Published by
NIGERIAN SOCIETY OF PHYSICAL SCIENCES
Available online @ https://asr.nsps.org.ng/index.php/asr

African Scientific Reports 3 (2024) 181

African Scientific Reports

Predicting the needs of people living with a disability using the two-level logit-skewed exponential power model

Abayomi Ajayi [a,*], Olaniyi Olayiwola [a], Fadeke Apantaku [a], Idowu Osinuga [b], Oluwaseun Wale-Orojo [a]

[a] Department of Statistics, Federal University of Agriculture, Alabata Road, Abeokuta 111101, Ogun State, Nigeria
[b] Department of Mathematics, Federal University of Agriculture, Alabata Road, Abeokuta 111101, Ogun State, Nigeria

Abstract

The impact of high cost of living and movement of medical personnel in Nigeria to other countries. has a major impact on families/households. These has affected Person's living with disabilities (PLWD) with special needs which require regular visitation to clinic and additional expenses. Therefore this study proposed predictive models with generalized distributed (combination of normal and non-normal) error term under Two Stage Sampling. Data from the Nigeria Living Standard Survey (NLSS) 2018 on frequency of doctor's visit and cost by households with difficulty in remembering were used. Logit Skewed Exponential Power (LSEP-II) were developed for two-level Random Effect Model using Bayesian framework. The parameters for LSEP-II were estimated using the Markov Chain Monte-Carlo (MCMC) algorithm with JAGS software. Cartograms were used to determine the spatial distribution for the proportion of doctor's visit and cost using the predicted values. R-hat showed that the posterior distribution converges. The study revealed that Logit Skewed Exponential Power for two-level Random Effect Model modelled and predicted doctor's visit and additional cost for PLWDs in two-Stage Stratified Random Sampling Design.

DOI:10.46481/asr.2024.3.2.181

Keywords: Logit-Skewed exponential power distribution, People living with a disability, Hierarchical model

Article History:
Received: 04 February 2023
Received in revised form: 24 May 2024
Accepted for publication: 08 June 2024
Published: 04 July 2024

1. Introduction

Planning and support for those person's living with disability (PLWDs) is very important in today's Nigeria economy. A survey carried out by National Bureau of Statistics (N.B.S.) [1] in 2018 called the Nigeria living standard survey was aimed at measuring aimed at measuring the living standard of citizens of Nigeria, which include information on person's living with or without disability. This survey was sponsored by the World Bank [1].The availability of data from this survey allows researchers to work on estimating, predicting or modelling the needs of these special people (PLWDs).

Several authors had worked on meeting the needs, finding supports, alleviating the pains, etc., of people living with a disability (PLWD), but there is dearth of knowledge in modelling and predicting their needs in Nigeria, this work intend to fill that gap. For example Onalu and Nwafor [2] advocated for social support for PLWDs so that they fully realize their full potentials. With

*Corresponding author: Tel.: +234-802-910-5483.
Email address: aoajayi@funaab.edu.ng (Abayomi Ajayi)

approximately 180 million citizens, it was reported by Thompson [3] that half of Nigeria's population lives in "poverty", while 30 percent lives in extreme poverty. Research has shown that where the general population is being affected by a problem, it is usually more severe with a sub-population who are more disadvantaged [4]. Also, Sanmartin *et al.* [5] affirmed that the odds of failure to receiving care is increased by 50 percent for PLWDs. Also they may experience barriers in in receiving healthcare.

It was asserted by Amadasun [6] that the impact of a disability in a culture marked by poverty can be extremely detrimental. According to Haruna [7], poverty can cause disability as it may limit a person's ability to receive quality medical care, for example illnesses like polio can leave a person permanently disabled if treatment is not received, also poverty has its effect on disability. A study by McColl *et al.* [8] which was done to determine the unmet needs for adults living with disabilities in Canada, it was concluded that PLWDs fail to report their unmet needs which include with the greatest being additional cost of receiving needed services.

One of the problems of estimating the proportion of PLWDs in a survey is either over-estimating or under-estimating them in households since survey respondents may not necessarily know how to detect these people except they are grown to the stage of exhibiting the symptoms. It was reported by Ken [9] that in other to decrease false-negative responses in disability survey, the yes/no responses were scaled to none/ a little/ a lot and Yes. This actually decrease false negatives but at the cost of increased false-positive. This scale was used by the N.B.S. [1] in the estimation PLWDs in households. Subsequently making modeling and prediction of the needs of PLWDs more precise for easy planning for the government. The Nigeria population is divided into different levels of senatorial district, states, local government till the last household level. This makes the Nigeria statistical system complex, thereby requiring a complex sampling technique, usually the multi-stage sampling is being employed as it was seen in N.B.S. [1].

Two-stage sampling is an easy example of multistage sampling. Two-stage sampling is a kind of multi-stage sampling in which samples are taken in two stages: the first stage, known as primary sampling units, or PSUs, and the second step, known as secondary sampling units, which makes use of PSU subsamples. Two stages of clustering are involved, with sampling done at each level. According to Refs. [10, 11], two-stage sampling, also known as subsampling, is the process of first picking clusters and then selecting a predetermined number of items from each selected cluster. Due of the two stages of sample, Mahalanobis [12] named this concept two-stage sampling. First, a sample of units—often referred to as primary units must be chosen. The second step is known as the (SSU). The criteria for selecting a unit at a particular level frequently rely on characteristics seen in the preceding stage, according to Refs. [13–16]. Sample units at multiple stages of study planning can be beneficial, according to Refs. [10, 17, 18]. The example that goes with it shows how effective two-stage sampling is compared to single-stage methods like cluster, systematic, and basic random sampling.

