SPATIAL DEPENDENCE AND SPATIAL NON-STATIONARITY IN THE FACTORS EXPLAINING UTILISATION OF MATERNAL HEALTH CARE IN KENYA

Susan Wambui Kamundia

University of Cape Town

ABSTRACT

Standard regression models assume that the measured attributes in one location are not related to measured attributes in another location. However, this does not hold in some cases. The distribution of individual, household and health facility covariates is also bound to vary according to the geographical location. As such, the effect of these variables on an outcome is also bound to differ depending on the location. This study aims at exploring whether spatial dependence exists in the utilization of deliveries by a skilled provider and if so, utilize a spatial model for analysis of the factors explaining the utilization when spatial dependence is controlled for. Spatial dependence is shown to exist in the utilisation of maternal health care and geographically weighted regression (GWR) models are used to analyze the factors explaining the utilization of maternal health care when spatial dependence is taken into account. GWR models are also used to explore the possibility of non-stationarity existing in the factors that explain the utilization of maternal health care. The largest positive effects on the utilization of maternal health care are in clusters with low maternal education levels, younger mothers and a lower alternative supply of health facilities. The largest reductions in utilization of maternal health care are in areas with mothers who have more children and in rural areas with a lower density of health facilities. Mothers are also more likely to utilize higher-level health facilities in areas with a higher density of higher-level health facilities. The study recommends that the presence of spatial dependence and spatial non-stationarity be explored before deciding on whether to utilize global models for analysis.

Table of Contents

List of Tables

List of Figures

1. Background

Spatial dependencies exist when measured attributes of entities in one area are dependent on the attributes of surrounding areas. This arises due to two phenomena. The first is spatial heterogeneity which implies that the measured attributes in one area are most likely different from those in other areas across a study space. Secondly, neighbouring areas tend to have very similar measured attributes. This is referred to as spatial clustering. If areas with a high measured attribute are close to other areas with high measured attributes, this is positive spatial autocorrelation. The same applies to areas with low measured attributes surrounded by low measured attributes. If an area with a high measured attribute is close to an area with a low measured attribute and vice versa, this is referred to as negative autocorrelation. Ignoring spatial dependencies, where they exist, leads to a wrong estimation of standard errors and incorrect inferences (Anselin, 1995; Harris, 2019; Ward and Gleditsch, 2019).

The socioeconomic, demographic and health facility characteristics are also bound to be different across space, especially for a large study area such as a country. As such, the effect of these characteristics on an outcome varies depending on the location of the individual. This results in spatial non-stationarity in the effect of these covariates on the outcome; i.e., the effect of a covariate on an outcome will vary depending on the underlying characteristics of the entities in that location (Brunsdon, Fotheringham and Charlton, 1996; Wheeler, 2014; Harris, 2019). This study, endeavours to explore whether spatial dependencies exist in the utilisation of maternal health care in Kenya and if so, use a spatial model to analyse the factors that explain utilisation when spatial dependency is controlled for. It also aims at determining whether spatial non-stationarity exists in the factors that explain the utilisation of maternal health care.

2. Data Sources

2.1. Demographic and Health Surveys (DHS)

DHS provide data on the socioeconomic and demographic characteristics of the individuals and the maternal health outcome of interest in the study; i.e., delivery by a skilled provider. They also provide data on the geographical coordinates of clusters interviewed in the survey. Data from the 2014 survey are utilised. The Kenya 2014 DHS data was conducted in 1612 clusters. 1594 of these clusters had the necessary sample for use in this study; i.e., births within 5 years of the 2014 DHS survey. Within this period 20931 children were born to 14924 women. Figure A1 in the appendix shows the distribution of the clusters interviewed in the Kenya 2014 DHS survey (Kenya National Bureau of Statistics *et al.*, 2015).

2.2. Kenya Master Health Facilities List (KMHFL)

While DHS are rich in demographic and socioeconomic characteristics of the women and households that make up the samples in these surveys, they are lacking in health facility information. This implies that most studies which use DHS data lack information on the accessibility and characteristics of health facilities which are pertinent in explaining utilisation and in turn maternal and child health outcomes. The KMHFL provides a list of health facilities currently providing antenatal care, postnatal care and maternity services. There were 10,505 operational health facilities in 2015. Out of these, 2103 offered maternity services (OpenAFRICA and Muthami, 2015; Kenya Ministry of Health, 2021). Data from KMHFL is appended to DHS data and used to construct the supply-side variables.

2.3. Kenya District Information System (DHIS2), Google Earth and ArcGIS

Data extracted from the KMHFL identifies the location of health facilities up to the ward level which is the lowest administrative level in Kenya. Precise locations are required to enable the calculation of the distance to health facilities variable. Therefore, GPS coordinates of health facilities are sourced from the DHIS2, Google Earth and ArcGIS. These data sources provide data on the latitudes and longitudes of the facilities offering maternal health care services in Kenya. (Kenya Ministry of Health, 2018; Google, 2019; Maina *et al.*, 2019; Environmental Systems Research Institute, 2020). Data with GPS coordinates for Kenya's country boundaries are sourced from Data World (2017).

3. Methodology and Results

3.1. Spatial autocorrelation in the utilisation of Maternal Health Care Utilisation in Kenya

Figure A2 in the appendix shows the utilisation of maternal health care across the country. However, the map does not show how similar areas are to their neighbours in terms of utilisation. To assess this, the Moran I statistic is used as a measure of whether spatial autocorrelation exists between the utilisation in a cluster and its neighbours. It also explores whether there is any patterning in the measured attributes (Anselin, 1995; Brunsdon and Comber, 2018).

The first step is to define neighbours to cluster i. Neighbours are defined using a bandwidth which can either be adaptive or fixed. An adaptive bandwidth considers a certain number of clusters close to cluster i as the neighbours while fixed bandwidth considers the observations within a certain distance of cluster i as neighbours. I utilise the adaptive bandwidth since the clusters interviewed are not uniformly distributed as shown in figure A1 in the appendix. As such, the adaptive bandwidth is more appropriate since it takes account of the differences in the distribution. Therefore a cluster will always have a neighbour even in sparsely distributed areas where the distance between the clusters is large.

The optimal number of neighbours is determined using a bisquare kernel bandwidth which employs a distance decay function where the observations on coordinate *i* are given the highest weight of one and have the largest influence on the local regression for coordinate *i*. The weights of the observations away from coordinate *i* reduce as the distance increases with the observations which are furthest away receiving the lowest weights. The use of a lower bandwidth means that the weights will reduce rapidly with increasing distance while the use of higher bandwidth will result in the weights being almost constant for all the observations used in the local regression (Brunsdon, Fotheringham and Charlton, 1996; Lu *et al.*, 2014; Gollini *et al.*, 2015; Hajarisman and Karyana, 2016). The more preferable option is to allow the statistical package that is being used to estimate the model, to determine the optimal bandwidth rather than producing one. The optimal bandwidth is calculated as 20[1]

I then use Moran's I statistic to test the hypothesis that the observations are spatially independent. It is calculated as:

$$I = \frac{\mathrm{n}}{\mathrm{S_0}} \frac{\sum_{\mathrm{i}=1}^{\mathrm{n}} \sum_{\mathrm{j}=1}^{\mathrm{n}} \mathrm{w_{i,j}} (\mathrm{y_i} - \bar{\mathrm{y}})(\mathrm{y_j} - \bar{\mathrm{y}})}{\sum_{\mathrm{i}=1}^{\mathrm{n}} (\mathrm{y_i} - \bar{\mathrm{y}})^2} \tag{3.4}$$

where:

- y_i are the individual observations, in this context clusters
- y_j are cluster i's neighbours
- $w_{i,j}$ is the i, j[th] element of the distance weight matrix. It is a binary indicator equal to 1 if i and j are neighbours and zero otherwise
- S_0 is equal to $\sum_{i=1}^{n} \sum_{j=1}^{n} w_{i,j}$

Table 1 presents the results of the Moran I test.

[1] Calculated using the GWmodel package in R (Gollini *et al.*, 2015).

Table 1: Moran I statistics

	Moran I statistic	*95% CI*		*Expectation*	*p-value*
Delivery assisted by a skilled provider	*0.3933*	*0.3787*	*0.4079*	*-0.0006*	*0.0000*

The presence of spatial autocorrelation in the utilisation of maternal health care is statistically significant as shown by the expected Moran I statistics being outside the 95% confidence interval of the estimated statistic for all the outcomes. The sign of the statistic is positive thus indicating the presence of positive spatial autocorrelation. This implies that areas with high utilisation of deliveries by a skilled provider have neighbours with high utilisation and vice versa. While this is true at the country level, this scenario might not be true for all the clusters examined.

Figure A3 in the appendix presents the Moran scatterplots which show the different associations that exist between the utilisation within cluster x_i and the weighted mean utilisation of their neighbours Wx_i (Anselin, 1995; Ward and Gleditsch, 2019). The fitted regression line confirms the presence of positive spatial autocorrelation for all the outcomes. However, not all clusters have neighbours with similar values. The plots are divided into four quadrants. The quadrants are determined by the mean utilisation of cluster i on the x-axis and the mean utilisation of cluster i's neighbours on the y-axis. The observations in the top right and bottom left quadrants of the plot display positive spatial autocorrelation; i.e., clusters with high utilisation are surrounded by clusters with high utilisation and vice versa. However, the clusters in the top-left quadrant and the bottom-right quadrants display negative autocorrelation; i.e., clusters with low utilisation are surrounded by clusters with high utilisation and vice versa. The clusters with the highest influence on the global Moran I statistic and the outliers are indicated by crossed diamond points.

These differences in spatial autocorrelation are further investigated using local Moran statistics which are calculated for the 1581 individual clusters. The aim here is to identify pockets of non-stationarity across the study space and assess the influence of individual observations on the spatial autocorrelation observed above (Anselin, 1995). The local Moran I statistic is calculated as:

$$I_i = \frac{\mathrm{n}}{\mathrm{S}_0} \frac{\sum_{\mathrm{i}=1}^{\mathrm{n}} \sum_{\mathrm{j}=1}^{\mathrm{n}} \mathrm{w}_{\mathrm{i,j}} (\mathrm{y}_\mathrm{i} - \bar{\mathrm{y}})(\mathrm{y}_\mathrm{j} - \bar{\mathrm{y}})}{\sum_{\mathrm{i}=1}^{\mathrm{n}} (\mathrm{y}_\mathrm{i} - \bar{\mathrm{y}})^2}$$

(3.5)

where the definitions remain as in equation 3.4.

