicate that doing so in this context would give increased correlation coefficients (rho) and could indicate stronger bivariate relationships between travel time and deprivation than is actually the case.

Table 5: Comparing the spearman correlation results from bivariate analysis of travel time to level III and Level IV health facilities and moderate and severe deprivation at the cluster level. Includes both displaced and non-displaced travel time estimates.

Deprivation type	Travel time types	Displaced or Non-displaced	S stat	P value	Rho (correlation)
moderate	Level IV	Non	30884774	$<2.2e^{-16}$	0.416
moderate	Level IV	Displaced	25840481	$<2.2e^{-16}$	0.511
moderate	Level III	Non	27373297	$<2.2e^{-16}$	0.485
Moderate	Level III	Displaced	21549175	$<2.2e^{-16}$	0.592
Severe	Level IV	Non	32731287	$<2.2e^{-16}$	0.384
Severe	Level IV	Displaced	32243825	$<2.2e^{-16}$	0.390
Severe	Level III	Non	36405102	$<2.2e^{-16}$	0.314

Analysing bivariate relationships between travel time and overall deprivation at the cluster level

There appeared to be a trend in the data (Figure 5). Travel times to level IV health facilities and level III health facilities were higher for clusters with higher proportions of children classed as moderately poor in at least two dimensions. . The boxplots also allow for comparison between displaced and non-displaced travel time estimations. The travel time estimations have similar patterns for Level IV health facilities when using the displaced or non-displaced GPS locations. However, there are some considerable differences between the travel times estimation using displaced and non-displaced GPS data for Level III health facilities. For example, the range of travel time values appears to be much larger for the poorest communities (>80%) in the displaced data compared with the non-displaced data. Kruskal-Wallis chi-square tests indicated that there were significant differences in each of the travel time datasets. However, Wilcox pairwise comparisons indicated that there were more complex patterns within the data (Table 6) as some pairwise comparisons had significant differences whilst others did not.

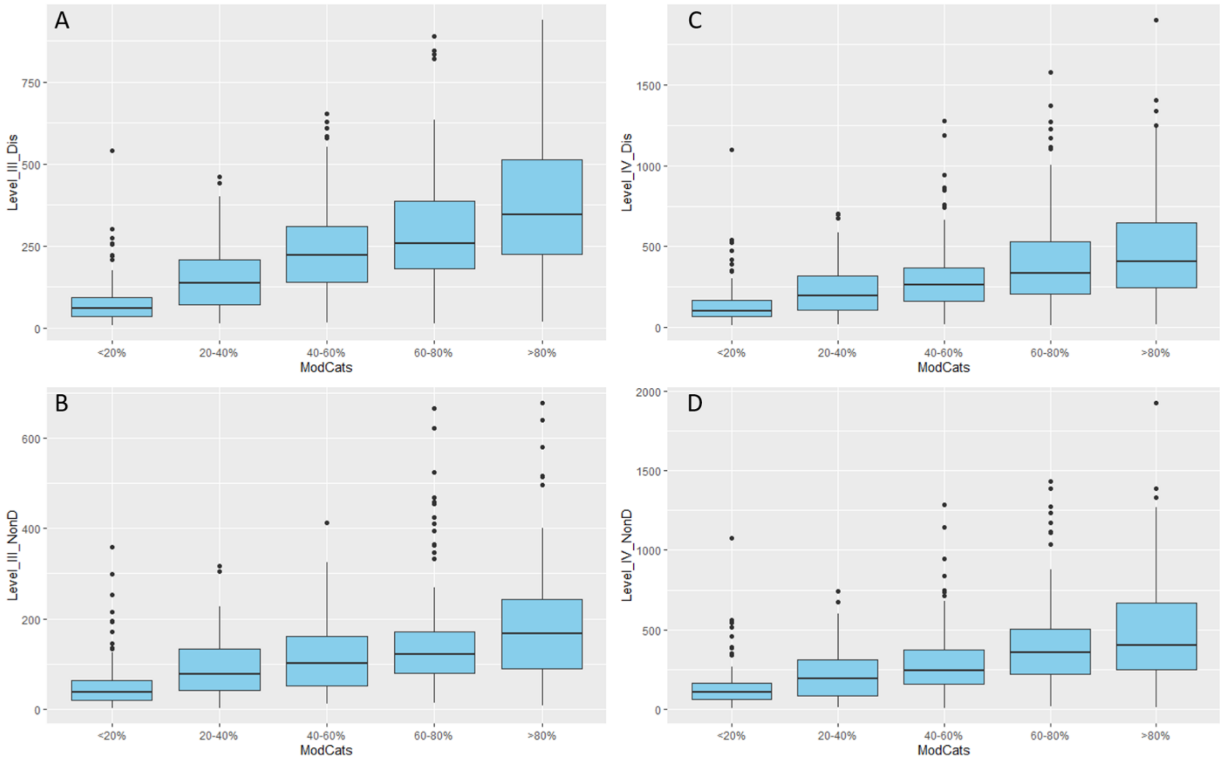

Figure 5: Boxplots of % of children in a cluster that are moderately poor vs travel time to (A) Level III health centres calculated using displaced GPS locations, (B) Level III health centres using non-displaced GPS locations, (C) Level IV health centres using displaced GPS locations and (D) level IV health centres using non-displaced GPS locations. The poorest clusters here are those with >80% of the children recorded as being moderately deprived.

Table 6 Wilcoxon pairwise comparisons of travel time and poverty groups using non-displaced GPS Cluster locations at the cluster level for moderately poor index

	<20%	20-40	40-60	60-80
Level IV				
20-40	0.001	-	-	-
40-60	1.9e-13	0.054	-	-
60-80	<2.2e16	1.8e-07	0.0004	-
>80	<2.2e-16	1.3e-09	3.9e-07	0.758
Level III				
20-40	8.6e-05	-	-	-
40-60	1.1e-11	0.196	-	-
60-80	<2.2e-16	0.0002	0.178	-
>80	<2.2e-16	2.3e-08	6.4e-06	0.05

Density plots were used to visualise how the travel times differed across the different deprivation categories/groups (Figure 6). The DHS clusters with the lowest percentage of moderately deprived children had the lowest travel times to level III and level IV health facilities (blue line in Figure 6). The blue line also indicates a smaller range of travel times indicating that the majority of clusters in this category have similar travel times to level III and Level IV facilities. There also appears to be little difference between the wealthiest DHS clusters between displaced and non-displaced estimations of travel time. The clusters with higher levels of poverty (Orange line in Figure 6) had larger travel times for both level III and Level IV health facilities. There also appeared to be some differences between displaced and non-displaced travel times especially for Level III health facilities. Using displaced DHS cluster locations appear to overestimate the travel times to level III facilities. Overall, Figure 6 indicates that clusters with >80% of children classed as moderately poor have larger travel times on average compared to all other clusters for both Level III and Level IV health facility types. There are also signs that a larger number of clusters have exceptional situations (vertical coloured lines on the plot are mostly orange, so for example, only orange lines appear above 500 minutes indicating that only the poorest clusters have 500 minutes or more to travel to a health facility with child facilities) when considering more specialised health services such as Level III and Level IV. These plots indicate the difficulties a regression/classification tree approach would have if only travel times are available for prediction as there were no clear patterns between most of the five classes of poverty.

