Comparing Frequentist and Bayesian Quantile Regression Models for Child Hypertension in South Africa

Anesu Gelfand Kuhudzai[1,2,*] and Kolentino N. Mpeta[1]

[1]*Business Statistics and Operations Research Department, Faculty of Economic and Management Sciences, North-West University, South Africa*

[2]*Statistical and Data Science Consultant, Statistical Consultation Services, University of Johannesburg, Johannesburg, South Africa*

Abstract: *Background*: Traditional approaches to modelling paediatric hypertension in South Africa have relied on descriptive or mean regression methods, which inadequately capture risk factors driving the distributional extremes of blood pressure. Quantile regression provides a flexible alternative, and Bayesian methods offer advantages in precision and uncertainty estimation, yet their comparative performance has not been assessed in this context.

Methods: Nationally representative cross-sectional data from 1,812 adolescents (15–17 years) in the South African National Income Dynamics Study (NIDS) Wave 5 (2017–2018) was analysed. Frequentist and Bayesian quantile regression models were fitted for systolic (SBP) and diastolic blood pressure (DBP) at the 75th and 95th percentiles. Model performance was compared in terms of parameter estimates, interval precision, and convergence diagnostics.

Results: BMI and gender were consistent predictors of both SBP and DBP across models. Bayesian quantile regression additionally identified age, race, and pulse rate as significant risk factors for upper quantiles. Bayesian credible intervals were consistently narrower than frequentist confidence intervals, indicating improved precision. Convergence diagnostics confirmed robust posterior inference.

Conclusion: Bayesian quantile regression provides more efficient inference than the frequentist alternative when modelling health outcomes concentrated in distributional extremes. This is the first study to apply Bayesian quantile regression to paediatric hypertension in South Africa, demonstrating both methodological value and empirical insights into adolescent health risks.

Keywords: Bayesian quantile regression, frequentist quantile regression, paediatric hypertension, South Africa, extreme values.

INTRODUCTION

Hypertension in children and adolescents has emerged as a growing public health concern worldwide. A recent meta-analysis estimated that approximately 4% of individuals under 19 years of age are affected globally [1]. In South Africa, reported prevalence rates vary substantially across regions and age groups, with studies documenting rates ranging from 5% to over 20% among school-going children and adolescents [1, 2]. Left undiagnosed and untreated, paediatric hypertension contributes to early vascular changes and increases the likelihood of adult cardiovascular disease [3].

Despite its importance, hypertension in South African children remains under-researched and often underestimated. Most existing studies have focused on descriptive statistics or mean regression models to investigate risk factors such as body mass index (BMI), gender, and lifestyle behaviours [4]. However, these approaches are limited because they fail to account for the distributional extremes of blood pressure (BP), where paediatric hypertension is most pronounced. Analysing mean effects may mask critical insights into risk factors that exert disproportionate influence at the upper quantiles of systolic and diastolic BP.

Quantile regression offers an alternative, enabling researchers to study how risk factors influence the tails of the BP distribution [5]. Within this framework, Bayesian quantile regression is particularly attractive, as it incorporates prior information, accounts for parameter uncertainty, and provides more precise inference compared to frequentist approaches [6, 7]. While both methods have been applied in other health contexts, their comparative performance in modelling South African child hypertension remains unexplored.

Previous South African studies on child hypertension have relied primarily on descriptive or mean regression approaches, which inadequately capture the effects of risk factors at the distributional extremes of blood pressure. This methodological gap limits understanding of the drivers of paediatric hypertension, particularly among adolescents who are at high risk. To the best of our knowledge, no study has

*Address correspondence to this author at the Statistical and Data Science Consultant, Statistical Consultation Services, University of Johannesburg, Johannesburg, South Africa; E-mail: gelfand9@yahoo.com

E-ISSN: 1929-6029/25

applied Bayesian quantile regression to model child hypertension in South Africa, nor compared its performance with the frequentist approach.

MATERIALS AND METHODS

This section describes the implementation of the Bayesian Approach to quantile regression framework to investigate the impact of risk factors on upper quantiles of blood pressure`s distribution (SBP and DBP). Thus, the study variables, data, theoretical model and data analysis techniques.

Data and Variables

The study used nationally representative cross-sectional data obtained from the South African National Income Dynamics Study (NIDS) Wave 5 Household survey conducted between 2017 and 2018. The purpose of NIDS was to evaluate the welfare of South Africans nationwide. The study utilised a clean data set of 1 812 respondents obtained after dropping 26 cases with missing data for the variables used in the study and participants aged above 18.

Systolic and diastolic blood pressure, as well as modifiable risk factors like body mass index (BMI), exercise, cigarette smoking, depression, and perceived health status, are the variables included in this study. Non-modifiable risk factors include age, gender, race and pulse rate. While age, BMI, gender, race, exercise, cigarette smoking, depression, subjective health status, and pulse rate are the independent variables, systolic and diastolic blood pressure are the outcome variables. Since the participants in this study are between 15 to 17 years, the normal blood pressure ranges are the following: SBP – 100 to 120 mmHg and DBP – 60 to 80 mmHg [8].

The University of Cape Town Faculty of Commerce Ethics Committee provided ethics approval for the NIDS study, and each study participant provided informed consent.