Figure A4 in the appendix shows the local Moran value groups which are statistically significant. The plot shows the utilisation of deliveries by a skilled provider in cluster i in relation to the weighted mean utilisation of the neighbouring clusters. The p-values are adjusted to take into account multiple testing since the local statistics are calculated over and over again for each cluster and are likely to result in false positives; i.e., an indication of spatial autocorrelation where it does not exist (Brunsdon and Comber, 2018). Clusters with high utilisation which have neighbours with high utilisation are mainly concentrated in the central and south-eastern parts of the country. The clusters with low utilisation which have neighbours with low utilisation are mainly in the northern part of the country.

3.2. Spatial heterogeneity in the utilisation of maternal health care in Kenya

The second aspect of spatial dependency seeks to explore spatial heterogeneity in the utilisation of maternal health care. This would imply that the local statistics at cluster i are statistically different from the global statistics (Brunsdon and Comber, 2018). I use the mean utilisation of maternal health care to assess whether spatial heterogeneity exists. The geographically weighted local means are calculated for each of the 1581 clusters. The calculation employs a distance decay function as before in section 3.1 where observations in cluster i have a weight of 1 and the weight reduces as one moves away from cluster i. Outside the optimal bandwidth, observations are given a weight of zero. Figure A5 shows the geographically weighted means calculated at the local level. While the local means in most of the clusters are not significantly different from the global mean, some of the clusters in the north of Kenya have significantly lower utilisation of deliveries by a skilled provider compared to the global averages and some of the clusters in central Kenya have significantly higher utilisations compared to the global mean.

Having established the presence of spatial dependencies; i.e., spatial autocorrelation and heterogeneity, in the utilisation of deliveries assisted by a skilled provider, I now proceed to characterise the relationship between utilisation and its determinants using a spatial model.

3.3. Geographically Weighted Regression (GWR) for Maternal Health Care Utilisation in Kenya

Unlike in traditional regression models where the coefficients are constant across a study area and observations are assumed to be homogenous, geographically weighted regressions allow for heterogeneity of observations that arise from regional and geographical differences. Administrative regions such as provinces and states have been used in past studies to show geographical differences. This forms the basis of utilising GWR models which use geographic positioning system (GPS) coordinates to indicate the spatial location of an individual which is more precise compared to traditional analysis that uses provinces to indicate spatial location. To the best of my knowledge, this kind of analysis has not been done for Kenya.

The individual socioeconomic and demographic characteristics controlled for in this study are well motivated by previous studies on maternal health care utilisation. These characteristics are the mother's education level, household wealth, region of residence, place of residence, mother's age when the child is born, parity and marital status (Omotayo, 2008; Ejiagha, Ojiako and Eze, 2012; Asamoah, Agardh and Cromley, 2014; Ganle *et al.*, 2014; McLaren, Ardington and Leibbrandt, 2014; Anafcheh *et al.*, 2018; Okosun, 2018). In addition, distance to the nearest health facility, size of the nearest health facility and alternative supply of health facilities within a 5 km radius of an individual are added as control variables to measure the access of individuals to quality health facilities[2]. Table A1 in the appendix shows the variables used and their definitions as used in the context of this study.

Figures A6.1 -A6.4 in the appendix show the average distribution of some of the covariates used in the GWR analysis except for marital status and place of residence. The average level of education for mothers in most of the clusters is primary education (44.91% of the clusters) and secondary education (45.98% of the clusters). On a lesser scale are clusters with an average of less than primary education (5.19%) and only 3.92% of the clusters have high average levels of education (higher than secondary education). Most of the clusters with low levels of education (lower than primary are located in the north and north-eastern areas of the country which are considered Arid and Semi-Arid Lands (ASAL) which are highly disadvantaged in terms of education due to the nomadic way of life in these areas in search of pasture (Commission on Revenue Allocation, 2012; Kenya National Bureau of Statistics, 2018). The highest proportion of clusters has households in the middle of the asset index distribution (33.40% of clusters are in the third asset quintile) followed by 20.87%, 17.90%, 17.84% and 9.99% of the clusters at quintiles 4,5,2 and 1, respectively. Same as in the case of education, the poorest of the clusters are mostly in the north and north-eastern parts of the country and are thus disadvantaged for the same reasons discussed above in the case of education. The clusters

[2] According to the Kenya Ministry of health, a health facility is accessible to an individual if it is within 5 km of their residence, (Ministry of Health, 2014).

in these areas also have the highest distance to cover to reach the nearest health facility offering maternal health care and as such, this further makes an already tough situation worse; i.e., the mothers have to contend with barriers to utilisation of maternal health care from both the demand and supply-side.

While the average mother's age at the time a child is born and the average number of children born to a mother do not seem to display any discernible patterns across the study space, a substantial proportion of the clusters (24.41%) on average have mothers who are in the lowest age quintile (15-22 years) and have 2-3 children (32.6% of the clusters). Most of the clusters (62.08%) are closest to a level two health facility. Most of the clusters (45.86%) also do not have an alternative supply of health facilities within a five km radius.

Given these observed differences across clusters, it is expected that the effect of these covariates on the outcome variable is bound to be different depending on the underlying cluster characteristics which differ depending on the location. GWR models, therefore, serve a twofold purpose. The first is to determine whether there is spatial non-stationarity in the factors that explain the outcome variable; i.e., does the standard regression model which assumes homogeneity of observations explain the relationship between the covariates and the outcome well, and second, to determine the covariates whose effect on the outcome varies spatially.

The estimation of the GWR models is accomplished in several steps. First, a standard regression also referred to as a global model, is estimated. The global model assumes that the associations between the outcome and covariates in question do not vary spatially.

$$y = \beta_0 + \sum_{k}^{\rho-1} \beta_k x_k + \varepsilon \quad (1)$$

Where:

- y is the outcome variable; i.e., delivery by a skilled provider
- β_0 is the intercept
- β_k is the coefficient for the k^{th} covariate; i.e., mother's education levels, wealth, place of residence, mother's age, parity, marital status, distance to the nearest health facility, size of nearest health facility, alternative supply of health facilities
- ε is the random error

- $\boldsymbol{\rho}$ is the number of regression coefficients to be estimated

Local regressions are then estimated for each of the geographical coordinates. The local regression estimates n*ρ number of coefficients; i.e., ρ coefficients for each of the n locations. The number of clusters with observations relevant to this study is 1581. The model is represented as:

$$y_i = \beta_{0i} + \sum_{k}^{\rho-1} \beta_{ki} x_{ki} + \varepsilon_i \tag{2}$$

Where x_{ki} is the value of k^{th} covariate of location i

The local regression is estimated over a specific bandwidth/ distance from each geographical coordinate. The bandwidth is calibrated to minimise:

$$\sum_{i=1}^{n} \left(y_i - \hat{y}_{\neq i}(B) \right)^2 \tag{3}$$

Where:

- y_i is the fitted value of the outcome variable from the global regression
- $\hat{y}_{\neq i}(B)$ is the fitted value of the outcome variable from the local regression excluding observations which are at coordinate *i*

The local regressions employ a distance decay function where the observations on coordinate *i* are given the highest weight of one and have the largest influence on the local regression for coordinate *i*. The weights of the observations away from coordinate *i* reduce as the distance increases with the observations which are furthest away receiving the lowest weights. The use of a lower bandwidth means that the weights will reduce rapidly with increasing distance while the use of higher bandwidth will result in the weights being almost constant for all the observations used in the local regression (Brunsdon, Fotheringham and Charlton, 1996; Lu *et al.*, 2014; Hajarisman and Karyana, 2016). The more preferable option is to allow the statistical package that is being used to estimate the model, to determine the optimal bandwidth rather than producing one. The optimal bandwidth is estimated at 0.02863[3, 4]. This is the proportion of nearest observations used per geographical coordinate to estimate the local regression. This

3 Estimated on Stata (StataCorp, 2021)) using gwr package, (Pearce, 1998).

4 Computations were performed using facilities provided by the University of Cape Town's ICTS High Performance Computing team: hpc.uct.ac.za translates to approximately 595 observations of the 20,873 available observations being used to estimate the local regressions at each geographical coordinate.

Table 2 presents a summary of the local regression coefficients and compares them to the global regression coefficients[5].

5 Estimated on RStudio using spgwr package (Bivand et al., 2022).

Table 2: Summary of coefficients for local and global regressions using geographically weighted regression

	Local regression					***Global regression***
	Min.	***1st Quartile***	***Median***	***3rd Quartile***	***Max.***	
Intercept	-2.80571	-1.90971	-1.43418	-0.74041	0.760351	-1.1424
Mother's years of education	0.064653	0.144125	0.165067	0.179586	0.216519	0.1567
Household asset index	0.012541	0.117969	0.192094	0.271606	0.623003	0.1865
Place of residence (Omitted category: Urban)						
Rural	-2.19194	-0.64525	-0.41404	-0.2232	0.102262	-0.6737
Mother's age at child's birth	-0.00404	0.026856	0.0399	0.061357	0.163521	0.0447
Number of children	-0.70871	-0.31173	-0.18958	-0.16623	-0.10537	-0.2181
Marital status (Omitted category: Single/not living together)						
Married/ living together	-0.42947	-0.04003	0.074251	0.199648	0.659717	0.0551
Size of nearest health facility (Omitted category: Level 2)						
Level 3	-0.61883	0.110196	0.246202	0.435624	0.917065	0.2367
Level 4	-1.2173	0.177445	0.331073	0.537879	1.015911	0.2992
Level 5	-7.01892	0.378279	0.604396	0.955847	14.50597	0.7431
Level 6	-4.57623	-0.23181	0.237001	4.34516	18.19502	-0.0774
Alternative supply	-0.06223	-0.00498	0.021853	0.044508	0.999656	0.0108
Distance	-0.07304	-0.02305	-0.00926	0.004729	0.079176	-0.0108

The signs of the coefficients in the global model are as expected. The probability of a woman being delivered by a skilled provider increases with an increase in education levels, household asset wealth, age at which they bear a child, alternative supply of health and size of nearest health facilities (except for level 6 health facilities) and reduces with increase in the number of children and distance to the nearest health facility. Women living in rural areas are less likely to be delivered by a skilled provider compared to their urban counterparts. Married women and those living with a partner are less likely to be delivered by a skilled provider compared to their single counterparts.

