GENERALIZED ADDITIVE MODELS FOR LOCATION, SCALE AND SHAPE WITH AN APPLICATION TO ANALYSE BLOOD PRESSURE CORRELATES IN NAMIBIA

Atanasia K P Bonge, Lawrence Kazembe and Lillian Pazvakawambwa

Abstract

The aim of this paper is to apply a Generalized Additive Models for Location Scale and Shape (GAMLSS) model, to analyse the distributive occurrences of blood pressure in Namibia. The model offers a flexible approach as it allows flexible modelling of the mean, variance, skewness and kurtosis. For the median parameter, blood pressure increased with age, BMI, glucose (SBP only) and haemoglobin (DBP only) levels. In relation to the latter finding, the frequency in blood pressure adults was higher among middle class, male adults residing in urban areas, with no level of education, and with a higher alcohol consumption rate. The findings presented a negative relationship between high blood pressure and weight, as well as height, the BMI and DBP observed in the scale parameter. With the shape parameter, high blood pressure was associated with no attainment of education (SBP), increased BMI and alcohol intake (DBP).

Keywords: GAMLSS, SBP, DBP, distribution

1. Introduction

Namibia is one of the sub-Saharan countries with concerning increases in hypertension cases. The country is experiencing a double burden of communicable and non-communicable diseases, coupled with an increasing urbanization, which further engenders the vulnerability of communities due to reduced health and wellbeing conditions (Ministry of Health and Social Services Namibia, 2017). Non-communicable diseases NCDs in Namibia have affected the health and economic sectors in terms of medication and treatment expenditures, and accounted for approximately 41% mortalities of the population to date (World Health Organization, 2020). The prevalence of hypertension is also high among adults aged 35 -64 , estimated to be 46% (Craig, Gage, & Thomas, 2018). Moreover, according to the NDHS for 2013 44% of women and 45% of men aged 35-64 were hypertensive (MoHSS and ICF International, 2014). Globally, efforts have been made to mitigate the frequent succumbing of patients to hypertension, however the occurrences of this disease continue to increase. Recent studies have indicated that if hypertension is not treated it weakens the body's immunity, making it susceptible to others dreadful diseases, such as coronary heart diseases, cardiovascular diseases, stroke, diabetes (World Health Organization, 2020) and the recent COVID 19 outbreak (Rodilla et al., 2020; Tadic et al., 2021). Therefore, it remains that a

continuous development of statistical models is necessary and important for the on-going prediction of risk factors contributing to high blood pressure and the on-going monitoring of such factors.

A significant challenge which is common in applied statistics is the selection of an appropriate model to estimate the relationship between a dependent variable and its determinants. Researchers in this field of study have developed various regression models to assist in health-related modelling studies and vast amounts of focus have been given to parametric and nonparametric models. Several statistical methods have been utilized to identify the relationship between blood pressure and its contributing factors. These includes the Generalized Linear Models (GLMs) (Nelder & Wedderburn, 1972) provide a parsimonious approach to modelling linear relationships, while Generalized Additive Models (GAMs) (Hastie & Tibshirani, 1990) extend this to account for nonlinear relationships. Additionally, linear mixed models incorporate random effects to capture more complex relationships. However, some of these models prone to inconsistent estimates and limited flexibility, particularly in capturing various distributions and restricted to modelling only the mean of predictors, with variance, skewness, and kurtosis being implicitly modelled through their dependence on μ. Rigby and Stasinopoulos (2005) introduced Generalized Additive Models for Location, Scale, and Shape (GAMLSS) to address these limitations. GAMLSS offers flexibility by accommodating a wide range of distributions and allows for the simultaneous modelling of location, scale, and shape parameters. This paper will explore several statistical approaches used to identify the relationship of blood pressure with its contributing factors and deploy the model based on AIC and BIC.

2. Material and method

This cross-sectional study utilizes secondary data obtained from the 2013 Namibian Demographic Health Survey (NDHS). A total of 3095 men and women from 35 to 64 years old were included in the analysis. The methodological approaches explored include linear model (LM), GLM, GAMs and the GAMLSS.

GAMLSS represents a semi-parametric regression model capable of accommodating a wider array of distributional assumptions for the response variable. It allows for the modelling of distribution parameters of the response variable as functions of explanatory variables, which can include linear, nonlinear, and smooth terms, as well as random effects. This model considers independent observations y_i, $i = 1,2,\dots,n,$ with conditional probability density function $f(y_i|\boldsymbol{\theta}^i)$ characterized by up to four distribution parameters $\boldsymbol{\theta}^{iT} = (\theta_{i1}, \theta_{i2}, \cdots \theta_{ip})$. These parameters

include location (μ), scale (σ), skewness (ν), and kurtosis (τ). The GAMLSS model is defined as follows:

$$(\boldsymbol{\theta}_k) = \boldsymbol{\eta}_k = \boldsymbol{X}_k\boldsymbol{\beta}_k + \sum_{j=1}^{J_k} \boldsymbol{Z}_{jk}\gamma_{jk} \ , \qquad k = 1,2,3,4$$

Which can be extended to modelling of the parameters under the following conditions:

$$g_1(\boldsymbol{\mu}) = \boldsymbol{\eta}_1 = \boldsymbol{X}_1\boldsymbol{\beta}_1 + \sum_{j=1}^{J_1} \boldsymbol{Z}_{j1}\gamma_{j1} \ ,$$

$$g_2(\boldsymbol{\sigma}) = \boldsymbol{\eta}_2 = \boldsymbol{X}_2\boldsymbol{\beta}_2 + \sum_{j=1}^{J_2} \boldsymbol{Z}_{j2}\gamma_{j2} \ ,$$

$$g_3(\boldsymbol{\nu}) = \boldsymbol{\eta}_3 = \boldsymbol{X}_3\boldsymbol{\beta}_3 + \sum_{j=1}^{J_3} \boldsymbol{Z}_{j3}\gamma_{j3} \ ,$$

$$g_4(\boldsymbol{\tau}) = \boldsymbol{\eta}_4 = \boldsymbol{X}_4\boldsymbol{\beta}_4 + \sum_{j=1}^{J_k} \boldsymbol{Z}_{j4}\gamma_{j4} \ ,$$

where μ, σ, ν, τ corresponding to the mean, variance, skewness, and kurtosis. $g_k(\cdot)$ denotes a known monotonic link function that connect the distribution parameters k –th to the predictor $\boldsymbol{\eta}_{ik}$, $\boldsymbol{\beta}_k^T = (\beta_{1k}, \beta_{2k,} \cdots \beta_{J_k k})$ represents a parameter vector of dimension of J_k, $\boldsymbol{X}_k$ (fixed effect) and $\boldsymbol{Z}_{jk}$ (random effect) are design matrices of order $n \times J_k$ and $n \times q_{jk}$ respectively. γ_{jk} is a random variable of dimension q_{jk} that follows a normal distribution $\mathcal{N}_{q_{jk}}\left(\mathbf{0}, \boldsymbol{G}_{jk}^{-\mathbf{1}}\right)$, where $\boldsymbol{G}_{jk}^{-\mathbf{1}}$ is the generalized inverse of the symmetric matrix $\boldsymbol{G}_{jk} = \boldsymbol{G}_{jk}(\lambda_{jk})$ of order a $q_{jk} \times q_{jk}$ and λ_{jk} is vector of hyperparameters.

