Predictive Risk Factors of Hypertension in Sub-Saharan Africa: A Fixed Effect Modelling Study in Burundi

Arnaud IRADUKUNDA (arnaudiradukunda5@gmail.com)

Emmanuel Nene Odjidja
Village Health Works https://orcid.org/0000-0003-3502-5120

Cheilla IZERE
Universite Clermont Auvergne UFR de Mathematiques

Nestor NTAKABURIMVO
Pathfinder International

Arlene AKIMANA
Universite de N'Djamena Faculte des Sciences de la Sante et du Developpment

---

Research article

Keywords: High blood pressure, hypertension, logistic regression modelling, Hoaglin criterion, Welsh-Kuh distance, Burundi, sub-Saharan Africa

Posted Date: May 18th, 2020

DOI: https://doi.org/10.21203/rs.3.rs-27871/v1

Predictive Risk Factors of Hypertension in sub-Saharan Africa: A Fixed Effect Modelling Study in Burundi

Arnaud IRADUKUNDA, BSc, MD (c) [1, 2, 3]*(arnaudiradukunda5@gmail.com), Emmanuel Nene Odjidja, MSc[3, 4] (emmaodjidja@gmail.com), Cheilla IZERE, BSc, MSc (c) [5] (cheilla61@gmail.com), Nestor NTAKABURIMVO, BSc[2, 6] (ntakaburimvonestor@gmail.com), Arlene AKIMANA, MD[6] (akimana2010@gmail.com)

[1] Department of Medicine, University of Burundi, PB1550, Bujumbura, Burundi

[2] Department of Statistics, Lake Tanganyika University, PB 5304, Mutanga, Burundi

[3]Royal Society of Tropical Medicine and hygiene, London, United Kingdom

[4]Village Health Works, Burundi

[5]Department of Computer Mathematics, Clermont Auvergne University,PB 63000, France

[6]Institut Universitaire des Sciences de la Santé et du Développement, Bujumbura, Burundi

* Correspondence: arnaudiradukunda5@gmail.com

Abstract

Background

Hypertension, signalled by persistently high systolic and diastolic blood pressure is a major threat to public health globally. Especially in sub-Saharan African countries, this coexists with high burden of other infectious diseases, creating a complex public health situation which is difficult to address. Tackling this will require targeted public health intervention based on evidence that well defines the at risk population. In this study, using retrospective data from two referral hospitals in Burundi, we model the risk factors associated with hypertension in Burundi

Methods

Retrospective data of a sample of 353 randomly selected from a population of 4,380 patients admitted in 2019 in two referral hospitals in Burundi: Military and University teaching hospital of Kamenge. The predictive risk factors were carried out by fixed effect logistic regression. Model performance was assessed by Area under Curve (AUC). Model was internally validated via bootstrapping with 2000 replications. All analysis were conducted in R.

Results

Overall, 16.7% of the patients were found to be hypertensive. After adjustment of the model for cofounding covariates, associated risk factors found were advanced age (40 years) AOR: 6.03, 95% CI: 1.86- 17.19) and above 60 years, (AOR: 12.76, 95% CI: 3.30 – 14.26). Patients comorbid with chronic kidney failure were 4.95 times more (95% CI: 1.83-15.82) to be hypertensive and among those with family history of hypertension, the adjusted risk were twice. Compared to non-smokers, smokers were 2.87 times more likely to develop hypertension (95% CI: 0.87 – 9.15). The highest probabilities are observed to patients who are at the same time smokers, overweight, with chronic kidney failure, family history with hypertension with secondary or university as highest educational level. The model had an excellent predictive performance (AUC), accurately predicting 88.71% (95% CI: 84.17%-92.5%) of all observations

Conclusion

The relatively high prevalence and associated risk factors of hypertension in Burundi raises a call for concern especially in this context where there exist an equally high burden of infectious diseases, other chronic diseases including chronic malnutrition. Targeting interventions based on these identified risk factors will allow judicious channel of resources and effective public health planning.

Keys words: High blood pressure; hypertension; logistic regression modelling; Hoaglin criterion; Welsh-Kuh distance; Burundi; sub-Saharan Africa

Background

Hypertension corresponds to a permanently raised blood pressure in arteries and arterioles. It is defined as a systolic blood pressure equal or above 140 mmHg and /or a diastolic blood pressure above 90 mmHg. Hypertension is a threat to global public health as it tires vessels, the heart and causes damage to artery walls [1]. It is a major risk factor for cardiovascular diseases [2] with high morbidity and mortality rate [3]. If not identified and treated early, arterial hypertension may result in serious complications including strokes, coronary artery, kidney and hypertensive heart diseases [4, 5, 6] which are among the leading causes of mortality in the world. Approximately, cardiovascular diseases account for 17.8 million death in 2017 [7], nearly 1/3 of total [8], of which more than three quarters were in low and middle-income countries (LMICs). Hypertension complications, cardiovascular diseases account 9.4 million (52.8%) every year [7]. It is responsible for 45% deaths due to heart disease and 51% stroke related deaths [8, 9]. Premature death and health care expenditure for treatments due the hypertension puts an economic toll on families and pushes many into poverty [10]. At the macro level, these high expenses and human losses significantly impacts on economic growth and reduces productivity [11, 12].

