Hindawi
Journal of Probability and Statistics
Volume 2022, Article ID 2706504, 10 pages
https://doi.org/10.1155/2022/2706504

Research Article

Tweedie Model for Predicting Factors Associated with Distance Traveled to Access Inpatient Services in Kenya

Ngugi Mwenda[1], **Mathew Kosgei**[1], **Gregory Kerich**[1], and **Ruth Nduati**[2]

[1]*Moi University, Eldoret, Kenya*
[2]*University of Nairobi, Nairobi, Kenya*

Correspondence should be addressed to Ngugi Mwenda; geeglm2020@gmail.com

Received 20 December 2021; Accepted 31 March 2022; Published 19 April 2022

Academic Editor: Xiaofeng Zong

Aim. This study aims to examine which factors influence the distance traveled by patients for inpatient care in Kenya. *Methods*. We used data from the fourth round of the Kenya Household Health Expenditure and Utilization survey. Our dependent variable was the self-reported distance traveled by patients to access inpatient care at public health facilities. As the clustered data were correlated, we used the generalized estimating equations approach with an exchangeable correlation under a Tweedie distribution. To select the best-fit covariates for predicting distance, we adopted a variable selection technique using the QIC_u and R^2 criteria, wherein the lowest (highest) value for the former (latter) is preferred. *Results*. Using data for 451 participants from 47 counties, we found that three-fifths were admitted between 1 and 5 days, two-thirds resided in rural areas, and 90% were satisfied with the facilities' service. Wealth quintiles were evenly distributed across respondents. Most admissions (81%) comprised < 15, > 65, and 25–54 years. Many households were of medium size (4–6 members) and had low education level (48%), and nine-tenths had no access to insurance. While two-thirds reported employment-based income, the same number reported not having cash to pay for inpatient services; 6 out of 10 paid over 3000 KES. Thus, differences in employment, ability to pay, and household size influence the distance traveled to access government healthcare facilities in Kenya. *Interpretation*. Low-income individuals more likely have large households and live in rural areas and, thus, are forced to travel farther to access inpatient care. Unlike the unemployed, the employed may have better socioeconomic status and possibly live near inpatient healthcare facilities, thereby explaining their short distances to access these services. Policymakers must support equal access to inpatient services, prioritize rural areas, open job opportunities, and encourage smaller families.

1. Introduction

Inpatient care is a case wherein an individual is hospitalized for more than 24 hours and reflects a more serious health complaint. An estimated 1.2 million Kenyans required these services in 2013. This number is predicted to increase exponentially in the upcoming decade [1].

Among those seeking care, various factors are key in predicting the distance traveled to access healthcare facilities. For example, differences in wealth may determine the distance traveled. Wealthier individuals are more likely to possess the capabilities to access any facility of their choice, compared to those in the lower quintiles. The former can better afford to pay the healthcare costs through out-of-pocket or pay insurance premiums that can assist in securing admission at the required facilities in case of inpatient care need. In contrast, those in the lower-income quintiles have fewer options regarding the type of healthcare facility they can access due to financial constraints.

Government healthcare facilities have previously been a choice for many who are seeking inpatient care because they offer the needed care at affordable cost. However, despite being preferred among the majority population, such inpatient facilities precisely for inpatient care remain inaccessible [2].

To mitigate this, the Kenyan government has made efforts to increase the number of inpatient care facilities. This includes building new hospitals or upgrading the lower cadre facilities that are currently offering outpatient care. This is achieved by purchasing equipment required for operating an inpatient service, as well as deploying personnel by hiring more medical and nonmedical staff to manage these facilities [3].

For a developing country like Kenya with a struggling economy and whose health budget, compared to the total government revenue (7%) [4], is still slightly below halfway in meeting the 15% Abuja declaration [5], this implies that the number of facilities will continue to decrease. One major challenge is expensive and sophisticated medical equipment that is often not locally available and must be imported [6]. Another challenge is ensuring that the machines get uninterrupted electricity and having a good road network to access such healthcare facilities [7].

A pandora box opens when we explore electricity distribution in Kenya: the latest numbers from the Kenya Power and Lighting Company [8] indicate that more urban areas (77.6%) have electricity connections than rural areas (48.36%). Moreover, urbanites have better road networks than ruralites [7]. Conversely, the rural population (15 million) was twice that of the urban (33 million), according to the 2019 Census [9]. This implies that road networks and electricity coverage in urban areas are denser than those in rural areas. Therefore, inpatient facilities are more likely to favor urban than rural areas [6]. Thus, ruralites will be forced to travel to urban areas to access the required healthcare services.

Facility categorization in Kenya is based on the type of service they provide, with those offering inpatient care categorized from levels 3 to 6 [10]. According to the 2020 Economic Survey report, there was only a 3.2% increase in the number of government-owned facilities with inpatient care from 2015 to 2019 [3], compared to the approximately 10% increase in the population (5 million Kenyans) within the same period [11]. Thus, the facilities are seemingly already overstretched. In fact, according to the 2019 Kenya Master Health Facility List, the national average inpatient bed density was 13.3, or half of the targeted density of 25 [12].

Long distances remain the most prominent physical obstacle that has led to the underutilization of inpatient care services in the country. Data suggest low access to any physical health facility coverage in Kenya at one facility per 50–200 km [13]. This has been accompanied by an increase in poor health outcome: more hospitalization days in a facility due to an increase in the severity of a condition and nonattendance at follow-ups [14], including at fatalities [15]. The influence of distance on access in such scenarios includes the following aspects: (1) the extensive distance to be traveled to access a facility itself and (2) extensive distances being costly for a patient in need.

Distance, in general, when accessing healthcare facilities has been investigated [16, 17]. However, the influence of distance in accessing inpatient care is unclear. Through a literature review, we found studies focusing on women seeking delivery services and children below the age of five years that need hospital admissions. Notably, there has been an increase in the number of women seeking such services (46% delivered within a public facility), with 6 out of 10 availing the services at the facilities. This reduced maternal mortality rate (362 maternal deaths per 100,000 live births) and neonatal mortality (21 deaths per 1000 live births) in Kenya in 2014. 22% of the women aged 15–49 reported that distance was the most serious problem in accessing health care [18].

Studies on distances for inpatient care have found that mothers in a rural coastal area in the country were still delivering at home due to long distances to access inpatient facilities [19]. The same was observed in a rural western setting in another study [20]. These sets of authors reported that long distances to facilities were a hindrance to access.

Kukla et al. [21] examined children suffering from diarrheal illness and requiring inpatient care in Kenya and found out that there were challenges in accessing public facilities due to long distances and high transport cost. This implies that, for those living far away from public facilities, their access to care is determined by their financial capabilities.