The local coefficients show variation with coefficients ranging from negative to positive except for coefficients for years of mother's education and household asset wealth which are consistently positive. The variation in the local coefficients alludes to the possibility that spatial non-stationarity might exist in the factors that explain the utilisation of maternal health care as the coefficient estimates vary depending on the location of the cluster across the study space. To ascertain this, Monte Carlo simulations are used to determine which, if any, of the variables display statistically significant differences in the local regression coefficients as compared to the global regression coefficients. The simulations assign observations to a different geographical coordinate and measure how this affects the local regression coefficients. The hypotheses being evaluated are:

> *H_0: there are no statistically significant differences between the coefficients from the global and local regressions; i.e., the coefficients of the covariates explaining the outcome are constant across a geographical location.*
>
> *H_a: there are statistically significant differences between the coefficients from the global and local regressions; i.e., the coefficients of the covariates explaining the outcome vary depending on geographical location.* (4)

Table 3 presents the results of the Monte Carlo simulation testing for spatial non-stationarity of the variables which affect maternal health care utilisation[6]. All the covariates considered display spatial non-stationarity except for the covariate representing marital status and also for level 5 and 6 health facilities.

---

6 *Estimated in RStudio (RStudio Team, 2020) using GWmodel package (Lu et al., 2014; Gollini et al., 2015).*

Table 3: p-values of spatial non-stationarity tests of factors determining utilisation of deliveries by a skilled provider using Monte Carlo simulations

	p-values
Intercept	0.00
Mother's years of education	0.00
Household asset index	0.00
Place of residence: omitted category (Urban)	
Rural	0.00
Mother's age at child's birth	0.00
Number of children	0.00
Marital status: omitted category (Not married/not living together)	
Married/ living together	0.12
Size of nearest health facility: omitted category (Level 2)	
Level 3	0.00
Level 4	0.01
Level 5	0.11
Level 6	0.11
Alternative supply of health facilities	0.00
Distance to the nearest health facility	0.00

Given the results of the Monte Carlo simulations, figures A7.1-A7.5 in the appendix are plotted to illustrate the spatial distribution of the coefficients for the local regressions of the variables that display statistically significant differences from the global regression coefficients. A correlation matrix (figure A8 in the appendix)[7] is also included to explore the relationship between the GWR coefficients and the underlying cluster characteristics. Due to the categorical nature of the place of residence and health facility level variables, correlations will not suffice and, therefore, the GWR coefficient data for the categorical variables are overlaid on a

7 Plotted using a Stata user written command heatplot (Jann, 2019).

choropleth map showing the probability of finding a health facility within a 5 km radius to aid in the interpretation of the observed GWR coefficients[8].

The increase in the probability of utilising maternal health care is highest in the clusters where mothers on average have lower years of education, mothers deliver children at a younger age and have a lower number of alternative supply of health facilities. The largest reductions in the probability of utilising maternal health care are associated with clusters where mothers have relatively more children and rural clusters in areas with a low density of health facilities. As expected, the higher the probability of finding a level 3 or level 4 health facility within a 5 km radius of a woman's dwelling, the more likely they are to utilise it compared to a level 2 health facility. The distance covariate also displays the expected effect on delivery by a skilled provider with the largest decreases in the probability of utilisation in the clusters which are furthest from a health facility. These results are further discussed in the next section.

3.4. Discussion of Results

The distribution of the covariates explaining the utilisation of maternal health care varies across Kenya thus presenting the possibility that their effect on the outcomes will vary depending on the location. One indicator of this possibility is the divergent signs of the coefficients of the GWR local regression estimates from positive to negative for covariates which present a strictly positive/negative coefficient in a standard global regression model except for the coefficient for the variables measuring the mother's years of education. The Monte Carlo simulations show the presence of spatial non-stationarity in the covariates. The effect of marital status does not show spatial non-stationarity. This means that the global model does estimate the effect of marital status on maternal health outcomes well. The effect of level 5 and 6 health facilities also do not display spatial non-stationarity compared to level 2 health facilities, possibly due to the small number of these facilities offering maternal health care in Kenya; i.e., only three level 6 health facilities and sixteen level 5 health facilities were offering maternal health care services as of 2015.

Most of the clusters which display high coefficients for mothers' education and wealth are located in the north and north-eastern parts of Kenya. These are areas which are arid and semi-arid lands (ASAL). As such, these areas tend to be marginalised due to low attendance at schools and have higher poverty levels (Kenya National Bureau of Statistics, 2018). Therefore,

8 Stata user-written commands spgrid (Pisati, 2011) used to generate the underlying grid and spkde (Pisati, 2009) used to calculate the probability density function.

being more educated in such areas offers a woman a higher advantage in the utilisation of maternal health care. This is similar to results by Ohonba, Ngepah and Simo-Kengne (2019) who found that the effect of maternal education on child health outcomes was higher in South Africa for sub-populations which have higher educational deficits; i.e., Black people and coloureds. Higher education coefficients are also associated with clusters which have wealthier households, fewer children per woman and a higher alternative supply of health facilities. More educated women in clusters with richer households, therefore, have a double advantage of being aware of the maternal health care services that they require and having the means to mitigate the barriers that are associated with utilisation such as cost and distance. Wealthier women can also afford to live in areas with enough amenities thus reducing the distance between them and the facilities where they need to seek maternal health care. More educated women are also expected to have fewer children due to having information on the dangers posed by having more children and thus they would be more open to utilising family planning to manage their family sizes (Liu and Raftery, 2020).

The negative effect of a higher number of children on the probability of utilising maternal health care is further exacerbated in clusters where women on average deliver at an older age and for clusters which are further away from health facilities. This fits the expectation that older women are more likely to have more children. The negative effect of distance is especially higher in rural areas which are normally disadvantaged in terms of accessibility due to the low density of health facilities. The clusters with the highest effect of the covariate alternative supply of health facilities also have the lowest distances to the nearest health facility thus augmenting the positive effect of the alternative health facilities. The increase in the probability of utilisation of level 3 and 4 health facilities compared to level 2 health facilities is higher for clusters in areas where the probability of finding a level 3 or 4 health facility is also higher. However, there exist clusters where living close to these higher-level health facilities compared to level two health facilities reduces the probability of utilising the higher-level facilities and an exploration of the reasons why this might be the case would be of added value. Women in rural areas are less likely to utilize maternal health care, especially in clusters with women with lower average levels of education, where mothers deliver at a younger age and with a lower alternative supply of health facilities.

4. Conclusion

This study explores whether there exists spatial dependence and spatial non-stationarity in the factors explaining the utilization of maternal health care. The results show that spatial autocorrelation and spatial heterogeneity exist in the utilization of deliveries assisted by a skilled provider. Additionally, for most of the covariates controlled for in the geographically weighted regression models, the coefficients are not constant across the study space as assumed by standard global regression models. The study recommends that an exploration of spatial non-stationarity be conducted before deciding whether to use standard global regression models or spatial models for analysis.

REFERENCES

Anafcheh, T. *et al.* (2018) 'Temporal and spatial distribution of under-five mortality and factors associated with multiple cases of under-five deaths within a family in the rural area of Khuzestan, Southern Iran', *Scientific Reports*, 8(1), pp. 1–8. doi: 10.1038/s41598-018-36438-5.

Anselin, L. (1995) 'Local Indicators of Spatial Association—LISA', *Geographical Analysis*, 27(2), pp. 93–115. doi: 10.1111/J.1538-4632.1995.TB00338.X.

Asamoah, B. O., Agardh, A. and Cromley, E. K. (2014) 'Spatial Analysis of Skilled Birth Attendant Utilization in Ghana', *Global Journal of Health Science*, 6(4), pp. 117–127. doi: 10.5539/gjhs.v6n4p117.

Bivand, R. *et al.* (2022) 'Package "spgwr"'. RStudio, PBC.

Brunsdon, C. and Comber, L. (2018) *An introduction to R for spatial analysis & mapping*. 2nd edn. Los Angeles: SAGE.

Brunsdon, C., Fotheringham, S. and Charlton, M. E. (1996) 'Geographically Weighted Regression: A Method of Exploring Spatial Nonstationarity', *Geographical Analysis*, 28(4), pp. 281–298. doi: 10.1016/B978-008044910-4.00447-8.

Commission on Revenue Allocation (2012) *Survey Report on Marginalized Areas/Counties in Kenya*. CRA Working Paper 2012/03. Available at: https://devolutionhub.or.ke/resource/survey-report-on-marginalized-areascounties-in-kenya (Accessed: 24 June 2021).

Data World (2017) *Kenya - Subnational Administrative Boundaries*. Available at: https://data.world/ocha-rosea/2c0b7571-4bef-4347-9b81-b2174c13f9ef (Accessed: 26 June 2021).

Ejiagha, I. R., Ojiako, J. C. and Eze, C. G. (2012) 'Accessibility Analysis of Healthcare Delivery System within Enugu Urban Area Using Geographic Information System', *Journal of Geographic Information System*, 4, pp. 312–321. doi: 10.4236/jgis.2012.44036.

Environmental Systems Research Institute (2020) *GIS Mapping Software, Location Intelligence & Spatial Analytics | Esri*. Available at: https://www.esri.com/en-us/home (Accessed: 22 June 2021).

Ganle, J. K. *et al.* (2014) 'A qualitative study of health system barriers to accessibility and utilization of maternal and newborn healthcare services in Ghana after user-fee abolition', *BMC Pregnancy and Childbirth*, 14. doi: 10.1186/s12884-014-0425-8.

Gollini, I. *et al.* (2015) 'Gwmodel: An R package for exploring spatial heterogeneity using geographically weighted models', *Journal of Statistical Software*, 63(17), pp. 1–50. doi: 10.18637/JSS.V063.I17.

Google (2019) *Google Earth*. Available at: https://earth.google.com/web/@0,0,0a,22251752.77375655d,35y,0h,0t,0r (Accessed: 22 June 2021).

Hajarisman, N. and Karyana, Y. (2016) 'Geographic Modeling on The Infant Mortality Rate in West Java', *MIMBAR,* 32(1), pp. 194–205. Available at: https://media.neliti.com/media/publications/159647-EN-geographic-modeling-on-the-infant-mortal.pdf (Accessed: 3 April 2019).