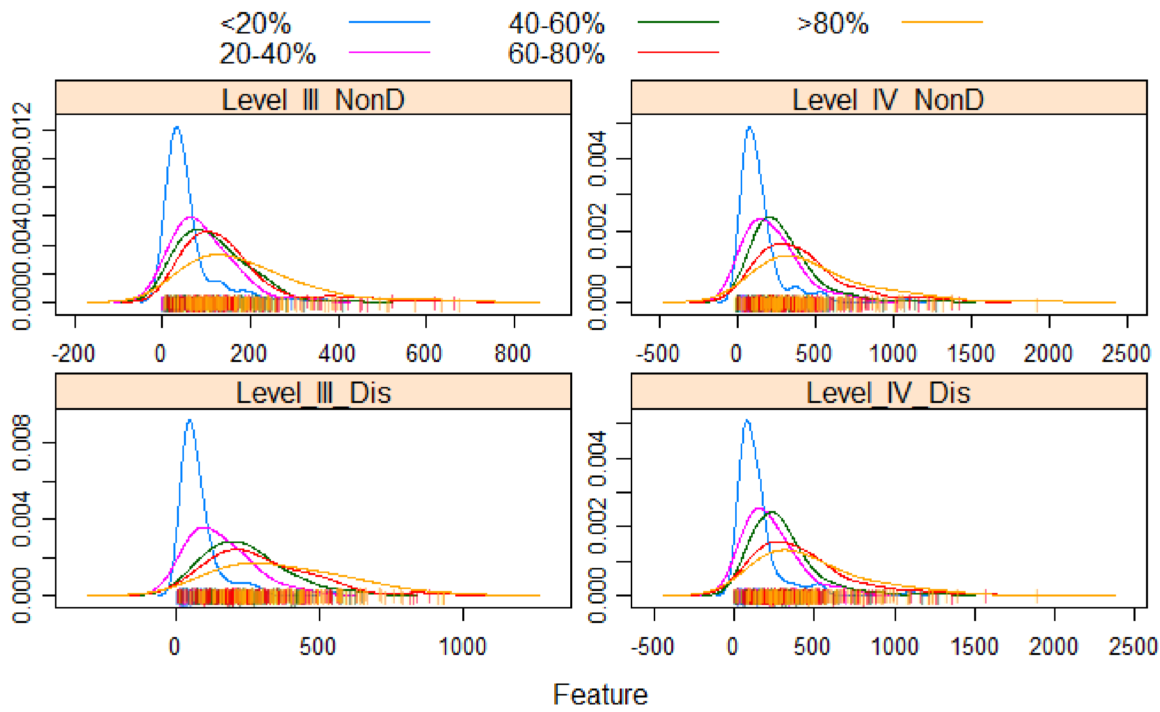

Figure 6 density plots showing how travel time to level III health facilities, level IV facilities, schools and any health facility differ for each poverty class/group

Severely deprived analysis method and results:

The same analysis was conducted for the severely deprived index, by taking the proportion of children in the cluster that were severely deprived in at least 1 dimension. Due to the limited number of clusters at the most severely deprived end of the spectrum, he data were split into 4 groups/categories instead of 5:

1. 0-25% of children were severely deprived (n=244)
2. 25-50% of children were severely deprived (n=236)
3. 50-75% of children were severely deprived (n=73)
4. >75% of children were severely deprived (n=50)

There appeared to be a trend in the data (Figure 7) as travel times to level IV health facilities and level III health facilities were higher for clusters with larger proportions of children severely deprived. There were some marked differences between displaced and non-displaced GPS cluster location. For example, the range of travel times for the poorest clusters (>75%) were 100 – 300 minutes in the level III non-displaced analysis but this increased to 200 – 600 minutes in the displaced analysis which indicates a doubling of the interquartile range between displaced and non-displaced travel time calculations. Kruskal-Wallis chi-square tests indicated that there were significant differences in each of the travel time datasets. However, Wilcox pairwise comparisons indicated that there were more complex patterns within the data (Table not shown but is similar to table 5).

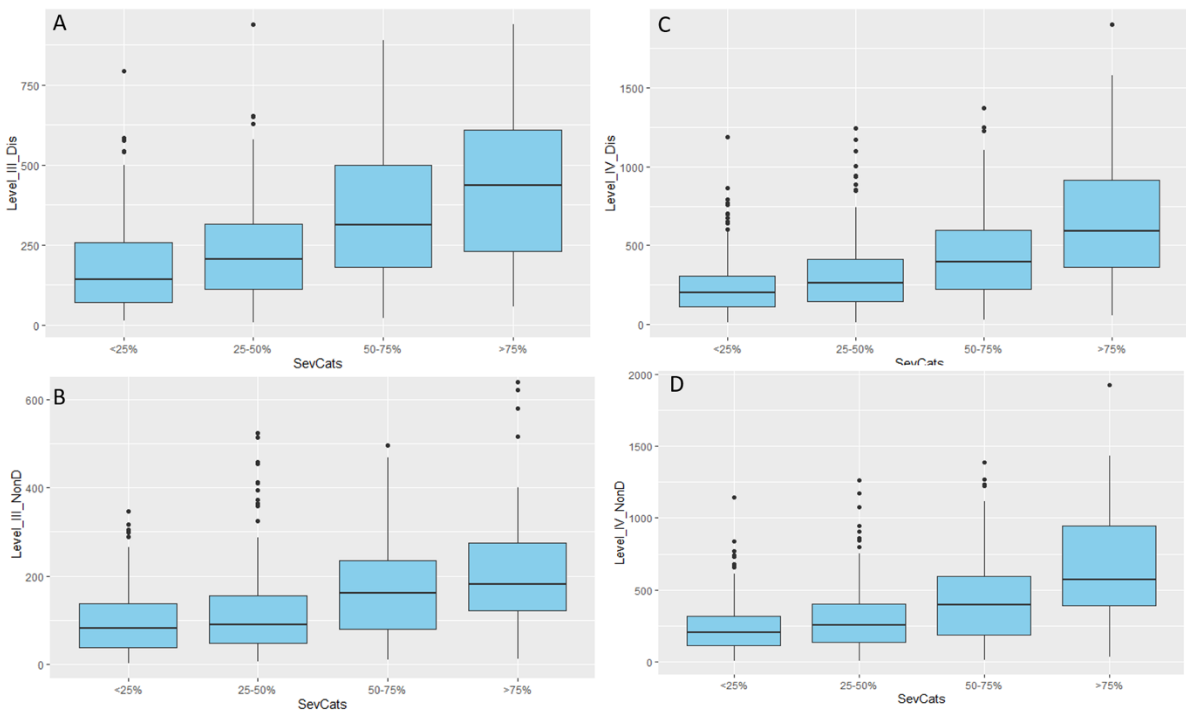

Figure 7 Boxplots of % of children in a cluster that are severely poor vs travel time to (A) Level III health centres calculated using displaced GPS locations, (B) Level III health centres using non-displaced GPS locations, (C) Level IV health centres using displaced GPS locations and (D) level IV health centres using non-displaced GPS locations. The poorest clusters here are those with >80% of the children recorded as being moderately deprived.