GAMLSS models were employed using the RS algorithm, as detailed by Rigby and Stasinopoulos (2005), aiming to maximize the penalized log likelihood function. The selection of the most appropriate model for the data used the Generalized Akaike Information Criterion (GAIC) (Akaike, 1983) mechanism, where the goodness of fit was assessed by adding the fitted global deviance to a fixed penalty κ for each effective degree of freedom utilised in the model. Thus, GAIC(κ) is defined as $GD + \kappa\, df$, where df represents the total effective number of degrees of freedom used in the model, and GD denotes the fitted global deviance. Additionally, diagnostic tools of residuals were employed to assess the adequacy of the model. Finally, a comparison between the GAMLSS

model and classical regression models (LM, GLM, and GAM) was conducted to ascertain the flexibility of each approach. This comparison utilised the Akaike Information Criterion (AIC) (Akaike, 1974), the Bayesian Information Criterion (BIC) (Schwarz, 1978) and Anderson darling test.

3. Results

3.1 Distribution of blood pressure

In our quest to find an appropriate model for predicting blood pressure, we began by examining its distribution. This provided insight into which distribution would be most suitable for modelling blood pressure. Figure 1 illustrates the distributions of both systolic blood pressure (SBP) and diastolic blood pressure (DBP). The histogram for SBP reveals a heavily right-skewed distribution, while DBP appears symmetric. Kurtosis is evident in both histograms. Furthermore, we utilized box plots and scatter plots to explore the relationships between SBP/DBP and explanatory variables. Among categorical variables, we observed variation and positive skewness.

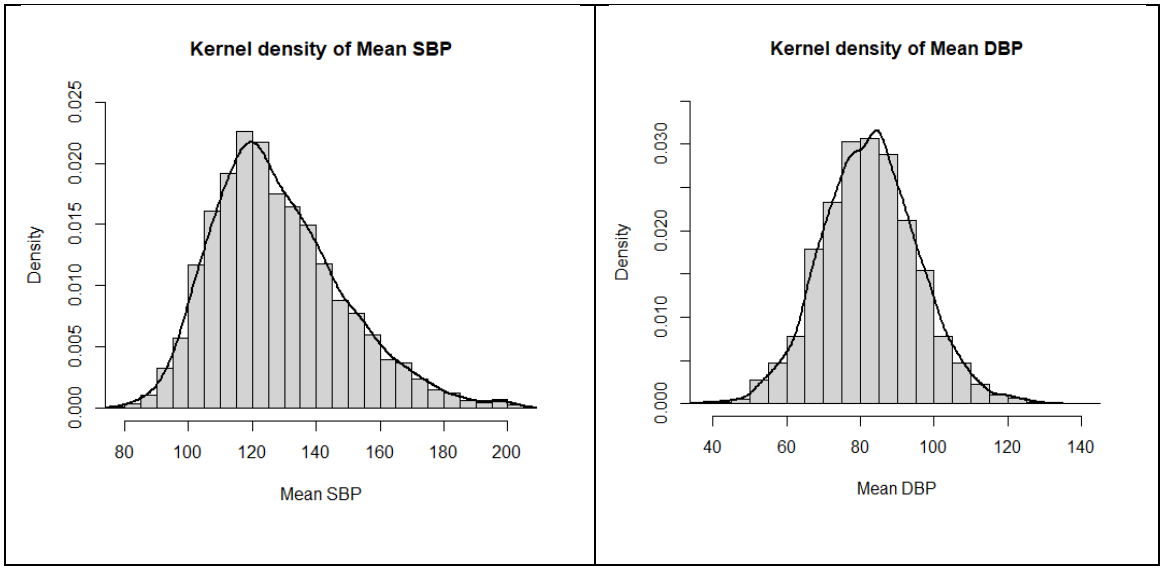


Figure 1 Histogram of SBP/ DBP

3.2 Application of GAMLSS modelling blood pressure

In this section, we showcased the application of the GAMLSS model in analysing blood pressure data from Namibia. Initially, we focused on selecting distributions suitable for the dataset, considering that blood pressure values are positive and continuous. Only positive continuous distributions were fitted to explore the effects of various explanatory variables. Selection of the

appropriate distribution involved fitting and diagnosing the model. Among the continuous positive models fitted, the Generalized Gamma (GG), Box-Cox Cole Green original (BCCGo), and Box Cox Power Exponential original (BCPEo) models demonstrated superior performance for modelling SBP. For estimating mean DBP, appropriate models included Box Cox Power Exponential (BCPE), BCPEo, and Box Cox t (BCT). Nonlinear relationships observed in the data prompted the inclusion of additive functions in the models to accommodate the additive predictor η_k. Penalized splines were employed to address nonlinear relationships.

Selection of the link function was confined to default options for each distribution. Terms for the six models were selected for each distribution parameter using a stepwise selection procedure to identify ideal terms for estimating μ, σ, ν and τ. The stepGAIC function in the gamlss package facilitated this selection. Additionally, the drop1 function in GAMLSS was utilized to identify and remove highly insignificant terms. The selection of suitable effective degrees of freedom for penalized spline functions and the hyperparameter λ employed an automatic procedure (Rigby & Stasinopoulos, 2004).

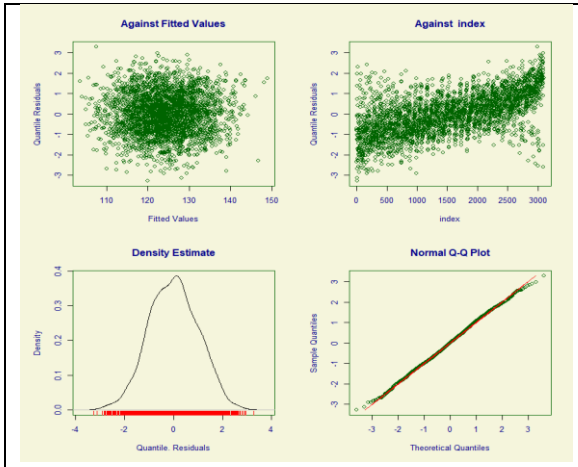
Evaluating model adequacy is crucial for assessing the strengths and weaknesses of the regression model. Diagnostic tools for the six fitted models were examined using normalized quantile residuals developed by Dunn and Smyth (1996). These residuals assume a standard normal distribution when the model is adequate, regardless of the distribution they follow. Based on Figure 2, Table 1, and a GAIC value of 26714.34, the BCPEo distribution demonstrated reasonable fit for approximating SBP. However, for DBP, GAIC favored the BCPE distribution, as shown in Table 1. Although the residuals of BCT seemed appropriate for DBP in terms of residuals, GAIC did not indicate superior fit compared to BCPE. The study observed that residuals of BCPEo and BCPE behaved well, with mean close to zero, variance approximately one, skewness coefficient close to zero, and kurtosis coefficient near 3.