Bayesian Quantile Regression

Quantile regression Bayesian Models are a solution to the drawbacks of the frequentist quantile regression models and provide accurate estimates by using information from quantile levels and locations [9]. Bayesian inference is one of the most popular approaches for regression analysis since it provides with an entire posterior distribution of a parameter of interest as well as inclusion of parameter uncertainty and prior information about data [10]. Bayesian Quantile Regression can improve model accuracy (in terms of interval estimates, bias/variance trade-offs and tail performance) compared to frequentist quantile regression especially in small samples, extreme quantiles, when the error distribution is skewed and clustered, missing or censored data contexts [9].

Considering the standard linear model where $x_i \in \Re^k$ and $\beta \in \Re^k$ are column vectors of size k and $y_i \in \Re$ is a scalar variable:

$$y_i = x_i^T \beta + \varepsilon_i \tag{1}$$

Assuming that $E(\varepsilon \mid x) = 0$, yields the conditional mean model, then

Med $(\varepsilon \mid x) = 0$, yields the conditional median model. In the conditional mean model, the regression coefficients by solving:

$$\hat{\beta} = \arg\min_{\beta \in \Re^k} \sum_{i=1}^{n} \left(y_i - x_i^T \beta \right)^2 \tag{2}$$

In median regression, the coefficients can be solved in the same way, but an estimate of β can be obtained by minimising the sum of the absolute deviations:

$$\hat{\beta} = \arg\min_{\beta \in \Re^k} \sum_{i=1}^{n} \left| y_i - x_i^T \beta \right| \tag{3}$$

Quantile regression proceeds by extending the median case to all other quantiles of interest:

$$\hat{\beta}_\tau = \arg\min_{\beta \in \Re^k} \sum_{i=1}^{n} \rho_\tau \left(y_i - x_i^T \beta \right) \tag{4}$$

Where $\rho_\tau(x) = x(\tau - 1(x < 0))$ and where $I(\cdot)$ denotes the indicator function. For any $\tau \in (0,1)$, the loss function ρ_τ assigns a weight of τ to positive residuals and a weight of τ to positive residuals and a weight of $(1-\tau)$ to negative residuals. The quantity $\widehat{\beta_\rho}$ is called the τth regression quantile. Note that the case where τ equals 0.5 corresponds to minimising the sum of absolute residuals, i.e., median regression. The objective function in Equation 4 can be minimised using the method of [11].

Considering that the Bayesian approach to quantile regression is normally performed by formulating a

likelihood function based on the asymmetric Laplace distribution (ALD), irrespective of the actual distribution of the data, [12] propose a three-parameter ALD with a skewness parameter that can be used directly to model the quantile of interest:

$$f\left(x \mid \mu,\sigma,\tau\right)=\frac{\tau\left(1-\tau\right)}{\sigma}\exp\left\{-\rho_{\tau}\left(\frac{x-\mu}{\sigma}\right)\right\} \quad (5)$$

Minimising equation 4 is equivalent to maximising a regression likelihood using ALD errors with $\mu = x_i^T\beta$.

Using the ALD, the quantile of interest $\left(\tau\right)$ has to be specified, and priors should be put on the model parameters β and σ. The resulting posterior distribution can be represented as follows:

$$\psi\left(\beta,\sigma \mid y,x,\tau\right)\alpha\,\pi\left(\beta,\sigma\right)\prod_{i=1}^{n}ALD\left(y_i \mid x_i^T\beta,\sigma,\tau\right) \quad (6)$$

Where $\pi\left(\beta,\sigma\right)$ is the joint prior on the regression parameters. Inference about the model parameters then follows Bayesian procedures.

Data Analysis

Descriptive statistics in the form of proportions for categorical variables were produced using IBM Statistical Package for the Social Sciences (SPSS) version 30. The classical quantile regression models were fitted using the quantreg R package [13].

MCMC methods from the R package MCMCpack were used to build the Bayesian quantile regression models [14]. The coda package can then summarize the coda MCMC objects that are returned by the models generated by MCMCpack [15]. In addition to diagnostic tests of convergence, the coda package has functions for charting and summarising the results of the MCMC simulations [16].

Two quantile models at the 75th and 95th percentiles are examined in this study. It makes more sense to model high systolic and diastolic blood pressure values, which correspond to the upper distribution of either SBP or DBP, when simulating hypertension [17].

In contemporary Bayesian computing, MCMC algorithms have become highly effective and well-liked techniques for fitting Bayesian statistical models. Sinharay (2003) asserts that the main factor contributing to the popularity and utility of MCMC algorithms is their ease of fitting complex models in contrast to more conventional methods like maximum likelihood estimation (MLE).

RESULTS

This section presents the study's empirical findings in the form of tables and figures.

Table **1** summarizes the distribution of blood pressure by demographic and lifestyle characteristics. Blood pressure status is categorised as normal or not normal, with normal SBP defined as ≤120 mmHg and normal DBP as ≤80 mmHg. The statistical significance of association was determined by the p-values, while Phi and Cramer's V values measure the strength of association. According to Rea & Parker (2014), effect size statistics of the Phi and Cramer`s V values are 0.00 to under 0.10 = very weak association, 0.10 to under 0.20 = weak association, 0.20 to under 0.40 = moderate association and 0.40 and above = strong association.

Males exhibited a significantly higher proportion (p-value < 0.05) of abnormal SBP (22.4%) compared to females (11.7%). DBP levels were similar between genders. Racial differences were statistically significant for DBP (p-value = 0.025), with Coloured and White participants showing higher rates of abnormal diastolic pressure. Across ages 15 to 17, the variation in blood pressure status was minimal. Age did not significantly influence BP levels within this adolescent group [SBP: p-value = 0.290 (Cramer's V = 0.037); DBP: p-value = 0.192 (Cramer's V = 0.043)].

