


# Modelling of South African Hypertension: Comparative Analysis of the Classical and Bayesian Quantile Regression Approaches

INQUIRY: The Journal of Health Care Organization, Provision, and Financing  
Volume 59: 1–9  
© The Author(s) 2022  
Article reuse guidelines:  
[sagepub.com/journals-permissions](https://sagepub.com/journals-permissions)  
DOI: 10.1177/00469580221082356  
[journals.sagepub.com/home/inq](https://journals.sagepub.com/home/inq)  


Anesu Gelfand Kuhudzai, PhD candidate<sup>1,2,\*</sup> , Guido Van Hal, PhD<sup>1,3</sup>, Stefan Van Dongen, PhD<sup>1,3</sup> and Muhammad Hoque, PhD<sup>1</sup>

## Abstract

Hypertension has become a major public health challenge and a crucial area of research due to its high prevalence across the world including the sub-Saharan Africa. No previous study in South Africa has investigated the impact of blood pressure risk factors on different specific conditional quantile functions of systolic and diastolic blood pressure using Bayesian quantile regression. Therefore, this study presents a comparative analysis of the classical and Bayesian inference techniques to quantile regression. Both classical and Bayesian inference techniques were demonstrated on a sample of secondary data obtained from South African National Income Dynamics Study (2017–2018). Age, BMI, gender male, cigarette consumption and exercises presented statistically significant associations with both SBP and DBP across all the upper quantiles ( $\tau \in \{0.75, 0.95\}$ ). The white noise phenomenon was observed on the diagnostic tests of convergence used in the study. Results suggested that the Bayesian approach to quantile regression reveals more precise estimates than the frequentist approach due to narrower width of the 95% credible intervals than the width of the 95% confidence intervals. It is therefore suggested that Bayesian approach to quantile regression modelling to be used to estimate hypertension.

## Keywords

hypertension, classical quantile regression, Bayesian quantile regression, confidence and credible intervals, South Africa

What do we already know about this topic?

Hypertension has become a major public health challenge and a crucial area of research due to its high prevalence across the world including the sub-Saharan Africa.

How does your research contribute to the field?

A comparative analysis of the classical and the Bayesian approaches to quantile regression in order to study the effect of blood pressure risk factors on the upper quantiles of blood pressure's distribution.

What are your research's implications towards theory, practice or policy?

To recommend statistical methods (models) for making reasonably robust and precise inferences about the risk factors of hypertension among South African adults.

losses, high blood pressure accounts for approximately 9.4 million deaths globally every year.<sup>2</sup> A recent study<sup>3</sup> has revealed that there is a high prevalence of hypertension and associated cardiovascular diseases in sub-Saharan Africa. In sub-Saharan Africa, South Africa has the highest prevalence of hypertension (between 42% and 54%) and also the largest

## Introduction

Worldwide, approximately 17 million deaths a year are caused by cardiovascular diseases.<sup>1</sup> Out of these human

<sup>1</sup>Department of Social Epidemiology and Healthy Policy, University of Antwerp, Belgium

<sup>2</sup>Statistical Consultation Services, University of Johannesburg, South Africa

<sup>3</sup>Department of Evolutionary Ecology and Biology, University of Antwerp, Belgium

### Corresponding Author:

\*Anesu Gelfand Kuhudzai, Department of Social Epidemiology and Healthy Policy, University of Antwerp, Prinsstraat 13, Antwerpen 2000, Belgium; Statistical Consultation Services, University of Johannesburg, 20 Chiselhurst Drive, Auckland Park 2006, South Africa.  
Email: [gelfand9@yahoo.com](mailto:gelfand9@yahoo.com)



Creative Commons Non Commercial CC BY-NC: This article is distributed under the terms of the Creative Commons Attribution-NonCommercial 4.0 License (<https://creativecommons.org/licenses/by-nc/4.0/>) which permits non-commercial use, reproduction and distribution of the work without further permission provided the original work is attributed as specified on the SAGE

and Open Access pages (<https://us.sagepub.com/en-us/nam/open-access-at-sage>).

number of people whose blood pressure is still not controlled, even whilst being on treatment.<sup>3</sup>

Due to the high prevalence of raised blood pressure across the world including the sub-Saharan Africa, it is crucial to research on possible risk factors of hypertension in order to minimise their effect in the lifestyle of individuals. Most studies have performed modelling of hypertension using binary and multiple logistic regression.<sup>4-9</sup>

Quantile linear regression has emanated as a useful additional technique to either binary logistic regression or classical linear regression. Basically, quantile regression is a natural extension of classical linear regression. Quantile regression models the impact of predictors on different specific quantiles (or percentiles) of the response distribution, and thus provides a more comprehensive picture of the relationship between predictor variables and the response variable.