Their normalized quantile residuals exhibited approximate normal distribution, indicating that both models provided excellent estimates for blood pressure. This analysis was further supported by the Anderson-Darling test, where the p-values for the BCPEo and BCPE models were 0.19 and 0.78, respectively. Since these p-values were greater than 0.05, it suggests that the normalized quantile residuals followed a normal distribution. Additionally, other diagnostic tools were employed to detect any potential misspecifications in the models. Worm plots, introduced by van Buuren and Fredriks (2001), were utilized to scrutinize residuals across different regions and identify any regions where the explanatory variables did not adequately fit the data.

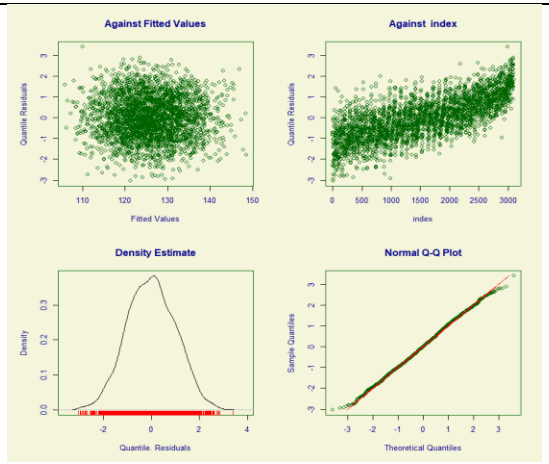
Table 1: Summary of quantile residuals for fitted SBP models and DBP models

	SBP			DBP		
Moments	GG	BCCGo	BCPEo	BCPE	BCPEo	BCT
Mean	0.00	0.00	0.00	0.00	0.00	0.00
Variance	1.00	1.00	1.00	1.00	1.00	1.00
Coef. of skewness	0.00	0.02	0.01	0.00	0.00	0.00
Coef. of kurtosis	2.78	2.76	2.99	2.97	2.97	3.00
Filliben correlation coef.	1.00	1.00	1.00	1.00	1.00	1.00
Anderson darling test p-value	0.01	0.01	0.19	0.78	0.81	0.97
GAIC	26718.09	26718.18	***26714.34***	***24187.95***	24188.80	24255.47

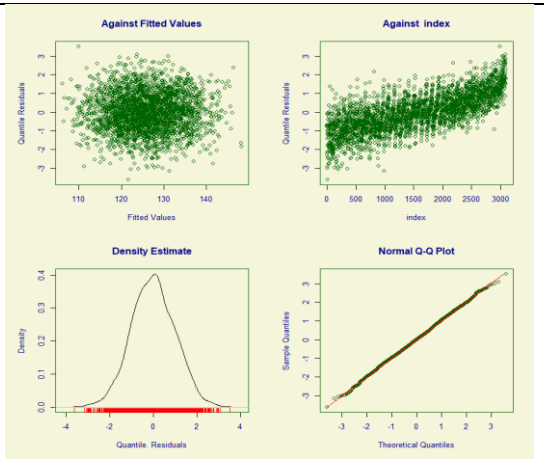
These detrended Q-Q plots illustrate potential deviations from normality. Figures 2 and 3 revealed that all points for the BCPEo and BCPE models fell within the 95% pointwise confidence interval (represented by two elliptical curves) and displayed a flat shape, indicating no evidence of model misspecification for these two models. Conversely, for the GA and BCCGo models, a few observations appeared inside the upper elliptical curve, suggesting less satisfactory performance compared to the BCPEo worm plot. Moreover, while the BCT model demonstrated a nearly perfect fit for residuals as depicted in Figure 3 and the worm plot, it was slightly inferior in terms of goodness of fit.



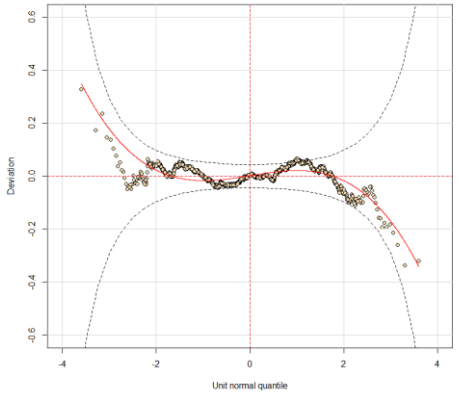
Residuals plots of SBP based on the GA



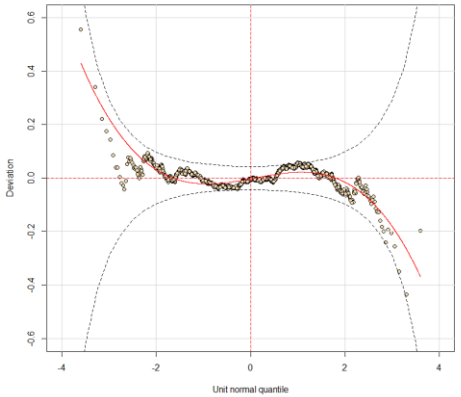
Residuals plots of SBP based on the BCCGo



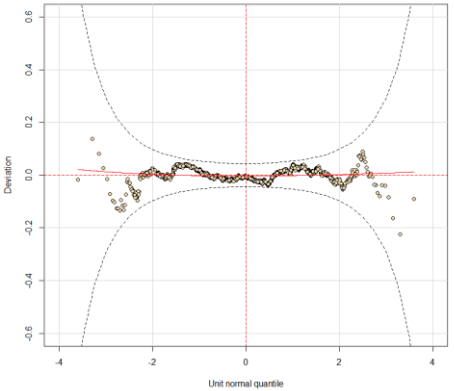
Residuals plots of SBP based on the BCPEo



Worm plots of SBP (GA)

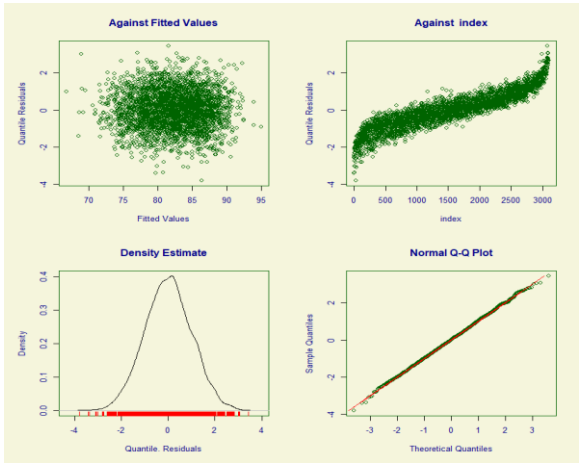


Worm plots of SBP (BCCGo)

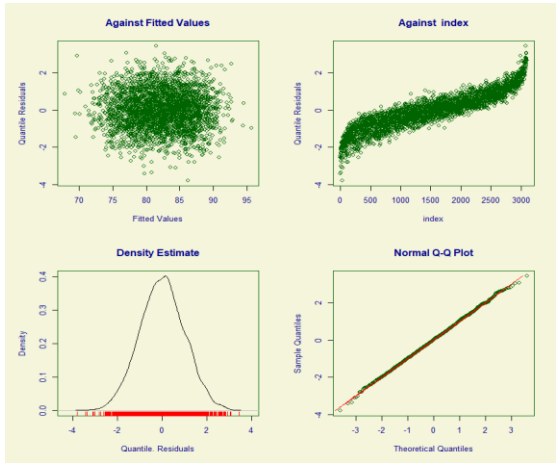


Worm plots of SBP (BCPEo)