Density plots (Figure 8) indicate that the wealthiest two categories (<25% and 25-50%) of clusters have lower travel times to both level III and Level IV health facilities as compared to the poorest two categories (50-75% and >75%). There is a shift in the travel times to Level III health facilities between displaced and non-displaced GPS cluster locations. But this shift is less clear in the Level IV health facilities which is likely because the displacement of cluster locations does not allow for clusters to be moved outside a district and Level IV facilities are often district level hospitals so the differences in travel times are likely to be less affected than Level III facilities which are more common and therefore more likely to be affected by displacement of GPS points. In the density plots for severely deprived clusters it could be suggested that a two category split of the data would work since the poorest and wealth two categories follow similar patterns but this would result in one category (<50%) having over 480 samples out of a total population of 603 clusters creating a unequal sample and subsequent problems for statistical analysis.

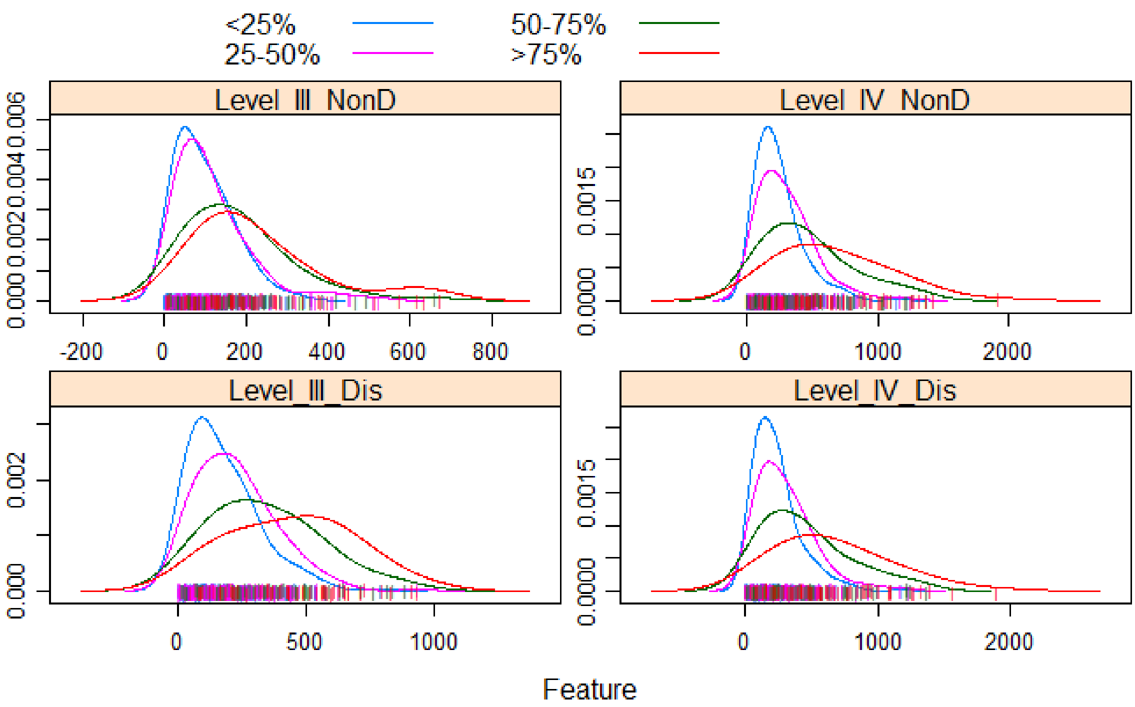

Figure 8 Density plots of travel times and severe poverty groups.

Summary of bivariate relationships

The first question we were interested in asking was: "is geographic access or physical travel time to key services related to multidimensional child poverty?" Using bivariate boxplots and examining data distributions across different deprivation classes the answer appeared to be that travel time is correlated with multidimensional childhood deprivation. Specifically for some health centre types with more specialised services higher levels of child deprivation were associated with longer travel times. A second question was to consider how displacement of GPS Cluster points impacts analysis of this type. It is common in the literature to see displaced GPS points being used to link geospatial data such as travel time grids to the household/cluster data of the DHS. We can see from this analysis that this can result in large overestimates of travel time to Level III health facilities with interquartile ranges often doubling between displaced and non-displaced GPS locations. Overestimating travel time is effectively underestimating physical access to health centres as the travel times, in reality, will not be as high as they appear when using displaced GPS locations.

Analysing the relationships that travel time has with deprivation over and above other socioeconomic parameters that are commonly linked with deprivation and poverty

The next research question was; "how much does geographic access/physical travel time contribute to child deprivation". We examined the relationship between geospatial data that have been associated with poverty/deprivation in the past[28,29]. For example:

- Growing season length – a longer season often associated with wealthier communities as more productive/higher yields can be seen in agricultural regions
- Slope – steeper slopes often associated with poorer communities as the soil is less deep and has fewer nutrients affecting yield.
- Wet Days – complex relationships, too little or too much rain can negatively impact agriculture
- Proximity to protected land – often poorer communities are closer to protected land as it is further from cities. But protected land also often prevents other uses so if communities are closer to protected land there may be limits on agricultural expansion and land shortages
- Proximity to water – often areas closer to water have opportunities for irrigation to extend cropping seasons, add additional harvest periods, and provide additional livelihoods such as fisheries, so communities with closer access to water may be wealthier

The density plots below broken down by moderate deprivation level (Figure 9) show some expected patterns. Wealthiest communities in Uganda (blue line in Figure 9) have wetter days, longer growing seasons lower maximum temperatures, are further from protected areas, are closer to water and are located on flatter ground. Poorest communities (orange line in figure 9) have a wide range of wet days but tend to be drier, have shorter growing seasons, higher maximum temperatures, are closer to protected areas and are further from water sources. These patterns in the density plots indicate that a regression model may be able to use some of these geospatial datasets to be able to differentiate/predict different levels of moderate poverty in Uganda.