In order to produce more accurate metrics, it is crucial to conduct statistical study over small areas. A small geographic region is typically meant when the word "small area" is used, for example. It can also be used to refer to a demographic group within a survey, such as a group of people from a same socio-economic background or a particular racial or ethnicity. When the domain-specific sample is too small to support direct estimates with a respectable level of statistical accuracy, a small region is typically defined. In this work PLWDs in different households for each state is being considered has a multilevel small area. Verbeke and Lesaffre [19] had questioned the validity of normal assumptions and how it leads to misleading inferences. Several authors had worked on the non-normality assumptions of the variable of interest and or random errors. Variables such as income, poverty rate, were considerably skewed with potential outliers. To produce better estimates in these regions, Moura *et al.* [20] assumed the Skew-normal distribution and Skew-t distributions for the error terms and variable of interest. Logit-Exponential Power model was developed by Liu and Lahiri [21] to capture the non-normality thereby having a better precision. The model was used in the prediction of low birth-weight using the 2002 natality survey in the United States of America. Small area model for poverty gap data and assumed the error term follows a skew-normal distribution was developed by Diallo and Rao [22], this actually increases the precision of the poverty data.

This work intends to predict the proportion of doctor's visit and additional cost due to having PLWD(s) in the family for each state in Nigeria. This work employ the use of small area estimation techniques and assumed that unit level data is available. The two-level logistic regression model where the random errors follows the Skewed Exponential power distribution were considered for modeling. Two stage sampling procedure was considered in N.B.S. [1], the first stage were the states (area) and households for the second stage. Section 2 presents the materials and methodology. Section 3 discusses the data analysis, results, and Section 4, the conclusion.

2. Materials and methods

2.1. Data source

Data were obtained from N.B.S. [1], the NLSS is nationally representative and collects data from 36 states which usually represents administrative units and the Federal Capital Territory using a stratified two-stage sampling procedure. Its objective is to gather information on the living condition among the Nigeria population, furthermore it include individual and household socio-demographics (marital status, gender, presence of person's living with disability) accesses to health services, assets, income, also used to measure poverty prevalence [1].

2.2. Sample size, inclusion criteria and interest variable

A total 22,110 households consisting of 116,320 individuals were interviewed, the primary sampling units which were enumeration areas (EA) in the 36 states of Nigeria and the federal capital territory was drawn from the National Integrated Household surveys (NISH2).

The inclusion criteria were households who reported having someone who has difficulty in remembering and concentrating. The interview question followed the recommendation by Ken [9], the response options include; Too young to determine, No difficulty, Yes some difficulty, Yes a lot of difficulty and Cannot remember and concentrate. These responses will be considered as response groups/area see Table 1. The variable of interest was if they had visited a doctor in the last month and if they incurred additional expenses due to having person's living with disability in the household. The household responses were analyzed based on the groups/areas and states which serves as the multilevel factors. No auxiliary/independent variables were considered in the analysis.

2.3. Development of multi-level model

Since the survey considered a two stage sample survey, an appropriate multi-level analysis will be the two-level binary logistic regression model. The groups/areas were considered as cluster while states were considered as stratum. Since variable y_{dik} is binary for $d = 1, \ldots, D, i = 1, \ldots, Md; k = 1, \ldots, N_{di}$, it can reasonably be assumed that

$$y_{dik}|\theta_{di} \overset{iid}{\sim} Bernoulli(\theta_{di}); \quad d = 1, \ldots, D, \quad i = 1, \ldots, M_d; \quad k = 1, \ldots, N_{di}. \tag{1}$$

The popular logistic mixed effect regression is usually assumed for the prior distribution of $\{\theta_{di} : d = 1, \ldots, D, j = 1, \ldots, M_d\}$ Liu and Lahiri [21]. That is,

$$logit(\theta_{di}) = x'_{di}\beta + v_d + u_{di}, \tag{2}$$

where independently for all d and i,

$$u_{di} \overset{iid}{\sim} N(0, \sigma_u^2) \; d = 1, \ldots, D, i = 1, \ldots, M_d, \tag{3}$$

also independently for all d,

$$V_{di} \overset{iid}{\sim} N(0, \sigma_v^2) \; d = 1, \ldots, D. \tag{4}$$

It should be noted that the effect u_{di} accounts for the two-stage sample design in the areas and v_d acounts for the random effect in the area. The models in equations (2)-(4) was referred to as Bernoulli-Logit Normal (LN) by Liu and Lahiri [21].

In order to improve LN model, this work assume that the random effects follows a Skewed Exponential Power distribution (SEP), which is a more robust generalized distribution as it contains the normal distribution and it accommodates skewed and kurtosis data sets. The Skewed Exponential Power distribution as given by Hutson [23] is: t

$$f_{sep}(z) = \frac{k}{\sigma} exp\left(-\frac{1}{2}(|z| + (2\alpha - 1)z)^{\frac{2}{(1+\lambda)}}\right), \tag{5}$$

where

$$k^{-1} = \frac{\Gamma\left(1 + \frac{1+\lambda}{2}\right) 2^{1+\frac{1}{2}(1+\lambda)}}{(4\alpha(1-\alpha))'}.$$

$0 < \alpha < 1, -1 < \lambda \leq 1, -\infty < \mu < \infty, \sigma > 0, z = \frac{(y-\mu)}{\sigma}$, and $\Gamma(.)$ is the Euler gamma function.