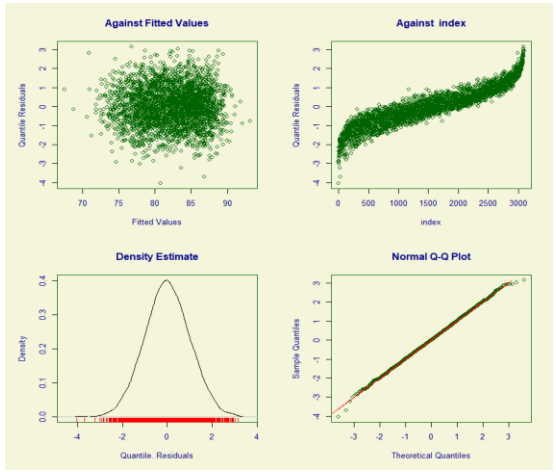
Figure 2: Model diagnostic of SBP



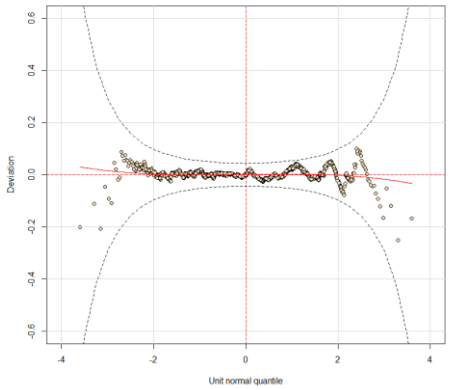
Residuals plots of DBP based on the BCPE



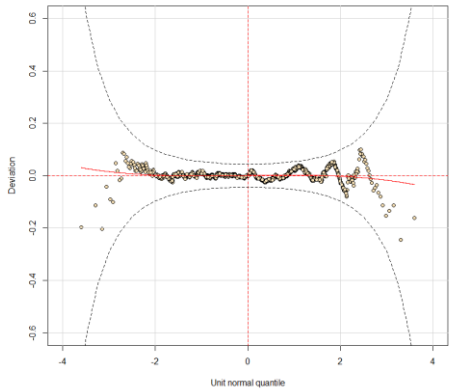
Residuals plots of DBP based on the BCPEo



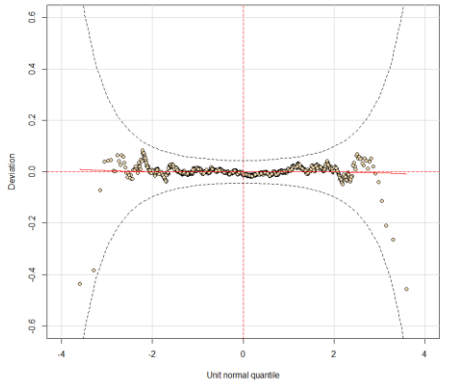
Residuals plots of DBP based on the BCT



Worm plots of DBP (BCPE)



Worm plots of DBP (BCPEo)



Worm plots of DBP (BCT)

Figure 3: Model diagnostic of DBP

3.3 Model selection

Statistical regression models were compared in this study to determine the most suitable approach for modeling blood pressure, with model selection criteria aimed at identifying the best asymmetric model among competing alternatives. Table 2 presents the goodness of fit for four different classes of regression models using AIC and BIC values as criteria for assessment. The linear model (LM) assumes a linear relationship and normal distribution, but in this study, the blood pressure did not follow a normal distribution, thereby violating the assumption of normality. This limitation was evident by p-value <0.01 and AIC and BIC values compared to other classical models. The GLM addressed normality, homoscedasticity, and skewness in the response variable (SBP/DBP), but it had its shortcomings, particularly in handling nonlinear structures in some continuous variables. The GAM was employed to model nonlinear predictors using smoothing functions but was constrained by its restriction to exponential family distributions.

Table 2 Comparing the Models for blood pressure estimates.

Class	SBP			DBP		
	AIC	BIC	P-value	AIC	BIC	P- value
LM	27147.33	27231.86	< 0.01***	24330.09	24396.50	< 0.01***
GLM	26908.9	26999.47	< 0.01***	24353.13	24431.62	< 0.01**
GAM	26837.81	26969.82	< 0.01***	24285.79	24390.23	0.04
GAMLSS	***26714.34***	***26942.22***	***0.19***	***24187.95***	***24380.44***	***0.78***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

According to the Anderson-Darling test, neither the LM, GLM, nor GAM models were deemed adequate, as the distribution of blood pressure did not align with exponential family distributions. Unlike LM, GLM, and GAM, the generalized additive models for location, scale, and shape (GAMLSS) approach considered the variance, skewness, and kurtosis in terms of predictors, offering greater flexibility. Relaxing some assumptions of the models and including smoothing functions, as seen in GAM, significantly improved AIC and BIC values. Furthermore, allowing the model to be more flexible by modeling scale and shape parameters (σ, ν, τ) and permitting blood pressure to follow any distribution resulted in a substantial decrease in AIC and BIC values.

The GAMLSS model emerged as the preferred approach for estimating SBP and DBP, with lower AIC and BIC values compared to other competing models. This preference was corroborated by residual plots, which appeared normal (Figure 2 and Figure 3). Additionally, the Anderson-Darling

test supported the adequacy of the GAMLSS model, while other models (LM, GLM, and GAM) violated their assumptions.

3.4 Estimates of blood pressure

As established in the preceding section, the GAMLSS model emerged as the appropriate regression model for predicting blood pressure, with accurate estimates based on the BCPEo (SBP) and BCPE (DBP) distributions. These distributions, BCPEo and BCPE, offer high flexibility by allowing modeling for location, scale, skewness, and various types of kurtosis (leptokurtosis, platykurtosis, and mesokurtosis).

Table 3 presents the estimates of the GAMLSS using BCPEo and BCPE distributions for the mean, dispersion, and shape parameters. Given the involvement of smoothers in our analysis, Table 3 may not provide accurate information for testing the significance of the terms. The t-test in Table 3 examines the linear part in the explanatory variables rather than testing the smoothing effects. To accurately test the significance of the covariates in the model, the generalized likelihood ratio test (GLRT) was employed as a more suitable criterion compared to the Wald test (t-value), which is useful for testing statistical significance in the linear term rather than the overall smoothing term for p-spline functions.

Table 4 indicated factors influencing SBP and DBP on different distribution parameters. For the median parameter μ, SBP was significantly associated with residence, wealth quantiles, education, sex, alcohol consumption, age, BMI, glucose, and hemoglobin. The scale parameter σ (SBP) was significantly linked with age, weight, BMI, height, and alcohol consumption. Additionally, skewness (SBP) parameter ν was influenced by educational attainment level. Similarly, for DBP, residence, wealth, education, alcohol consumption, age, BMI, and hemoglobin significantly contributed to the median, while scale parameter σ of DBP was significantly associated with BMI, wealth, age, and height, and skewness parameter ν was linked with BMI. Education attainment level had an effect on DBP based on the kurtosis parameter τ.