In 2015 only, the prevalence of hypertension in adults was 40% with an estimated 1.13 billion people living with different forms of hypertension. [11]. Data from the World Health Organisation Global Health Observatory Repository [12] found the highest prevalence of hypertension in the Africa region (46%) followed by the Americas (35%) and other regions, majority of whom, were undiagnosed and untreated [13]. In sub-Saharan Africa (SSA), just like other settings, hypertension has been associated with lifestyles, diets, physical inactivity urbanization and socio economic status [14, 15, 16]. More than 125 million people with hypertension are expected by the year 2025 in SSA alone [17, 18]. By year 2030, hypertension and other non-communicable diseases are projected to surpass communicable diseases as the top of mortality causes on the continent [1, 17]. From 2011 to 2025, the cumulative lost output with non-communicable diseases is projected to be US$7.28 trillion in low and middle income countries which is approximately a loss of US$500 billion per year [19]. Cardiovascular diseases including hypertension account for nearly half this cost [20]. Despite this, SSA faces a major problem of early screening, timely treatment and control of hypertension [21].

Burundi at present ranks 16th worldwide and 12th in sub-Saharan Africa on age standardized hypertension and related mortality. Yet, studies to understand the epidemiology and associated in the context of Burundi are lacking, prompting the conduct of this study [22]. Therefore, in this study, we determine the overall prevalence of hypertension, evaluate predictive risk factors, and predict their probabilities. Knowing these factors could support effective public health planning and facilitate policy makers to formulate plausible policies towards the fight against hypertension and its complications.

Methods

Sources of data and sampling methods

Data used in this study were collected in 2018 in two referral hospitals in Burundi: University teaching hospitals of Kamenge and Kamenge Military hospital in different departments. Cross sectional population of 4380 patients were stratified in 2 groups in both hospitals. Random sampling were done with proportional allocation to the admitted patients by service and by hospital.

Inclusion Probability and Sample Size Calculation

Inclusion probability p_i is the same for patients admitted in the same i service of the selected hospital and is calculated as:

$$pi = \frac{N_i}{N} \tag{1}$$

With N_i total number of patients in i service and N population size.

The minimal size of sample is calculate as

$$n = \frac{t_p^2 \; P(1-P)N}{t_p^2 \; P(1-P) + y^2(N-1)} \tag{2}$$

t_p The quintile of the normal low with at 95% of confidence (1.96), N population size, n minimal size sample, p prevalence of hypertension and y acceptable margin (5%). As the prevalence of hypertension is unknown, a value of $p = 0.5$ was selected. According to these parameters, the minimal size of the sample is 353 patients. Basically, a respondent was selected if he had of the following criteria: Admitted in internal medicine service in the period of 2018, measured diastolic and systolic blood pressure three times.

Outcome and Independent Variables

In this study, quantitative and qualitative variables were used. Hypertension was considered as outcome variable and was defined as systolic blood pressure $\geq$ 140 mmHg and/or diastolic blood pressure $\geq$ 90 mmHg. Body mass index (BMI) was calculated as the weight in kilograms to square of height in meters and was categorized in into; underweight (BMI<18.5kg/m^2), normal (BMI 18.5-24.9 kg/m^2), overweight (BMI 25.0-29.9 kg/m^2) and obesity (BMI $\geq$ 30kg/m^2). Others independents variables were classified as: sex (Men, women), residence (urban, rural), educational level (Primary or less, secondary, university), alcohol consumption (yes/no), smoking (yes/no), chronic kidney failure (yes /no), diabetes (yes/no), cardiovascular comorbidity (yes/no) and familial history of hypertension (yes/no).

Data Analysis

Data analysis were undertaken in different steps: descriptive statistic, binary logistic modelling with fixed effects, power predictive evaluation of final model and probabilities prediction. Hypertension associated risk factors were done firstly via univariate and multivariate logistic regressions secondary. We calculated the odds ratios (ORs) at 95% confidence level for each covariate to identify predictors of hypertension. The risk estimate equation for multiple logistic regression is as follows:

$$p = \frac{e^{\beta_0 + \beta_1 X_1 + \ldots + \beta_k X_k + \varepsilon}}{1 + e^{\beta_0 + \beta_1 X_1 + \ldots + \beta_k X_k + \varepsilon}} \quad (3)$$

$$\operatorname{logit}(p) = \log\left(\frac{p}{1-p}\right) = \beta_0 + \beta_1 X_1 + \ldots + \beta_k X_k + \varepsilon \quad (4)$$

Where p is outcome realization probability, β_0 intercept, β_i coefficients, X_i independents variables and ε error. Significant variables on 15% threshold were introduced in multivariate logistic modelling to determine a combined effect on the outcome. Finally, the predictor variables of the model were manually selected step by step using decreasing method on a 5% threshold. The likelihood ration test, the score test and the Wald test were used to determine significance of independent variables on the outcome [23].To select the best model for this study, the Akaike Information Criterion (AIC) based on adjustment [24] were used in a given equation as:

$$\mathrm{AIC} = 2k - 2\ln(L) \quad (5)$$

Where L refers to the maximum value of the likelihood function of the model, k model parameters number. The best model is one with low AIC value. The relevance of the final model to make prediction was assessed by Pearson residuals test. Receiving Operating characteristics (ROC) and Area under Curve (AUC) were respectively used to compare and evaluate performance and predictive power of the model. Furthermore, the ROC was used to determine the discriminatory performance of the model, determining the false positive and false negative rates. The Mann Whitney statistics method showed that the two distributions were offset: normotensive people had an average higher scores than hypertensive people.