Mochida et al. [22] investigated access for inpatient care for children suffering from malaria, the most lethal killer disease in Sub-Saharan Africa [23]; the delay in seeking treatment affects disease management, progression, and outcomes. The authors found that distance played a critical role in deciding on whether to utilize a healthcare service. Further, they established that those who need to travel longer distances were less likely to use the facilities. This situation contributed to morbidities and worse-case mortalities among children. Such studies provide evidence of a distance decay in the utilization of healthcare facilities [24], with those far away associated with under-use and those closer with better use.

In summary, most studies have focused on mothers' and children's access to healthcare facilities and others focused on general access, combining both inpatient and outpatient care. Thus, a research gap exists on inpatient access to public facilities in the general population in Kenya.

2. Materials and Methods

2.1. Methodology. Tweedie distributions are members of the exponential dispersion models (EDMs) [25] defined as follows:

$$f(y|\theta,\phi) = c(y,\phi)\exp\left\{\frac{y\theta-\gamma(\theta)}{\phi}\right\}, \tag{1}$$

where $\phi > 0$ is the dispersion parameter and θ is the natural parameter. The mean and variance for the random variable Y are given by $E(Y) = \mu = \gamma'(\theta)$ and $\operatorname{Var}(y) = \gamma''(\theta)\phi$, respectively. Tweedies, as a special case of EDMs, have a mean-power variance relationship expressed as $\operatorname{Var}(\mu) = \phi\mu^p$ such that $p \notin (0,1)$.

Tweedie distributions have been widely applied in modeling nonnormal response data with a discrete mass of zero because they can incorporate skewness without any data transformation. Most methods suggested in the literature for the analysis of such data mainly consider data transformation [26], two-way analysis [27], and Bayesian methods [28].

However, these methods are not efficient for our approach because of the correlational nature of our data within clusters. Clustered data naturally have depedence due to the reason that they share common property [29]; for example, people who come from the same region have shared culture and traditions and mostly share common medical facilities.

Approaches for analyzing nonnormal data in the generalized linear model framework may ignore the correlation among subjects who belong to the same cluster. Our data were clustered by counties; thus, an exchangeable correlation structure is the most flexible in modeling the association through the generalized estimation equations (GEE) approach. Notably, both independent [30] and autoregressive (AR) (1) [31, 32] correlation structures under nonnormal responses have been considered. The independent structure assumes a weak correlation among the subjects in a cluster, and the number of clusters is small. Meanwhile, the AR (1) assumes a time dependence order for association in a repeated measure. Such measurements taken far apart have a weaker correlation compared to those that are taken closer.

Previously mentioned evidence alluded to the influence of covariates on distance to healthcare facilities. However, a research gap on the selection of the best-fit covariates remains. This study aims to determine the combination of covariates that influence the distance that a Kenyan citizen will travel to seek inpatient care. This also extends the application of the Tweedie distribution to understand the influence of a given set of covariates on distance.

2.2. Data. A cross-sectional survey to determine healthcare expenditure for both inpatient and outpatient for Kenyans was conducted from 9 April to 19 May 2018 across the 47 counties. A multistage sampling design was employed such that the stratified cluster sampling selected 1500 clusters spread evenly through the country. In the second stage, 25 households were uniformly sampled within each cluster, with 527 (923) located in urban (rural) areas.

The Kenya Household Health Expenditure and Utilization Survey (KHHEUS) questionnaire [2] was designed by a technical working committee comprising members drawn from the Ministry of Health (MoH), the World Bank (WB), and the Kenya National Bureau of Statistics (KNBS). This was then administered to all sampled household through computer-assisted personal interviews designed using CSpro software [33].

A total of 37,500 households were sampled, with an overall response rate of 95%. There were 4822 individuals who sought inpatient care in the country, with 55.49% seeking care in government-owned facilities. Complete information was available for only 451 individuals and is used here. The missing mechanism was mostly Missing Not at Random (MNAR); therefore, no imputation was performed to the missing values as it may result in biased estimates [34].

Inpatient health care is defined as services provided to patients admitted at least for 24 hours in a facility following a decision by a doctor [2]. The main causes of hospital admissions, in descending order, were the following: malaria/fever (14.1%), normal delivery (9%), surgery (7.1%), pneumonia (5.3%), accidents and injuries (4.8%), hypertension (4.5%), caesarean delivery (4.1%), and diarrhea (3.8%).

2.3. Selection of Variables. The inpatient data from the KHHEUS were obtained, cleaned, and coded. We investigated several covariates against our dependent variable, that is, the distance to a health facility traveled by Kenyans seeking inpatient services. Variable selection was based on the descriptive report of KHHEUS 2018 [2] and the Demographic and Health Survey 2014 [18], as well as published literature [17, 35–41].

2.4. Dependent Variable. Our dependent variable was the distance traveled by people to access inpatient care facilities. The values ranged from minimum of 0 km (for those living within reach of the facilities) up to a maximum of 700 km (those living far from the facilities). Zero, which is a nonnegative response, is often common because of nontravelers, while the positive responses, which are right skewed, are due to users.

2.5. Independent Variables. To predict the distance traveled for inpatient care, we considered days admitted to the facility: few (1–5), medium (6–20), and more (21 and above); place of residence; and rural or urban. Meanwhile, we computed five wealth index quintiles ranging from richest to poorest constructed from the ownership of different household assets using the principal component analysis, as described by Filmer and Pritchett [42].

Education level was categorized using the methods from Rippin et al. [43]: those who never went to school (those under 3 years of age and who responded "Never went to school") or have low-level education (pre-primary, primary, and informal (madrassa)), intermediate-level education (secondary, vocational, and college), and higher-level education (university degree or higher). Age groupings for employment followed those defined by the Organisation for Economic Cooperation and Development (OECD): those aged below 15 and above 65 who are considered as unable to work; those aged 15–24 who are entering the labor market after an education; those aged 25–54 who are in their prime working years; and those aged 55–64 who are passing the peak of their career and approaching retirement [44].

We divided household sizes into three categories: small (1–3 members), medium (4–6 members), and large (7 + members). Access to insurance was classified as yes or no, while employment status (i.e., formal, informal/casuals, or self-employed) was classified as either employed or unemployed.

Amounts paid for healthcare were categorized into three groups: low (1–3,000 KES), medium (3,001–10,000 KES), and high (10,000 + KES). The cost included the amount paid for the service received at a facility. Further, satisfaction was categorized into two: satisfied and not satisfied. Finally, if they had cash to pay for the service, it was categorized into either yes or no.