Higher BMI categories (Obese, Very Obese, Morbidly Obese) were strongly associated with increased rates of abnormal BP [(SBP: p < 0.05 (Cramer's V = 0.122);

DBP: p < 0.05 (Cramer's V = 0.147)]. Individuals with poor perceived health had the highest proportion of abnormal SBP (57.1%) and were significantly associated with SBP (p-value = 0.021). DBP rates did not vary as strongly.

Exercise frequency is significantly associated with BP levels [SBP: p-value = 0.006 (Cramer's V = 0.090), p-value = 0.003 (Cramer's V = 0.094)], though the trend suggests complexity (e.g., reverse causation). Minimal difference in BP levels between smokers and non-smokers was observed. No statistically significant association was found between smoking and blood pressure [SBP: p-value = 0.623 (Phi = 0.011);

Table 1: Blood Pressure among South African Adolescents by Demographics and Life Style Characteristics

		SBP		DBP	
Variables	Categories	Normal BP (≤ 120 mmHg)	Not Normal BP (>120 mmHg)	Normal BP (≤ 80 mmHg)	Not Normal BP (> 80 mmHg)
Gender	Male	669 (77.6%)	193 (22.4%)	756 (87.7%)	106 (12.3%)
	Female	839 (88.3%)	111 (11.7%)	831 (87.5%)	119 (12.5%)
p-value (Phi value)		p-value < 0.05 (0.143)		p-value = 0.887 (0.003)	
Race	African	1281 (84.0%)	244 (16.0%)	1351 (88.6%)	174 (11.4%)
	Coloured	180 (77.9%)	51 (22.1%)	189 (81.8%)	42 (18.2%)
	Asian/Indian	20 (90.9%)	2 (9.1%)	19 (86.4%)	3 (13.6%)
	White	27 (79.4%)	7 (20.6%)	28 (82.4%)	6 (17.6%)
p-value (Cramer`s V value)		p-value = 0.086 (0.060)		p-value = 0.025 (0.072)	
Age	15 years	226 (86.6%)	35 (13.4%)	222 (85.1%)	39 (14.9%)
	16 years	627 (82.6%)	132 (17.4%)	660 (87.0%)	99 (13.0%)
	17 years	655 (82.7%)	137 (17.3%)	705 (89.0%)	87 (11.0%)
p-value (Cramer`s V value)		p-value = 0.290 (0.037)		p-value = 0.192 (0.043)	
BMI	Underweight	350 (90.9%)	35 (9.1%)	340 (88.3%)	45 (11.7%)
	Healthy	952 (81.7%)	213 (18.3%)	1 029 (88.3%)	136 (11.7%)
	Overweight	147 (81.2%)	34 (18.8%)	164 (90.6%)	17 (9.4%)
	Obese	45 (76.3%)	14 (23.7%)	42 (71.2%)	17 (28.8%)
	Very Obese	8 (61.5%)	5 (38.5%)	7 (53.8%)	6 (46.2%)
	Morbidly Obese	6 (66.7%)	3 (33.3%)	5 (55.6%)	4 (44.4%)
p-value (Cramer`s V value)		p-value < 0.05 (0.122)		p-value < 0.05 (0.147)	
Perceived Health Status	Excellent	702 (82.9%)	145 (17.1%)	741 (87.5%)	106 (12.5%)
	Very Good	511 (82.3%)	110 (17.7%)	533 (85.8%)	88 (14.2%)
	Good	263 (86.8%)	40 (13.2%)	276 (91.1%)	27 (8.9%)
	Fair	29 (85.3%)	5 (14.7%)	31 (91.2%)	3 (8.8%)
	Poor	3 (42.9%)	4 (57.1%)	6 (85.7%)	1 (14.3%)
p-value (Cramer`s V value)		p-value = 0.021 (0.080)		p-value = 0.230 (0.056)	
Exercises	Never	828 (84.7%)	149 (15.3%)	868 (88.8%)	109 (11.2%)
	Less than once a week	127 (87.6%)	18 (12.4%)	134 (92.4%)	11 (7.6%)
	Once a week	117 (87.3%)	17 (12.7%)	105 (78.4%)	29 (21.6%)
	Twice a week	138 (79.3%)	36 (20.7%)	149 (85.6%)	25 (14.4%)
	Three or more times a week	298 (78.0%)	84 (22.0%)	331 (86.6%)	51 (13.4%)
p-value (Cramer`s V value)		p-value = 0.006 (0.090)		p-value = 0.003 (0.094)	
Cigarette Consumption	No	1 452 (83.3%	291 (16.7%)	1 526 (87.6%)	217 (12.4%)
	Yes	56 (81.2%)	13 (18.8%	61 (88.4%)	8 (11.6%)
p-value (Phi value)		p-value = 0.623 (0.011)		p-value = 0.833 (0.005)	
Depression	Rarely or none of the time (Less than 1 day)	1 117 (84.0%)	213 (16.0%)	1 159 (87.1%)	171 (12.9%)
	Some or Little of the time (1-2 days)	301 (79.6%)	77 (20.4%)	335 (88.6%)	43 (11.4%)
	Occasionally or a Moderate amount of time (3-4 days)	70 (86.4%)	11 (13.6%)	71 (87.7%)	10 (12.3%)
	All of the time (5-7 days)	20 (87.0%)	3 (13.0%)	22 (95.7%)	1 (4.3%)
p-value (Cramer`s V value)		p-value = 0.181 (0.052)		p-value = 0.574 (0.033)	