Each individual's score was ranked in ascending order. Thus, the AUC which determined the number of observations accurately predicted was calculated as:

$$\mathrm{AUC} = \frac{\sum_{i:yi=1}^{n} r_i - \frac{N_+(N_+ + 1)}{2}}{N_+ N_-} \quad (6)$$

Where AUC refers to the Area under Curve, N_+ number of normotensive people, N_- hypertensive people, r_i the ranks of i normotensive individual. If AUC=0.5 no discrimination, $0.7 \leq \mathrm{AUC} \leq 0.8$ acceptable discrimination, $0.8 \leq \mathrm{AUC} < 0.9$ excellence discrimination, $\mathrm{AUC} \geq 0.9$ exceptional discrimination [25]. Influents points of the model was analysed by Hoaglin and Welsh criterion:

$$\frac{2*(p+1)}{n} \quad (7)$$

Where p is parameters of the model and n the sample size.

The R software (3.5.0) Foreign and forest model packages were used to carry out results in this study [26].

Results

More than 75% of the patients had high normal blood pressure according the World Health Organisation's (WHO) classification (Table 1). Overall, the prevalence of hypertension is 16.71% (Table 1).This prevalence 2 times higher in overweight patients and 3 times more in diabetic patients. The high proportions above the overall prevalence are observed in people with cardiovascular comorbidity, married, people aged between 40-60 years or 60 and above and over, chronic kidney failure, men, smokers, obese people, with secondary and university level. These prevalence are 16.9%, 19.3%, 21.4 %, 33.3%, 29.3%, 17.7%, 28.6%, 39.1%, 18.2% and 29.4% respectively.

Table 1: Descriptive Characteristics of patients included in the study

Individual characteristics	**Modalities**	**N**	**H^-**	**H^+**	**P+**	***95% CI***
Sex	Men	206	173	33	16.0	11.29 - 21.75
	Women	147	121	26	17.7	11.89 - 24.83
Age	15-39	138	135	3	2.2	00.45 - 06.22
	40-59	131	103	28	21.4	14.70 - 29.39
	≥60	84	56	28	33.3	23.42 - 44.46
Marital status	Married	89	81	8	9.0	03.96 - 16.94
	Unmarried	264	213	51	19.3	14.74 - 24.60
Educational level	Primary	137	123	14	10.2	05.70 - 16.55
	Secondary	165	135	30	18.2	12.62 - 24.93
	University	51	36	15	29.4	17.49 - 43 .83
Alcohol	No	173	145	28	16.2	11.03 - 22.53
	Yes	180	149	31	17.2	12.01 - 23 .55
Smoking	No	325	274	51	15.7	11.91 - 20.11
	Yes	28	20	8	28.6	13.22 - 48.67
Cardiovascular comorbidity	No	28	24	4	14.3	04.03 - 32.66
	Yes	325	270	55	16.9	13.01 - 21.45
Family History with Hypertension	No	289	262	27	9.3	06.25 - 13.30
	Yes	64	32	32	50.0	37.23 - 62.77
Diabetes	No	273	247	26	9.5	06.32 - 13.64
	Yes	80	47	33	41.3	30.35 - 52.82
Chronic kidney failure	No	169	164	5	3.0	00.98 - 06.77
	Yes	184	130	54	29.3	22.88 - 36.49
Obesity	No	330	280	50	15.2	11.46 - 19.48
	Yes	23	14	9	39.1	19.71 - 61.46
Overweight	No	271	237	34	12.5	08.85 - 17.08
	Yes	82	57	25	30.5	20.80 - 41.67

Logistic Regression Modelling of Predictive Risk factors of Hypertension

The table 2 shows the univariate logistic regressions where hypertension is modeled by each explanatory variable.