[28] Watmough et al. (2013) J. LandUse Sci 8(3) https://doi.org/10.1080/1747423X.2012.667447
[29] Watmough et al. (2016) World Dev. 78; https://doi.org/10.1016/j.worlddev.2015.10.031

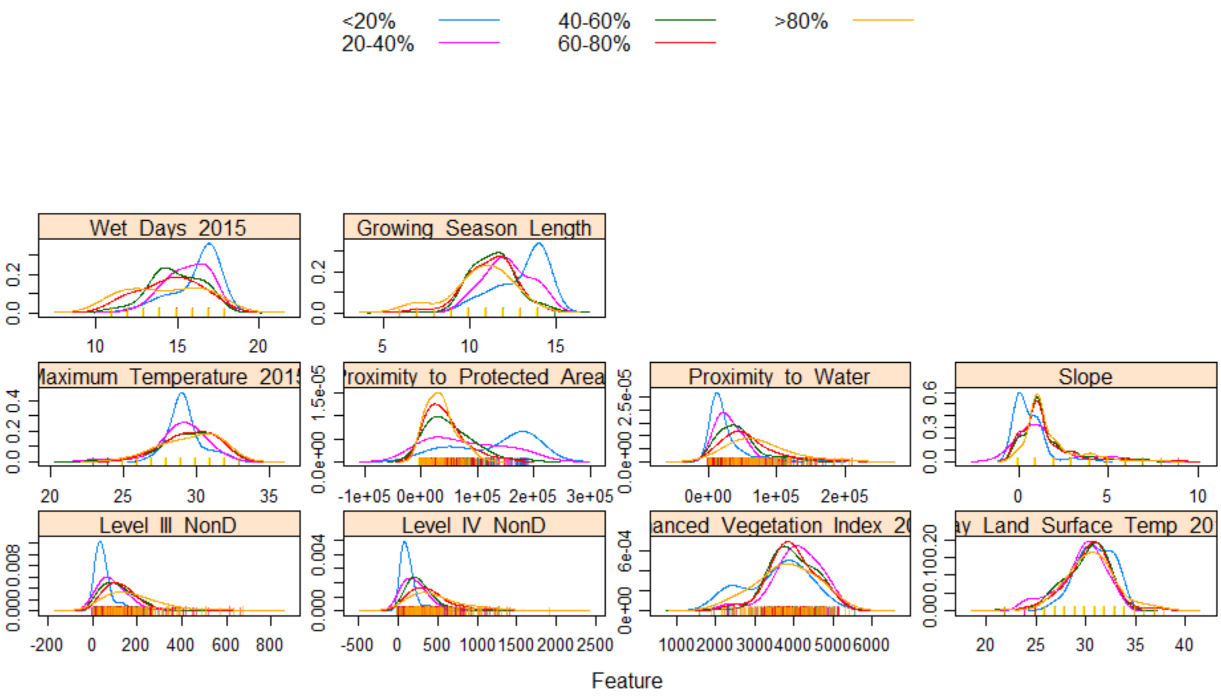

Figure 9 Density plots for moderate deprivation broken down by class level deprivation categories at the cluster level

We used a random forest approach with classification trees to explore how travel time to health centres were related to moderate and severe deprivation when considering other geospatial data. A null model was run at the cluster level that used the two measures of access (travel time to level IV, level III). Overall accuracy when using the percent of children within a cluster classed as severely deprived was 50% and 44% when using the percent of children within a cluster that were moderately deprived. Geospatial data and travel time to health centre were used for these models as aggregating the demographics and socioeconomic data from household level to cluster did not make sense here when a multi-level model would be used in future. These accuracies are arguably too low to allow for reliable interpretation of the results from the models. When looking at the confusion matrix the model did a reasonable job of predicting the least poor clusters but a poor job for the other categories. The confusion matrix does though indicate where some of the problems or class confusion was coming from. For example, there was confusion between <25% and 25-50% groups, confusion between 50-75% and 25-50% group (Table 6). Therefore, the conclusion was that the way in which we split the poverty data up into four categories was not appropriate as the pattern of travel times vs poverty in the boxplots was not captured.

Table 6 confusion matrix for the random forest model used to predict poverty using travel time only

OOB	<25%	25-50	50-75	75%	Class error
<25%	125	113	6	0	49%
25-50	123	95	10	8	60%
50-75	22	34	7	10	90%
>75	6	22	11	11	78%

The class level errors were lowest in the wealthiest category for the moderate deprivation (red line <20% in Figure 10) and wealthiest category for the severe deprivation (red line <25% in Figure 11).

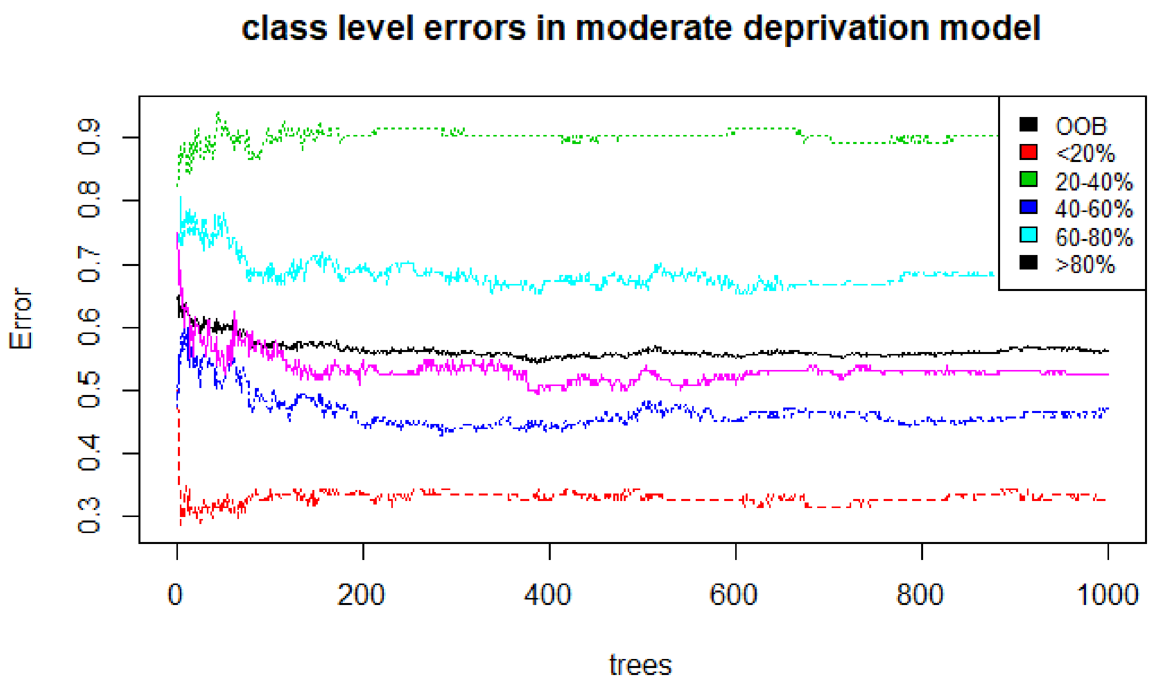

Figure 10 Class level out of bag errors for the moderate deprivation Random forest model