μ and σ serve as the location and scale parameter respectively, and α and λ serve as the shape parameters noting the skewness and kurtosis of the distribution.

Some symmetric distributions are sub models of the skew exponential power distribution for specific parameter values they include exponential power distribution, Skew normal distribution Uniform distribution, Normal distribution amongst others.

The PDF of the random variable $Y = \mu + \sigma Z$ is

$$f_{sep}(y|\mu, \sigma, \alpha, \lambda) = \frac{k}{\sigma} exp\left(-\frac{1}{2}\left(|\frac{y-\mu}{\sigma}| + (2\alpha - 1)\left(\frac{y-\mu}{\sigma}\right)\right)^{\frac{2}{1+\lambda}}\right). \tag{6}$$

Let $Y \sim SEP(\mu, \sigma, \alpha, \lambda)$, then $E(Y) = \mu + \sigma E(z)$ with,

$$E(z) = \frac{(\beta+1)(1-2\alpha)2^{\frac{\beta-3}{2}}\Gamma(\beta+1)}{(1-\alpha)\alpha\Gamma\left(\frac{\beta+3}{2}\right)}.$$

Hence, the distribution $Y \sim SEP(-\sigma E(z), \sigma, \alpha, \lambda)$ will have zero mean and therefore it can be used to generate random numbers in the nested error model.

Given that the random effects of the model in equation (2) follows a skewed exponential power distribution then equation (2) becomes

$$logit(\theta_{di}) = x'_{di}\beta + v_d + u_{di}, \tag{7}$$

$$u_{di} \overset{iid}{\sim} SEP(-\sigma_u E(z)_u, \sigma_u, \alpha_u, \lambda_u) \; d = 1, \ldots, D, i = 1, \ldots, M_d, \tag{8}$$

$$v_d \overset{iid}{\sim} SEP(-\sigma_v E(z)_v, \sigma_v, \alpha_v, \lambda_v). \tag{9}$$

tThe random errors are independent and $E(v_d) = E(u_{di}) = 0$.

The hyperparameters are assumed unknown $(\sigma_v, \alpha_v, \lambda_v, \sigma_u, \alpha_u, \lambda_u)$ this model shall be referred to as LSEP-II. By implication from the SEP distribution, it should be noted that to ensure the means of random errors are equal to zero, $\mu_v = -\sigma_v E(z)_v; \mu_u = -\sigma_u E(z)_u$; the sub-script u and v representing the area and state random level effect.

2.4. Parameter estimation using hierarchical Bayesian methods

Denoting $\theta = (\theta_{11}, ..., \theta_{1m}, ..., \theta_{D1}, ..., \theta_{DmD})'$ and area effect $v = (v_1, ..., v_D)'$ $\omega_{SEP} = (\beta, \sigma_u, \sigma_v, \alpha_u, \alpha_v, \lambda_u, \lambda_v)'$ denote the hyperparameters for LSEP model.

$$\begin{aligned} f(\theta, v, \lambda_{SEP} | y_S) &\propto f(y_S | \theta, v, \lambda_{SEP}) f(\theta | v, \lambda_{SEP}) f(v | \sigma_v) f(\lambda_{SEP}) \\ &\propto f(\lambda_{SEP}) \prod_{d=1}^{D} \left\{ \begin{array}{l} \prod_{i=1}^{m_d} \theta_{di}^{y_{di}-1} (1-\theta_{di})^{n_{di}-y_{di}-1} \\ \times \frac{k_u}{\sigma_u} \exp\left[-\frac{1}{2} \left(\begin{array}{l} \left| \frac{\text{logit}(\theta_{di}) - x'_{di}\beta - v_d}{\sigma_u} \right| \\ + (2\alpha_u - 1) \left(\frac{\text{logit}(\theta_{di}) - x'_{di}\beta - v_d}{\sigma_u} \right) \end{array} \right)^{2/(1+\lambda_u)} \right] \end{array} \right\} \\ &\qquad \times \frac{k_v}{\sigma_v} \exp\left[-\frac{1}{2} \left(\left| \frac{v_d}{\sigma_v} \right| + (2\alpha_v - 1) \left(\frac{v_d}{\sigma_v} \right) \right)^{2/(1+\lambda_v)} \right], \end{aligned} \tag{10}$$

where

$$k_v^{-1} = \Gamma\left(1 + \tfrac{1+\lambda_v}{2}\right) 2^{1+\frac{1}{2}(1+\lambda_v)} \Big/ (4\alpha_v (1-\alpha_v)) \text{ and } k_u^{-1} = \Gamma\left(1 + \tfrac{1+\lambda_u}{2}\right) 2^{1+\frac{1}{2}(1+\lambda_u)} \Big/ (4\alpha_u (1-\alpha_u)).$$

2.5. Prediction of small area proportion using LSEP-II

The aim is to find the Bayes estimator Pd which is the mean of the posterior distribution P_d Considering areas that contain at least one sampled PSU, i.e., $m_d > 0$. P_d can then be written as

$$P_d = \frac{1}{N_d} \left(\sum_{j \in S_d} \sum_{k \in S_{di}} y_{dik} + \sum_{j \in S_d} \sum_{k \in S^c_{dij}} y_{dik} + \sum_{j \in S^c_d} \sum_{k=1}^{N_{di}} y_{dik} \right), \tag{11}$$

where S^c_{di} is the set of non-sampled elements in the jth sampled PSUs, S^c_{di} is the set of non-sampled PSUs, and $N_d = \sum_{i=1}^{M_d} N_d i$ is the area level population total.