Harris, R. J. (2019) 'Not just nuisance : spatialising social statistics', in Whitworth, A. (ed.) *Towards a Spatial Social Policy : Bridging the Gap Between Geography and Social Policy*. Bristol: Policy Press, pp. 149–168.

Jann, B. (2019) 'HEATPLOT: Stata module to create heat plots and hexagon plots', *Statistical Software Components*. Boston College Department of Economics.

Kenya Ministry of Health (2018) *Kenya Health Information Systems (KHIS)*. Available at: https://hiskenya.org/dhis-web-commons/security/login.action (Accessed: 22 June 2021).

Kenya Ministry of Health (2021) *Kenya Master Health Facility List*. Available at: http://kmhfl.health.go.ke/#/home (Accessed: 24 June 2021).

Kenya National Bureau of Statistics *et al.* (2015) *Kenya Demographic and Health Survey 2014*. Available at: https://dhsprogram.com/publications/publication-FR308-DHS-Final-Reports.cfm (Accessed: 20 October 2022).

Kenya National Bureau of Statistics (2018) *BASIC REPORT: 2015/2016 Kenya Integrated Household Budget Survey (KIHBS)*.

Liu, D. H. and Raftery, A. E. (2020) 'How Do Education and Family Planning Accelerate Fertility Decline?', *Population and Development Review*, 46(3), pp. 409–441. doi: 10.1111/PADR.12347.

Lu, B. *et al.* (2014) 'The GWmodel R package: Further topics for exploring spatial heterogeneity using geographically weighted models', *Geo-Spatial Information Science*, 17(2), pp. 85–101. doi: 10.1080/10095020.2014.917453/SUPPL_FILE/TGSI_A_917453_SM1923.R.

Maina, J. *et al.* (2019) 'A spatial database of health facilities managed by the public health sector in sub Saharan Africa', *Scientific Data*, 6, p. 134. doi: 10.1038/s41597-019-0142-2.

McLaren, Z. M., Ardington, C. and Leibbrandt, M. (2014) 'Distance decay and persistent health care disparities in South Africa', *BMC Health Services Research*, 14. doi: 10.1186/s12913-014-0541-1.

Ministry of Health (2014) *Kenya Health Policy 2014-2030*. Available at: https://www.afidep.org/?wpfb_dl=80 (Accessed: 3 April 2019).

Ohonba, A., Ngepah, N. and Simo-Kengne, B. (2019) 'Maternal education and child health outcomes in South Africa: A panel data analysis', *Development Southern Africa*, 36(1), pp. 33–49. doi: 10.1080/0376835X.2018.1456908.

Okosun, O. S. (2018) 'Health Facilities Locational Analysis in Makurdi Local Government Area of Benue State-Nigeria Using GIS Application', *Geosfera Indonesia*, 3(3), pp. 26–42. doi: 10.19184/geosi.v3i3.8830.

Omotayo, O. (2008) 'Spatial Distribution and Accessibility to Healthcare Facilities in Akure South Local Government Area of Ondo State, Nigeria', *Annals of the University of Oradea, Geography Series*, pp. 7–18. Available at: https://www.researchgate.net/publication/327317706.

OpenAFRICA and Muthami, D. (2015) *Health Facilities in Kenya*. Available at: https://africaopendata.org/dataset?q=kenya (Accessed: 9 August 2021).

Pearce, M. S. (1998) *Geographically weighted regression : A method for exploring spatial nonstationarity*, *Stata Technical Bulletin*. Edited by H. J. Newton et al.

Pisati, M. (2009) 'SPKDE: Stata module to perform kernel estimation of density and intensity functions for two-dimensional spatial point patterns', *Statistical Software Components*. Boston College Department of Economics.

Pisati, M. (2011) 'SPGRID: Stata module to generate two-dimensional grids for spatial data analysis', *Statistical Software Components*. Boston College Department of Economics.

RStudio Team (2020) 'RStudio: Integrated Development for R'. Boston, MA: RStudio, PBC,.

Available at: http://www.rstudio.com/.

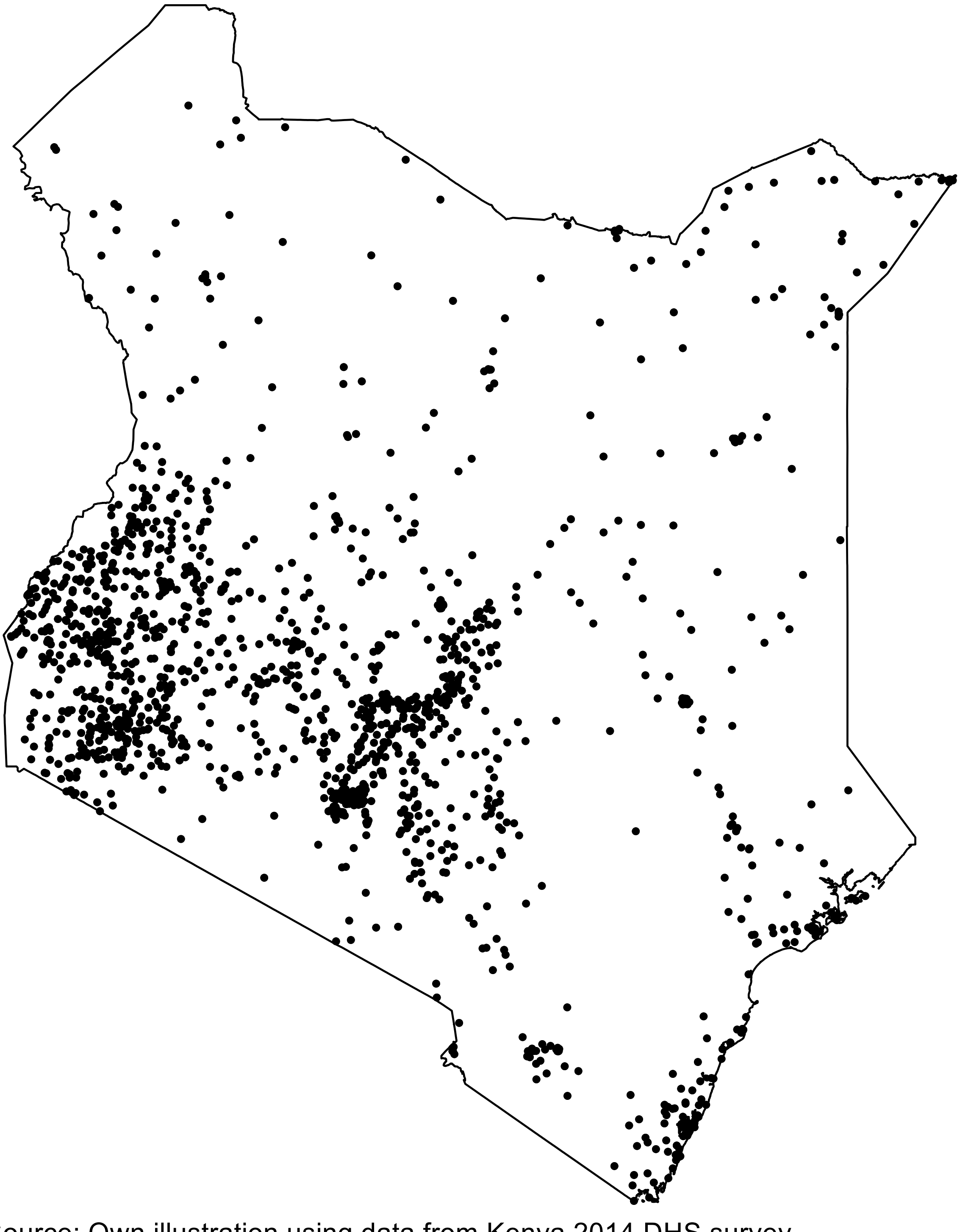
StataCorp (2021) ‘Stata Statistical Software’. College Station, TX: StataCorp LLC.

Ward, M. D. and Gleditsch, K. S. (2019) *Spatial Regression Models*. 2nd edn. Thousand Oaks: SAGE Publications, Inc. (Quantitative Applications in the Social Sciences). doi: 10.4135/9781071802588.

Wheeler, D. C. (2014) ‘Geographically weighted regression’, in Fischer, M. M. and Nijkamp, P. (eds) *Handbook of Regional Science*, pp. 1435–1459. doi: 10.1007/978-3-642-23430-9_77.

APPENDIX

Figure A1: Clusters interviewed in the DHS Kenya 2014 survey



Source: Own illustration using data from Kenya 2014 DHS survey

Table A1: Variable Definition

Variables	***Definition***
Outcome Variable	
Delivered by a skilled provider	It refers to a woman being assisted during delivery by a doctor or a nurse. It is measured by a binary variable with 1 representing deliveries done by a skilled provider and zero otherwise.
Individual (mother) level variables	
Mother's education	Measured by a categorical variable with 0 =no education, 1 = primary education, 2 = secondary education and 3 = higher than secondary education
Province of residence	Measured by a categorical variable with 0=Nairobi, 1=Central, 2=Coast, 3=Eastern, 4=Nyanza, 5=Rift valley, 6=Western and 7=North eastern province
Age of the mother at the child's birth	Measures the age of the mother at the time a child was born. Represented by a discrete variable. Ranges between 15 and 49 years.
Parity	Measures the number of children that have ever been born to a woman. Measured by a binary variable with 0 = less than four children and 1 = four or more children
Marital status	Measured by a binary variable with 1 =women who are married or living with a partner and 0 = women living alone.
Household-level variables	
Wealth	Measured by an asset index that is represented by a discrete variable with 0 = 1st quintile, 1=2nd quintile, 2=3rd quintile, 3=4th quintile and 4=5th quintile.
Cluster level variables	

Place of residence	Measured by a binary variable with 0 =urban dwellers and 1 = rural dwellers.
Distance	It measures how far a DHS cluster is from the nearest health facility in kilometres. It is represented by a continuous variable.
Size of health facility	It is represented by a discrete variable measuring the level of the nearest health facility where 0=level 2, 1=level 3, 2=level 4, 3=level 5 and 4=level 6.
Alternative supply	It is measured by the number of health facilities within a 5 km radius of the cluster.