Table 2: Univariates logistic regressions

Variables	Modalities	OR	95% CI	p	AIC
Educational level (EL)	At most primary	Reference			
	Secondary	1.95	[1.01,3.96]	0.045	314.64
	University	3.66	[1.61,8.38]	0.002	
Smoking	No	Reference			
	Yes	2.15	[0.85,04.98]	0.085	319.95
Diabetes	No	Reference			
	Yes	6.67	[3.67, 12.27]	<0.001	284.15
Age	15-39 years	Reference			
	40-59 years	12.23	[4.19, 22.14]	<0.001	277.78
	60 and over	22.49	[17.58,26.78]	<0.001	
Currently working	No	Reference			
	Yes	0.980	[0.56, 1.72]	0.944	322.63
Residence	Rural	Reference			
	Urban	1.05	[0.60, 1.84]	0.868	322.6
Obesity	No	Reference			
	Yes	3.60	[1.43, 8.66]	0.005	315.51
Overweight (OW)	No	Reference			
	Yes	1.13	[0.64, 1.98]	<0.001	322.46
Sex	Women	Reference			
	Men	0.78	[0.40,1.48]	0.451	261.81
Marital status	unmarried	Reference			
	married	2.42	[1.16, 5.72]	0.027	316.95
Cardiovascular comorbidity	No	Reference			
	Yes	1.22	[0.45, 4.28]	0.720	322.5
Alcohol	No	Reference			
	Yes	1.08	[0.61, 1.89]	0.794	322.56
Chronic Kidney Disease	No	Reference			
	Yes	13.62	[5.81, 19.96]	<0.001	271.78
Familial History of	No	Reference			
	Yes	9.70	[5.20, 18.4]	<0.001	272.13

Eight variables were significantly associated to hypertension in univariate analysis (**Table 2**). Those variables were: Educational level{University (OR=3.66;95% CI=1.61,8.38;p<0.002), Secondary(OR=1.95; 95% CI =1.01-3.96; p=0.045), Diabetes (OR=6.67;95% CI = 33.67-

12.27;p<0.001), Age (40-59 years (OR=12.23;95% CI=4.19-22.14; p<0.001), 60 and above (22.49; 95% CI =7.58- 26.78, <0.001), Obesity (OR=3.60; 95% CI =1.43- 8.66; 0.005), Overweight (OR=1.13; 95% CI = 0.64-1.98; p<0.001), Marital status (OR=2.42; 95% CI =1.16-5.72; p=0.027), Chronic kidney disease (OR=13.62, 95% CI =5.81-19.96; p<0.001) and family hypertension history (OR=9.70, 95% CI =5.20-18.4, p<0.001). All these variables were introduced in a multivariate analysis. In addition, other relevant variables were introduced in the multivariate model as well, namely: sex, smoking, alcohol and cardiovascular comorbidity. The identification of variables with significant effects shows six variables are significantly associated with hypertension. The table 3 shows the variables with significant association on hypertension.

Table 3: Significantly associated variables on hypertension

Variables	**Df**	**Deviance**	**AIC**	**LRT**	***p***
Educational level	2	225.12	239.12	18.91	<0.001
Smoking	1	209.25	225.25	3.04	0.081
Age	2	221.62	235.62	15.41	<0.001
Overweight	1	213.38	229.38	7.17	0.007
Chronic kidney disease	1	216.67	232.67	10.47	0.001
Hypertension familial history	1	212.87	228.87	6.6632	0.018

A saturated equation model was derived from the above multivariate model denoted as:

$$\operatorname{logt}(p) = \beta_0 + \beta_1 \times EL + \beta_2 \times Smoking + \beta_3 \times Age + \beta_4 \times OW + \beta_5 \times CKF + \beta_6 \times FHHTN + \varepsilon \quad (6)$$

With $\beta_0 = -6.23$, $\beta_1 = (1.49, 2.09), \beta_2 = 1.06, \beta_3 = (1.79, 2.54)$, $\beta_4 = 1.03, \beta_5 = 1.59$, $\beta_6 = 1.08$

EL: Educational level, OW: Overweight, CKF: Chronic kidney failure and FHHTN: familial history of hypertension. After controlling the cofounding variables as indicated above, an adjusted odds ratio, their lower and upper bound along with a corresponding p-value were derived. Table 4 details these results.

Table 4: Results of multivariate saturated logistic model

Variables	Modalities	N	N^-	N^+	AOR	95% CI	p
Educational level	Primary or less	137	123	14	Reference		
	Secondary	165	130	35	4.46	[1.91-11.15]	<0.001
	University	51	36	15	8.09	[2.75-25.47]	<0.001
Smoking	No	325	274	51	Reference		
	Yes	28	20	8	2.87	[0.87- 9.15]	0.076
Age	15-39 years	138	135	3	Reference		
	40-59 years	131	103	28	6.03	[1.86-17.19]	0.007

	60 years and over	84	56	28	12.76	[3.30-14.26]	<0.001
Overweight	No	271	237	34	Reference		
	Yes	82	57	25	2.79	[1.32-5.97]	0.007
CKF	No	169	164	5	Reference		
	Yes	184	130	54s	4.95	[1.83-15.82]	0.003
Familial history of HTN	No	289	262	27	Reference		
	Yes	64	32	32	2.93	[1.29- 6.87]	0.011