2.6. Tweedie Distributions in Modeling Distance for Inpatient Care. We used a Tweedie distribution to model the predictors of distance for inpatient care. The justification to use the Tweedie distribution is provided in Figure 1: distance, as the dependent variable, has several characteristics that are interesting to model including (1) positive skewness and nonnormality of the distance, (2) the range of positive distance, and (3) excess outliers for those who covered more distance than the average (Table 1).

We analyzed the distance with the covariates to determine the best combinations to explain any existing association.

Let R_i be the distance recorded for a Kenyan traveling to seek inpatient care. We can assume that the distances traveled within a county during the survey period N follow a Poisson distribution with mean λ so that the following holds: if there is no distance covered, then $N = 0$; Y represents total distance covered and equals the Poisson sum of the gamma random variables such that $Y = R_1 +, \ldots, +R_N$. The resulting distribution is called the Poisson gamma and is explained as follows.

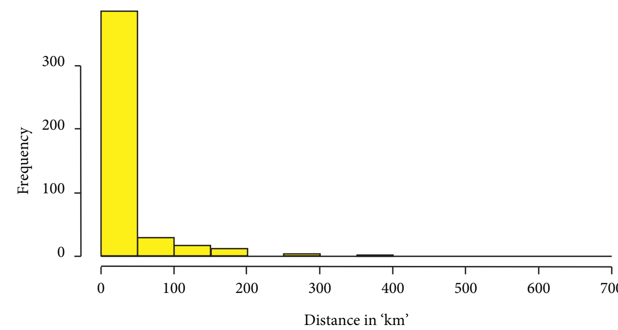

Figure 1: Histogram of distance traveled to seek inpatient care in Kenya.

Table 1: Descriptive analysis of distance for inpatient care.

Statistic	Distance traveled (km)
Minimum	0
1st quarter	3
Median	10
Mean	32.1
3rd quarter	30
Maximum	700
Shape	Right skewed
Skewness	4.8

2.7. Notations. Let $\mathbf{y}_{\text{it}}$ be a vector of responses with a set of corresponding r covariates $\mathbf{X}_{\text{it}}$, where i indexes K units of analysis $i = 1, 2, \ldots, K$; and t indexes the time points $t = 1, 2, \ldots, n_i$ for each unit. Thus, the number of clusters observed is K. Further, $N = \sum n_i$ and is the total number of observations across all units. For inclusion of an intercept, the first element of $\mathbf{X}_{\text{it}}$ is set to 1.

Further, let $\mathbf{y}_i = [y_{i1}, y_{i2}, \ldots, y_{\text{in}_i}]$ denote the corresponding column vector of observations on the response variables for unit i and $\mathbf{X}_i = X_{i1}, X_{i2}, \ldots, X_{\text{in}_i}$ indicate the $n_i \times r$ matrix covariates for unit i.

In our case application to model distance data, we define the following:

(i) The response variable y_{it} which is the distance traveled in a given county by any given individual seeking inpatient care.

(ii) With data available for all 47 counties, then $K = 47$ for $i = 1, \ldots, 47$.

(iii) The linear predictor.

(iv) A link function used to relate the response to the linear combination of the covariates as $g(\mu_{\text{it}}) = \eta_{\text{it}}$.

2.8. Exchangeable or Symmetrical Correlation for Imbalanced Data. Accounting for correlation in any model is a key principle to avoid biased results. Observations from an individual tend to be correlated when referring to people in the same cluster. Therefore, to let our results be more efficient, we need to specify the correlation structure described by Liang and Zeger [45]. Our preferred correlation structure is the exchangeable or symmetrical correlation structure which assumes that there is a common correlation $(R_i(\alpha))$ within the observation in a given county, such that $R_{u,v} = \begin{cases} 1 & u = v \\ \alpha & \text{otherwise} \end{cases}$. Following Hardin and Hilbe [46], the exchangeable correlation structure uses the Pearson residuals from the current fit of the model to estimate the exchangeable correlation parameter, $\widehat{r}_{\text{it}} = (y_{\text{it}} - \widehat{\mu}_{\text{it}})/\sqrt{V(\widehat{\mu}_{\text{it}})}$ with α estimated using

$$\widehat{\alpha} = \frac{1}{\widehat{\phi}} \sum_{i=1}^{n} \left\{ \frac{\sum_{u=1}^{n_i} \sum_{v=1}^{n_i} \widehat{r}_{iu} \widehat{r}_{iv} - \sum_{u=1}^{n_i} \widehat{r}_{iu}^2}{n_i (n_i - 1)} \right\}. \quad (2)$$

An important property of Tweedie under GEE is its ability to accommodate both the correlation and the right skewness, which is a characteristic of our continuous data. We use this approach here. This complements the work of Swan [31], who used the AR (1) correlation structure. Tweedie regression models allow relation of the mean of distance to the selected covariates. This allows the mean of distance to be modeled as a linear function of covariates using the log link given by

$$\log(\mu_i) = \beta_0 + \beta_1 x_i +, \ldots, +\beta_n x_n, \quad (3)$$

where β_j vectors are regression coefficients that correspond to the x_j vectors of covariates, all fitted based on Tweedie EDM. The means are calculated to assess the relationship between covariates and the distance.

To fit the models, we need to estimate the index parameter p and the βs from the Tweedie distribution using the GLM framework. This can be computationally difficult, but the R packages Tweedie [47] and statmod [48] facilitate a flexible fit. The estimated index parameter with the 95% CI, using a set of covariates by employing cubic spline smoothing interpolation, is shown in Figure 2.

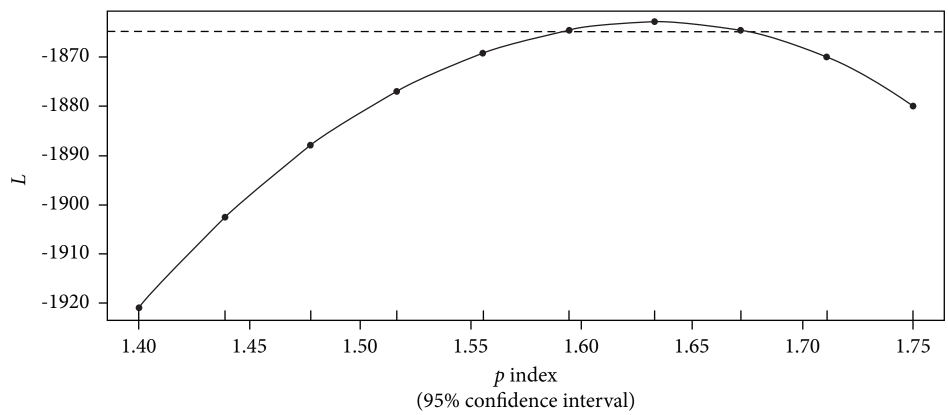

Figure 2: The profile log-likelihood plots for the distance traveled for inpatient care in Kenya using the gender of the household head, household size, and education as covariates. The points on the line show the computed likelihood values for different p estimates, while the solid line represents the cubic-spline smooth interpolation through the points. This plot was used to estimate the maximum likelihood value of $\hat{p}$ as 1.63 with the 95% confidence interval (C.I.) between 1.59 and 1.67.