Table 3: Estimates of blood pressure using GAMLSS model

Covariates of SBP	Estim	Std. E	t value	Pr(>|t|)	Covariates of DBP	Estim	Std.E	t value	Pr (>|t|)
μ Coefficient using a log link function					μ Coefficient using a log link function				
(Intercept)	4.48	0.03	141.88	< 0.01 ***	(Intercept)	53.93	2.37	22.80	< 0.01 ***
Residence (Rural vs Urban)	-0.02	0.01	-3.64	< 0.01 ***	Residence (Rural vs Urban)	-2.00	0.50	-4.04	< 0.01 ***
Wealth (Middle vs Poor)	0.01	0.01	1.58	0.11	Wealth (Middle vs Poor)	0.98	0.61	1.60	0.11
Wealth (Rich vs Poor)	-0.01	0.01	-1.65	0.10.	Wealth (Rich vs Poor)	-0.86	0.65	-1.33	0.18
Education (Pri. vs No Ed.)	-0.02	0.01	-1.57	0.12	Education (Pri. vs No Ed.)	0.00	0.69	-0.01	0.10
Education (Sec. vs No Ed.)	-0.03	0.01	-3.34	< 0.01 ***	Education (Sec. vs No Ed.)	-0.58	0.72	-0.81	0.42
Education (Hig. vs No Ed.)	-0.06	0.01	-4.02	< 0.01 ***	Education (High. vs No Ed.)	-3.29	1.00	-3.30	< 0.01 ***
Sex (Female vs Male)	-0.03	0.01	-5.52	< 0.01 ***	Alcohol(Yes vs No)	1.33	0.43	3.11	< 0.01**
Alcohol(Yes vs No)	0.02	0.01	3.15	< 0.01**	pb(Age)	0.08	0.03	2.67	0.01**
pb(Age)	0.00	0.00	9.59	< 0.01 ***	pb(Glucose)	0.52	0.04	13.36	< 0.01 ***
pb(BMI)	0.00	0.00	10.21	< 0.01 ***	pb(Hemoglobin)	0.93	0.11	8.27	< 0.01 ***
pb(Glucose)	0.00	0.00	1.91	0.06.	σ Coefficient using a log link function				
pb(Hemoglobin)	0.01	0.00	4.72	< 0.01 ***	< 0.01 ***	-1.13	0.27	-4.18	< 0.01 ***
σ Coefficient using a log link function					pb(BMI)	-0.01	0.00	-6.17	< 0.01 ***
Intercept)	-4.65	1.01	-4.62	< 0.01 ***	Wealth (Middle vs Poor)	-0.12	0.04	-3.44	< 0.01 ***
pb(Age)	0.01	0.00	6.27	< 0.01 ***	Wealth (Rich vs Poor)	-0.04	0.03	-1.36	0.18
pb(Weight)	-0.02	0.01	-3.22	< 0.01**	pb(Age)	0.00	0.00	2.36	0.02 *
pb(BMI)	0.05	0.02	2.77	< 0.01**	pb(Height)	-0.35	0.15	-2.35	0.02 *
pb(Height)	1.52	0.60	2.52	0.01*	ν Coefficient using an identity link function				
Education (Pri. vs No Ed.)	-0.08	0.04	-2.08	0.04 *	(Intercept)	0.90	0.42	2.17	0.03*
Education (Sec vs No Ed.)	-0.12	0.04	-3.00	< 0.01**	pb(BMI)	-0.02	0.02	-0.98	0.33
Education (Hig. vs No Ed.)	-0.09	0.06	-1.50	0.13	τ Coefficient using a log link function				
					(Intercept)	0.57	0.06	9.71	< 0.01 ***
ν Coefficient using an identity link function					Alcohol(Yes vs No)	0.22	0.08	2.62	0.01 **
(Intercept)	0.02	0.24	0.08	0.94					
Education (Pri. vs No Ed.)	-0.76	0.29	-2.62	< 0.01**					
Education (Sec. vs No Ed.)	-0.99	0.29	-3.44	< 0.01 ***					
Education (Hig.vs No Ed.)	-1.08	0.47	-2.28	0.02 *					
τ Coefficient using a log link function									
(Intercept)	0.82	0.05	15.95	< 0.01 ***					

Table 4 Significance testing of SBP/DBP using GAMLSS model

SBP	AIC	LRT	Pr(Chi)	DBP	AIC	LRT	Pr(Chi)
μ terms				**μ terms**			
Residence	26726	13.25	< 0.01 ***	Residence	24202	14.71	< 0.01 ***
Wealth	26722	12.39	< 0.01**	Wealth	24193	9.44	0.01 **
Education	26729	20.95	< 0.01 ***	Education	24197	15.29	< 0.01**
Sex	26740	27.11	< 0.01 ***	Alcohol	24196	10.08	< 0.01**
Alcohol	26722	9.50	< 0.01**	pb(Age)	24201	18.26	< 0.01 ***
pb(Age)	26836	130.62	< 0.01 ***	pb(BMI)	24369	184.59	< 0.01 ***
pb(BMI)	26844	137.20	< 0.01 ***	pb(Hemoglobin)	24252	66.50	< 0.01 ***
pb(Glucose1)	26716	4.29	0.05 *	**σ terms**			
pb(Hemoglobin)	26738	29.92	< 0.01 ***	pb(BMI)	24234	54.07	< 0.01 ***
σ terms				Wealth	24195	10.47	< 0.01**
pb(Age)	26752	41.73	< 0.01 ***	pb(Age)	24194	10.34	0.01 **
pb(Weight)	26719	9.61	0.01 *	pb(Height)	24191	5.30	0.03 *
pb(BMI)	26721	14.07	0.01**	**ν terms**			
pb(Height)	26717	8.34	0.03 *	pb(BMI)	24191	7.05	0.03*
Alcohol	26717	8.43	0.03*	**τ terms**			
ν terms				Alcohol	24192	5.94	0.01*
Education	26720	11.27	0.01 **				

Nonparametric smoothing functions cannot be described simply in mathematical form, hence the effect of the smooth function is depicted in Figure 4 and Figure 5. Term plots in the GAMLSS package were used to illustrate the effects of covariates on each distributional parameter. These plots provide a clearer depiction of how the coefficients β change against the covariates. Figure 4 and Figure 5 highlight the impact of different covariates on SBP and DBP with respect to their distributional parameters. Blood pressure increases with age, BMI, glucose (SBP only), and hemoglobin (DBP only) for the median parameter. Additionally, increased blood pressure is noted among adults living in urban areas, with no education, middle-class individuals, alcohol consumers, and male adults (SBP). The scale parameter shows that blood pressure decreases with weight (Figure 4, BMI, and height (Figure 5). With the shape parameter, blood pressure is associated with education (SBP), BMI, and alcohol intake (DBP). Weight and height were not significant with respect to the location parameter μ, but significant with the scale parameter, possibly due to dispersion in the response variable. Skewness parameter has a higher effect on education level and SBP, indicating that blood pressure reduces with education level. The skewness parameter of DBP suggests that adults who consume alcohol are at greater risk of hypertension. Smoking or marital status did not influence blood pressure in Namibia. Furthermore, nonlinear relationships can be observed in BMI, hemoglobin, and age.