Given the Bernoulli component of the model $E(y_{dik}|\theta_{di}) = \theta_{dik}$, and $V(y_{dik}|\theta_{di}) = \theta_{di}(1 - \theta_{di}), t$.

The Bayes Estimator for posterior mean of P_i is

$$E(P_d|y_s) = E\left[E\left(P_d|\theta_{di}, y_s\right)|y_s\right], \tag{12}$$

$$\begin{aligned} &= \frac{1}{N_d} E\left(\left[\sum_{i \in S_d} \sum_{k \in S_{di}} y_{dik} + \sum_{j \in S_i} (N_{di} - n_{di})\theta_{di} + \sum_{i \in S^c_d} N_{di}\theta_{di} \right] | y_s \right), \\ &= \frac{1}{N_d} \left[\sum_{i \in S_d} \sum_{k \in S_{di}} y_{dik} + \sum_{i \in S_d} (N_{di} - n_{di}) E(\theta_{di}|y_s) + \sum_{i \in S^c_d} N_{di} E(\theta_{di}|y_s) \right]. \end{aligned} \tag{13}$$

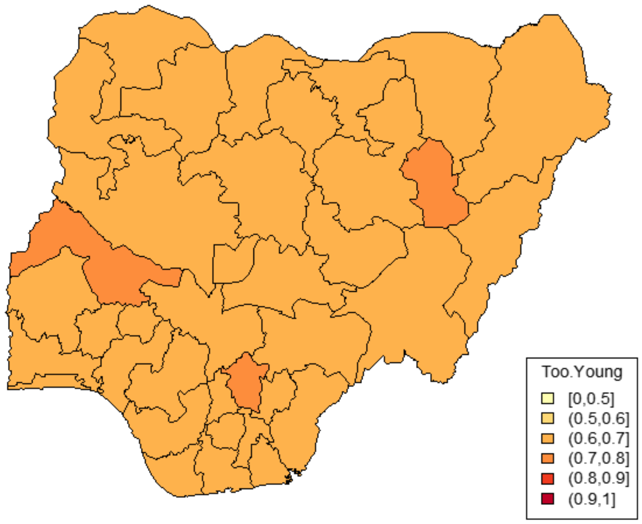

Figure 1: Doctor's visit: response group "Too young to determine".

Table 1: Brief statistics of the areas.

Response Groups/cluster/Area	Does (NAME) have difficulty remembering or concentrating?	Frequency	Percent
1	TOO YOUNG TO DETERMINE	3595	3.1
2	No, no difficulty	109896	94.5
3	Yes, some	2005	1.7
4	Yes, a lot	547	0.5
5	Cannot do (Yes)	277	0.2
	Total	116320	100

And the posterior variance of P_i is

$$\begin{aligned} V(P_d|y_s) &= E[V(P_d|\theta_{di}, y_s)] + V[E(P_{di}|\theta_{di}, y_s)|y_s], \\ &= \frac{1}{N_d^2} E\left(\left[\sum_{i \in S_{di}} (N_{di} - n_{di})\theta_{di}(1-\theta_{di}) + \sum_{i \in S_d^c} N_{di}\theta_{di}(1-\theta_{di})\right]|y_s\right) \\ &+ \frac{1}{N_d^2} V\left(\left[\sum_{i \in S_d}\sum_{k \in S_{di}} y_{dik} + \sum_{i \in S_d}(N_{di} - n_{di})\theta_{di} + \sum_{i \in S_d^c} N_{di}\theta_d\right]|y_s\right), \\ &= \frac{1}{N_d^2}\left[\sum_{i \in S_{di}} (N_{di} - n_{di}) E(\theta_{di}(1-\theta_{di})|y_s) + \sum_{i \in S_d^c} N_{di} E(\theta_{di}(1-\theta_{di})|y_s)\right] \\ &+ \frac{1}{N_d^2} V\left(\left[\sum_{i \in S_d} (N_{di} - n_{di})\theta_{di} + \sum_{i \in S_d^c} N_{di}\theta_{di}\right]|y_s\right), \end{aligned} \tag{14}$$

where $\theta_{dik} = \frac{exp(\beta x'_{dik} + v_d + u_{di})}{1 + exp(\beta x'_{dik} + v_d + u_{di})}$, for $i \in s_d$ and $\theta_{di} = \frac{exp(\beta x'_{dik} + v_d)}{1 + exp(\beta x'_{dik} + v_d)}$, for $i \in s_d^c$.

Considering areas that no sample PSU were taking, Liu and Lahiri [21] considered using a synthetic estimator $P_d = \frac{\sum_{i=1}^{M_d}\sum_{k=1}^{N_{di}} y_{dik}}{N_d}$. This is based on the assumption that information on each PSUs is available which is a special case of the posterior mean and posterior variance. Another solution maybe to consider closely related PSUs to the one not sample.