To estimate the βs, we apply the approach of McLuggar and Nelder [49] referred to as the iterative reweighted least square (IRLS) method in the GLMs. These β s are then updated under the GEE framework.

Following Hardin and Hilbe [46], the quasi-likelihood estimating equations for GLM without any preference for mean and variance function with 47 clusters (counties) are expressed as follows:

$$\Psi(\beta) = \sum_{i=1}^{n} \left\{ x'_{ik} T \frac{1}{V(\mu_i)} \frac{y_i - \mu_i}{a(\phi)} \right\}, \tag{4}$$

for $k = 1, 2, \ldots, 47$, where T is a diagonal matrix of derivatives $\partial\mu_i/\partial\eta_i$.

$V(\mu_i)$ is an $n_i \times n_i$ diagonal matrix which can be decomposed into

$$V(\mu_i) = T\left[V(\mu_i)^{1/2} I_{(n_i \times n_i)} V(\mu_i)^{1/2}\right]. \tag{5}$$

The observations in the clusters are treated as independent, but our focus is on the population average for which both the mean and the variance are averaged over all clusters. Following Liang and Zeger [45], the identity matrix in the above is replaced with a more general correlation matrix, since the variance matrix for data which are correlated does not have a diagonal form, as follows:

$$V(\mu_i) = T\left[V(\mu_i)^{1/2}\right] CM_i(\alpha) T\left[V(\mu_i)^{1/2}\right], \tag{6}$$

where the correlation matrix $CM_i(\alpha)$ is estimated by vector α. Proper specification of the $CM_i(\alpha)$ means that the $\hat{\beta}$ are consistent and asymptotically normal.

To achieve more efficiency, it is necessary to include a hypothesized correlation structure within the clusters. The structure that is suitable for our clustered data without any time dependence is the exchangeable given by the following:

$$CM_{ik}(\alpha) = \begin{cases} 1 & \text{if } i = k \\ \alpha & \text{if } i \neq k \end{cases}.$$

We explored the regression model fit to select the best model using the quasi-likelihood based information criterion (QIC_u) proposed by Hardin and Hilbe [46], which is an extension of the QIC proposed by Pan [50].

The QIC_u imposes a penalty based on model complexity to ensure that only a few covariates are used to achieve model parsimony. This criterion is used as a guide for selecting models when no scientific knowledge is available for the researcher to rely on. Once the values are computed for different models, the one with the lowest values is usually considered.

Finally, Table 2 shows the deviance and degrees of freedom for the residuals, considering logarithmic and canonical link functions. The logarithm link function produces the lowest residual deviance (5251.198) against canonical deviance (5280.799). This indicates that it was the most suitable link function to use when modeling distance to access inpatient care in Kenya.

R statistical software version 3.6.3 [51] was used in the analysis.

3. Results

Exploratory analysis of the data is presented in Table 3, from which we select the covariates that predict distance traveled to access inpatient care. Overall, complete information was available for 451 patients across 47 counties in Kenya who visited public health facilities in 2018 and were admitted more than 24 hours.

Furthermore, 3 out of 5 patients were admitted between 1 and 5 days, two-thirds resided in rural areas, and 90% were satisfied with the service received at the facility. The wealth quintiles were evenly distributed across the respondents. Most admissions (81%) comprised < 15, > 65, and 25–54 years. The majority of households were of medium size (4–6 members) and had lower education (48%), while 9 out of 10 did not have access to insurance.

Two-thirds of members reported income through employment. The same number reported not having cash available to pay for the inpatient service, with 6 out of 10 paying over 3000 KES for the hospital bills and more than

Table 2: The residual deviance and degrees of freedom for a Tweedie GLM with differing link functions for the selected model.

Link function	Deviance	DF
Logarithm	5251.198	455
Canonical	5280.799	455

Table 3: Exploratory data analysis of the covariates considered in the analysis.

Variable	Number	Percentage
Days admitted		
1–5 days	258	57.21
6–20 days	120	26.61
21 days and above	73	16.19
Residence		
Rural	299	66.30
Urban	152	33.70
Satisfied		
Yes	402	89.14
No	49	10.86
Wealth quintiles		
Poorest	130	28.82
Poor	111	24.61
Middle	106	23.50
Rich	66	14.63
Richest	38	8.43
Age group		
< 15 and > 65	140	31.04
15–24	20	4.43
25–54	228	50.55
55–64	63	13.97
Household size group		
1–3 members (small)	179	39.69
4–6 members (medium)	186	41.24
7 + members (large)	86	19.07
School category		
Never went to school	140	31.04
Lower education	216	47.89
Intermediate education	89	19.73
Higher education	6	1.33
Employment category		
Employed	307	68.07
Not employed	144	31.93
Access to health insurance		
Yes	58	12.86
No	393	87.14
Had cash		
Yes	148	32.82
No	303	67.18
Paid category		
1–3000 KES (low)	190	42.13
3001–10,000 KES (medium)	137	30.38
Above 10,0001 (high)	124	27.49
Facility admitted		
Levels 5 and 6	57	12.63
Level 4	356	78.93
Level 3	38	8.42

three-quarters of all inpatient cases sought services in level 4 hospitals.

Table 4 presents 10 competing models for distance that demonstrates the best-fitting model with the lowest QIC_u. We added the covariates into the model and computed their QIC_u and R^2. We then removed the covariates one by one and checked whether the changes improved the model fit.