Binary logistic regression and multiple logistic regression require the dependent variable to be binary and ordinal, respectively, limiting the accuracy of the results as compared to quantile regression which uses continuous dependent variables.<sup>10</sup>

Classical linear regression provides a simple way of exploring how the mean of a response variable changes with the effect of predictor variables whilst quantile regression does focus on estimating families of conditional quantile functions. If quantile regression estimates are run simultaneously for  $\tau = 0.05$  to  $\tau = 0.95$  in regular intervals, then the complete relationship between the explanatory variable(s) and response variable along the entire distribution of the response variable can be detected.<sup>11</sup>

Quantile regression models have become popular because estimates are more robust against outliers in the response measurements than classical regression models. Also, quantile regression makes no distributional assumption about the error term in the model and thus enables it to accommodate non-normal errors which are common in many applications. Another main advantage of quantile regression as compared to ordinary least-squares regression is its ability to model data with heterogeneous conditional distributions such that it is now being applied to model panel data, time series data, conditional extreme value, nonlinear models, binary response models and duration models.<sup>12</sup>

However, the Bayesian approach to quantile regression may lead to exact inference in estimating the influence of potential risk factors on the upper quantiles (75% and 95%) of the conditional distribution of hypertension as opposed to the asymptotic inference of the classical or frequentist quantile regression.<sup>13</sup> Furthermore, Bayesian quantile regression does provide estimations and predictions which take into account parameter uncertainty.<sup>14</sup> In Bayesian inference, population parameters are associated with a posterior probability or distribution which quantifies the value of the parameter of interest.

A comparative study between Bayesian and frequentist approaches in the analysis of risk factors for female

cardiovascular disease (CVD) patients in Malaysia revealed that the Bayesian approach was a better one due to smaller standard errors obtained from the Bayesian approach than the frequentist approach.<sup>15</sup>

Therefore, the aim of this study is to conduct a comparative analysis of the classical and the Bayesian approaches to quantile regression in order to study the effect of blood pressure risk factors on the upper quantiles of blood pressure's distribution. Thus, systolic blood pressure (SBP) and diastolic blood pressure (DBP).

## Materials and Methods

This section gives an account of how the Bayesian approach to quantile regression framework was carried out in order to study the effect of blood pressure risk factors on different quantiles of blood pressure's distribution. Thus, the data, study variables, theoretical model and data analysis techniques.

### Data and Variables

This was a retrospective study performed on a nationally representative sample obtained from the South African National Income Dynamics Study (NIDS) Wave 5 Household survey conducted between 2017 and 2018. NIDS was embarked in order to assess the welfare of South African individuals across the entire country.

The study employed multi-stage sampling to randomly select a sample of 30 110 adults aged 18 years and above. Trained fieldworkers were instructed to collect the data. A total of 21 180 cases were found valid after data cleaning.

The study variables included systolic blood pressure and diastolic blood pressure as the response variables. Predictor variables were age, body mass index (BMI), gender, race, exercises, cigarette consumption, depression and employment status.

The ethics approval to conduct the NIDS study was granted by the University Of Cape Town Faculty Of Commerce Ethics Committee and informed consent was attained from each study participant.

### Bayesian Quantile Regression

Quantile regression seeks to estimate models for the conditional quantile functions.<sup>16</sup> Quantile regression is particularly useful in applications where extremes are important, such as blood pressure where upper quantiles (tails) of systolic and diastolic blood pressure levels are critical from a public health perspective.

Bayesian approach to quantile regression is normally carried out by formulating a likelihood function based on the asymmetric Laplace distribution irrespective of the actual distribution of the data.<sup>14</sup> Generally, any prior can be chosen for each of the quantile regression parameters, but it has been shown that the use of improper uniform priors produces a

proper joint posterior distribution.<sup>14</sup> Bayesian quantile regression approach produces exact inference and accommodates missing, clustered or censored data.<sup>17</sup>

Since the Bayesian framework is formulated on a likelihood function based on the asymmetric Laplace distribution,<sup>14</sup> a random variable  $U$  follows the asymmetric Laplace distribution if its probability density function is given by

$$f_p(u) = p(1-p) \exp\{-\rho_p(u)\} \quad (1)$$

where  $0 < p < 1$

$$\begin{aligned} \text{and } \rho_p(u) &= u(p-1)(u < 0) \\ &= u(pI(u > 0) - (1-p)I(u < 0)) \\ &= \frac{|u| + (2p-1)u}{2} \end{aligned} \quad (2)$$

when  $p = \frac{1}{2}$  then  $f_p(u) = \frac{1}{4} \exp(-|u|/2)$  which is the density function of a standard symmetric Laplace distribution.

For all other values of  $p$ , the probability density function in (1) is asymmetric.

The mean of  $U$  is  $\frac{1-2p}{p(1-p)}$  and it is positive only for  $p > \frac{1}{2}$ . And the variance is given by  $\frac{(1-2p+2p^2)}{p^2(1-p)^2}$ .

If the location and scale parameters  $\mu$  and  $\sigma$  are inserted into the probability density function (1), the following function is obtained

$$f_p(u; \mu, \sigma) = \frac{p(1-p)}{\sigma} \exp\left\{-\rho_p\left(\frac{u-\mu}{\sigma}\right)\right\} \quad (3)$$

Now given the observations,  $y = (y_1, y_2, y_3, \dots, y_n)$ , the posterior distribution of  $\beta$ ,

$\pi(\beta|y)$  is given by

$$\pi(\beta|y) \propto L(y|\beta) p(\beta) \quad (4)$$

where  $p(\beta)$  is the prior distribution of  $\beta$  and  $L(y|\beta)$  is the likelihood function written as

$$L(y|\beta) = p^n (1-p)^n \exp\left\{-\sum_i \rho_p(y_i - x'_i \beta)\right\} \quad (5)$$

with a location parameter  $\mu_i = x'_i \beta$ .

However, any prior can be used for  $p(\beta)$ . In case where there is no realistic information, improper prior distributions can be used for all components of  $\beta$ .<sup>14</sup> Alternatively, Markov chain Monte Carlo (MCMC) methods can be used to approximate the posterior distributions of the unknown parameter(s).<sup>18</sup>

## Data Analysis

IBM Statistical Package for the Social Sciences (SPSS) version 27 was used to generate descriptive statistics in form of proportions for categorical variables. The quantreg R package<sup>19</sup> was employed to fit the classical quantile regression models.