Table 4 shows that advanced age and education level are the most associated risk factor of hypertension with $p<0.001$. It follows chronic kidney failure, overweight, hypertension familial history and smoking with probabilities equal to 0.003, 0.007, 0.011 and 0.076 respectively. Patients with university level have twice hypertension risk than those with secondary level people and more than 8 times the risk to become hypertensive than those with none or primary level. Smokers have more than twice the risk (AOR: 2.87, 95% CI: 0.87- 9.15) of becoming hypertensive than non-smokers. People aged between 40-59 years have more than 6 times (AOR: 6.03, 95% CI: 1.86- 17.19) the risk to become hypertensive than people aged under **40years old**. This risk is 12.76 times (AOR: 12.76, 95% CI: 3.30-14.26) in people over 60 years old. People with a family history of hypertension and overweight people have more than twice the risk of hypertension. People with chronic kidney failure have approximately 5 times (AOR: 4.95, 95% CI: 1.83-15.82) the risk of hypertension than normal people. The Wald test ($X^2 = 86.7$, df = 8, $p<0.001$) rejects the null hypothesis and therefore to confirm the alternative hypothesis stating that there is at least one coefficients significantly different to zero. Pearson residuals test of ($X^2=$ 266.17, df= 344) was determined with a p-value of 0.99 which shows the model was well adjusted on the observations. A McFadden statistic (R^2: 0.35) also indicates that this model has a good fit.

The influential points' analysis based on Hoaglin and Welsh criterion shows that only 9 points are influential. Also, Cook's distance shows that 3 points (108,114 and 199) are outliers, it means the influential points are not numerous. Studentized residue analysis (Figure 1) shows that 97% (343/353) are between -2 and 2. Observations with residues (Figure 2) greater than 2 are ten (9, 34,105, 108, 114,199, 212, 232,265,273). No observation with studentized residue less than -2, indicating that the number of outliers are negligible.

Cross validation and Probabilities predictions

The figures 3 and 4 show respectively the ROC curve and complexities parameters from the decision tree. A bootstrap method with 2000 replications was used to determine an AUC of 88.71% (95% CI: 84.17% - 92.5%) which suggests an excellent discrimination. This implies that saturated model has an excellent predictive power and probabilities accurately determine people with hypertension based on these identified characteristics. On the decision tree, Root node error: 59/353 = 0.167. Substitution Error is equal to 0.1331.

The table 5 shows the predicted probabilities of becoming hypertensive based on different scenarios of having either a risk factor or a combination of risk factors. The first individual is the one with no risk factors which was considered as the reference individual. Fifteen predictions were generated from reference individual to whom with all hypertension risk factors.

Table 5: Predicted probabilities

Individuals	EL	Smoking	Age	Overweight	CKF	FHHTN	p
1*	Primary or less	no	15-39	no	no	no	0.002
2	Primary or less	**yes**	15-39	no	no	no	0.006
3	Primary or less	no	40-59	no	no	no	0.011
4	Primary or less	no	60 and over	no	no	no	0.062
5	Primary or less	no	15-39	**yes**	no	no	0.011
6	Primary or less	no	15-39	no	**yes**	no	0.009
7	Primary or less	no	15-39	no	no	**yes**	0.005
8	Primary or less	**yes**	40-59	no	no	no	0.031
9	Primary or less	**yes**	40-59	**yes**	no	no	0.031
10	Primary or less	**yes**	60 and over	**yes**	no	no	0.161
11	Primary or less	**yes**	40-59	**yes**	**yes**	**yes**	0.136
12	Primary or less	**no**	60 and over	**yes**	**yes**	**yes**	0.492
13	Secondary	**yes**	40-59	**yes**	**yes**	**yes**	0.850
14	Secondary	**yes**	60 and over	**yes**	**yes**	**yes**	0.972
15	University	**yes**	60 and over	**yes**	**yes**	**yes**	0.999

*Reference individual, **p** probability of becoming hypertensive based on combination of risk factors

The reference individual have 0.002 as probability to become hypertensive. This probability goes from simple to the triple when the last one is smoker. Individual aged between 40 and 60 old and the overweight individual have the same probability to become hypertensive (p=0.011).This probability is six times higher to individual with 60 years old and over. If the individual aged between 40 and 60 who is at the same time smoker and overweight, his probability to become hypertensive is 0.031. If this last one is 60 years old or over, that probability is 5.2 times higher the last one (p=0.161). If the overweight individual born in the family of hypertensive people have chronic kidney disease, his probability to become hypertensive is around 50% (p=0.492). That probability double when that preceding is smoker. Individuals with all risk factors have at least 0.850 as probability. The highest probabilities are observed firstly to patients who are at the same time smokers, secondary level, chronic kidney failure, born in the hypertensive family and secondly to the last one but with university as highest educational level. Their probabilities are 0.972 and 0.999 respectively.

General cardiovascular risk evaluation based on high blood pressure

Table 6 shows different levels of cardiovascular risk according to association of blood pressure stages with hypertension risk factors. Risk are divided in five groups: Low risk (<15%), Moderate risk (15-20%), High risk (20-30%) and Very High Risk (>30%).