The same covariates identified for the location parameter μ of the GAMLSS model for SBP were also identified for LM. Additionally, the nonlinear effect on the location parameter μ for both SBP and DBP assumed a similar form as that obtained for GAM.

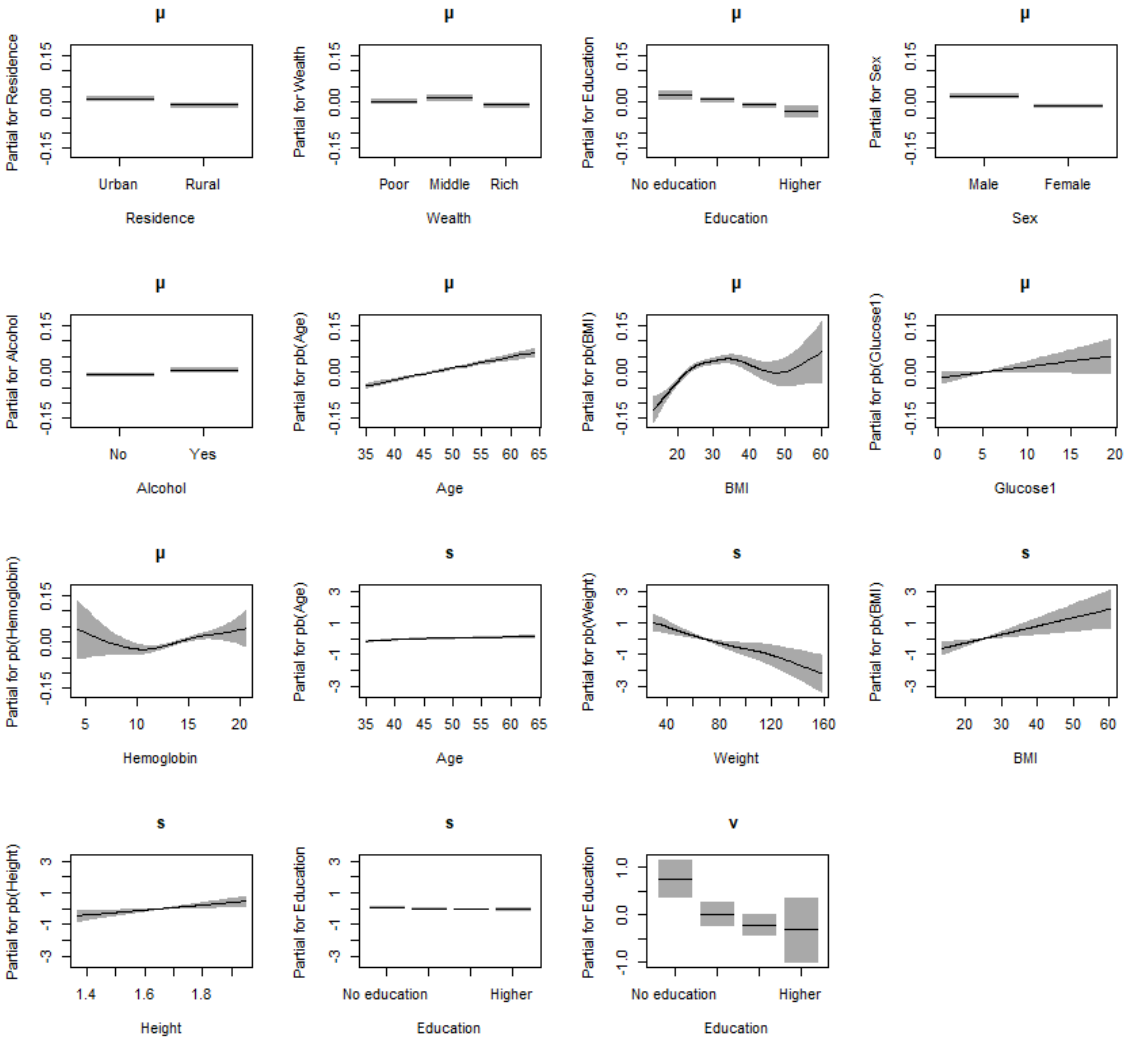


Figure 4: Estimated effect of smoothing function of the BCPEo

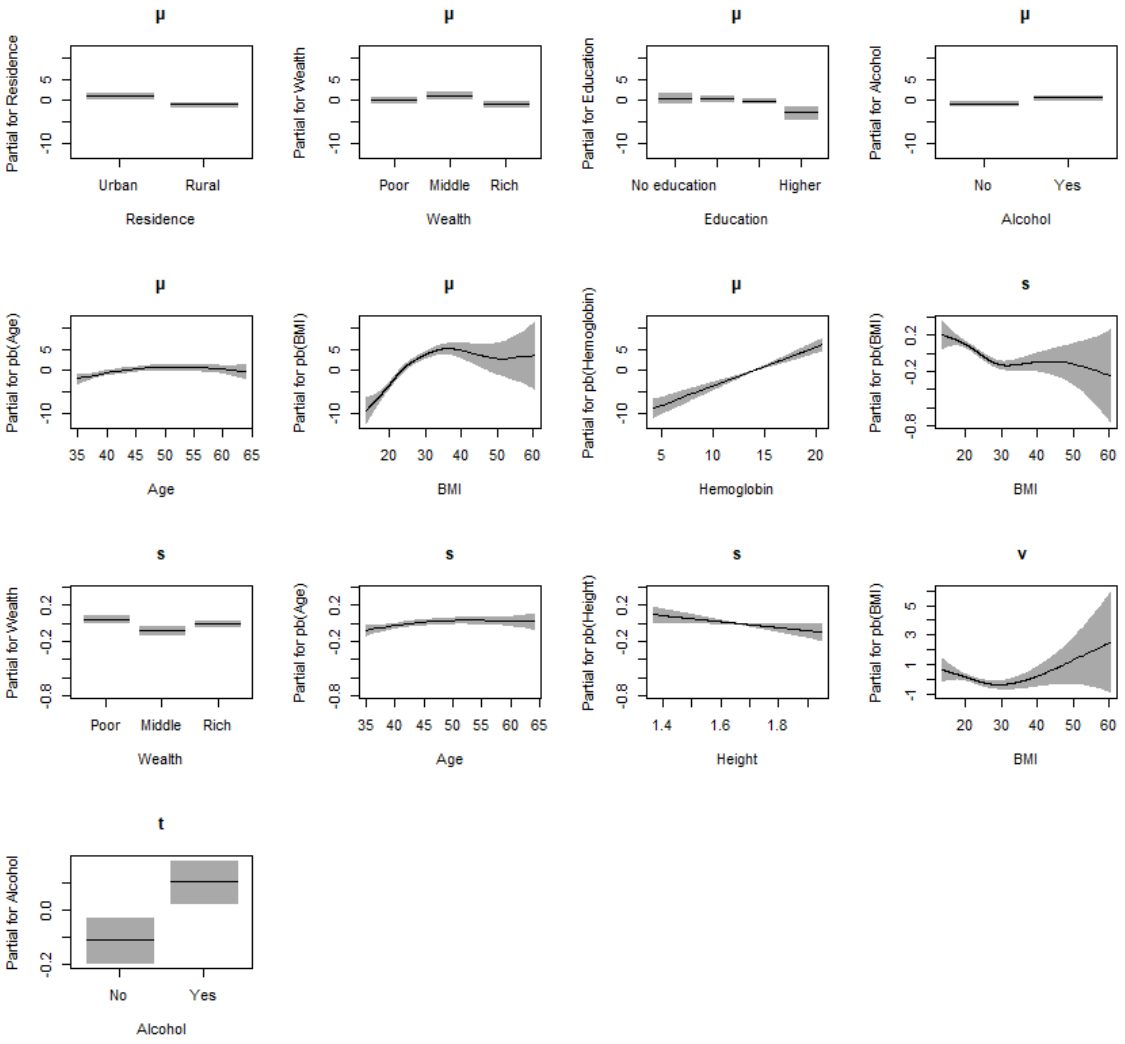


Figure 5: Estimated effect of smoothing function of the BCPE

4. Discussion

A multitude of regression models have been developed since the 19th century to address prediction uncertainties. However, some of these models may prove inadequate in representing the data, leading to flawed predictions and interpretations of results. This study delved into how LM, GLM, GAM, and GAMLSS were employed in predicting blood pressure in Namibia, aiming to identify the most suitable model for estimating blood pressure. It became evident that LM and GLM were not adequate, as they violated their model assumptions. Additionally, GAM lacked flexibility in terms of goodness of fit, and the chosen distribution for modeling was unsatisfactory, as depicted in Figure 1, where kurtosis in the response variable was not accounted for. Consequently, based on

the study's findings, the GAMLSS model emerged as the only model that adhered to its statistical model assumptions. Through judgment mechanisms employing AIC, BIC, and diagnostic tools, GAMLSS was deemed the best fit for estimating blood pressure, effectively overcoming limitations associated with other models.

The estimated GAMLSS model revealed significant associations between blood pressure and various factors, including residence, wealth quintiles, education, sex, alcohol consumption, age, BMI, glucose, height, weight, and hemoglobin. Notably, adults residing in urban areas were more prone to increased blood pressure compared to rural residents, a trend consistent with findings from studies conducted in other geographical contexts such as studies conducted in Malawi (Price et al., 2018), Nigeria (Omisore et al., 2018) and Ethiopia (Abebe et al., 2015). Elevated blood pressure in urban areas could be attributed to unhealthy lifestyle behaviors, socio-economic stresses linked with urban living, environmental changes, and social stressors (Craig et al., 2018; Ibrahim & Damasceno, 2012).

Additionally, this study observed higher blood pressure among middle-class individuals and poorer adults aged between 35-64 years, contrary to some previous studies. The research conducted by Antignac et al. (2018) indicated that in low-income areas of sub-Saharan Africa, individuals with lower economic status faced a greater risk of hypertension compared to those in middle- and higher-income brackets. Conversely, Sharma et al. (2021), in their study in Mthatha town (South Africa), found a positive association between higher income levels (≥R1000) and hypertension. Furthermore, individuals with no educational background exhibited higher blood pressure levels, aligning with findings by Chen and Tan's study (2013) attributing this to a lack of awareness and understanding of hypertension risks and preventive measures.

Elevated SBP was found to be more prevalent among males than females, consistent with several other studies by Saka et al. (2020), Alhawari et al. (2018) and Craig et al. (2018). However, females showed higher diastolic blood pressure compared to males. This discrepancy could be attributed to gender-specific lifestyle choices, work nature, and familial responsibilities. Moreover, alcohol consumption and higher BMI were identified as risk factors for hypertension, consistent with the study by Craig et al. (2018).