Bayesian approach to quantile regression was implemented by adopting MCMC algorithms contained in the R package called MCMCpack.<sup>20</sup> Models produced by MCMCpack return coda MCMC objects that can then be summarised by the coda package. The coda package provides functions for summarising and plotting the output from the MCMC simulations, as well as diagnostic tests of convergence.<sup>21</sup> This study considers 2 quantile models at the 75<sup>th</sup> and 95<sup>th</sup> percentiles. When modelling hypertension, it makes more sense to model high values of systolic and diastolic blood pressure which corresponds to the upper distribution of either SBP or DBP.<sup>22</sup>

MCMC algorithms have emerged as very useful and popular tools for fitting Bayesian statistical models in modern Bayesian computing. According to Sinharay (2003), the key reason why MCMC algorithms have become useful and popular is that the algorithms can fit quite complex models easily as compared to standard techniques such as maximum likelihood estimation (MLE).

## Results

Empirical results of this study are presented in this section, in form of tables and figures.

Table 1 presents systolic blood pressure and diastolic blood pressure proportions among South African Adults by demographic and life style characteristics. Also, presented in Table 1 is the significance and magnitude of association between SBP, DBP, demographic and life style characteristics of the study participants. The magnitude of the association is measured by the Cramer's V value and then compared with guidelines outlined by Ref. 24, .00 to under .10 = very weak association, .10 to under .20 = weak association, .20 to under .40 = moderate association and .40 and above = strong association.

It can be seen from Table 1 that hypertension was more prevalent in men than women for both BP measures. Concerning race, Table 1 illustrates that elevated high blood pressure was most prevalent among coloured participants when looking at both SBP and DBP proportions.

It is evident from Table 1 that the prevalence of hypertension increased with age, with the 50 years and above age group recording the highest proportions for both blood pressure measures. The same trend was observed with BMI, whereby the proportions of raised blood pressure were increasing with the level of BMI. Underweight and healthy respondents had the lowest prevalence whilst the very and morbidly obese had the highest prevalence.

In regard to exercises, it is apparent that the study participants who do not participate in physical activities recorded the highest proportions of high blood pressure. There was not much difference in elevated blood pressure among the 3 levels of depression. It is apparent from Table 1 that participants who do smoke recorded the highest proportions of hypertension for both BP measures.

Mixed results were observed on employment status, unemployed participants recorded a higher proportion of hypertension

**Table I.** Blood Pressure among South African Adults by Demographic and Life Style Characteristics.

		SBP (n = 21 180)			DBP (n = 21 180)		
		Normal BP (<120 mmHg)	Pre- hypertension (120– 139 mmHg)	Hypertension (140 mmHg and above)	Normal BP (<80 mmHg)	Pre- hypertension (80–89 mmHg)	Hypertension (90 mmHg and above)
Gender	Male	3728 (43.3%)	3409 (39.6%)	1479 (17.2%)	4749 (55.1%)	2279 (26.5%)	1588 (18.4%)
	Female	7260 (57.8%)	3463 (27.6%)	1841 (14.7%)	7266 (57.8%)	3121 (24.8%)	2177 (17.3%)
P-value (Cramer's V value)		P-value < .05 (.148)			P-value < .05 (.030)		
Race	African	9295 (54.7%)	5295 (31.1%)	2409 (14.2%)	10 056 (59.2%)	4135 (24.3%)	2808 (16.5%)
	Coloured	1157 (41.4%)	1021 (36.6%)	614 (22.0%)	1273 (45.6%)	814 (29.2%)	705 (25.3%)
	Asian/Indian	165 (48.8%)	112 (33.1%)	61 (18.0%)	170 (50.3%)	96 (28.4%)	72 (21.3%)
	White	371 (35.3%)	444 (42.2%)	236 (22.5%)	516 (49.1%)	355 (33.8%)	180 (17.1%)
		P-value < .05 (.073)			P-value < .05 (.064)		
Age	18–29 years	5266 (68.8%)	2079 (27.1%)	313 (4.1%)	5569 (72.7%)	1547 (20.2%)	542 (7.1%)
	30–39 years	2592 (58.5%)	1430 (32.3%)	412 (9.3%)	2484 (56.0%)	1202 (27.1%)	748 (16.9%)
	40–49 years	1496 (46.9%)	1164 (36.5%)	532 (16.7%)	1475 (46.2%)	895 (28.0%)	822 (25.8%)
	50 years and above	1634 (27.7%)	2199 (37.3%)	2063 (35.0%)	2487 (42.2%)	1756 (29.8%)	1653 (28.0%)
		P-value < .05 (.237)			P-value < .05 (.166)		
BMI	Underweight	902 (68.8%)	302 (23.0%)	107 (8.2%)	947 (72.2%)	238 (18.2%)	126 (9.6%)
	Healthy	5075 (59.0%)	2604 (30.3%)	929 (10.8%)	5656 (65.7%)	1907 (22.2%)	1045 (12.1%)
	Overweight	2488 (48.8%)	1732 (34.0%)	880 (17.3%)	2727 (53.5%)	1378 (27.0%)	995 (19.5%)
	Obese	1417 (42.9%)	1160 (35.1%)	727 (22.0%)	1540 (46.6%)	994 (30.1%)	770 (23.3%)
	Very obese	691 (40.4%)	648 (37.9%)	370 (21.7%)	722 (42.2%)	528 (30.9%)	459 (26.9%)
	Morbidly obese	415 (36.1%)	426 (37.1%)	307 (26.7%)	423 (36.8%)	355 (30.9%)	370 (32.2%)
		P-value < .05 (.099)			P-value < .05 (.111)		
Exercises	Never	7513 (51.5%)	4590 (31.4%)	2492 (17.1%)	8010 (54.9%)	3734 (25.6%)	2851 (19.5%)
	Once or two times a week	2057 (53.3%)	1314 (34.0%)	490 (12.7%)	2317 (60.0%)	991 (25.7%)	553 (14.3%)
	Three or more times a week	1418 (52.1%)	968 (35.5%)	338 (12.4%)	1688 (62.0%)	675 (24.8%)	361 (13.3%)
		P-value < .05 (.051)			P-value < .05 (.053)		
Depression	Rarely or none of the time	6336 (52.1%)	3952 (32.5%)	1864 (15.3%)	7022 (57.8%)	3060 (25.2%)	2070 (17.0%)
	Some or little of the time	3209 (51.2%)	2044 (32.6%)	1018 (16.2%)	3441 (54.9%)	1630 (26.0%)	1200 (19.1%)
	Occasionally or all of the time	1443 (52.3%)	876 (31.8%)	438 (15.9%)	1552 (56.3%)	710 (25.8%)	495 (18.0%)
		P-value = .380 (.014)			P-value < .05 (.024)		
Cigarette consumption	No	9162 (53.5%)	5358 (31.3%)	2614 (15.3%)	9909 (57.8%)	4287 (25.0%)	2938 (17.1%)
	Yes	1826 (45.1%)	1514 (37.4%)	706 (17.4%)	2106 (52.1%)	1113 (27.5%)	827 (20.4%)
		P-value < .05 (.069)			P-value < .05 (.048)		
Employment status	No	7612 (52.8%)	4437 (30.8%)	2359 (16.4%)	8420 (58.4%)	3581 (24.9%)	2407 (16.7%)
	Yes	3376 (49.9%)	2435 (36.0%)	961 (14.2%)	3595 (53.1%)	1819 (26.9%)	1358 (20.1%)
		P-value < .05 (.066)			P-value < .05 (.056)		