Model 7, with covariates ability to pay, employed, and household size, was selected as the most parsimonious as it had the lowest value for QIC_u. Considering the intercept, the model shows that, on average, a Kenyan seeking inpatient care traveled 22.04 km in 2018. The model with covariates of choice can be expressed as follows:

$$\begin{aligned}\log(\mu) &= 3.093 + 0.222\text{paidMed} + 1.226\text{paidHigh} \\ &\quad - 0.523\text{Employed} + 0.092\text{hMed} + 0.471\text{hLarge},\end{aligned} \tag{7}$$

where μ is the expected distance traveled to access inpatient care. In this model, the amount paid for healthcare (medium and high) equals 0 or 1 depending on which category is being assessed. The low group is the reference category. Employment equals 1 if the respondent is employed. Finally, household size (medium and large) equals 0 or 1 depending on what is being assessed. This model resulted in $\phi = 6.12$, $\alpha = 0.045$, $R^2 = 9.96\%$, and $QIC_u = 13158.23$, with $p = 1.64$, 95% CI (1.59, 1.68). To interpret the coefficients which are captured in logarithmic form, we need to take the exponential.

Compared to those who paid the lowest amounts for healthcare (1–3,000 KES), citizens in the middle pay category (3,001–10,000 KES) traveled 1.24 times the distance to a healthcare facility, whereas those who paid the most traveled 3.40 times the distance. The employed traveled half the distance to a healthcare facility for inpatient care than the unemployed (0.59 times). Compared to small households (1–3 members), medium-sized households (4–6 members) traveled 1.096 times the distance to a healthcare facility, while the largest (7+) traveled 1.60 times the distance compared to small households.

We calculated probabilities based on Dunn and Smyth [52] to demonstrate the applicability of the Tweedie distributions in modeling the distance traveled to access inpatient care. When $1 < p < 2$, then the Tweedie parameters (μ, p, ϕ) can be parameterized to Poisson and gamma parameters $(\lambda, \gamma, \alpha)$ which we use for estimation. These are given in the following equation:

$$\lambda = \frac{\mu^{(2-p)}}{\phi(2-p)}, \tag{8}$$

$$\gamma = \phi(p-1)\mu^{(p-1)}, \tag{9}$$

$$\alpha = \frac{(p-2)}{(1-p)}, \tag{10}$$

where λ is the average distance traveled, γ is the shape, and $\alpha\gamma$ is the mean.

Considering our best fitting model, the parameter index p is 1.64, $\mu = 22.04$, and $\phi = 6.12$. When we reparametrize to gamma and Poisson, it gives the predicted mean cost spend per month calculated as

Table 4: Model selection using QIC_u and R^2.

Model number	Covariates	QIC_u	R^2	Variance power p(95% CI)	Number of covariates
10	Ability to pay, employed, household size, wealth index, education level, age	13304.7	10.39	1.63 (1.58, 1.67)	6
9	Ability to pay, employed, household size, wealth index, education level	13306.16	10.41	1.63 (1.59, 1.67)	5
8	Ability to pay, employed, household size, wealth index	13317.38	9.7	1.62 (1.59, 1.67)	4
7	**Ability to pay, employed, household size**	**13158.23**	**9.96**	**1.64 (1.54, 1.68)**	**3**
6	Ability to pay, employed	13280.7	9.5	1.64 (1.59, 1.68)	2
5	Have insurance, place of residence	12733.1	0.19	1.67 (1.63, 1.71)	2
4	Ability to pay	13066.33	8.38	1.64 (1.60, 1.68)	1
3	Place of residence	12773.2	0.17	1.67 (1.63, 1.71)	1
2	Household size	12755.65	0.54	1.67 (1.63, 1.71)	1
1	Employed	12698.31	1.4	1.67 (1.63, 1.71)	1

$$\lambda = \frac{22.04^{(2-1.64)}}{6.12\,(2-1.64)} = 1.38, \tag{11}$$

and

$$\gamma = 6.12\,(1.64-1)22.04^{(1.64-1)} = 28. \tag{12}$$

Finally,

$$\alpha = \frac{1.64-2}{1-1.64} = 0.56. \tag{13}$$

The mean distance to access inpatient care is $\alpha\gamma = 0.56 * 28 = 15.75$.

Following Dunn and Smyth [52], the probability of incurring zero cost on outpatient care by the household (in other words, the probability of not seeking outpatient care) is given by the following:

$$\Pr(Y=0) = \exp(-\lambda) = \exp\left[-\frac{\mu^{2-p}}{\phi\,(2-p)}\right]. \tag{14}$$

The probability of households covering zero distance is given by $\exp(-1.38) = 0.25$, meaning that 25% of household is within reach to an inpatient facility.

4. Discussion

This study identified factors associated with the distance travelled for inpatient care in Kenya. The results indicate that only one-quarter (25%) of the Kenyan population is within reach of inpatient services. Healthcare costs at the facility, household sizes, and being employed were predictors of distances. These results should be interpreted with caution because the data were MNAR, wherein no sufficient imputation method would be appropriate for imputing such a problem [34].

The amount paid at the facility can be a good indicator for how far a person traveled for inpatient care. Here, higher costs at a facility were associated with longer distances traveled. This implies seeking a service that may not be available at a closer facility. When a healthcare provider at a low-level hospital refers a patient to a higher-level hospital due to a lack of skills or equipment, this is referred to as a referral [53]. Complex cases call for referrals, and the higher facilities are few and typically found in major urban areas. A study in Asembo, Kenya, found that more children with severe illnesses traveled further [24] as facilities to treat their illness were not available.

Most complex cases become worse with time. This implies that if a case is handled early before it progresses, then it can be averted. A case example is cervical cancer, which is preventable through vaccination and screening once detected early [54, 55]. However, acquiring diagnostics and treatments in Kenya is impossible due to long travel distances [56]. This indicates that conditions may deteriorate, cost more in later treatment, and lead to impoverishment of the already poor households [57]. In Kenya, most referrals to far facilities were observed in rural (10.4%) versus urban areas (7.1%) [2]. This is because the facilities that can handle these complexities are mostly found in urban areas, wherein infrastructure such as roads [7] and services like electricity [6] exist.

However, being employed helps reduce the hindrance to access inpatient service, similar to the finding by Allin et al. [58]. The majority who have jobs likely live in urban areas or places where social amenities such as water and electricity are typically available [59]. In these areas, one may expect inpatient facilities. Studies have shown that health insurance, which cushions people against catastrophic spending, is mostly acquired by the employed. This indicates that the employed enjoy the double benefits of living close to the facilities and incurring minimum expenses when they seek care. They can also afford the fee as they have income, which indirectly influences access to inpatient care as reported in Kenya [19, 20], Ghana [60], Zambia, and Malawi [61].

Members from large families are more likely to travel long distances to access inpatient health facilities. These facilities are mostly found in urban areas [62, 63]. However, large families are more associated with poverty [64, 65], which influences healthcare utilization due to access capability. Further studies have shown that large family sizes are poor purchasers of insurance [66], thereby exposing large families to catastrophic spending.