Figure A2: Average utilisation of deliveries assisted by a skilled provider in the clusters interviewed in the Kenya 2014 DHS survey

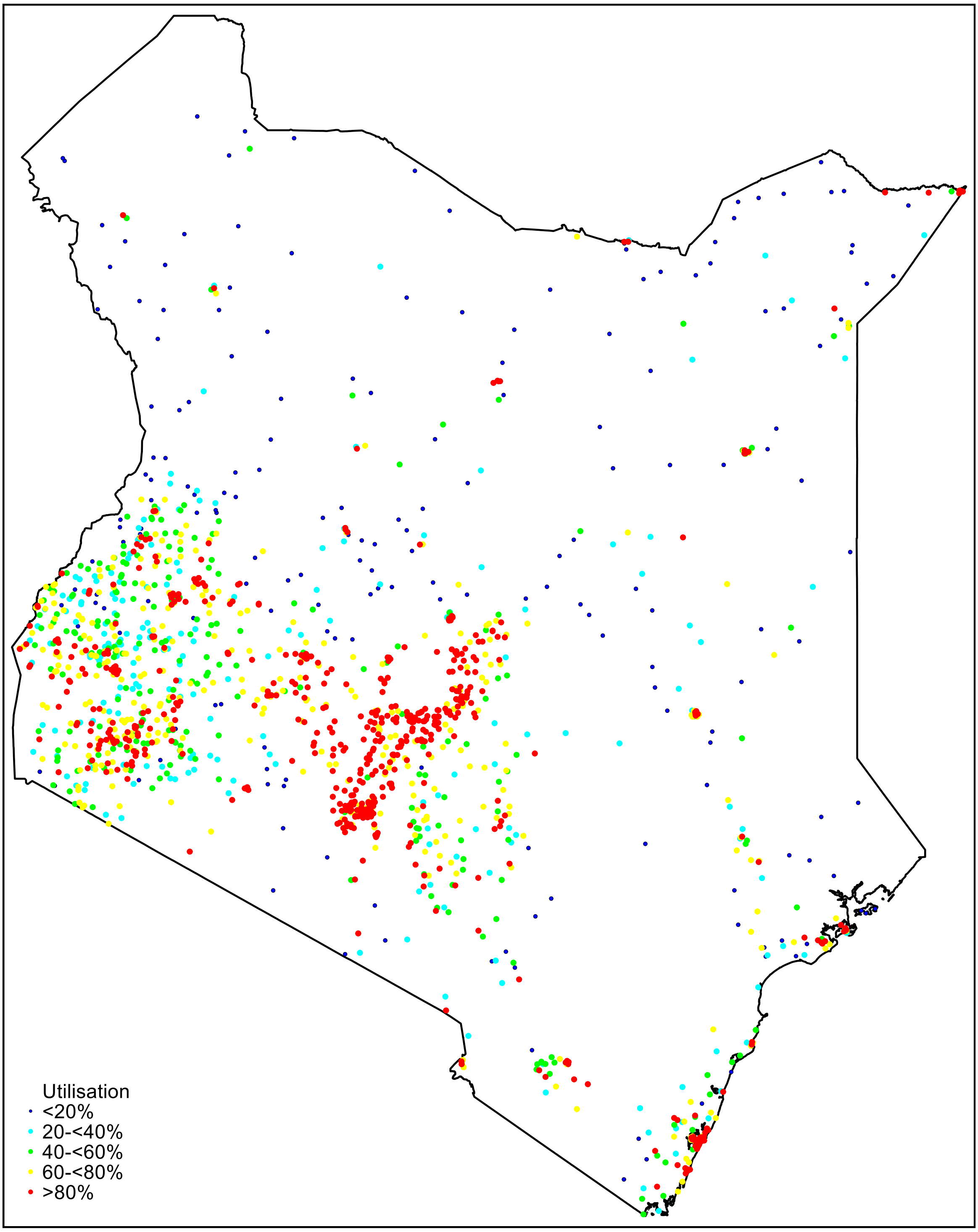


Figure A3: Moran scatterplot

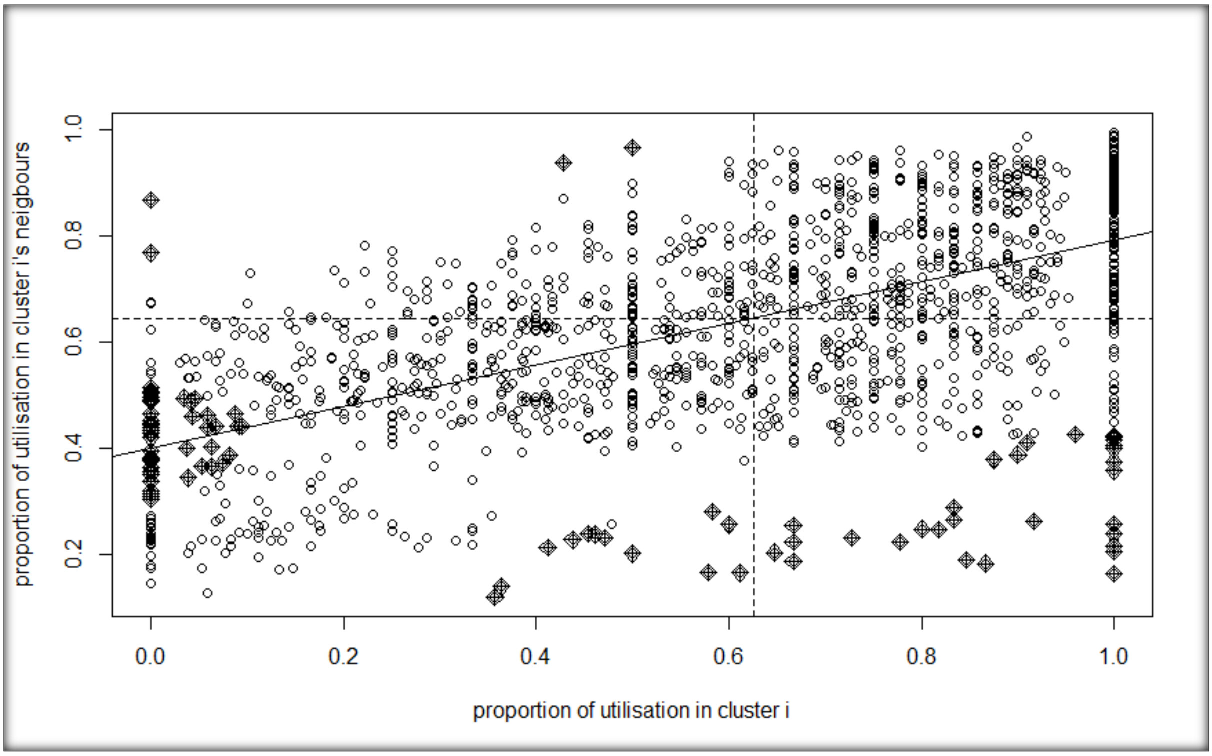


Figure A4: Local Moran value groups

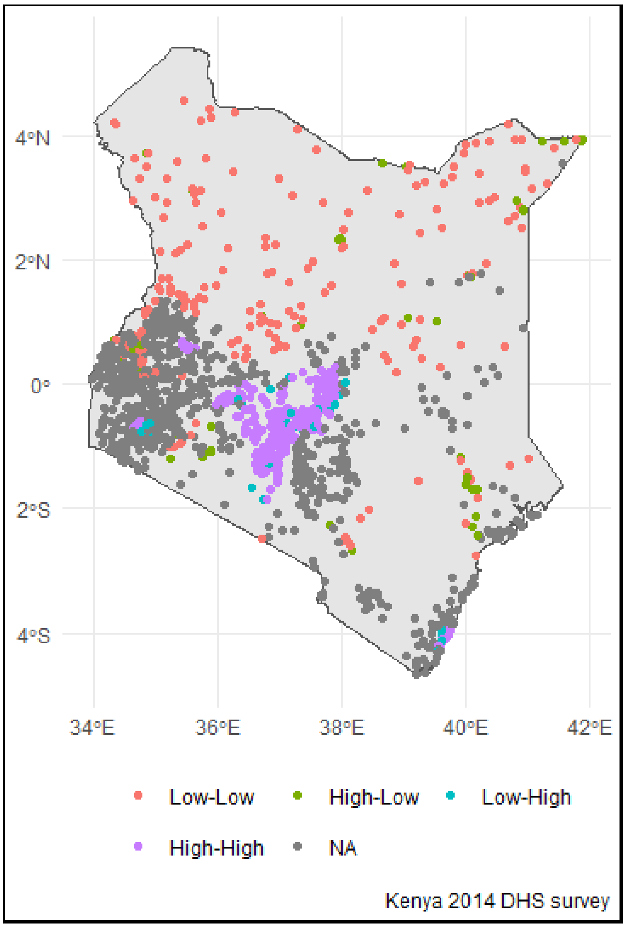


Figure A5: Geographically weighted mean utilisation of delivery assisted by a skilled provider