Table 2: Frequentist and Bayesian Quantile Regression Estimates for SBP`s Risk Factors

	Frequentist Quantile Regression		Bayesian Quantile Regression	
τ	Q (0.75)	Q (0.95)	Q (0.75)	Q (0.95)
Age	-0.45 (-1.61, 0.70)	0.77 (-1.17,2.73)	0.50 (0.44, 0.86)	0.89 (0.07,1.73)
BMI	0.86 (0.66, 1.06)	0.72 (0.38, 1.06)	0.84 (0.77, 0.91)	0.73 (0.60, 0.86)
Pulse Rate	0.03 (-0.04, 0.10)	0.002 (-0.11,0.12)	0.03 (-0.01, 0.05)	0.002 (-0.03, 0.04
Gender	-8.08 (-9.93, -6.23)	-7.24 (-10.37, -4,11)	-8.01 (-8.57, -7.45)	-7.52 (-8.94, -6.15)
Race	1.37 (-0.13, 2.88)	2.23 (-0.31, 4.77)	1.19 (0.69, 1.65)	2.34 (1.48, 3.31)
Exercises	0.28 (-0.23, 0.80)	0.67 (-0.21, 1.54)	0.30 (-0.14, 0.45)	0.61 (-0.20, 1.00)
Cigarette Consumption	0.06 (-4.27, 4.41)	-9.45 (-16.78, 2,14)	-0.01 (-1.61, 1.65)	-8.10 (-11.15, 4.29)
Depression	-0.36 (-1.67, 0.95)	0.85 (-1.39,3.04)	-0.32 (-070, 0.09)	0.99 (-0.19, 2.20)
Perceived Health Status	0.01 (-0.97, 1.00)	0.20 (-1.46,1.86)	-0.04 (-0.37, 0.27)	0.04 (-0.61, 0.64)

DBP: p-value = 0.833 (Phi = 0.005)]. No clear pattern in BP levels across depression frequency was noted. Even those depressed "all the time" had relatively normal BP levels. Depression frequency did not significantly associate with BP status [SBP: p-value = 0.181 (Cramer's V = 0.052); DBP: p-value = 0.574 (Cramer's V = 0.033)].

Table **2** displays the upper frequentist quantile regression coefficients along with the related 95% confidence intervals for SBP`s risk factors. Also, Bayesian posterior means and the related 95% credible intervals in parentheses for each SBP`s risk factor are illustrated. BMI and Gender seem to have the most consistent and statistically significant effects on SBP across both high quantiles (i.e. 0.75^{th} and 0.95^{th}) and models (frequentist and Bayesian) as evidenced by the 95% confidence and credible intervals containing no zero value. Race and Age had statistically significant credible intervals on both upper quantiles of SBP`s distribution for only the Bayesian approach. Pulse rate, Exercises, Cigarette Consumption, Depression, and Perceived Health Status did not present statistically significant relations with SBP on both the 75^{th} and 95^{th} quantiles of the frequentist and Bayesian approaches.

Applying the 95% confidence and credible intervals, Table **3** illustrates that BMI, Pulse Rate, Gender, and Race were consistently and significantly associated with higher DBP at both upper quantiles. Age, Exercises, Cigarette Consumption, Depression and Perceived Health Status had statistically insignificant associations with DBP across all the upper quantiles estimated as shown by the 95% confidence and credible intervals containing the zero value.

Convergence of the Bayesian Quantile Regression Approach

In this study convergence of the Bayesian Quantile Regression approach was examined using the Gelman-Rubin Brooks Plots and the autocorrelation plots. According to [18], convergence happens when the generated Markov chain converges in distribution to the posterior distribution of interest. According to [20], the convergence diagnostics aim to confirm the accuracy of the posterior summary measures and the stationarity of the Markov chain.

According to [21], the Gelman-Rubin diagnostic also known as the estimated potential scale reduction factor (PSRF) or $\widehat{R}$ is a formal means of assessing convergence. It involves running multiple MCMC chains with over dispersed starting positions to compute an estimate of the posterior distribution. The Gelman-Rubin statistic is calculated as:

$$\widehat{R} = \sqrt{\frac{\widehat{V}}{W}}$$

Table 3: Frequentist and Bayesian Quantile Regression Estimates for DBP`s Risk Factors

	Classical Quantile Regression		Bayesian Quantile Regression	
τ	Q (0.75)	Q (0.95)	Q (0.75)	Q (0.95)
Age	-0.81 (-1.62, 0.002)	-1.75 (-3.28, 0.23)	-0.85 (-1.20, 0.50)	-1.89 (-2.59, 1.15)
BMI	0.37 (0.23, 0.51)	0.37 (0.10, 0.63)	0.37 (0.31, 0.43)	0.37 (0.29, 0.47)
Pulse Rate	0.10 (0.05, 0.14)	0.10 (0.01, 0.19)	0.09 (0.07, 0.12)	0.09 (0.06, 0.13)
Gender	-1.30 (-2.60, -0.01)	-3.48 (-5.93, -1.04)	-1.34 (-1.90, -0.77)	-3.66 (-4.80, -2.48)
Race	0.84 (0.22, 1.90)	2.13 (0.14, 4.11)	0.84 (0.33, 1.38)	1.92 (0.88, 2.89)
Exercises	0.11 (-0.25, 0.47)	0.56 (-0.12, 1.25)	0.11 (-0.04, 0.27)	0.55 (-0.22, 0.89)
Cigarette Consumption	0.55 (-2.50,3.60)	1.91 (-3.81, 7.63)	0.42 (-1.08, 1.69)	0.57 (-2.51, 2.99)
Depression	-0.55 (-1.48,0.37)	0.25 (-1.48, 1.98)	-0.51 (-0.89, 0.14)	0.17 (-0.72, 1.08)
Perceived Health Status	-0.44 (-1.13, 0.25)	0.34 (-0.95, 1.64)	-0.45 (-0.76, 0.17)	0.39 (-0.17, 1.0)

Where $\widehat{V}$ is the estimated variance of the target distribution (a weighted average of within- and between-chain variance).