Table 6: Cardiovascular risk evaluation

BP / RF	Normal	HN	Stage1	Stage2	Stage3
	0 risk	0 risk	Low risk	Moderate risk	High risk
0 RF	**0**	**15**	**0**	**0**	**0**
1-2 RF	Low risk	Low risk	Moderate risk	Moderate risk	High risk
	1	**143**	**18**	**6**	**6**
≥3 RF	Moderate risk	High risk	High risk	High risk	Very high risk
	1	**88**	**24**	**8**	**11**
≥3RF+ Clinical symptoms	Very high risk	Very high risk	Very high risk	Very high risk	Very high risk
	0	**13**	**8**	**6**	**5**

In this study, only 15 patients had zero risk of cardiovascular diseases. More than a 1/3 had low risk (<0.15), 25 had moderate risk between 0.15 and 0.20,126 patients a high risk less than 0.30 and 35 patients had very high risk more than 30%. The more the blood pressure is high, more the risk increased. Also, more the risk factors increase at the same individual, more the risk to develop cardiovascular diseases. Finally the combination of the two parameters to the same person increase his probability of becoming patients with cardiovascular diseases.

Discussion

In this study, we determined the prevalence of hypertension, identified principal predictive risk factors of hypertension and predictive probabilities to become hypertensive based on a combination of risk factors. Overall, the prevalence of hypertension was 16.7%. Considering women only, the prevalence was 17.7% (95%CI 17.24% - 23.31%). This prevalence is similar to results of a recent study conducted in Lesotho which showed a prevalence of 17.3% among women [37]. In another study conducted in Saudi Arabia in 2014, the prevalence was 15.2% among those aged 15 years old and above had different levels of hypertension [28]. This study did not show the significant difference of the prevalence (X^2=0.080, df =1, p= 0.778) between women (16%) and men (17.7%) [28]. This finding is consistent in a study conducted in Benin where there was no significant difference between men (32.8%) and women (33.0%) [29]. The highest prevalence of hypertension observed in diabetic patients (41.2%) and the lowest in youngest patients (2.2%) aged under 40 years [29].

Literature on marital status and hypertension is inconclusive and mostly compares never married to currently married persons [30]. In congruence to this, our study did not show association between marital status and hypertension. This is contrary to what had previously reported on this association [31]. After adjusting hypertension on other covariates via logistic regression model, high educational level, smoking, advanced age, overweight, chronic kidney failure and familial history of hypertension are the associated factors with hypertension. Similar findings were found in several previous studies in developing countries: Malawi, Uganda, Northwest of Ethiopia and Birmania in 2018, 2015, 2015, and 2016 respectively which showed that the factors associated in odds of hypertension were overweight, smoking, education level and older age [32, 33, 34, 35]. The association between advanced age and high risk of hypertension could be due to the biological effect of increased arterial resistance which increases with old age [36]. Our study did not find the association between residence and marital status. Furthermore, as in this study, alcohol was also not associated with hypertension which is also consistent in two studies conducted among Europeans countries and Beninese [29, 37].

This study shows that predicted probabilities to become hypertensive is low in young people, aged under 40 years. High probabilities are observed in oldest people with many risk factors (Table 5). The highest probabilities, more than 60% were observed in people aged 40 years and above, with presence of all others risk factors as shown in 13th, 14th and 15th individual with 85.0%, 97.2% and 99.9% as probabilities respectively. Also, more than a 1/3 had low risk (<0.15), 25 had moderate risk between 0.15 and 0.20, 126 patients a high risk less than 0.30 and 35 patients had very high risk more than 30%, underscoring that the higher the blood pressure, the higher the cardiovascular risk. Similar probabilities were found in recent Chinese study conducted in 2020 which shows that <20% cumulative risk of hypertension for 57.62% of participants, 20-40% risk for 27.24%,40-60% risk for 12.19% ,and > 60% risk for 2.96% of participants [38].

One strength of our study is the ability to study hypertensive and normotensive people at the same time, combining descriptive and inferential statistics (logistic regression with fixed effects, Wald test, deviance test) to build the ROC curve and complexities parameters using decision tree. Another strength of the study is the ability to estimate area under curve and build bootstrap

AUC interval confidence using Bootstrap method to analyse model's residuals using Welsh-Kuh's distance to predict probabilities of becoming hypertensive given a combination of risk factors. However, despite these strengths, some limitations should be noted during interpretation and policy formulation. First, our study used secondary data and as such, we were unable to measure quantities and type of alcohol and tobacco consumed as well as obtain information on physical activities which have been found to be associated with hypertension. Caution should be taken when generalizing findings on high blood pressure as data used were only reported from two hospitals. To validate findings, additional studies should be conducted in other hospitals in the country and take into others characteristics including more biomarkers. A random effect logistic regression or Bayesian regression based on Markov chain Hamiltonian Monte Carlo simulates and Langevin algorithms could give precision in the estimation of model's parameters and Bayesian credibility intervals as such these methods are recommended for future research. The main interest of this study is to identify predictive risk factors which allows prediction probabilities of hypertension and further evaluating cardiovascular risks controlling possible cofounders.

Conclusion

This study showed that the hypertension prevalence was 16.7% with insignificant differences between men and women. Predictive risk factors of hypertension were advanced age (40-59 years, 60 years and over), smoking, presence of chronic kidney failure, cardiovascular comorbidity, educational level and overweight.