Age emerged as a significant contributor to elevated blood pressure, especially among adults aged 35-64, in line with findings of Craig et al. (2018) and Hosseini et al. (2015). Aging-related physiological changes, including arterial stiffness and neurohumoral dysfunction, were implicated in the escalation of blood pressure.

Additionally, the study revealed positive associations between glucose levels and both SBP and DBP, consistent with prior research linking hyperglycemia and metabolic disorders to increased blood pressure. Moreover, hemoglobin levels were positively associated with DBP, with a negative association observed for SBP at lower hemoglobin levels, aligning with findings from other studies. However, smoking showed no association with blood pressure levels in the final model, suggesting that further investigation with a larger sample size may be warranted to elucidate the potential impact of smoking on blood pressure in Namibia.

5. Conclusion and Recommendations

This study introduced the GAMLSS model as a robust regression framework for estimating blood pressure in Namibia. Unlike traditional regression models, GAMLSS offers versatility by accommodating a wide array of distributions for the response variable (blood pressure) and explicitly modeling its mean, dispersion, skewness, and kurtosis. This model allows for the consideration of both linear and nonlinear relationships, as evidenced by the non-linear association between BMI and blood pressure depicted in Figure 4 and Figure 5. Comparisons with other regression classes (LM, GLM, and GAM) revealed that GAMLSS outperformed them, as evidenced by superior results indicated by the AIC criteria method and residuals plot. Moreover, GAMLSS facilitates model comparison and the detection of data misspecification, thereby enhancing our understanding of which model best fits the data.

The study identified several factors associated with blood pressure among Namibian adults, including demographic (age, sex, gender, and residence), physiological (glucose, hemoglobin, height, and BMI), socio-economic (education and wealth), and lifestyle factors (alcohol intake). Additionally, it was observed that blood pressure tends to increase with age, highlighting the importance of age as a contributing factor.

To further explore the spatial dynamics of blood pressure in Namibia, the study recommends conducting a comprehensive spatial analysis to investigate potential regional variations in blood pressure levels. Such an analysis could provide valuable insights into the geographical determinants of blood pressure and inform targeted interventions aimed at addressing disparities across different regions of Namibia.

References

Abebe, S. M., Berhane, Y., Worku, A., & Getachew, A. (2015). Prevalence and associated factors of hypertension: A crossectional community based study in Northwest Ethiopia. *PLoS ONE*, *10*(4), 1–11. https://doi.org/10.1371/journal.pone.0125210

Adeloye, D., & Basquill, C. (2014). Estimating the prevalence and awareness rates of hypertension in Africa: A systematic analysis. *PLoS ONE*, *9*(8), 1–17. https://doi.org/10.1371/journal.pone.0104300

Akaike, H. (1974). A New Look at the Statistical Model Identification. *IEEE Transactions on Automatic Control*, *19*(6), 716–723.