W is the average within-chain variance.

According to [20], the $\widehat{R_c}$ is expected to be smaller than 1.1 for the chains to mix well and reach convergence of the posterior distribution and an $\widehat{R_c}$ more than 1.1 means the chains may not have mixed well, suggesting more iterations may be needed. The Gelman-Rubin Brooks plot is then meant to visually confirm the shrink factor convergence.

Figure **1** displays the Gelman-Rubin Brooks plots for SBP`s risk factors obtained after running an MCMC algorithm for 20 000 iterations with a burn-in period of 5 000 on two parallel chains. It can be seen from Figure **1** that the two MCMC parallel chains assessing each parameter reflect convergence as the chains show a decreasing trend across the iterations applied.

Figure **2** illustrates the Gelman-Rubin Brooks plots for DBP`s risk factors obtained after running an MCMC algorithm for 20 000 iterations with a burn-in period of 5 000 on two parallel chains. The two parallel chains utilised reflect evidence of settling together, suggesting stable convergence across parameters.

Convergence diagnostics confirmed that the Bayesian models reached stationarity. Gelman–Rubin plots for both SBP and DBP showed shrink factors approaching 1.0, indicating well-mixed chains and reliable posterior estimates.

Model Comparisons

It is apparent from Tables **2** and **3** that a notable methodological difference was that Bayesian quantile regression produced narrower credible intervals than the frequentist confidence intervals, indicating greater precision.

DISCUSSION

The present study applied frequentist and Bayesian quantile regression models to nationally representative data on South African adolescents to investigate the drivers of paediatric hypertension. By focusing on the upper quantiles of systolic and diastolic blood pressure, this analysis captured risk factors that exert stronger influence at the extremes of the BP distribution, where hypertension is most pronounced. This focus is clinically relevant, as individuals in these upper quantiles are at the highest risk of developing long-term cardiovascular complications, including left ventricular hypertrophy and early onset of atherosclerosis.