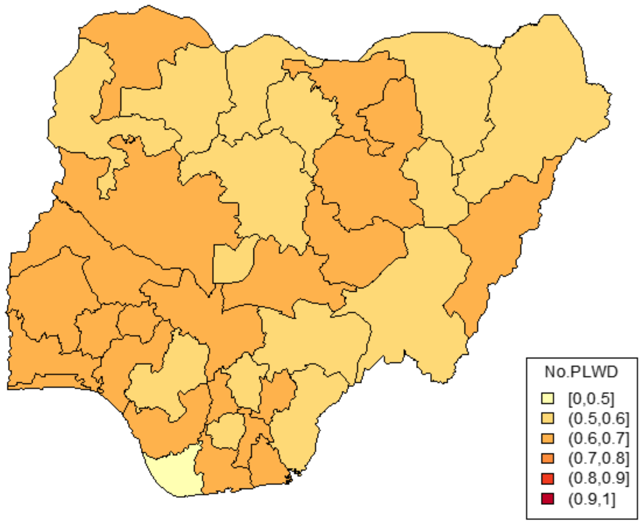

Figure 2: Doctor's visit: response group "No PLWD".

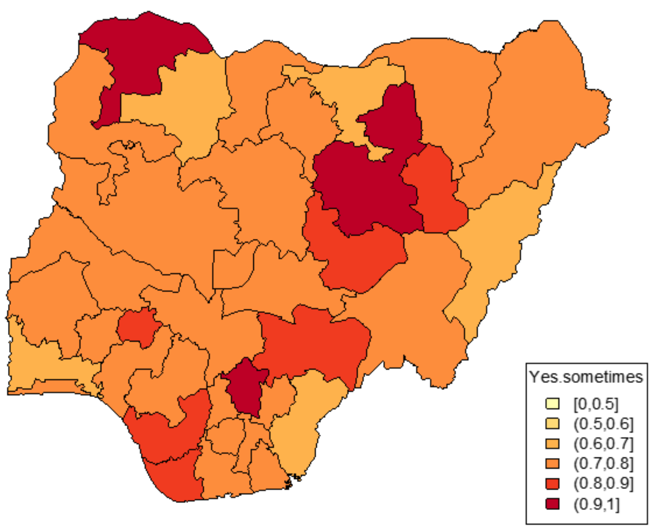

Figure 3: Doctor's visit: response group "Yes sometimes".

2.6. Hierarchical Bayesian (HB)-implementation

HB estimates were obtained for the finite population proportion using the full Bayesian method.. The prior distribution considered for the hyperparameters

i Flat prior for mean (mu) and beta (regression) parameters
ii Gamma prior for the variance component
iii Uniform prior for the Kurtosis (alpha) and Skewness (parameters)

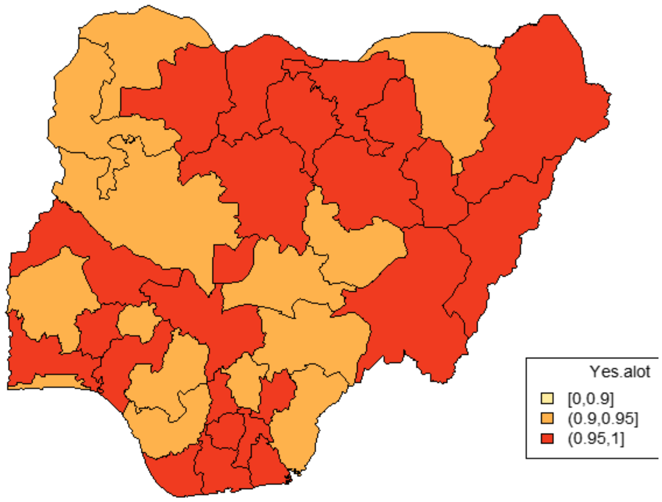

Figure 4: Doctor's visit: response group "Yes a lot".

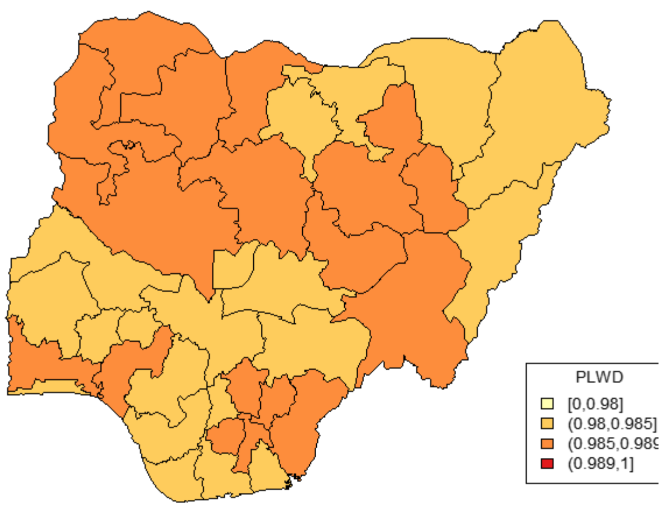

Figure 5: Doctor's visit: response group "Yes".

The HB models were implemented using JAGS (Just Another Gibbs Sampler), 3 independent chains were used for the models, For each with 50,000 iterations (first 25,000 discarded), to reduce autocorrelation the burn-in were thinned by a factor of 10 in the MCMC results. The resultant 25,000 MCMC samples were used to compute the mean and variance. The potential scale reduction factor R Gelman and Rubin [24] were used as the measure of convergence.