The lowest predicted probability is observed to young people with no risk factors. More than 85% predicted probabilities to become hypertensive are observed to people with all risk factors. Resources in Burundi are scarce, therefore, the tackling the high burden of cardiovascular diseases should be based on instituting systems for early detection and prompt treatment especially those identified as high risks. At the community level, efforts should be channelled towards intensifying innovative and inclusive health promotion aimed at behaviour change. At the health system, creating a risk-based nomogram based on these identified risks factors could allow those at high risks to be identified early and well targeted with the needed treatment. Finally, provision of long term care for those identified cases will depend on not just consistent treatment but also on the overall health systems' strengthening. This will ensure sustainability and effectiveness of public health interventions aimed at tackling chronic diseases along with other high burden infectious diseases.

List of Abbreviations

Abbreviation	**Full Meaning**
AIC	Akaike Information Criterion
AOR	Adjusted Odds Ratio
AUC	Area Under Curve
BMI	Body Mass Index
BP	Blood Pressure
CI	Confidence Interval
CKF	Chronic Kidney failure
Df	Degrees of freedom
FHHTN	Family History with Hypertension

HN	High Normal
HTN	Hypertension
LRT	Likelihood Ratio Test
N-	Normotensive people
N+	Hypertensive people
OR	Odds Ratio
OW	Overweight
RF	Risk Factors
ROC	Receiver Operating Characteristic
SSA	Sub-Saharan Africa
WHO	World Health Organisation

Declarations

Ethics approval and consent to participate

Written permission was acquired from ethical committees of the University teaching hospitals of Kamenge and Kamenge Military hospital to use this retrospective data for this study. In an effort to secure patient identity, patient data were anonymised and replaced with unique codes.

Consent for publication

N/A

Availability of data and material

Anonymised primary dataset are available upon reasonable request from the corresponding author via arnaudiradukunda5@gmail.com

Competing interests

The authors declare no competing interest

Funding

N/A

Authors' contributions

AI conceptualised and designed the study and wrote the first draft. ENO provided technical support, reviewed the first draft and wrote some sections of the manuscript. CI, NN supported in data cleaning, assisted in writing first draft and also provided reviews for subsequent versions of this manuscript. AA supervised the entire study and provided technical guidance at every stage from conceptualization to manuscript writing. All authors read and approved the final version of this manuscript to submission to BMC Public Health.

Acknowledgements

The authors are thankful to all frontline workers and administration of both hospitals

References

1. World Health Organization, A global brief on hypertension. WHO 2013.
2. World Health Organisation, 2020. *WHO | The Atlas Of Heart Disease And Stroke*. [online]Who.int. Available at: https://www.who.int/cardiovascular_diseases/resources/atlas/en/ [Accessed 2 May 2020].
3. Wang, H., Dwyer-Lindgren, L., Lofgren, K.T., Rajaratnam, J.K., Marcus, J.R., Levin-Rector, A., Levitz, C.E., Lopez, A.D. and Murray, C.J., 2012. Age-specific and sex-specific mortality in 187 countries, 1970–2010: a systematic analysis for the Global Burden of Disease Study 2010. *The Lancet*, *380*(9859), pp.2071-2094.
4. Messerli, F.H., Williams, B. and Ritz, E., 2007. Essential hypertension. *The Lancet*, *370*(9587), pp.591-603.
5. K Singh, T., Arya, V. and Navaratnarajah, N., 2014. Chronic kidney disease and cardiovascular disease: a focus on primary care. *Cardiovascular & Haematological Disorders-Drug Targets (Formerly Current Drug Targets-Cardiovascular & Hematological Disorders)*, *14*(3), pp.212-218.
6. J. He and P. K. Whelton, "Elevated systolic blood pressure as a risk factor for cardiovascular and renal disease," Journal of Hypertension, vol. 17, no. 2, pp. S7–S13, 1999.
7. Roth GA, Abate D, Abate KH, et al. Global, regional, and national age-sex-specific mortality for 282 causes of death in 195 countries and territories, 1980–2017: a systematic analysis for the Global Burden of Disease Study 2017. Lancet 2018; 392: 1736–88.
8. World Health Organisation, 2008. *Causes Of Death*. [online] Who.int. Available at: https://www.who.int/healthinfo/global_burden_disease/cod_2008_sources_methods.pdf [Accessed 2 May 2020].
9. Lim SS, Vos T, Flaxman AD, Danaei G, 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. Lancet. 2012; 380 (9859): 2224-60
10. World Health Organization. Impact of out-of-pocket payments for treatment of non-communicable diseases indeveloping countries : A review of literature WHO Discussion Paper 02/2011. Geneva, World Health Organization.
11. Zhou B, Bentham J, Di Cesare M, Bixby H, Danaei G, Cowan MJ, Paciorek CJ, Singh G, Hajifathalian K, Bennett JE, Taddei C. Worldwide trends in blood pressure from 1975 to 2015: a pooled analysis of 1479 population-based measurement studies with 19.1 million participants. Lancet (London, England). 2017;389(10064):37–55.
12. World Health Organization. Global Health Observatory Data Repository [online database]. Geneva, World Health Organization, 2008 (http://apps.who.int/gho/data/view.main) Accessed 11th February 2013.
13. World Health Organization, 2011. *Global Status Report On Non-communicable Diseases 2010*. [online] World Health Organization. Available at: https://www.who.int/nmh/publications/ncd_report2010/en/ [Accessed 2 May 2020].