**Table 2.** Classical and Bayesian Quantile Regression Estimates for SBP's Risk Factors.

$\tau$	Classical Quantile Regression		Bayesian Quantile Regression	
	Q (.75)	Q (.95)	Q (.75)	Q (.95)
Age	.58 (.56,0.60)	.93 (.88,0.97)	.58 (.57,0.59)	.93 (.91,0.94)
BMI	.64 (.59,0.69)	.72 (.61,0.84)	.64 (.62,0.66)	.71 (.68,0.75)
Gender_Male	10.86 (10.11, 11.62)	11.02 (9.32, 12.73)	10.86 (10.64, 11.08)	10.97 (10.51, 11.41)
Race	.25 (-.20,070)	-.74 (-1.75,0.27)	.28 (.13,0.44)	-.73 (-1.04,-.44)
Exercises	-.15 (-.62, .33)	-1.42 (-2.49, -.35)	-.13 (-.28, .00)	-1.41 (-1.71, -1.10)
Cigarette consumption	1.91 (1.03,2.79)	2.50 (.52, 4.48)	1.90 (1.62, 2.16)	2.44 (1.79, 3.04)
Depression	.12 (-.33,0.56)	-.13 (-1.13,0.88)	.11 (-.04,0.25)	-.09 (-.38,0.22)
Employment status	-1.23 (-1.92, -.54)	-1.49 (-3.05, .07)	-1.23 (-1.44, -1.02)	-1.49 (-1.91, -1.04)

**Table 3.** Classical and Bayesian Quantile Regression Estimates for DBP's Risk Factors.

$\tau$	Classical Quantile Regression		Bayesian Quantile Regression	
	Q (.75)	Q (.95)	Q (.75)	Q (.95)
Age	.21 (.20,0.23)	.31 (.28,0.33)	.21 (.21,0.22)	.31 (.29,0.32)
BMI	.47 (.43,0.50)	.48 (.41,0.55)	.47 (.45,0.48)	.48 (.45,0.51)
Gender_Male	3.64 (3.11,4.12)	3.12 (2.11, 4.14)	3.63 (3.44, 3.83)	3.06 (2.66, 3.45)
Race	-.14 (-.45,0.18)	-.74 (-1.34,-.14)	-.14 (-.26,-.02)	-.75 (-1.00,-.47)
Exercises	-.86 (-1.19, -.53)	-1.47 (-2.11, -.84)	-.87 (-.99, -.75)	-1.49 (-1.74, -1.24)
Cigarette consumption	2.66 (2.05,3.28)	2.87 (1.70, 4.04)	2.64 (2.38, 2.88)	2.84 (2.42, 3.24)
Depression	.19 (-.12,0.50)	.03 (-.57,0.62)	.18 (.07,0.30)	.08 (-.17,0.32)
Employment status	.86 (.38, 1.35)	.76 (-.17, 1.68)	.91 (.72, 1.10)	.79 (.43, 1.16)

when looking at SBP whilst employed respondents had a higher prevalence of hypertension when viewing the DBP figures.

Finally, Table 2 revealed statistically significant ( $P$ -values  $< .05$ ) very weak associations to weak associations between SBP, DBP, demographic and life style characteristics of the study respondents.