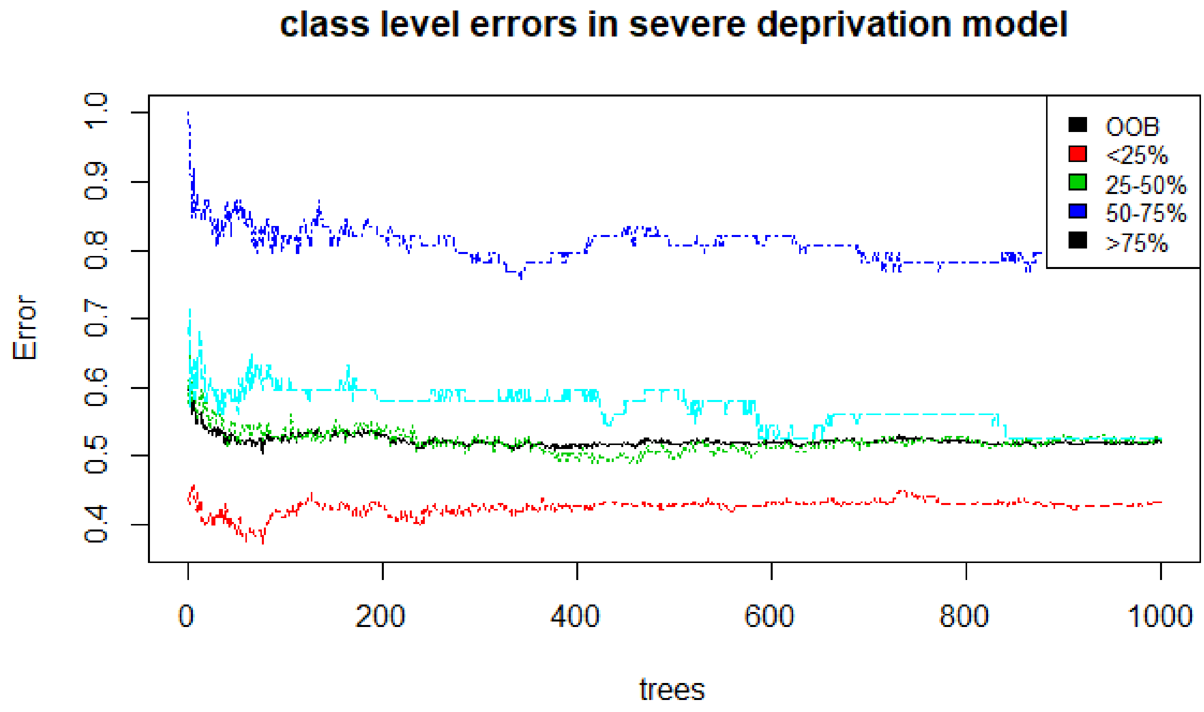

Figure 11 Class level out of bag errors for the severe deprivation random forest model

Variable importance plots (Figure 12) show which variables had the largest impact on prediction for each of the five deprivation categories in the moderate deprivation model. For the clusters with over 80% of children classed as moderately deprived the most important variable was the travel time to Level III health facilities, followed by proximity to protected areas, proximity to water and travel time to level IV health facilities. For the wealthiest clusters the most important variables were proximity to protected areas, travel time to level III health facilities, travel time to level IV health facilities and proximity to water. Indicating that although the overall accuracy was around 50% and the class level errors were high the travel time to health facilities calculated by the CPAS software were contributing significantly to the prediction.

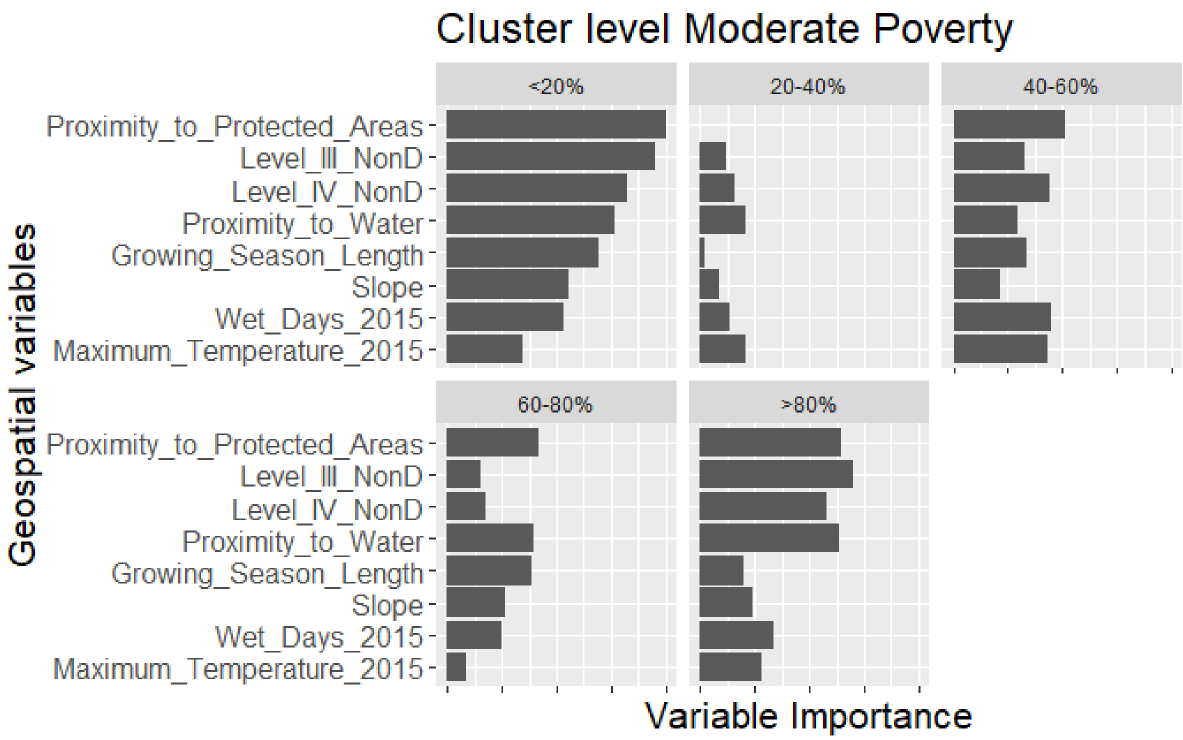

Figure 12 Variable Importance plot for the moderate deprivation cluster level model.