Alhawari, H. H., Al-Shelleh, S., Alhawari, H. H., Al-Saudi, A., Aljbour Al-Majali, D., Al-Faris, L., & Alryalat, S. A. (2018). Blood Pressure and Its Association with Gender, Body Mass Index, Smoking, and Family History among University Students. *International Journal of Hypertension*, *2018*, 1–5. https://doi.org/10.1155/2018/4186496

Antignac, M., Diop, I. B., De Terline, D. M. Q., Kramoh, K. E., Balde, D. M., Dzudie, A., … Jouven, X. (2018). Socioeconomic status and hypertension control in sub-saharan Africa the multination EIGHT study (evaluation of hypertension in Sub-Saharan Africa). *Hypertension*, *71*(4), 577–584. https://doi.org/10.1161/HYPERTENSIONAHA.117.10512

Chen, X., & Tan, X. (2013). Associations of Education with Blood Pressure in Hypertensive Patients: A Chinese Community Survey. *Chinese Medicine*, *04*(03), 97–100. https://doi.org/10.4236/cm.2013.43014

Craig, L. S., Gage, A. J., & Thomas, A. M. (2018). Prevalence and predictors of hypertension in Namibia : A national-level cross-sectional study. *PLoS ONE*, *13*(9), 1–14.

Dunn, P. K., & Smyth, G. K. (1996). Randomized Quantile Residuals. *Journal of Computational AndGraphical Statistics*, *5*(April), 236–244.

Hosseini, M., Baikpour, M., Yousefifard, M., Fayaz, M., Koohpayehzadeh, J., Ghelichkhani, P., … Gouya, M. M. (2015). Blood pressure percentiles by age and body mass index for adults. *EXCLI Journal*, *14*, 465–477. https://doi.org/10.17179/excli2014-635

Ibrahim, M. M., & Damasceno, A. (2012). Hypertension in developing countries. *The Lancet*, *380*(9841), 611–619. https://doi.org/10.1016/S0140-6736(12)60861-7

Mills, K. T., Bundy, J. D., Kelly, T. N., Reed, J. E., Kearney, P. M., Reynolds, K., … He, J. (2016). Global Disparities of Hypertension Prevalence and Control: A Systematic Analysis of Population-Based Studies From 90 Countries. *Circulation*, Vol. 134, pp. 441–450. https://doi.org/10.1161/CIRCULATIONAHA.115.018912

Ministry of Health and Social Services Namibia. (2017). *National Multisectoral Strategic Plan for Prevention and Control of Non-Communicable Diseases (NCDs) in Namibia 2017/18 – 2021/22*.

MoHSS and ICF International. (2014). *Namibia Demographic Health Survey 2013*. Retrieved from https://dhsprogram.com/pubs/pdf/fr298/fr298.pdf

Nyirenda, M. J. (2016). Non-communicable diseases in sub-Saharan Africa: Understanding the drivers of the epidemic to inform intervention strategies. *International Health*, *8*(3), 157–158. https://doi.org/10.1093/inthealth/ihw021

Omisore, A. G., Omisore, B., Abioye-Kuteyi, E. A., Bello, I. S., & Olowookere, S. A. (2018). In-school adolescents' weight status and blood pressure profile in South-western Nigeria: Urban-rural comparison. *BMC Obesity*, *5*(1), 1–9. https://doi.org/10.1186/s40608-018-0179-3

Price, A. J., Crampin, A. C., Amberbir, A., Kayuni-Chihana, N., Musicha, C., Tafatatha, T., … Nyirenda, M. (2018). Prevalence of obesity, hypertension, and diabetes, and cascade of care in sub-Saharan Africa: a cross-sectional, population-based study in rural and urban Malawi. *The Lancet Diabetes and Endocrinology*, *6*(3), 208–222. https://doi.org/10.1016/S2213-8587(17)30432-1

Rigby, R. A., & Stasinopoulos, M. D. (2005). Generalized additive models for location , scale. *Applied Statistics*, *54*(Part 3), 507–554.

Rigby, & Stasinopoulos, D. M. (2004). Smooth centile curves for skew and kurtotic data modelled using the Box-Cox power exponential distribution. *Statistics in Medicine*, *23*(19), 3053–3076. https://doi.org/10.1002/sim.1861

Rodilla, E., Saura, A., Jimenez, I., Mendizabal, A., Pineda-Cantero, A., Lorenzo-Hernandez, E., … Rabadan, E. (2020). Association of Hypertension with All-Cause Mortality among Hospitalized Patients with COVID-19. *Journal of Clinical Medicine*, *9*(3136), 1–14.

Ryan, M. J. (2013). An update on immune system activation in the pathogenesis of hypertension. *Hypertension*, Vol. 62, pp. 226–230. https://doi.org/10.1161/HYPERTENSIONAHA.113.00603

Saka, M., Shabu, S., & Shabila, N. (2020). Prevalence of hypertension and associated risk factors

in older adults in Kurdistan, Iraq. *Eastern Mediterranean Health Journal*, *26*(3), 265–272. https://doi.org/10.26719/emhj.19.029

Schwarz, G. (1978). Estimating the Dimension of a Model. *The Annals of Statistics*, *6*(2), 461–464.

Sharma, J. R., Mabhida, S. E., Myers, B., Apalata, T., Nicol, E., Benjeddou, M., … Johnson, R. (2021). Prevalence of Hypertension and Its Associated Risk Factors in a Rural Black Population of Mthatha Town , South Africa. *International Journal of Environmental Research and Public Health*, *18*(1215), 1–17.

Stanaway, J. D., Afshin, A., Gakidou, E., Lim, S. S., Abate, D., Abate, K. H., … Murray, C. J. L. (2018). Global, regional, and national comparative risk assessment of 84 behavioural, environmental and occupational, and metabolic risks or clusters of risks for 195 countries and territories, 1990-2017: A systematic analysis for the Global Burden of Disease Stu. *The Lancet*, *392*(10159), 1923–1994. https://doi.org/10.1016/S0140-6736(18)32225-6

Tadic, M., Saeed, S., Grassi, G., Taddei, S., & Mancia, G. (2021). Hypertension and COVID-19 : Ongoing Controversies. *Frontiers in Cardiovascular Medicine*, *8*(February), 6–11. https://doi.org/10.3389/fcvm.2021.639222

Van Buuren, S., & Fredriks, M. (2001). Worm plot: A simple diagnostic device for modelling growth reference curves. *Statistics in Medicine*, *20*(8), 1259–1277. https://doi.org/10.1002/sim.746

World Health Organization. (2020). *Non-communicable diseases Progress Monitor 2020*. https://doi.org/10.5005/jp/books/11410_18

Yan, Q., Sun, D., Li, X., Chen, G., Zheng, Q., Li, L., … Feng, B. (2016). Association of blood glucose level and hypertension in Elderly Chinese Subjects : a community based study. *BMC Endocrine Disorders*, *16:40*(2016), 1–8. https://doi.org/10.1186/s12902-016-0119-5

Zhou, B., Bentham, J., Di Cesare, M., Bixby, H., Danaei, G., Cowan, M. J., … Eggertsen, R. (2017). Worldwide trends in blood pressure from 1975 to 2015: a pooled analysis of 1479 population-based measurement studies with 19·1 million participants. *The Lancet*, *389*(10064), 37–55. https://doi.org/10.1016/S0140-6736(16)31919-5