Table 2 presents the upper classical quantile regression coefficients and the related 95% confidence intervals for SBP's risk factors. Also, Bayesian posterior means and the associated 95% credible intervals in parentheses for each SBP's risk factor are shown. The 95% credible interval is the range of values in which the researcher is 95% certain that the population mean,  $\mu$ , falls, based on the sample data of size  $n$ . The 95% confidence interval is the range of values such that, if all possible samples of the same size  $n$  are taken, 95% of them include the true population mean somewhere within the interval around their sample means, and only 5% of them do not.

On the Bayesian quantile regression analysis, a sample of 20 000 iterations was drawn from each Markov chain, of which 5000 samples were discarded as burn in. Independent improper uniform priors were assigned for all coefficients estimated and each of the parameters was run using a random walk Metropolis–Hastings algorithm (MH algorithm).

It can be seen from Table 2 that age, BMI, gender and cigarette consumption revealed statistically significant associations with SBP across all the upper quantiles as indicated

by both the 95% confidence and 95% credible intervals which do not include 0. Depression did not present significant relations with SBP on both the 75<sup>th</sup> and 95<sup>th</sup> quantiles. Finally, exercises presented a significant coefficient with only the 95<sup>th</sup> quantile of SBP's distribution.

Using the 95% confidence and credible intervals in parentheses, Table 3 suggests that age, BMI, gender male, exercises and cigarette consumption displayed statistically significant associations with DBP across all higher quantiles (ie .75<sup>th</sup> and .95<sup>th</sup>). Race and employment status had significant coefficients on both upper quantiles of DBP's distribution for only the Bayesian approach.

### Diagnostic Tests of Convergence

Convergence occurs when the generated Markov chain converges in distribution to the posterior distribution of interest.<sup>23</sup> Convergence in Bayesian inference is critical because it deals with the accuracy with which the integrals are computed.<sup>25</sup> Convergence of the MCMC algorithm enables the output of the Bayesian inference or posterior simulation results to be reported accurately.

In this paper, convergence was assessed using the trace plots or time-series plots and the density plots. It is of utmost importance to apply several diagnostic tools for assessing the convergence.<sup>23</sup> The convergence diagnostics are meant to

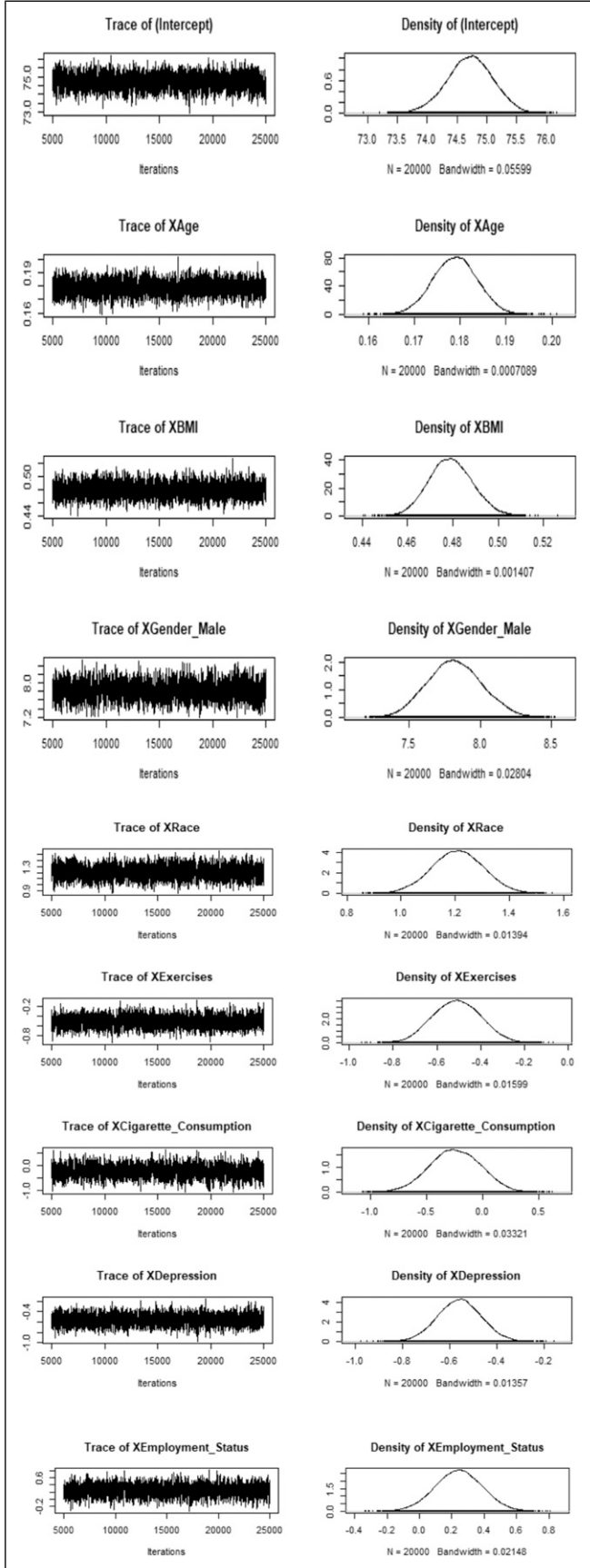


Figure 1. Trace and density plots for SBP's risk factors.