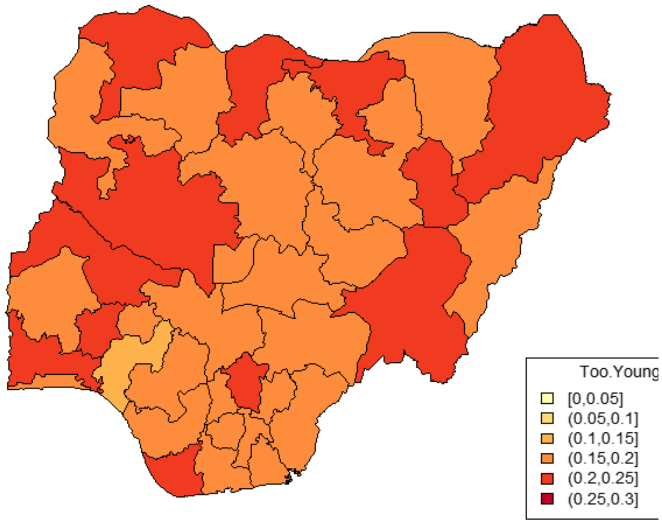

Figure 6: Additional cost: response group “Too young to determine”.

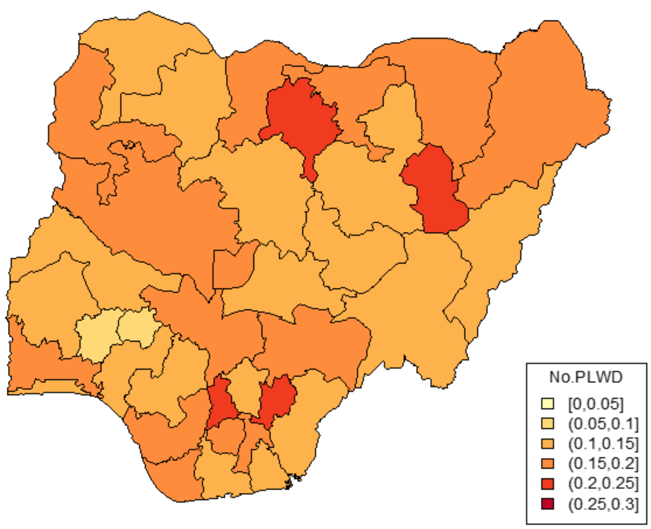

Figure 7: Additional cost: response group “No PLWD”.

3. Results and Discussion

Two-level logistic SEP model and one-level logistic model were fitted to the data, the parameters were estimated using Bayesian methods, R-hat was used as a measure of convergence, it shows that the posterior distribution converges (see Tables 2 and 3). The variance for the one-level logistic regression model for doctor’s visit and additional cost were found to be 48.662 and 50.623 respectively. The intra class correlation coefficient (ICC) was computed to determine the cluster and strata effect on the model. The two level SEP model has a lower ICC therefore, it is better for prediction. The focus is to estimate/predict the proportion of PLWDs

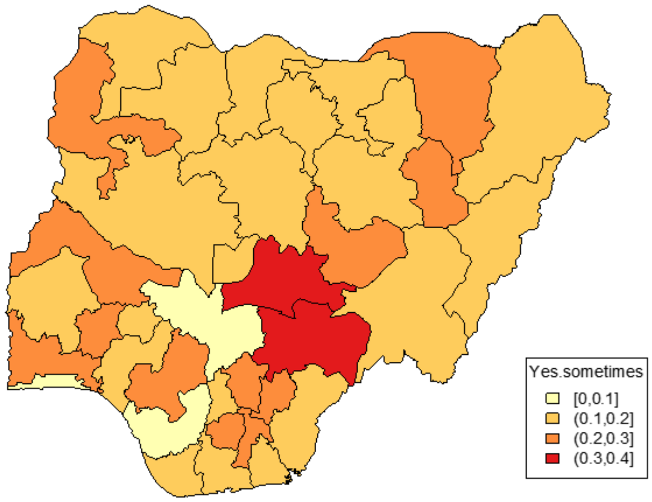

Figure 8: Additional cost: response group "Yes Sometimes".

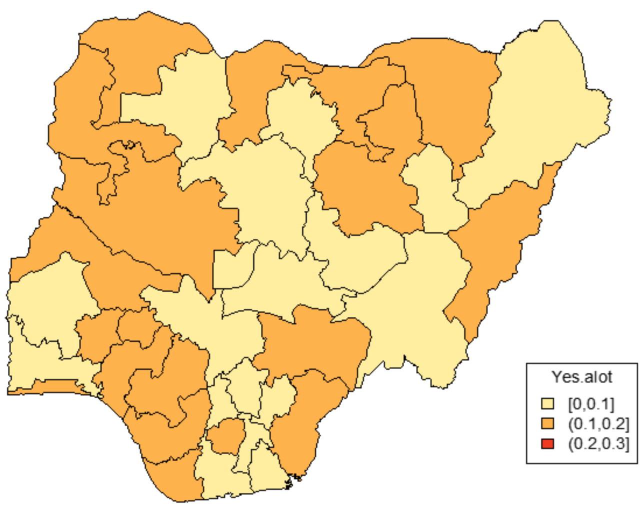

Figure 9: Additional cost: response group "Yes a lot".

who had visited a doctor and additional cost they incurred across Nigeria in various groups of responses.