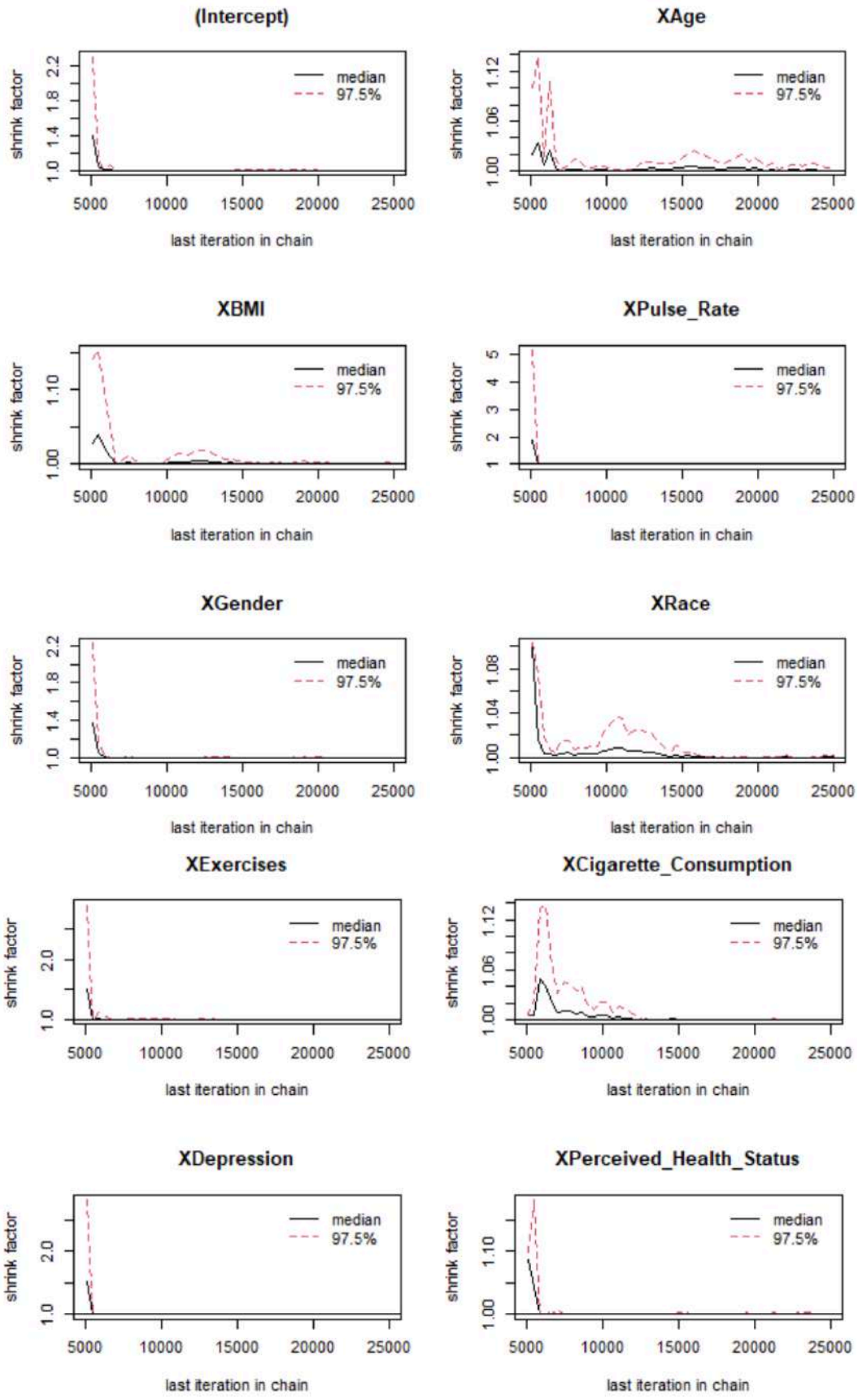

Figure 1: Gelman-Rubin Brooks Plots For SBP`s Risk Factors.

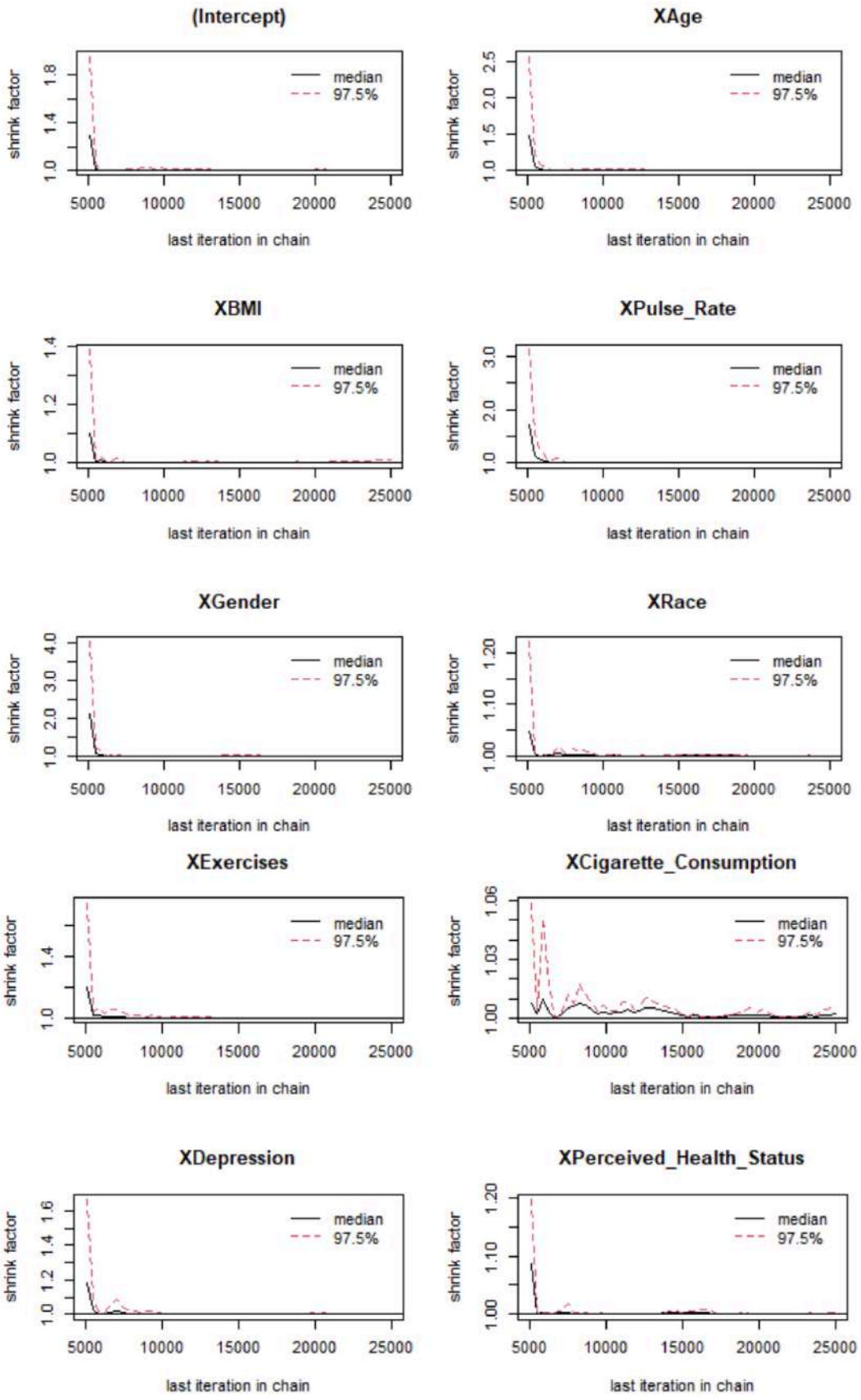

Figure 2: Gelman-Rubin Brooks Plots For DBP`s Risk Factors.

The results confirm that BMI and gender are consistent predictors of elevated blood pressure across both modelling approaches. This aligns with previous South African studies showing that overweight and obese children are at substantially higher risk of hypertension [1, 22], and that boys tend to have higher BP than girls during adolescence [23]. The persistence of BMI and gender across all quantiles suggests that these remain the most critical non-pharmacological targets for early intervention and prevention. Specifically, childhood obesity prevention and gender-sensitive screening protocols should be prioritized in public health strategies.