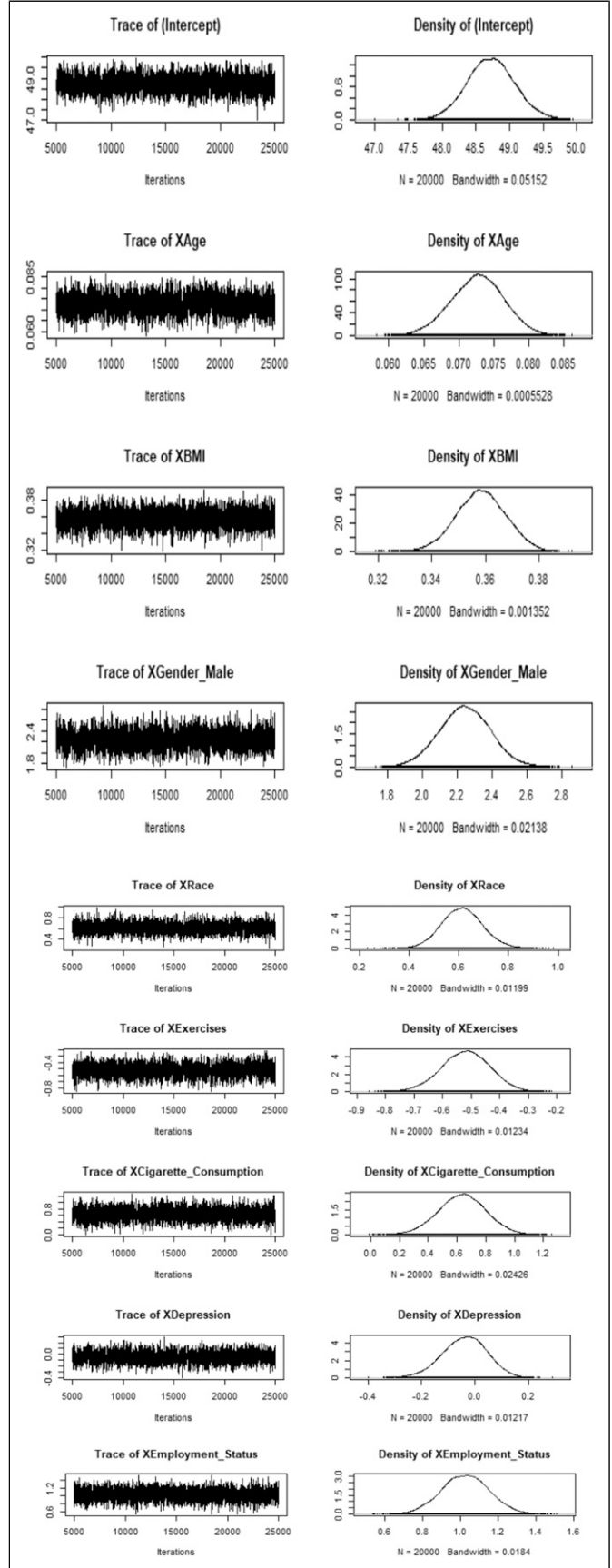


Figure 2. Trace and density plots for DBP's risk factors.

check stationarity of the Markov chain and verify the accuracy of the posterior summary measures.<sup>25</sup>

Figure 1 illustrates both the trace and density plots for each predictor variable of SBP obtained after running the MH algorithm for 20 000 iterations discarding 5000 samples as burn in. A trace plot is a time series plot showing the generated values of a parameter for each iteration in a chain.<sup>23</sup> Trace plots are most popular in checking convergence of an MCMC algorithm. If the chain has reached stationarity, the trace plot should appear as a horizontal strip and the individual moves are hardly discernable.<sup>25</sup> This is the foundation of the thick pen test.<sup>26</sup> The thick pen test does check whether the trace plot is covered by a thick pen. The mean and the variance of the trace plot should be relatively constant when stationarity occurs. Also, a trace plot shows the mixing rate of the Markov chain.

All the trace plots in Figure 1 do pass the thick pen test. No obvious trend is shown on all the time-series plots. The beginning of the runs looks almost similar to the end, implying that the chains mixed well and reached stationarity.

A density plot is a summary of the sampled values that define the stationary distribution of values, which approximates the posterior distribution of interest.<sup>27</sup> The peak of a density plot is known as the maximum a posteriori estimate. Basically, it is the mode of the distribution. Strange and unexpected peaks of a density plot can be a sign of poor convergence.

It is evident from Figure 1 that all the kernel density plots reflect convergence, showing that the Markov chain was able to find a smooth distribution.

The trace plots for DBP's risk factors illustrated in Figure 2 suggest that the chain is wandering through the same region of the parameter space and has found the stationary distribution. To ensure convergence, a burn-in of 5000 iterations was adopted. All time-series plots were generated through running the MH algorithm for 20 000 samples after the burn-in period. Also shown in Figure 2 are the smooth density plots for all DBP's risk factors.

## Model Comparisons

Based on the classical quantile regression 95% confidence intervals and Bayesian quantile regression 95% credible intervals presented in Tables 2 and 3, it can be seen that the width of the 95% credible intervals is narrower than the width of the 95% confidence intervals. This finding suggests that the Bayesian approach to quantile regression reveals more precise estimates than the frequentist approach.

## Discussion

In this article, Bayesian approach to quantile regression has been implemented by use of MCMC algorithms contained in the R package called MCMCpack. Formal and informal diagnostic tests of convergence have revealed that the Markov chain has reached the stationary distribution. Since good

convergence has been achieved, it implies that the calculated posterior summary measures are accurate and reliable. Therefore, the results can be safely used for inference.