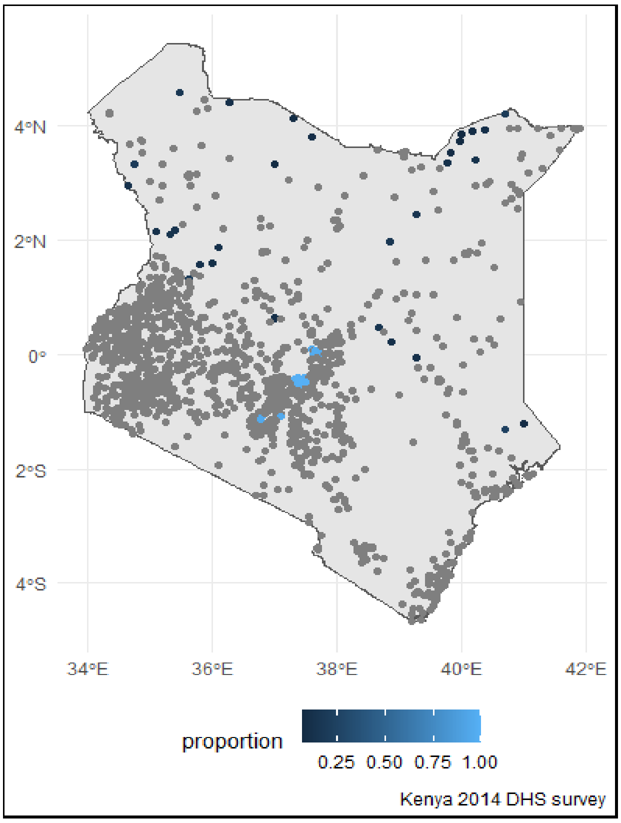
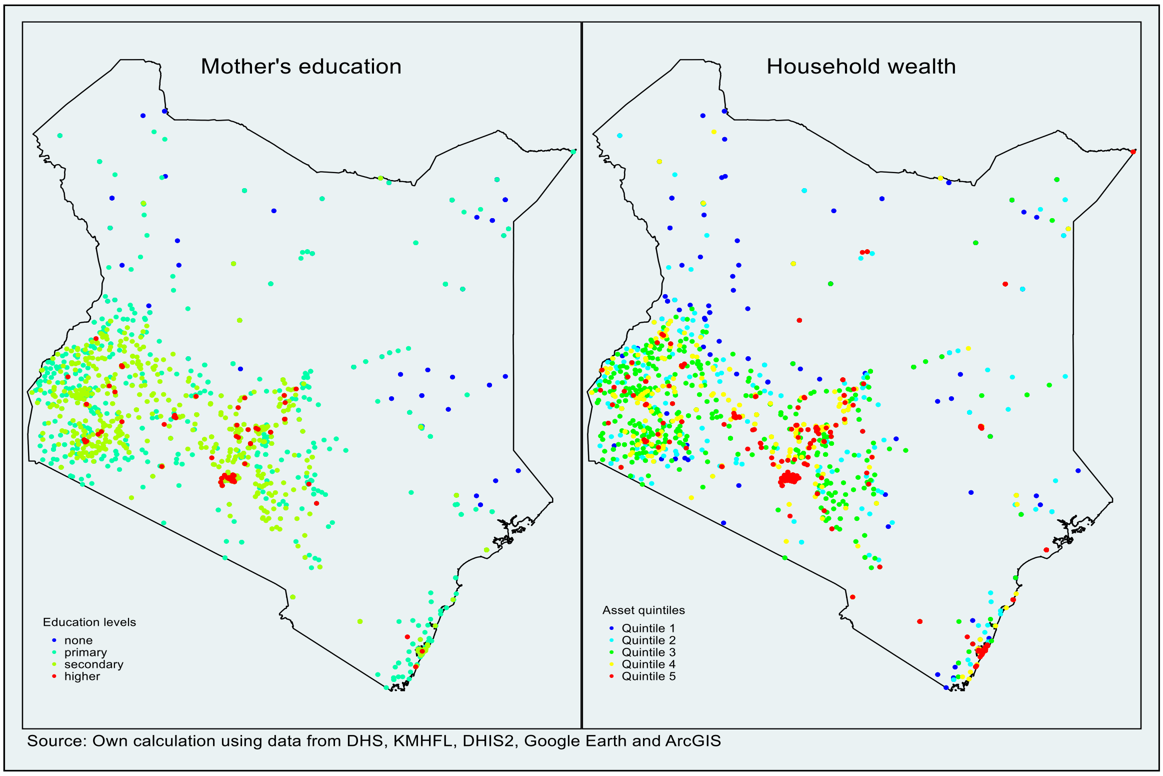
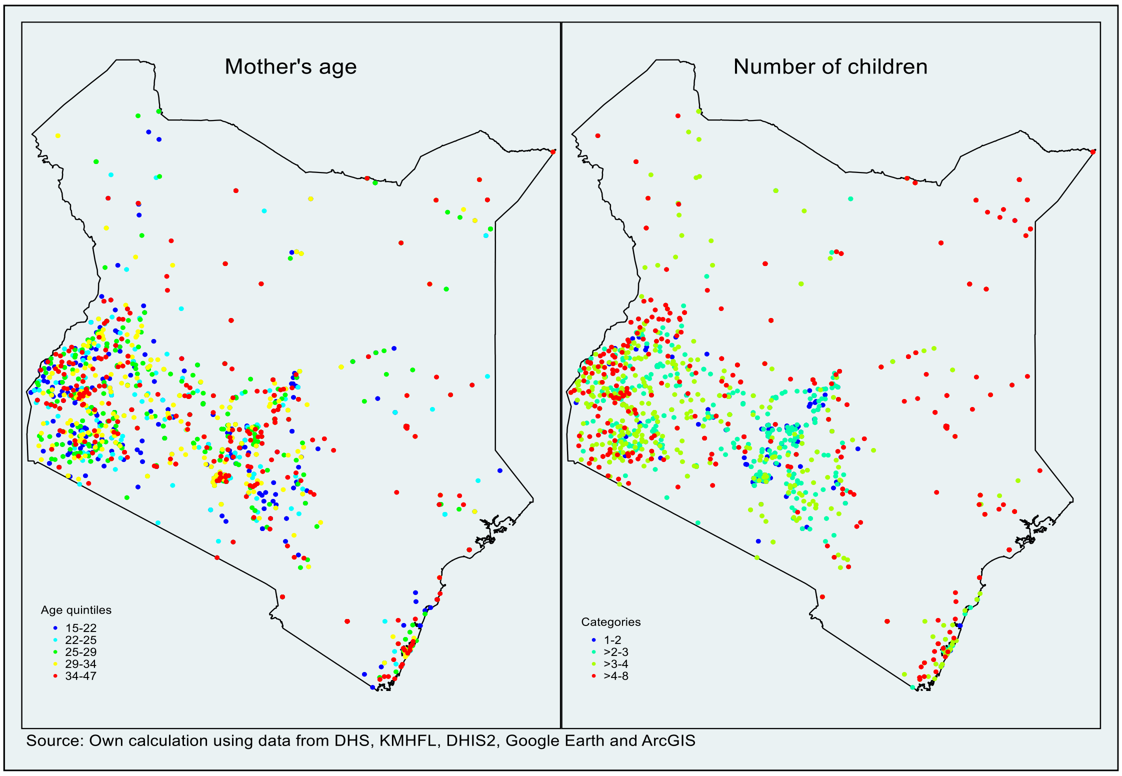


Figure A6.1: Average cluster characteristics



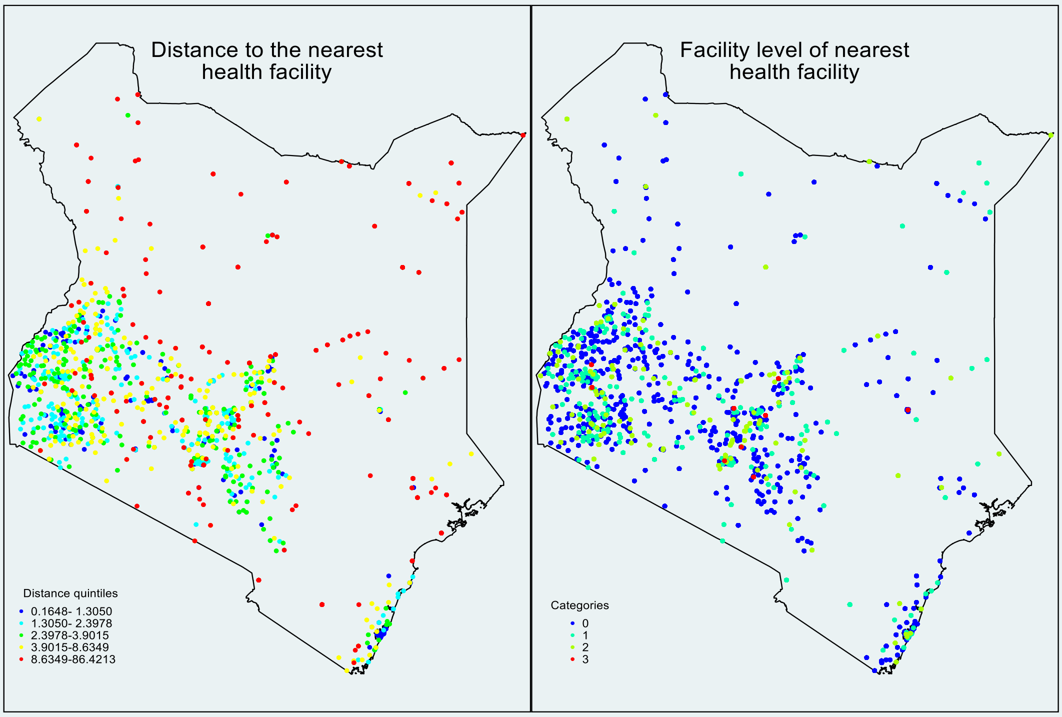
Source: Own calculation using data from DHS, KMHFL, DHIS2, Google Earth and ArcGIS

Figure A6.2:Average cluster characteristics



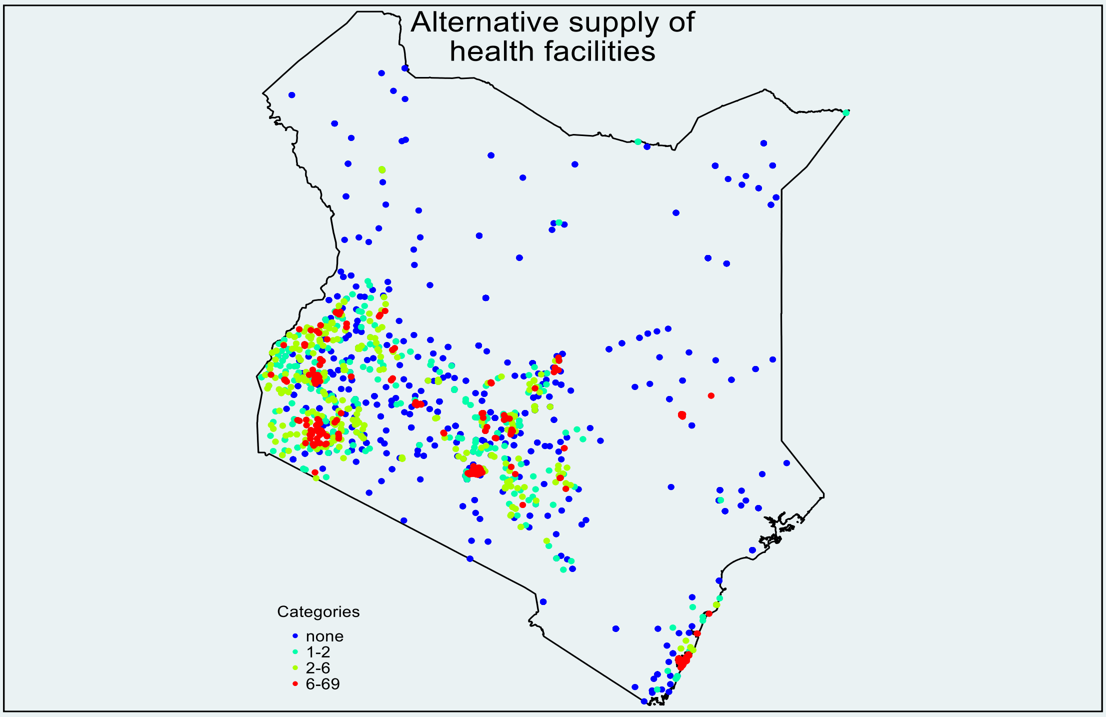
Source: Own calculation using data from DHS, KMHFL, DHIS2, Google Earth and ArcGIS