The Bayesian models additionally identified age and race as significant predictors of systolic blood pressure, while pulse rate showed strong associations with diastolic blood pressure. These findings are partly consistent with international studies linking age and race to paediatric hypertension risk [24, 25]. The identification of race as a determinant in the South African context highlights the interplay of genetic predisposition, socio-economic disparities, dietary habits, and access to healthcare, factors that may contribute to observed racial differences in hypertension prevalence. Clinically, this underscores the need for culturally and contextually tailored screening programs.

In contrast, factors such as smoking, exercise, depression, and perceived health status showed weak or inconsistent associations with blood pressure in this dataset. While other studies have reported significant associations [26-28], the lack of strong evidence here may be attributable to measurement limitations in self-reported lifestyle variables, or the relatively young age of participants. From a medical point of view, this suggests that the impact of these variables may emerge more prominently in adulthood, or may require more precise, objective measures (e.g., biochemical markers or activity trackers) to detect their true effect.

Methodological Insights

A key methodological contribution of this study is the demonstration that Bayesian quantile regression yields narrower credible intervals than frequentist quantile regression, implying greater precision and robustness in inference. This result echoes findings from previous methodological comparisons in cardiovascular research [10, 29]. From a medical research perspective, this suggests that Bayesian approaches may be preferable when investigating health outcomes concentrated in the distributional tails. This precision supports more confident clinical decision-making, particularly when tailoring interventions for high-risk groups.

CONCLUSION

This study examined the risk factors of systolic and diastolic blood pressure among South African adolescents using both frequentist and Bayesian quantile regression approaches. By focusing on the upper quantiles of the blood pressure distribution (75th and 95th percentiles), the analysis provided deeper insights into the drivers of paediatric hypertension compared to traditional mean-based models.

Across both modelling approaches, BMI and gender consistently emerged as the most important predictors of elevated blood pressure. Additional associations were observed with race, pulse rate, and age, particularly in the Bayesian models. Importantly, the Bayesian quantile regression produced narrower credible intervals than the frequentist models, indicating more precise estimates and greater inferential reliability.

The findings highlight the methodological value of Bayesian quantile regression in studying health outcomes that are concentrated at distributional extremes. To the best of our knowledge, this is the first application of Bayesian quantile regression to child hypertension in South Africa, and the comparative evidence generated offers new insights for both methodological research and public health practice.

From a policy perspective, the results underscore the need for interventions that target obesity prevention and management among children and adolescents, with particular attention to high-risk groups identified by gender and race. School-based screening programs and community health initiatives that monitor BMI and blood pressure could play an important role in early detection and prevention.

Future research should extend this work by applying longitudinal designs, exploring additional behavioural and environmental risk factors, and testing interventions aimed at reducing hypertension risk in children. By strengthening both methodological approaches and practical understanding, this study contributes to the growing evidence base needed to combat paediatric hypertension in South Africa.

POLICY IMPLICATIONS

The findings highlight the urgent need for targeted interventions addressing obesity among adolescents, particularly given its consistent role across all models. School-based screening for blood pressure and BMI, coupled with community-level awareness campaigns, may provide an effective strategy for early detection and prevention. In addition, the identification of race and gender as differential risk factors suggests that public health interventions should be culturally sensitive and tailored to high-risk groups.

ACKNOWLEDGEMENTS

The authors are quite grateful to the research team of the South African National Income Dynamics Study 2017-2018 (NIDS) for their permission to use their data.

ETHICAL CONSIDERATION

The South African National Income Dynamics Survey was conducted after the University of Cape Town, Faculty of Commerce Ethics Committee, granted ethical approval. Informed consent was obtained from each study participant.

FUNDING

There was no organisation that sponsored this study.

COMPETING INTERESTS

The authors declare that they have no competing interests.

AVAILABILITY OF DATA AND MATERIALS

The dataset analysed during the current study are available from the corresponding author on reasonable request.

REFERENCES

[1] Gomwe H, Seekoe E, Lyoka P, Marange CS. Blood pressure profile of primary school children in Eastern Cape province, South Africa: prevalence and risk factors. BMC Pediatr 2022; 22: 207. https://doi.org/10.1186/s12887-022-03221-5

[2] Nkeh-Chungag BN, Sekokotla AM, Sewani-Rusike C, Namugowa A, Iputo JE. Prevalence of Hypertension and Pre-hypertension in 13-17 Year Old Adolescents Living in Mthatha - South Africa: a Cross-Sectional Study. Cent Eur J Public Health 2015; 23: 59-64. https://doi.org/10.21101/cejph.a3922

[3] Kagura J, Adair LS, Musa MG, Pettifor JM, Norris SA. Blood pressure tracking in urban black South African children: birth to twenty cohort. BMC Pediatr 2015; 15: 78. https://doi.org/10.1186/s12887-015-0402-z

[4] Joubert N, Walter C, Du Randt R, Aerts A, Adams L, Degen J, *et al*. Hypertension among South African children in disadvantaged areas and associations with physical activity, fitness, and cardiovascular risk markers: A cross-sectional study. J Sports Sci 2021; 39: 2454-67. https://doi.org/10.1080/02640414.2021.1939964

[5] Koenker R, Hallock KF. Quantile Regression An Introduction. J Econ Perspect 2001; 15: 143-56.