Descriptive statistics revealed that hypertension was more prevalent in men than women for both BP measures. Both the Bayesian and classical quantile regression results corroborated these results, when gender was found to be positively significant with both BP measures, suggesting that males are more likely to suffer from hypertension than women. These results are consistent with those of the Tehran Lipid and Glucose Study (TLGS) conducted in Iran.<sup>6</sup>

It was established in Table 1 that the prevalence of uncontrolled hypertension increases with age, a finding consistent with quantile regression results. It was found that that age had positive statistically significant coefficients with both SBP and DBP, respectively. The age effect on both SBP and DBP is bigger on the 95<sup>th</sup> quantile than the 75<sup>th</sup> quantile, implying that the effect of age on SBP is stronger at the most extreme quantile of both BP measures. These results are in line with past studies that suggest that the prevalence of arterial stiffening and hypertension increases with age,<sup>7,9,28,29</sup> even though all these studies have utilised multiple logistic regression models in deriving their findings.

Highest proportions of raised blood pressure were evident in very and morbidly obese participants, an outcome confirmed by the quantile regression results that BMI had significant relations across the upper quantiles for both blood pressure measures. A study conducted among hypertensive patients on treatment in Lupane District, Zimbabwe, supported the current findings of this study.<sup>9,30</sup>

Cigarette consumption displayed positive statistically significant associations with both DBP and SBP across the higher quantiles. The highest proportions of elevated blood pressure recorded on respondents who do smoke supported this finding. These results were in agreement with earlier studies, which state that nicotine in cigarette smoke is a big part of the problem because it raises the blood pressure, and heart rate, narrows arteries and hardens their walls.<sup>4</sup>

The Bayesian and classical quantile regression results detected that exercises was negatively significant with both BP measures, implying that individuals who do not exercise (reference group) are vulnerable to high blood pressure. This result was corroborated by the high prevalence of high blood pressure among participants who do not engage in physical exercises (Table 1). A study conducted among South African adult residents of Mkhondo municipality showed that non-adherence to physical activity was related with high blood pressure.<sup>7,8</sup>

Based on the classical quantile regression 95% confidence intervals and Bayesian quantile regression 95% credible intervals presented in Tables 2 and 3, it can be seen that the width of the 95% credible intervals is narrower than the width of the 95% confidence intervals. This finding suggests that the Bayesian approach to quantile regression reveals more precise estimates than the frequentist approach. These findings

are consistent with a comparative study on the Bayesian and frequentist methods for prevalence estimation under misclassification which suggested that Bayesian prevalence estimation should be preferred over traditional frequentist methods.<sup>31</sup>

## Conclusion

This study was aimed at conducting a comparative analysis of the classical and the Bayesian approaches to quantile regression in order to study the effect of blood pressure risk factors on the upper quantiles of blood pressure's distribution. The study results suggest that the Bayesian approach to quantile regression reveals more precise estimates than the frequentist approach due to narrower width of the 95% credible intervals than the width of the 95% confidence intervals. Age, BMI, gender male, cigarette consumption and exercises presented statistically significant associations with both SBP and DBP across all the upper quantiles ( $\tau \in \{0.75, 0.95\}$ ). Basing on the study results, it is therefore suggested that Bayesian approach to quantile regression modelling to be used in estimating hypertension.

## Areas of Further Research

Panel quantile regression (Panel QR) could be another statistical technique which could be useful in the analysis of risk factors in hypertension. Panel QR has the capability to identify heterogeneous covariates effects and describe differences in longitudinal changes at different quantiles of the outcome, and provides more robust estimates when heavy tails and outliers exist.<sup>32</sup>

## Limitation of the Study

Power calculation was not done for estimation of the sample size because the researchers used secondary data which was then cleaned to yield the study sample size.

## 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.

## Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

## Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

## 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.

## Availability of Data and Materials

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

## ORCID iD

Anesu Gelfand Kuhudzai  <https://orcid.org/0000-0003-2867-6865>

## References

1. World Health Organisation. *A Global Brief on Hypertension: Silent Killer, Global Public Health Crisis*. World Health Organisation; 2013:1-39.
2. Lim SS, Vos T, Flaxman AD, et al. A comparative risk assessment of burden of disease and injury attributable to 67 risk factors and risk factor clusters in 21 regions, 1990–2010: A systematic analysis for the Global burden of disease study 2010. *The Lancet*. 2012;380(9859):2224-2260. doi:10.1016/S0140-6736(12)61766-8
3. Gómez-Olivé FX, Ali SA, Made F, et al. Regional and sex differences in the prevalence and awareness of hypertension: An H3Africa AWI-gen study across 6 sites in Sub-Saharan Africa. *Global Heart*. 2017;12(2):81. doi:10.1016/j.ghart.2017.01.007
4. Gao K, Shi X, Wang W. The life-course impact of smoking on hypertension, myocardial infarction and respiratory diseases. *Sci Rep*. 2017;7(1):4330. doi:10.1038/s41598-017-04552-5
5. Choi HM, Kim HC, Kang DR. Sex differences in hypertension prevalence and control: Analysis of the 2010-2014 Korea national health and nutrition examination survey. Spracklen CN, ed. *PLoS One*. 2017;12(5):e0178334. doi:10.1371/journal.pone.0178334
6. Kalantari S, Khalili D, Asgari S, et al. Predictors of early adulthood hypertension during adolescence: a population-based cohort study. *BMC Publ Health*. 2017;17(1):915. doi:10.1186/s12889-017-4922-3
7. Princewiel F, Cumber SN, Kimbi JA, et al. Prevalence and risk factors associated with hypertension among adults in a rural setting: the case of Ombe, Cameroon. *Pan Afr Med J*. 2019;34. doi:10.11604/pamj.2019.34.147.17518
8. Masilela C, Pearce B, Ongole JJ, Adeniyi OV, Benjeddou M. Cross-sectional study of prevalence and determinants of uncontrolled hypertension among South African adult residents of Mkhondo municipality. *BMC Publ Health*. 2020;20(1):1069. doi:10.1186/s12889-020-09174-7
9. P Opreh O, O Olajubu T, J Akarakiri K, et al. Prevalence and factors associated with hypertension among rural community dwellers in a local government area, South West Nigeria. *Afr H Sci*. 2021;21(1):75-81. doi:10.4314/ahs.v21i1.12
10. Pallant J. *SPSS Survival Manual: A Step by Step Guide to Data Analysis Using IBM SPSS*. 7th edition. Open University Press; 2020.