Tables 4 and 5 show the descriptives of the predicted values for household who had visited the doctor and additional cost due to PLWD respectively. It can be seen in Table 5 that the response group "Yes" which indicate that there is at least someone living with disability in the household had lower proportion for additional cost that is households don't have additional cost due to PLWD whereas there is higher proportion of households do visit doctors (Table 4).

Figure 1 shows the doctor's visit of the response group "Too young to determine" and it can be seen that Kwara, Gombe and

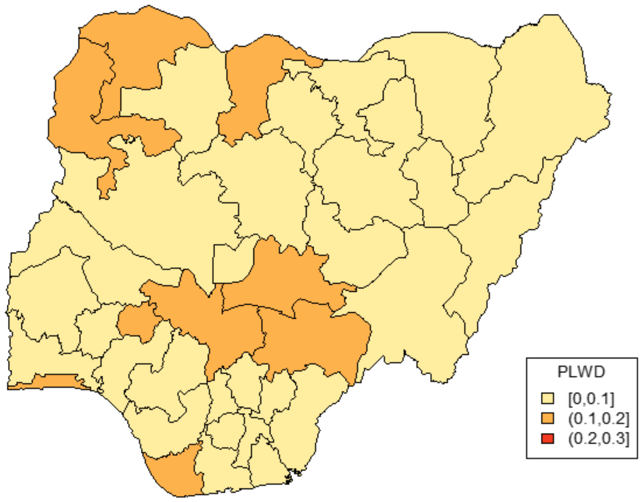

Figure 10: Additional cost: response group "Yes".

Table 2: Parameter estimation using two-level SEP model for visit to medical practitioner.

			Credible Interval		
Parameters	Estimates	SD	2.50%	97.50%	Rhat
Alpha	0.519	0.287	0.035	0.978	1.005
Lambda	0.506	0.283	0.027	0.973	1.008
Sigmau	49.201	28.564	3.117	98.053	1.004
Sigmav	51.135	28.433	2.769	97.165	1
DIC					33420

Table 3: Parameter estimation using two-level SEP model for additional cost.

			Credible Interval		
parameters	Estimates	SD	2.50%	97.50%	Rhat
alpha	0.492	0.287	0.025	0.968	1.011
lamda	0.488	0.283	0.025	0.97	1.006
sigmau	50.031	29.018	1.834	97.897	1.008
sigmav	49.48	28.234	2.617	95.654	1.002
DIC					2165

Table 4: Descriptive statistics of predicted values for households who had visited the Doctor.

Response Groups	N	Minimum	Maximum	Mean	Std. Deviation
Too young	37	0.6842	0.7025	0.69435	0.00407
No PLWD	37	0.4785	0.6805	0.60491	0.04403
Yes some	37	0.6118	0.91	0.76513	0.07025
Yes a lot	37	0.91	0.985	0.94917	0.0318
Yes	37	0.9817	0.988	0.98476	0.00165

Enugu states has a higher proportion. Figure 2 shows that South-West has the highest proportions of response group "No PLWDs" who had visited the doctor, with some few states in the South-South and Norths. Figures 3 and 4 showed that households who has

Table 5: Descriptive statistics of predicted values for additional cost due to PLWD.

Response Groups	N	Minimum	Maximum	Mean	Std. Deviation
Too Young	37	0.144	0.235	0.1881	0.01988
No	37	0.089	0.206	0.15041	0.03036
Yes some	37	0.045	0.36	0.18868	0.06928
Yes a lot	37	0.01	0.2	0.11519	0.08281
Yes	37	0	0.2	0.07237	0.07451

individual who behave sometimes or a lot PLWD has increase in visiting of medical practitioner, this may be due to not being sure if the individual is PLWD. Figure 5 shows that all the states of Nigeria has households with PLWD who visited the medical practitioner in the last one month, however it is higher in the North-Western, North-Central and North-Eastern Nigeria, also Ogun and Ondo states has households with PLWDs, it could be seen that South-Eastern with Cross-River state has households with PLWDs, it can be said that this states will require more medical practitioners in order to cater for the needs of this PLWDs. Figures 6-10 show the additional cost incurred by households in Nigeria across various response group. It can be seen that the proportion who incurred additional cost was low. Further studies could be done to ascertain the reasons for having lower proportion of additional cost and higher proportion of visit to the doctor. Figure 10 shows that the North central states, Lagos, Ekiti, Bayelsa, Kebbi, Sokoto and Katsina were predicted to having additional cost due to PLWDs.

4. Conclusion

The use of stratified two stage sampling techniques allows for the prediction of each responses/area per states for the needs of people living with a disability. The assumption that the random effects being normally distributed was relaxed in the formed two-parts model to accommodate a more robust distribution; skewed exponential power distribution (SEP) which also has Normal distribution as its member. Fitting the full two-part model accounts for the correlations between the random effects of the two parts. This actually increase prediction than using one-fold model. The use of HB method of estimation implemented using Monte-Carlo Markov chain (MCMC) simulations in R-JAGS, which eliminated the complexity of the two parts hierarchical model, thereby enabling the prediction of each area per states in Nigeria. The use of this approach requires specifying prior distributions, which can affect the inference/variance, particularly with a small number of sampled areas even when specifying noninformative priors. Methods like bootstrap or jacknife could be considered as alternative methods for the variance estimation.