Figure A6.3:Average cluster characteristics



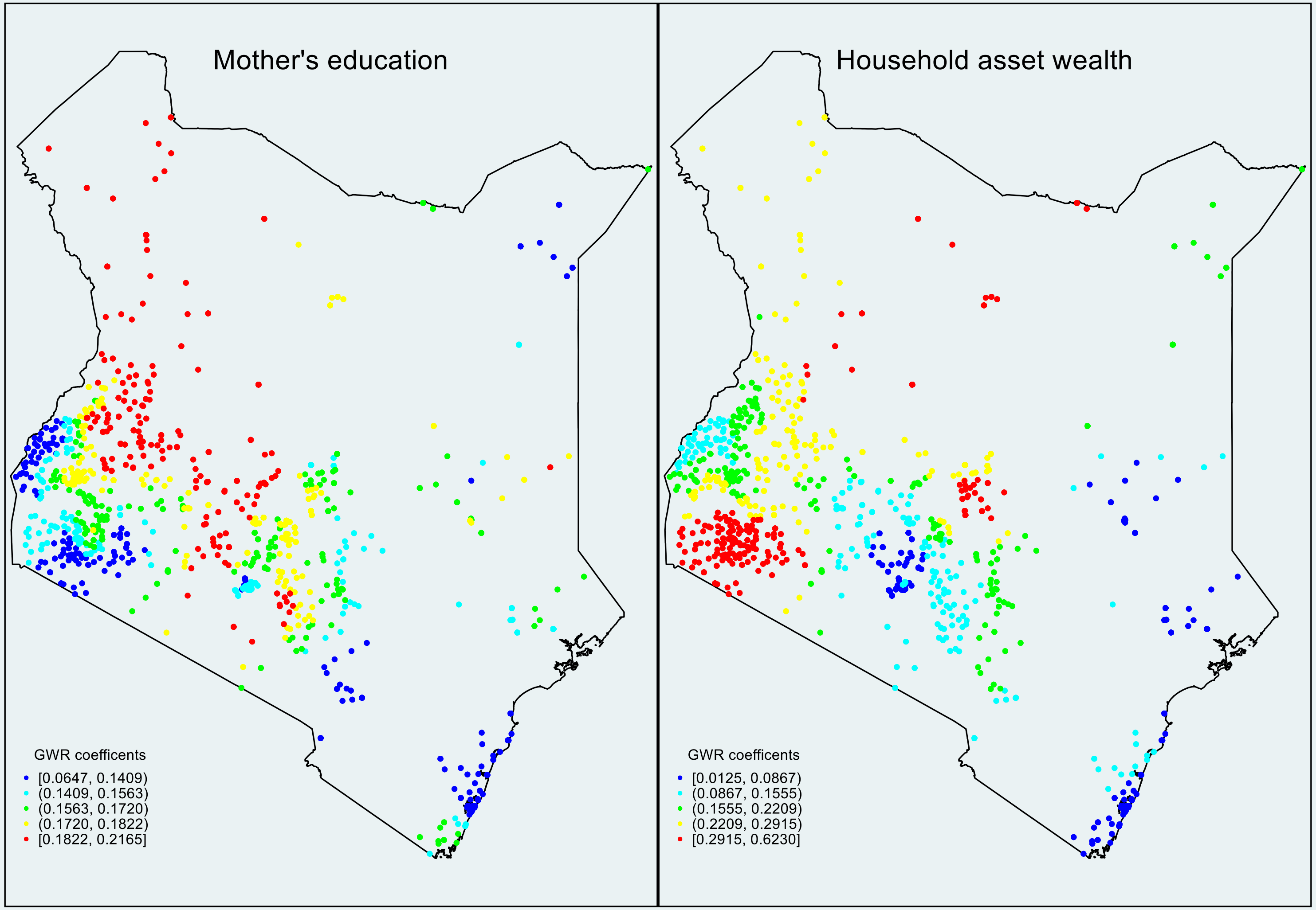
Source: Own calculation using data from DHS, KMHFL, DHIS2, Google Earth and ArcGIS

Figure A6.4:Average cluster characteristics



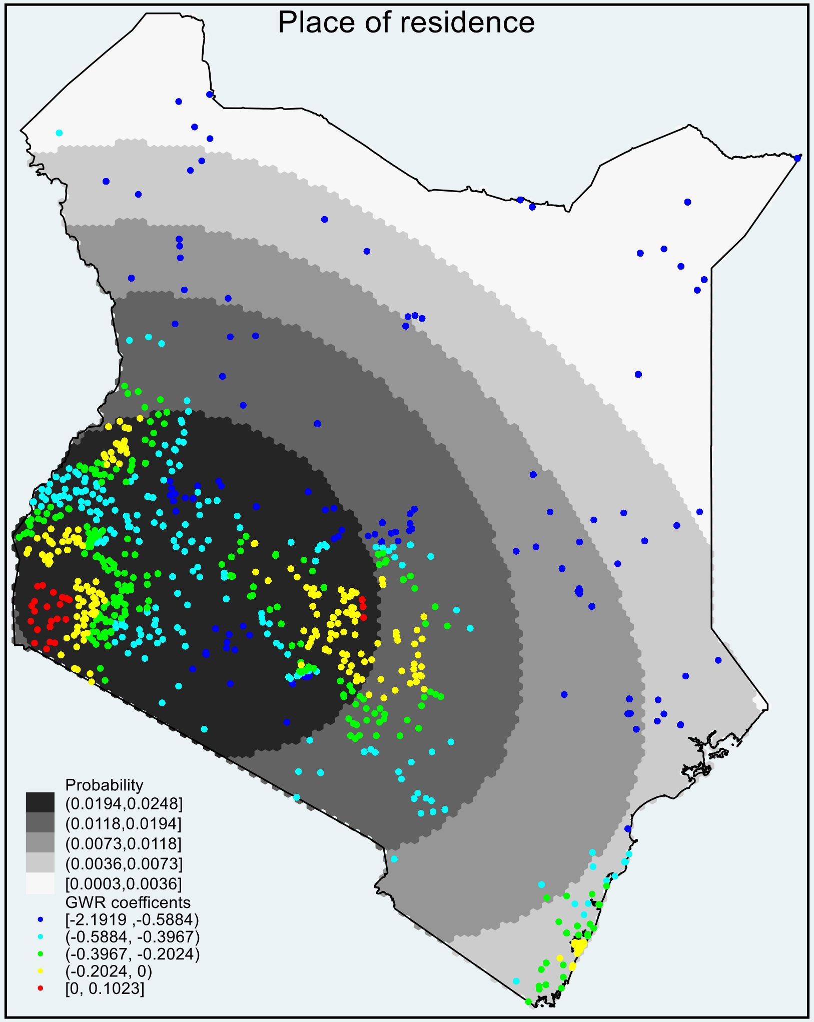
Source: Own calculation using data from DHS, KMHFL, DHIS2, Google Earth and ArcGIS

Figure A7.1: Spatial distribution of the effects of education and household asset wealth on utilisation of deliveries assisted by a skilled provider



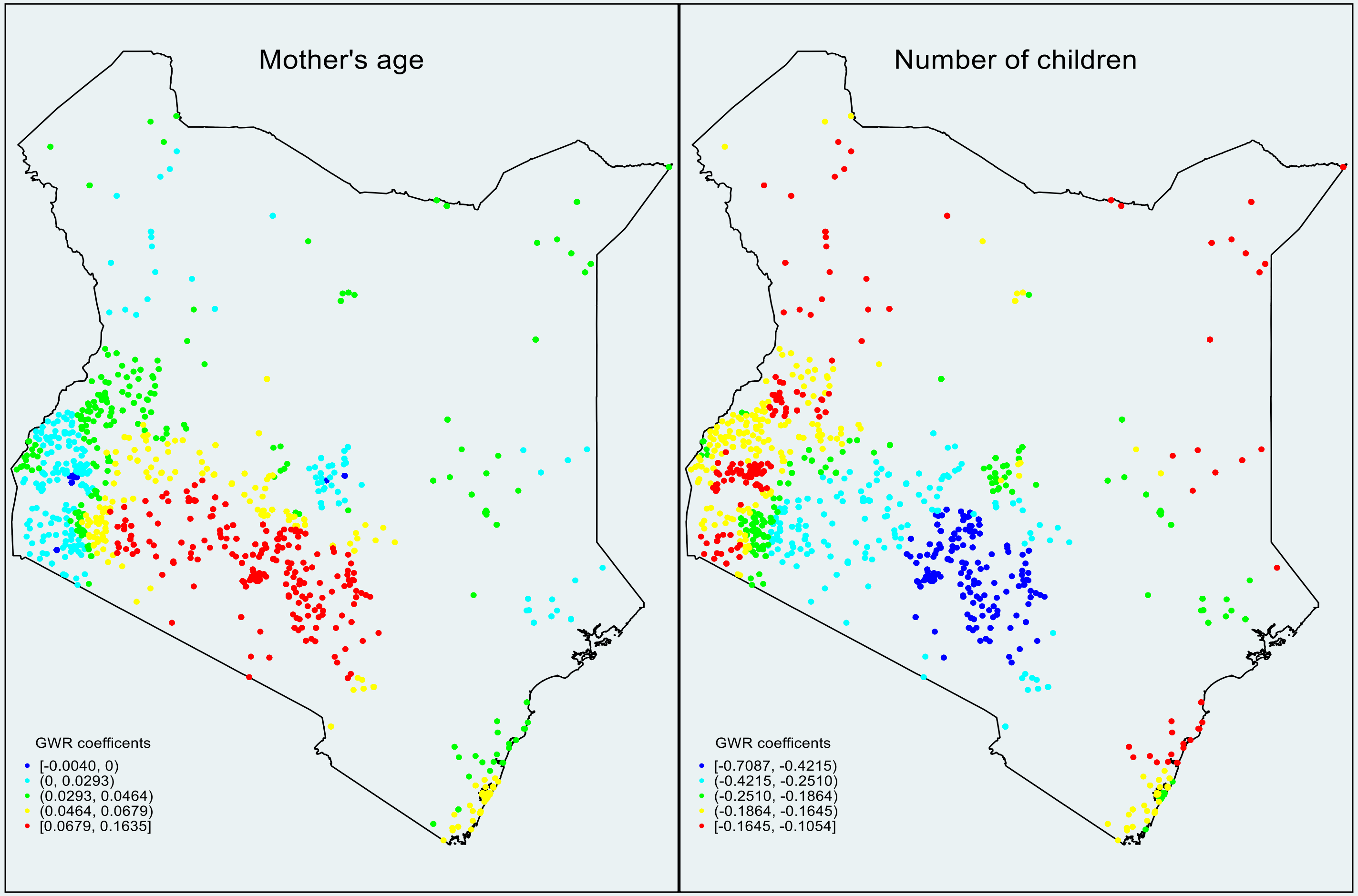
Source: Own calculation using GWR estimates from RStudio and data from KMHFL, DHIS2, Google Earth and ArcGIS