[6] Benoit DF, Van den Poel D. A Bayesian Approach to Quantile Regression. J Stat Softw 2017; 76. https://doi.org/10.18637/jss.v076.i07

[7] Yu K, Moyeed RA. Bayesian quantile regression. Stat Probab Lett 2001; 54: 437-47. https://doi.org/10.1016/S0167-7152(01)00124-9

[8] National High Blood Pressure Education Program Working Group on High Blood Pressure in Children and Adolescents. The Fourth Report on the Diagnosis, Evaluation, and Treatment of High Blood Pressure in Children and Adolescents. Pediatrics 2004; 114: 555-76. https://doi.org/10.1542/peds.114.S2.555

[9] Hu Y, Wang HJ, He X, Guo J. Bayesian joint-quantile regression. Comput Stat 2021; 36: 2033-53. https://doi.org/10.1007/s00180-020-00998-w

[10] Soomro S, Yu K. Bayesian fractional polynomial approach to quantile regression and variable selection with application in the analysis of blood pressure among US adults. J Appl Stat 2024: 1-22. https://doi.org/10.1080/02664763.2024.2359526

[11] Barrodale I, Roberts FDK. An Improved Algorithm for Discrete $l_1 $ Linear Approximation. SIAM J Numer Anal 1973; 10: 839-48. https://doi.org/10.1137/0710069

[12] Yu K, Zhang J. A Three-Parameter Asymmetric Laplace Distribution and Its Extension. Commun Stat - Theory Methods 2005; 34: 1867-79. https://doi.org/10.1080/03610920500199018

[13] Koenker R, Chernozhukov V, He X, Peng L, editors. Handbook of Quantile Regression. 1st ed. Chapman and Hall/CRC; 2017. https://doi.org/10.1201/9781315120256

[14] Martin AD, Quinn KM, Park JH. MCMCpack: Markov Chain Monte Carlo in R. J Stat Softw 2011; 42. https://doi.org/10.18637/jss.v042.i09

[15] Hamra G, MacLehose R, Richardson D. Markov Chain Monte Carlo: An introduction for epidemiologists. Int J Epidemiol 2013; 42: 627-34. https://doi.org/10.1093/ije/dyt043

[16] Plummer M, Best N, Cowles K, Vines K, Sarkar D, Bates D, *et al*. Output Analysis and Diagnostics for MCMC 2016.

[17] Fenske N, Kneib T, Hothorn T. Identifying Risk Factors for Severe Childhood Malnutrition by Boosting Additive Quantile Regression. J Am Stat Assoc 2011; 106: 494-510. https://doi.org/10.1198/jasa.2011.ap09272

[18] Sinharay S. Assessing Convergence of the Markov Chain Monte Carlo Algorithms: A Review. Princeton, NJ: Educational Testing Service; 2003.

[19] Rea LM, Parker RA. Designing and conducting survey research: a comprehensive guide. Fourth edition. San Francisco, CA: Jossey-Bass; 2014.

[20] Lesaffre E, Lawson A. Bayesian Biostatistics. Chichester: Wiley; 2012.

[21] Gelman A, Rubin DB. Inference from Iterative Simulation Using Multiple Sequences. Stat Sci 1992; 7: 457-72. https://doi.org/10.1214/ss/1177011136

[22] Matizirofa L, Kuhudzai AG. Predictors of High Blood Pressure in South African Children: Quantile Regression Approach. Int J Stat Med Res 2017; 6: 84-91. https://doi.org/10.6000/1929-6029.2017.06.02.4

[23] Raphadu TT, Staden MV, Dibakwane WM, Monyeki KD. A Non-Invasive Investigation into the Prevalence of Higher than Normal Blood Pressure, Hypertension and the Association between Blood Pressure and Body Weight in Male and Female Adolescents in the Polokwane Local Municipality, Limpopo-South Africa: A Cross-Sectional Study. Children 2020; 7: 18. https://doi.org/10.3390/children7030018

[24] Christofaro DGD, Casonatto J, Vanderlei LCM, Cucato GG, Dias RMR. Relationship between Resting Heart Rate, Blood Pressure and Pulse Pressure in Adolescents. Arq Bras Cardiol 2017. https://doi.org/10.5935/abc.20170050

[25] Muntner P. Trends in Blood Pressure Among Children and Adolescents. JAMA 2004; 291: 2107. https://doi.org/10.1001/jama.291.17.2107

[26] Chen J, Wang Y, Li W, Zhang Y, Cao R, Peng X, *et al*. Physical activity and eating behaviors patterns associated with high blood pressure among Chinese children and adolescents. BMC Public Health 2023; 23: 1516. https://doi.org/10.1186/s12889-023-16331-1

[27] Levy RV, Brathwaite KE, Sarathy H, Reidy K, Kaskel FJ, Melamed ML. Analysis of Active and Passive Tobacco Exposures and Blood Pressure in US Children and Adolescents. JAMA Netw Open 2021; 4: e2037936. https://doi.org/10.1001/jamanetworkopen.2020.37936

[28] Olive LS, Abhayaratna WP, Byrne D, Telford RM, Berk M, Telford RD. Depression, stress and vascular function from childhood to adolescence: A longitudinal investigation. Gen Hosp Psychiatry 2020; 62: 6-12. https://doi.org/10.1016/j.genhosppsych.2019.10.001

[29] Juhan N, Zubairi YZ, Mohd Khalid Z, Mahmood Zuhdi AS. A Comparison Between Bayesian and Frequentist Approach in the Analysis of Risk Factors for Female Cardiovascular Disease Patients in Malaysia. ASM Sci J 2020: 1-7. https://doi.org/10.32802/asmscj.2020.sm26(1.1)

Received on 08-10-2025 Accepted on 07-11-2025 Published on 24-11-2025

https://doi.org/10.6000/1929-6029.2025.14.66