Acknowledgment

The authors wish to thank Mr. Baba Tunde for helping to sort out the errors in R-code, and Miss Toluwanimi Owoeye for helping with the LaTex documents. We also appreciate the journal editors and reviewers for their constructive reviews and criticism. There was no special grants/funds for this research and no areas of conflict have been recorded to the best knowledge of the authors.

References

[1] Nigeria Bureau of Statistics, *Living standards survey (NLSS) 2018/19*. https://doi.org/10.48529/8gbe-z155.

[2] C. Onalu & N. Nwafor, "Social supports available to persons with disabilities in Nigeria", in *Evolutionary Psychology Meets Social Neuroscience*, R. Morese, V. Auriemma & S. Palermo, IntechOpen, 2022. http://dx.doi.org/10.5772/intechopen.97790.

[3] S. Thompson, *Nigeria situational analysis. Inclusion Works*, 2022. https://opendocs.ids.ac.uk/opendocs/handle/20.500.12413/15487.

[4] M. D. Brownell, N. P. Roos & L. L. Roos, "Monitoring health reform: a report card approach", Soc Sci Med. **52** (2001) 657. https://doi.org/10.1016/s0277-9536(00)00168-4.

[5] C. Sanmartin, C. Houle, S. Tremblay & J. M. Berthelot, "Changes in unmet health care needs", Health Rep. **13** (2002) 15. https://pubmed.ncbi.nlm.nih.gov/12743957.

[6] S. Amadasun, "Social work services for persons with disabilities in Nigeria: a Qualitative Inquiry", International Journal of Social Sciences Perspectives **6** (2020) 59. https://doi.org/10.33094/7.2017.2020.62.59.67.

[7] M. A. Haruna, "The problems of living with disability in Nigeria", Journal of Law, Policy and Globalization **6** (2017) 103. https://www.iiste.org/Journals/index.php/JLPG/article/view/39027/40128.

[8] M. A. McColl , Anna Jarzynowska & S.E.D. Shortt, "Unmet health care needs of people with disabilities: population level evidence", Disability & Society **25** (2010) 205. https://doi.org.10.1080/09687590903537406.

[9] B. Ken, "Testing a disability question for the census", in *International Views on Disability Measures: Moving Toward Comparative Measurement*, B. M. Altman & S. N. Barnartt, Emerald Publishing Limited, 2015, pp. 179-223. https://doi.org/10.1016/S1479-3547(05)04011-X.

[10] W. G. Cochran, *Sampling techniques* (3rd edn), John Wiley & Sons Inc., New York, United States of America, 1977. http://gauss.stat.su.se/master/statmet1/Cochran%20page%2039.annotated.pdf.

[11] J. Neyman, "Contributions to the theory of sampling human populations", Journal of the American Statistical Association **33** (1938) 101. https://doi.org/10.1080/01621459.1938.10503378.

[12] P. C. Mahalanobis, "On large-scale sample surveys", in *Phil. Trans. R. Soc. B*, R. A. Fisher, London **23** (1944) 1329. http://doi.org/10.1098/rstb.1944.0002.

[13] L. Kish, *Survey sampling*, John Wiley and Sons, New York, NY. United States of America, 1967.
[14] M. H. Hansen, W. N. Hurwitz, & W. G. Madow, *Sample survey methods and theory* Wiley and Sons, New York, NY, United States of America, 1975. https://archive.org/details/SampleSurveyMethodsAndTheory/page/n5/mode/2up.
[15] A. Fink, *How To Sample In Surveys*, Sage Publications, Thousand Oaks, CA. United States of America, 2002. https://doi.org/10.4135/9781412984478.
[16] J. E. Tate & M. G. Hudgens, "Estimating population size with two-stage and three-stage sampling designs", American Journal of Epidemiology **165** (2007) 1314. https://doi.org/10.1093/aje/kwm005.
[17] G. Kalton, *Introduction to survey sampling*, Sage Publications: Los Angeles, CA, United States of America, 1983. https://doi.org/10.4135/9781412984683.
[18] F. C. Okafor, *Sample survey theory with applications*, Afro-Orbis Publications Ltd., Nsukka, Nigeria, 2002.
[19] G. Verbeke, E. Lesaffre, "A linear mixed-effect model with heterogeneity in the random-effects population", Journal of American Statistical Association **91** (1997) 217. https://api.semanticscholar.org/CorpusID:121803659.
[20] F. A. S. Moura, A. F. Neves & D. N. Silva, "Small area models for skewed Brazilian business survey data", Journal of Royal Statistical Society A **180** (2017) 1039. http://www.jstor.org/stable/44682663.
[21] B. Liu & P. Lahiri, "Adaptive hierarchical Bayes estimation of small proportions", Calcutta Statistical Association Bulletin **69** (2007) 1. https://doi.org/10.1177/0008068317722293.
[22] M. S. Diallo, & J. N. K. Rao, "Small area estimation of complex parameters under unit-level models with skew-normal errors", Scandinavian Journal of Statistics **45** (2018) 1. https://doi.org/10.1111/sjos.12336.
[23] A. D. Hutson,"An alternative skew exponential power distribution formulation", Communication in Statistics - Theory and Methods **48** (2019) 3005. http://doi.org/10.1080/03610926.2018.1473600.
[24] Gelman, A.& D. B. Rubin, "Inference from iterative simulation using multiple sequences", Statistical Science **7** (1992) 457. http://www.jstor.org/stable/2246093.