Healthcare system performance can be assessed according to the healthcare service distribution, access, and utilization [67]. Access is mostly determined by cost and distance. Thus, irrespective of the availability of a service in a hospital, if a service is not utilized by the target group, its full utility cannot be actualized. Kenya's health system continues to suffer due to inadequate infrastructure and personnel in its healthcare system. Due to an insufficient number of personnel [68], specialized doctors from Cuba [69] were invited to fill in this gap in main county referral hospitals (level 4). This will strengthen primary health care and facilitates achieving Universal Health Care (UHC). However, these facilities may not fully meet the need of the majority as they are located in major towns; thus, those who are very far have limited access. This is specific for the rural population as reported by similar studies [70, 71].

Strengthening a healthcare system is a step-by-step process which requires a good framework provided by the government through the Ministry of Health [12]. The aim is to accelerate the achievement of UHC by providing a structure to ensure that infrastructure will be improved and increase the number of healthcare personnel to serve the entire population. In addition, an improvement in the health structural development was reported by [72, 73], who pointed out proper management at lower levels through devolution. However, they also cited challenges emanating from sources of finance. Thus, if UHC has to be attained anytime soon, then funding should not be an issue.

One of the key aspects toward strengthening the healthcare system is upgrading lower-level facilities and providing them with materials necessary to offer inpatient services. Admission by facilities that are more able to offer outpatient care has been reported, such as level 3 hospitals [2] wherein 8.42% were admitted in 2018. Although they are not able to handle extreme cases, they are well-equipped to handle many cases such as childbirth. This is likely to have a positive change in terms of facility deliveries and, thus, reduce the maternal and neonatal mortality rates. However, during childbirth complications, such as the need for caesarean services, patients may need to be referred to a larger facility. Therefore, major conditions still need referrals to large hospitals, and access for inpatient care for these services at lower-level hospitals remains a challenge.

A reasonable distribution of health facilities at a facility density of 2.2 per 10,000 population, which is above the World Health Organization's target of 2 per 10,000 [12], has been found as well. However, one may still argue that these facilities are out of reach for the majority of Kenyans. Therefore, the distance issue must be addressed if we ever hope to achieve Sustainable Development Goal 3.

A drawback of previous studies is the analysis of distance using summary statistics because researchers did not dig deeper into the data and only reported averages. Using Table 1 as a guide, for example, we could have reported a median of 10 km, which may be misleading, as we did not factor in skewness and correlations that exist in the data, which could provide more insight. This shows that advanced statistical analysis provides a more meaningful interpretation of the data by factoring in both skewness and correlations.

Finally, regarding policy implications, policies that target having more government healthcare facilities in rural areas and slums should be undertaken because large populations without adequate access to healthcare exist in these areas. However, setting up these facilities will require improved infrastructure (such as roads) and services (such as water and electricity), which are crucial components when setting up an inpatient facility.

In addition, policies that advocate smaller families should be encouraged to ensure that people can afford better healthcare. Job availability will also increase flexibility in terms of selecting a facility. Sophisticated services should be brought closer to low-income families to ensure that they do not travel long distances for much-needed services.

5. Limitations

This study has some limitations. First, the data were missing a significant amount of information, which made it difficult to use them as they were MNAR. Therefore, we only used complete data as opposed to the available imputation techniques. This aspect can be examined in future studies to determine if there may be any differences in the findings. Furthermore, we only considered public facilities in the analysis. We hope other researchers will consider private facilities and examine the differences in the findings. Lastly, our research does not provide for statistical significance of the variables, but rather focused on overall model selection. This is a limitation that other researchers can consider, to further research on analysis of nonnormal clustered data.

Data Availability

The datasets are freely available from the Kenya National Bureau of Statistics by registering at https://statistics.knbs.or.ke/nada. The authors confirm that others can access these data in the same manner as the authors and that they did not have any special access privileges. The Reproducible R code can be accessed in the GitHub repository https://github.com/samwenda/Tweedie-with-Exchengable-Correlation.

Ethical Approval

The study did not require any approvals as it was a secondary analysis.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

References

[1] MoH, *Kenya Household Health Expenditure and Utilization Survey*, Ministry of Health, Nairobi, Kenya, 2014.

[2] MoH, *Kenya Household Health Expenditure and Utilization Survey*, Ministry of Health, Nairobi, Kenya, 2018.

[3] KNBS, *Economic Survey*, Kenya National Bureau of Statistics, Nairobi, Kenya, 2020.

[4] S. Nganga, *Exploring the Fiscal Space of the Health Sector in Kenya*, KPMG East Africa, Nairobi, Kenya, 2021.

[5] WHO, *Abuja Declaration: Ten Years on*, Worla Health Organization, Geneva, Switzerland, 2010.

[6] D. R. Turin, “Health care utilization in the Kenyan health system: challenges and opportunities,” *Inquiries Journal/ Student Pulse*, vol. 2, 2010.

[7] P. M. Macharia, E. Mumo, and E. A. Okiro, “Modelling geographical accessibility to urban centres in Kenya in 2019,” *PLoS One*, vol. 16, no. 5, Article ID e0251624, 2021.

[8] K. Power, *Kenya Leads East Africa Peers in Access to Electricity*, Kenya Power and Lightning Company (KPLC), Nairobi, Kenya, 2018.

[9] KNBS, *2019 Kenya Population and Housing Census; Distribution of Population by Socio-Economic Characteristics*, Kenya National Bureau of Statistics, Nairobi, Kenya, 2019.

[10] A. Mariita, *Kenya's Health Structure and the Six Levels of Hospitals -RoGGKenya*, Transparency International Kenya, Nairobi, Kenya, 2019.

[11] KNBS, *2009 Kenya Population and Housing Census; Analytical Report on Population Projections*, Kenya National Bureau of Statistics, Nairobi, Kenya, 2012.

[12] MoH, *Kenya Harmonized Health Facility Assessment*, Ministry of Health, Nairobi, Kenya, 2019.

[13] MoH, *National Health Sector Strategic Plan of Kenya*, Ministry of Health, Nairobi, Kenya, 2005.

[14] C. Kelly, C. Hulme, T. Farragher, and G. Clarke, “Are differences in travel time or distance to healthcare for adults in global north countries associated with an impact on health outcomes? a systematic review,” *BMJ Open*, vol. 6, Article ID e013059, 2016.

[15] M. Karra, G. Fink, and D. Canning, “Facility distance and child mortality: a multi-country study of health facility access, service utilization, and child health outcomes,” *International Journal of Epidemiology*, vol. 46, no. 3, pp. 817–826, 2017.