11. Hohl K. *Beyond the Average Case: The Mean Focus Fallacy of Standard Linear Regression and the Use of Quantile Regression for the Social Sciences*. Methodology Institute, London School of Economics and Political Science (LSE). Published online December 2009:1-21.
12. Roger K, Hallock KF. Quantile regression an introduction. *J Econ Perspect*. Published online December 2000. Accessed November 1, 2017. [https://www.researchgate.net/publication/247312065\\_Quantile\\_Regression\\_An\\_Introduction](https://www.researchgate.net/publication/247312065_Quantile_Regression_An_Introduction)
13. Yu K, van Kerm P, Zhang J. Bayesian quantile regression: An application to the wage distribution in 1990s Britain. *Sankhya*. 2005;67(2):359-377.
14. Yu K, Moyeed RA. Bayesian quantile regression. *Stat Probab Lett*. 2001;54(4):437-447. doi:10.1016/S0167-7152(01)00124-9
15. 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. *ASMSJ*. Published online April 10, 2020:1-7. doi:10.32802/asmscj.2020.sm26(1.1)
16. Koenker R. *Quantile Regression*. Cambridge University Press; 2005. doi:10.1017/CBO9780511754098
17. Reich BJ, Bondell HD, Wang H. *Flexible Bayesian Quantile Regression for Independent and Clustered Data*. North Carolina State University. Published online 2010: 1-27.
18. Statistical Analysis System Institute. Bayesian quantile regression - SAS support. Published 2017. <https://support.sas.com/rnd/app/stat/examples/BayesQuantile/quantile.pdf>. Accessed October 25, 2017.
19. Koenker R, Chernozhukov V, He X, Peng L, eds. *Handbook of Quantile Regression*. 1st ed. Chapman and Hall/CRC; 2017. doi:10.1201/9781315120256
20. Martin AD, Quinn KM, Park JH, Vieilledent G, Malecki M, Blackwell M. Markov chain Monte Carlo (MCMC) package; 2017. <https://cran.r-project.org/web/packages/MCMCpack/MCMCpack.pdf>. Accessed November 8, 2017.
21. Plummer M, Best N, Cowles K, et al. *Output Analysis and Diagnostics for MCMC*; 2016. <https://cran.r-project.org/web/packages/coda/coda.pdf>. Accessed November 8, 2017.
22. 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):494-510. doi:10.1198/jasa.2011.ap09272
23. Sinharay S. *Assessing Convergence of the Markov Chain Monte Carlo Algorithms: A Review*. Educational Testing Service; 2003:1-39. <http://www.ets.org/Media/Research/pdf/RR-03-07-Sinharay.pdf>. Accessed November 23, 2017.
24. Rea LM, Parker RA. *Designing and Conducting Survey Research: A Comprehensive Guide*. 4th ed. Jossey-Bass; 2014.
25. Lesaffre E, Lawson A. *Bayesian Biostatistics*. Wiley; 2012.
26. Gelfand AE, Smith AFM. Sampling-based approaches to calculating marginal densities. *J Am Stat Assoc*. 1990;85(410): 398-409. doi:10.1080/01621459.1990.10476213
27. Hamra G, MacLehose R, Richardson D. Markov chain Monte Carlo: An introduction for epidemiologists. *Int J Epidemiol*. 2013;42(2):627-634. doi:10.1093/ije/dyt043
28. Sun Z. Aging, arterial stiffness, and hypertension. *Hypertension*. 2015; 65(2):252-256. doi:10.1161/HYPERTENSIONAHA.114.03617
29. AlGhatrif M, Strait JB, Morrell CH, et al. Longitudinal trajectories of arterial stiffness and the role of blood pressure: The Baltimore longitudinal study of aging. *Hypertension*. 2013; 62(5):934-941. doi:10.1161/HYPERTENSIONAHA.113.01445
30. Goverwa T, Masuka N, Tshimanga M, et al. Uncontrolled hypertension among hypertensive patients on treatment in Lupane District, Zimbabwe, 2012. *BMC Res Notes*. 2014;7(1): 703. doi:10.1186/1756-0500-7-703
31. Flor M, Weiß M, Selhorst T, Müller-Graf C, Greiner M. Comparison of Bayesian and frequentist methods for prevalence estimation under misclassification. *BMC Publ Health*. 2020;20(1):1135. doi:10.1186/s12889-020-09177-4
32. Huang Q, Zhang H, Chen J, He M. Quantile regression models and their applications: A review. *J Biom Biostat*. 2017;08(03). doi:10.4172/2155-6180.1000354