Figure A7.2: Spatial distribution of the effects of place of residence on utilisation of deliveries assisted by a skilled provider



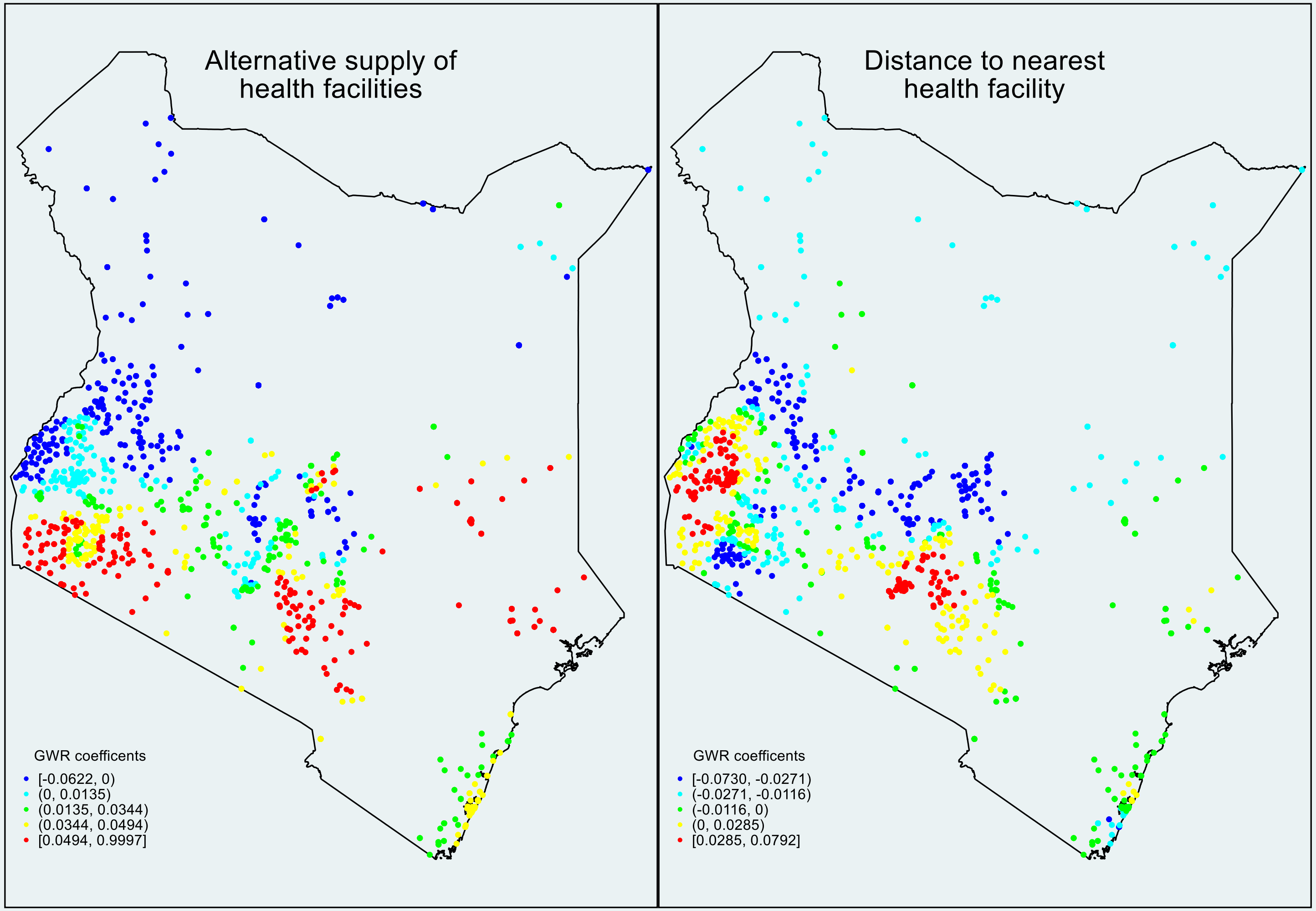
Source: Own calculation using GWR estimates from RStudio and data from KMHFL, DHIS2, Google Earth and ArcGIS

Figure A7.3: Spatial distribution of the effects of the number of children and mother's age on utilisation of deliveries by a skilled provider



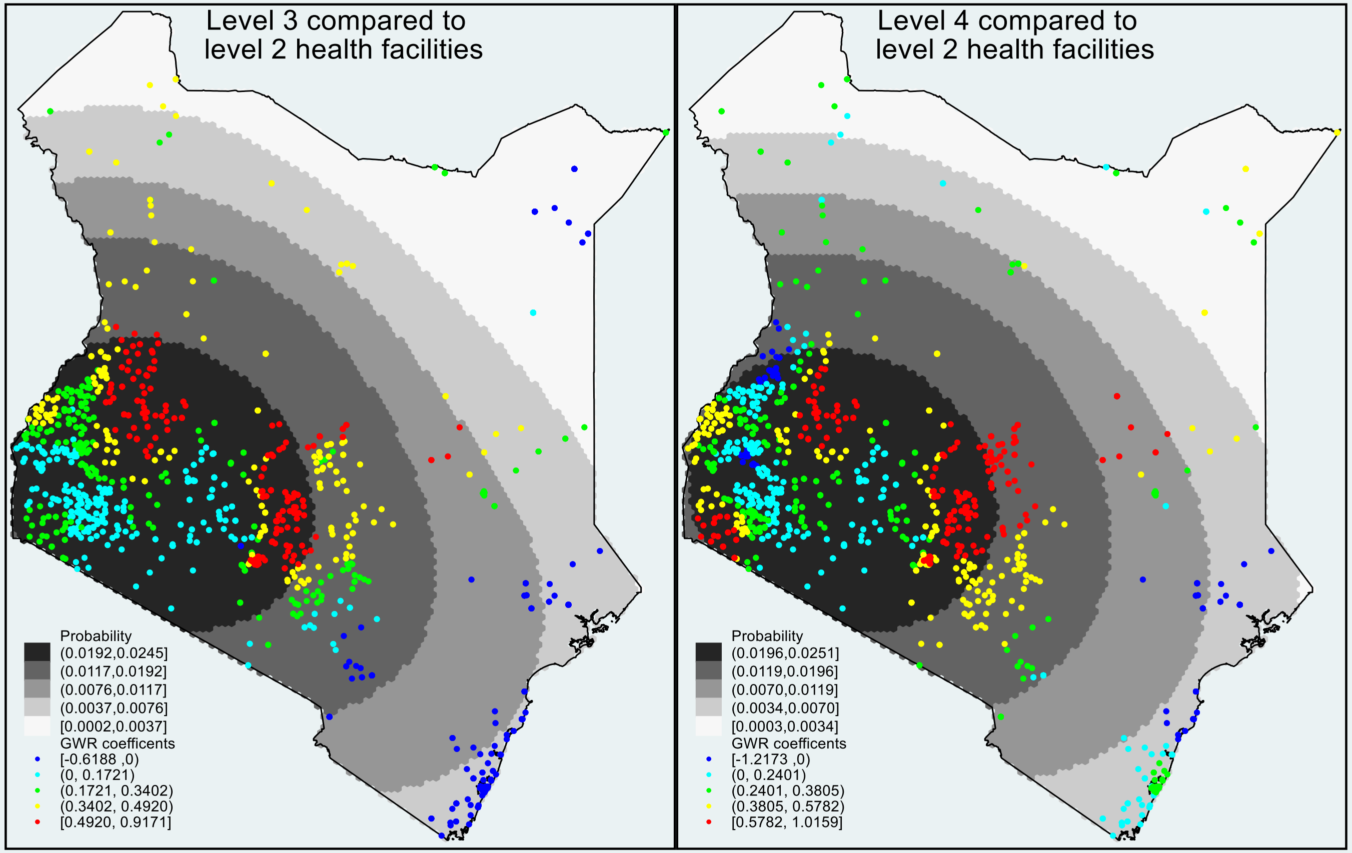
Source: Own calculation using GWR estimates from RStudio and data from KMHFL, DHIS2, Google Earth and ArcGIS

Figure A7.4: Spatial distribution of the effects of alternative supply and distance on utilisation of deliveries by a skilled provider



Source: Own calculation using GWR estimates from RStudio and data from KMHFL, DHIS2, Google Earth and ArcGIS

Figure A7.5: Spatial distribution of the effects of proximity to level 3 and level 4 compared to level 2 health facilities on utilisation of deliveries by a skilled provider



Source: Own calculation using GWR estimates from RStudio and data from KMHFL, DHIS2, Google Earth and ArcGIS

Figure A8: Correlation coefficients for covariates explaining deliveries assisted by a skilled provider and the underlying covariate characteristics