[16] A. M. Noor, D. Zurovac, S. I. Hay, S. A. Ochola, and R. W. Snow, “Defining equity in physical access to clinical services using geographical information systems as part of malaria planning and monitoring in Kenya,” *Tropical Medicine and International Health*, vol. 8, no. 10, pp. 917–926, 2003.

[17] V. Escamilla, L. Calhoun, J. Winston, and I. S. Speizer, “The role of distance and quality on facility selection for maternal and child health services in urban Kenya,” *Journal of Urban Health*, vol. 95, no. 1, pp. 1–12, 2018.

[18] KNBS, *Kenya Demographic and Health Survey 2014*, Kenya National Bureau of Statistics, Rockville, MD, USA, 2015.

[19] R. O. Moindi, M. M. Ngari, V. C. S. Nyambati, and C. Mbakaya, “Why mothers still deliver at home: understanding factors associated with home deliveries and cultural practices in rural coastal Kenya, a cross-section study,” *BMC Public Health*, vol. 16, no. 1, p. 114, 2016.

[20] E. Mwaliko, R. Downing, W. O'Meara et al., ““Not too far to walk”: the influence of distance on place of delivery in a western Kenya health demographic surveillance system,” *BMC Health Services Research*, vol. 14, no. 1, p. 212, 2014.

[21] M. Kukla, N. McKay, R. Rheingans et al., “The effect of costs on Kenyan households' demand for medical care: why time and distance matter,” *Health Policy and Planning*, vol. 32, no. 10, pp. 1397–1406, 2017.

[22] K. Mochida, D. Nonaka, J. Wamulume, and J. Kobayashi, “Supply-side barriers to the use of public healthcare facilities for childhood illness care in rural zambia: a cross-sectional study linking data from a healthcare facility census to a household survey,” *International Journal of Environmental Research and Public Health*, vol. 18, no. 10, p. 5409, 2021.

[23] C. Kahabuka, G. Kvåle, and S. G. Hinderaker, “Care-seeking and management of common childhood illnesses in Tanzania—results from the 2010 demographic and health survey,” *PLoS One*, vol. 8, no. 3, Article ID e58789, 2013.

[24] D. R. Feikin, L. M. Nguyen, K. Adazu et al., “The impact of distance of residence from a peripheral health facility on pediatric health utilisation in rural western Kenya,” *Tropical Medicine & International Health*, vol. 14, no. 1, pp. 54–61, 2009.

[25] B. Jørgensen, *The Theory of Dispersion Models*, Chapman and Hall/CRC, Boca Raton, FL, USA, 1st edition, 1997.

[26] S. Manikandan, “Data transformation,” *Journal of Pharmacology and Pharmacotherapeutics*, vol. 1, no. 2, pp. 126-127, 2010.

[27] S. Su, N. C. Dzupire, P. Ngare, and L. Odongo, “A poisson-gamma model for zero inflated rainfall data,” *Journal of Probability and Statistics*, vol. 2018, Article ID 1012647, 12 pages, 2018.

[28] B. Swallow, S. T. Buckland, R. King, and M. P. Toms, “Bayesian hierarchical modelling of continuous non-negative longitudinal data with a spike at zero: an application to a study of birds visiting gardens in winter,” *Biometrical Journal*, vol. 58, no. 2, pp. 357–371, 2016.

[29] B. Y. Ding, W. Gao, S. Dai, S. E. Abhadiomhen, W. He, and X. Yin, “Low rank correlation representation and clustering,” *Scientific Programming*, vol. 2021, Article ID 6639582, 12 pages, 2021.

[30] N. Mwenda, M. Kosgei, G. Kerich, and R. Nduati, “Predictors of household spending on out-patient expenses in Kenya,” 2020, https://www.preprints.org/manuscript/202012.0374/v1.

[31] T. Swan, *Generalized Estimating Equations when the Response Variable Has a Tweedie Distribution: An Application for Multi-Site Rainfall Modelling*, The University of Southern Queensland, Toowoomba, Australia, 2006.

[32] N. Mwenda, R. Nduati, M. Kosgei, and G. Kerich, “Skewed logit model for analyzing correlated infant morbidity data,” *PLoS One*, vol. 16, no. 2, Article ID e0246269, 2021.

[33] U. C. Bureau, “International programs—census and survey processing system overview—people and households,” 2013, https://www.r-project.org/.

[34] S. Fielding, P. M. Fayers, A. McDonald, G. McPherson, M. K. Campbell, and The RECORD Study Group, “Simple imputation methods were inadequate for missing not at random (MNAR) quality of life data,” *Health and Quality of Life Outcomes*, vol. 6, no. 1, p. 57, 2008.

[35] S. Gabrysch, S. Cousens, J. Cox, and O. M. R. Campbell, “The influence of distance and level of care on delivery place in rural Zambia: a study of linked national data in a geographic information system,” *PLoS Medicine*, vol. 8, no. 1, Article ID e1000394, 2011.

[36] D. Kadobera, B. Sartorius, H. Masanja, A. Mathew, and P. Waiswa, “The effect of distance to formal health facility on childhood mortality in rural Tanzania, 2005–2007,” *Global Health Action*, vol. 5, no. 1, 2012.

[37] A. Schoeps, S. Gabrysch, L. Niamba, A. Sié, and H. Becher, “The effect of distance to health-care facilities on childhood mortality in rural burkina faso,” *American Journal of Epidemiology*, vol. 173, no. 5, pp. 492–498, 2011.

[38] R. Stock, “Distance and the utilization of health facilities in rural Nigeria,” *Social Science & Medicine*, vol. 17, no. 9, pp. 563–570, 1983.

[39] T. T. Awoyemi, O. A. Obayelu, and H. I. Opaluwa, “Effect of distance on utilization of health care services in rural kogi state, Nigeria,” *Journal of Human Ecology*, vol. 35, no. 1, pp. 1–9, 2011.