Variable importance plots (Figure 13) show which variables had the largest impact on prediction for each of the five deprivation categories in the severe deprivation model. For the clusters with over 75% of children classed as severely deprived the most important variable was growing season length, followed by the number of wet days, proximity to protected areas and the travel time to level IV health facilities. For the wealthiest clusters (<25%) the most important variables were wet days, maximum temperature, proximity to water and travel time to level IV health facilities. Indicating that although the overall accuracy was around 50% and the class level errors were high the travel time to health facilities calculated by the CPAS software were contributing significantly to the prediction.

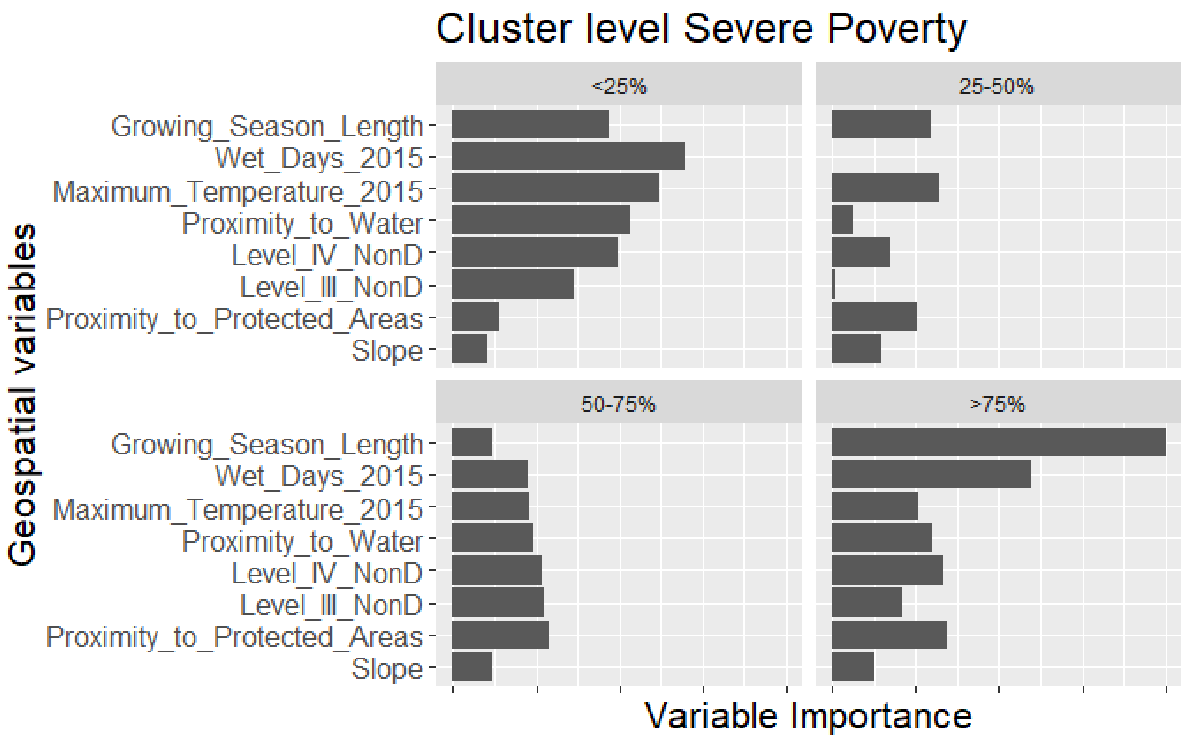

Figure 13 Variable importance plots for the severe deprivation cluster level model

Partial dependence plots can be used to examine how the probability of being within a class varies with the variable values. Figure 14 indicates that clusters classed as the poorest (Group 5) had shorter growing seasons, were closer to protected areas, were further from level IV health facilities and further from level III health facilities. For the severe deprivation model (Figure 15) partial dependence plots indicated that clusters with shorter growing seasons, fewer wet days and longer travel times to level IV health centres were more likely to be in group 4 (poorest).

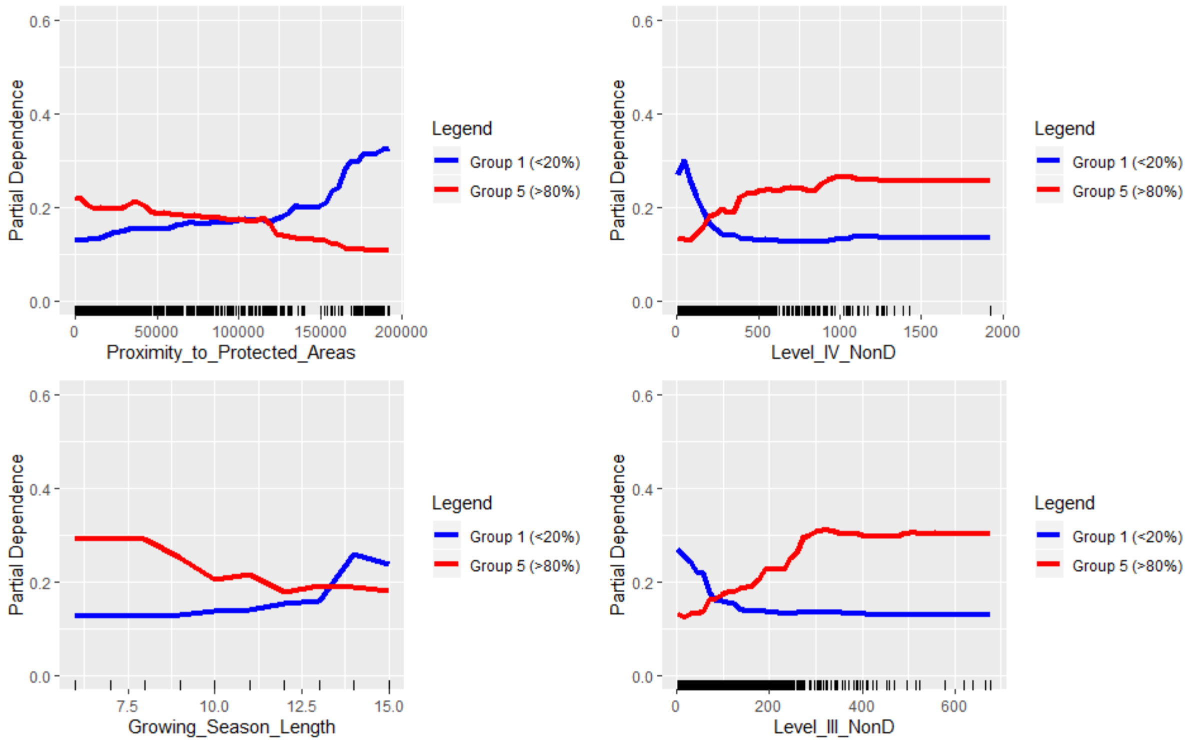

Figure 14 Partial Dependence Plots for the wealthiest (Group 1) and poorest (Group 5) clusters from the moderate deprivation models. Groups 2, 3, 4 were dropped due to high error levels.