[40] R. K. Biswas and E. Kabir, “Influence of distance between residence and health facilities on non-communicable diseases:

an assessment over hypertension and diabetes in Bangladesh," *PLoS One*, vol. 12, Article ID e0177027, 2017.
[41] N. Nic Carthaigh, B. De Gryse, A. S. Esmati et al., "Patients struggle to access effective health care due to ongoing violence, distance, costs and health service performance in Afghanistan," *International Health*, vol. 7, no. 3, pp. 169–175, 2014.
[42] D. Filmer and L. H. Pritchett, "Estimating wealth effects without expenditure data-or tears: an application to educational enrollments in states of India," *Demography*, vol. 38, no. 1, pp. 115–132, 2001.
[43] H. L. Rippin, J. Hutchinson, D. C. Greenwood et al., "Inequalities in education and national income are associated with poorer diet: pooled analysis of individual participant data across 12 European countries," *PLoS One*, vol. 15, no. 5, Article ID e0232447, 2020.
[44] OECD, *Employment Rate by Age Group (indicator)*, Organisation for Economic Co-operation and Development, Paris, France, 2020.
[45] K.-Y. Liang and S. L. Zeger, "Longitudinal data analysis using generalized linear models," *Biometrika*, vol. 73, no. 1, pp. 13–22, 1986.
[46] J. W. Hardin and J. M. Hilbe, *Generalized Estimating Equations*, CRC Press, Boca Raton, FL, USA, 2nd edition, 2012.
[47] P. K. Dunn and G. K. Smyth, "Evaluation of Tweedie exponential dispersion model densities by Fourier inversion," *Statistics and Computing*, vol. 18, no. 1, pp. 73–86, 2008.
[48] G. Giner and G. K. Smyth, "Statmod: probability calculations for the inverse Gaussian distribution," *The R Journal*, vol. 8, no. 1, pp. 339–351, 2016.
[49] P. McCullagh and J. A. Nelder, *Generalized Linear Models*, 2nd edition, CRC Press, Boca Raton, FL, USA, CRC Press, 1989.
[50] W. Pan, "Akaike's information criterion in generalized estimating equations," *Biometrics*, vol. 57, no. 1, pp. 120–125, 2001.
[51] U. S. C. Bureau, "R: a language and environment for statistical computing," 2017, https://www.R-project.org/.
[52] P. K. Dunn and G. K. Smyth, "Series evaluation of tweedie exponential dispersion model densities," *Statistics and Computing*, vol. 15, no. 4, pp. 267–280, 2005.
[53] A. A. R. Mahfouz and A. M. Hamid, "An epidemiologic study of primary health care service utilization of summer visitors to Abha, Asir, Saudi Arabia," *Journal of Community Health*, vol. 18, no. 2, pp. 121–125, 1993.
[54] E. Mwaliko, G. Van Hal, H. Bastiaens et al., "Early detection of cervical cancer in western Kenya: determinants of healthcare providers performing a gynaecological examination for abnormal vaginal discharge or bleeding," *BMC Family Practice*, vol. 22, no. 1, p. 52, 2021.
[55] M. Rosen-Zvi, L. Shpigelman, A. Kalton et al., "Estimating the impact of prevention action: a simulation model of cervical cancer progression," *Studies in Health Technology and Informatics*, vol. 205, pp. 288–292, 2014.
[56] L. K. Makau-Barasa, S. B. Greene, N. A. Othieno-Abinya, S. Wheeler, A. Skinner, and A. V. Bennett, "Improving access to cancer testing and treatment in Kenya," *Journal of Global Oncology*, vol. 4, 2018.
[57] J. Chuma and T. Maina, "Catastrophic health care spending and impoverishment in Kenya," *BMC Health Services Research*, vol. 12, no. 1, p. 413, 2012.
[58] S. Allin, C. Masseria, and E. Mossialos, "Measuring socioeconomic differences in use of health care services by wealth versus by income," *American Journal of Public Health*, vol. 99, pp. 1849–1855, 2009.
[59] M. A. Kuddus, E. Tynan, and E. McBryde, "Urbanization: a problem for the rich and the poor?" *Public Health Reviews*, vol. 41, no. 1, p. 1, 2020.
[60] P. W. Gething, F. A. Johnson, F. Frempong-Ainguah et al., "Geographical access to care at birth in Ghana: a barrier to safe motherhood," *BMC Public Health*, vol. 12, no. 1, p. 991, 2012.
[61] T. J. Lohela, O. M. R. Campbell, and S. Gabrysch, "Distance to care, facility delivery and early neonatal mortality in Malawi and Zambia," *PLoS One*, vol. 7, no. 12, Article ID e52110, 2012.
[62] KNBS, *Kenya Population and Housing Census*, Kenya National Bureau of Statistics, Nairobi, Kenya, 2019.
[63] P. O. Okwi, G. Ndeng'e, P. Kristjanson et al., "Spatial determinants of poverty in rural Kenya," *Proceedings of the National Academy of Sciences*, vol. 104, no. 43, pp. 16769–16774, 2007.
[64] J. O. Awiti, "Poverty and health care demand in Kenya," *BMC Health Services Research*, vol. 14, no. 1, p. 560, 2014.
[65] D. F. Meyer and R. Nishimwe-Niyimbanira, "The impact of household size on poverty: an analysis of various low-income townships in the northern free state region, South Africa," *African Population Studies*, vol. 30, no. 2, 2016.
[66] S. Ozawa, S. Grewal, and J. F. P. Bridges, "Household size and the decision to purchase health insurance in cambodia: results of a discrete-choice experiment with scale adjustment," *Applied Health Economics and Health Policy*, vol. 14, no. 2, pp. 195–204, 2016.
[67] S. Thaddeus and D. Maine, "Too far to walk: maternal mortality in context," *Social Science & Medicine*, vol. 38, pp. 1091–1110, 1994.
[68] M. H. Miseda, S. O. Were, C. A. Murianki, M. P. Mutuku, and S. N. Mutwiwa, "The implication of the shortage of health workforce specialist on universal health coverage in Kenya," *Human Resources for Health*, vol. 15, no. 1, p. 80, 2017.
[69] G. Kuria, "Cuba to send more than 100 doctors to kenya as part of medical exchange program," 2021, https://africa.cgtn.com/2021/06/09/cuba-to-send-more-than-100-doctors-to-kenya-as-part-of-medical-exchange-program/.
[70] J. Farmer, A. Clark, and S.-A. Munoz, "Is a global rural and remote health research agenda desirable or is context supreme?" *Australian Journal of Rural Health*, vol. 18, no. 3, pp. 96–101, 2010.
[71] R. Strasser, "Rural health around the world: challenges and solutions," *Family Practice*, vol. 20, no. 4, pp. 457–463, 2003.
[72] B. B. Masaba, J. K. Moturi, J. Taiswa, and R. M. Mmusi-Phetoe, "Devolution of healthcare system in Kenya: progress and challenges," *Public Health*, vol. 189, pp. 135–140, 2020.
[73] M. W. Moses, J. Korir, W. Zeng et al., "Performance assessment of the county healthcare systems in Kenya: a mixed-methods analysis," *BMJ Global Health*, vol. 6, no. 6, Article ID e004707, 2021.