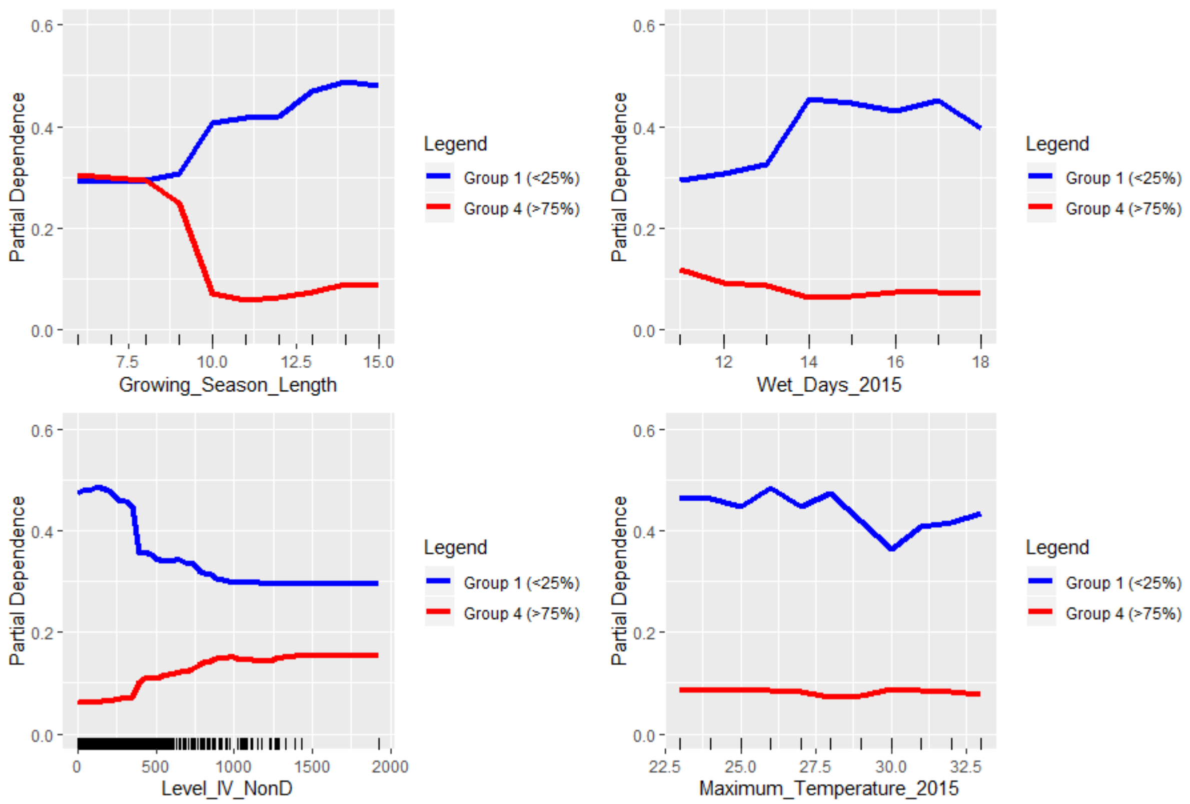

Figure 15 Partial dependence plots for the wealth and poorest clusters from the severe deprivation models. Groups 2 and 3 were dropped due to higher error levels.

Summary

None of the random forest models had good overall accuracies and therefore the interpretation was difficult and unreliable. However, using the bivariate relationships in the Spearman rank correlations, boxplots and random forest models there are clear patterns in the data. Specifically, that poorest clusters in Uganda were associated with longer travel times to both level III and level IV health facilities. The lower accuracies from the random forest models are likely down to the use of geospatial data only. Thus, we need to consider other options for the analysis. The reason for not using Household level analysis in the random forest analysis was that the travel time did not vary at the household level and since this is what we were trying to understand and say something about then working at the household level was less helpful since multiple households in a cluster will have the same travel time but different poverty scores. However, a multi-level model may help here as it would allow us to aggregate child poverty scores to the household and then consider other cluster level data. Blanford et al. (2012) in a study looking at hospitals and health clinics in Mali used a multi-level model and DHS data. Level 1 was the household data on things that determine treatment seeking behaviour and level 2 was the cluster level travel time. They ran a model with binary logistic regression (response variable was if a child had been fully vaccinated by 12 months or not). There is also literature suggesting that travel time can be categorised to above or below 1 hour as this is the threshold that appears to determine if treatment is sought or not. Overall, therefore, the analysis shows that:

1. Using displaced GPS points to calculate travel time and link to deprivation can lead to overestimates of travel time to level III health facilities.
2. Travel times to Level III and Level IV health facilities are longer for DHS clusters with higher levels of severe and moderate deprivation.

Analysing bivariate relationships between health deprivation and travel time to health facilities

As well as overall deprivation (above), we also ran bivariate statistical analyses between travel time estimations to different types of health facilities and the proportion of children deprived in the health dimension of the multidimensional deprivation index. We repeated the analysis for moderately deprived and severely deprived. Overall, there were 50472 children in the Uganda dataset. Of these 1802 children were classed as moderately poor in the health dimension and 404 were classed as severely poor.

We aggregated the deprivation scores to the cluster level, calculating the proportion of children within each cluster that were deprived in the health dimension and then compared these proportions with travel times to the following types of health centre: (1) level IV; (2) level III and; (3) any health facility. Spearman rank correlations (Table 8) were calculated along with boxplots. Only two correlation models were significant (moderate deprivation vs travel time to level III health facilities and severe deprivation vs travel time to level III health facilities) and the rho indicated a weak but positive relationship between travel time and health deprivation which suggests that clusters with higher proportions of deprivation in the health dimension having higher travel times to level III health facilities (Figure 16 and Figure 17).

Table 8 Spearman rank correlation results for models comparing moderate and severe deprivation in health dimension vs travel times to health facility

Model	S stat	P value	Rho (correlation)
Mod vs level III	46228090	0.0006*	0.13
Mod vs level IV	49967712	0.12	0.06
Mod vs any facility	49010039	0.99	0.00006
Sev vs level III	47568906	0.0064*	0.104
Sev vs Level IV	49127119	0.05	0.07
Sev vs any facility	48756769	0.89	0.005

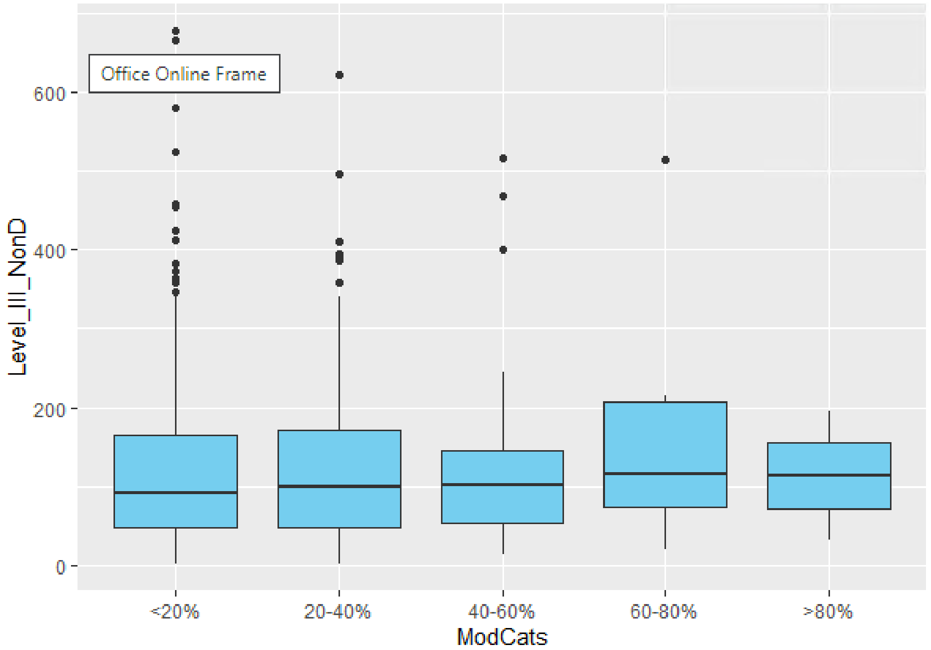

Figure 16 boxplot showing the relationship between moderate deprivation and travel time to level III health facilities

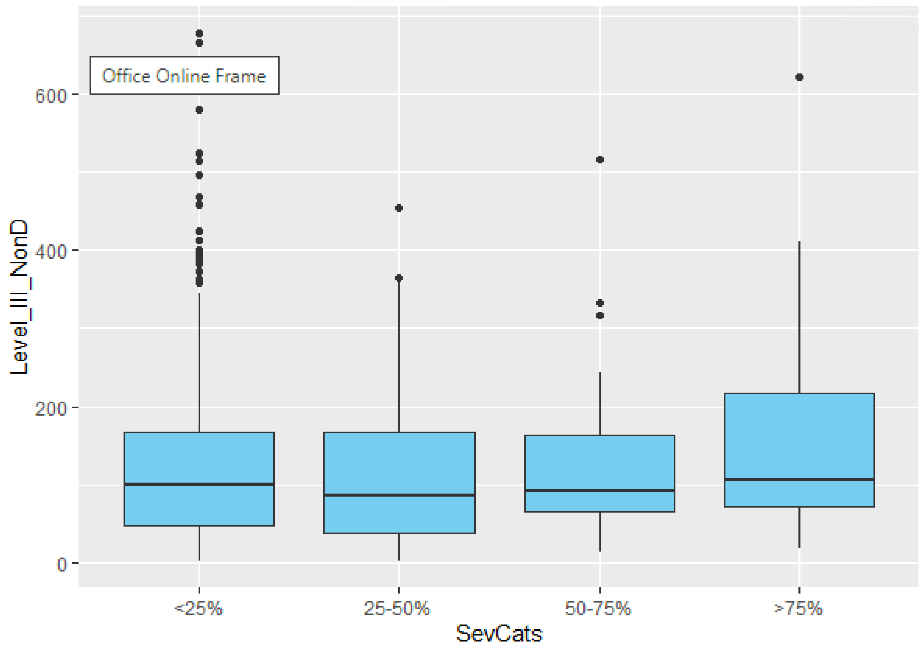

Figure 17 boxplot showing the relationship between severe deprivation in the health dimension and travel time to level III health facilities

The boxplots and the correlation results indicate that there is a small but potentially significant relationship between travel time and deprivation in the health dimension. However, further analyses are required to confirm this. Given the very small number of children that are moderately deprived (1802/50472) and severely deprived (404/50472) in the health dimension a multi-level model will be employed in the future to explore these relationships.

Summary

The analysis using health deprivation measures was inconclusive. Using the health deprivation was problematic for correlation analyses since there were very small proportions of individual children that were moderately deprived and severely deprived. A multi-level model will help with analyses where the sample is so unbalanced and this will form the next steps of analyses.

Appendix A

Table 5 Country MCA results - full

Country	Clinics types	HRSL	Schools	Sch completeness?	DHS Yr 1	DHS Yr 2	Land Cover type # 1	land cover within survey	land cover type # 2	land cover 2 within surve	OSM RD Available	OSM Speeds	OSM No rd types	OSM barrier to travel	electricity gr	OSM office	SUM
Uganda	1	1	1	3	1	1	3	4	2	3	1	1	1	1	1	1	26
Malawi	1	1	1	1	1	1	3	3	2	4	1	1	1	1	0	0	22
Ethiopia	1	1	1	1	1	1	3	4			1	1	1	1	1	0	18
Burundi	1	1	1	1	1	1	3	4			1	1	1	1	0	0	17
Tanzania					1	1	3	3	2	4						1	15
Zambia					1	1	3	3	2	3						0	13
Rwanda					1	1	3	1	2	4						0	12
Lesotho					1	1	3	1	2	3						0	11
South Africa					1	0	3	4								0	8
Zimbabwe					1	1	3	3								0	8
Benin					1	1	3	3								0	8
Angola					1	0	3	3								0	7
Guinea					1	1	3	1								0	6
Mali					1	1	3	1								0	6
Nigeria					1	1	3	1								0	6
Namibia					1	0	3	0	2	0						0	6
Kenya					1	0	3	1								0	5
Ghana					1	0	3	1								0	5
Chad					1	0	3	1								0	5
Liberia					1	0	3	0								0	4
Sierra Leone					1	0	3	0								0	4
Burkina Faso					1	0	3	0								0	4
Togo					1	0	3	0								0	4
Mozambique					1	0	3	0								0	4
Senegal					1	0	3	0								0	4
Cameroon					1	0	3	0								0	4
Cote d'Ivoire					1	0	3	0								0	4
Gabon					1	0	3	0								0	4
DRC					1	0	3	0								1	5
Madagascar					0	0	3	0								0	3
Comoros					1	0	0	0								0	1

Data for Children Collaborative WITH UNICEF

Child Poverty and Access to Services Analysis Summary

Combining household surveys and geospatial data to examine if physical access to basic services is a determinant of multi-dimensional childhood poverty?

THE UNIVERSITY of EDINBURGH

unicef for every child

CONTACT US:

@dataforchildren

Data for Children Collaborative

Hello@dataforchildren.ed.ac.uk

Supported by

THE UNIVERSITY of EDINBURGH | Data-Driven Innovation

THE UNIVERSITY of EDINBURGH

unicef for every child

The Scottish Government