14. Chow CK, Teo KK, Rangarajan S, Islam S, Gupta R, Avezum A, et al. Prevalence, awareness, treatment, and control of hypertension in rural and urban communities in high-, middle-, and low-income countries. Jama. 2013; 310(9):959–68.
15. Hendriks ME, Wit FWNM, Roos MTL, Brewster LM, Akande TM, de Beer IH, et al. Hypertension in sub-Saharan Africa: cross-sectional surveys in four rural and urban communities. Atashili J, editor. PLoS One. 2012;7(3):e32638.
16. Nahimana, M.R., Nyandwi, A., Muhimpundu, M.A., Olu, O., Condo, J.U., Rusanganwa, A., Koama, J.B., Ngoc, C.T., Gasherebuka, J.B., Ota, M.O. and Okeibunor, J.C., 2018. A population-based national estimate of the prevalence and risk factors associated with hypertension in Rwanda: implications for prevention and control. *BMC public health*, *18*(1), p.2.
17. Whitworth JA. World Health Organization, International Society of Hypertension Writing Group. 2003 World Health Organization (WHO)/International Society of Hypertension (ISH) statement on management of hypertension. J Hypertens 2003;21:1983–92
18. Institute for Health Metrics and Evaluation (IHME) 2015. http://www. healthdata.org.
19. World Health Organization and World Economic Forum. From Burden to "Best Buys": Reducing the Economic Impact of Non-Communicable Diseases in Low- and Middle-Income Countries. Geneva, World Health Organization and World Economic Forum, 2011. (http ://www.who.int/nmh/publications/best_buys_summary)
20. World Economic Forum and the Harvard School of Public Health, 2020. *Global Economic Burden of Non-Communicable Diseases*. [online] Www3.weforum.org. Available at: http://www3.weforum.org/docs/WEF_Harvard_HE_GlobalEconomicBurdenNonCommunicableDiseases_2011.pdf [Accessed 2 May 2020].
21. Ataklte, F., Erqou, S., Kaptoge, S., Taye, B., Echouffo-Tcheugui, J. B., & Kengne, A. P. (2014). Burden of Undiagnosed Hypertension in Sub-Saharan Africa: A Systematic Review and Meta-Analysis. Hypertension, 65(2), 291–298.
22. BARANKANIRA, E., IRADUKUNDA, A. and NTAKABURIMVO, N., 2019. Facteurs déterminants du mauvais contrôle glycémique au Burundi. *Afrique SCIENCE*, *15*(5), pp.74-87.
23. Kamble, T.S., Kashid, D.N. and Sakate, D.M., 2019. Consistent and robust variable selection in regression based on Wald test. *Communications in Statistics-Theory and Methods*, *48*(8), pp.1981-2000.
24. P. K. Josephat and A. Ame, "Effect of Testing Logistic Regression Assumptions on the Improvement of the Propensity Scores", Int J Stat Appl.; 8(1) (2018) 9-17
25. D.W. Hosmer, S. Lemeshow, Applied Logistic Regression, Second Edition, Wiley, 2000
26. R Development Core Team, "R: A language and environment for statistical computing", R Foundation for Statistical Computing, (2016), Vienna, Austria. ISBN 3-900051-07-0, https://cran.rproject.org/ , (Accessed,January 29, 2020)
27. Yaya, S., Ekholuenetale, M. and Bishwajit, G., 2018. Differentials in prevalence and correlates of metabolic risk factors of non-communicable diseases among women in sub-Saharan Africa: evidence from 33 countries. *BMC public health*, *18*(1), p.1168.

28. Charbel El Bcheraoui,Ziad A. Memish, Marwa Tuffaha, Hypertension and Its Associated Risk Factors in the Kingdom of Saudi Arabia, 2013: A National Survey, International Journal of Hypertension, Vol. 2014, Article ID 564679, 8 pages,2014.
29. Ileana Desormais, Salimanou Ariyoh ,Amidou, The prevalence, awareness, management and control of hypertension in men and women in Benin, West Africa: the TAHES study, BMC Cardiovascular Disorders (2019) 19:303
30. F.R.M.Causland, F. M. Sacks, and J. P. Forman, "Marital status,dipping and nocturnal blood pressure: results from the dietary approaches to stop hypertension trial," Journal of Hypertension, vol. 32, no. 4, pp. 756–761, 2014.
31. N. Al-Hamdan, A. Saeed, A. Kutbi, A. J. Choudhry, and R.Nooh, "Characteristics, risk factors, and treatment practices of known adult hypertensive patients in Saudi Arabia," International Journal of Hypertension, vol. 2010, Article ID 168739, 7pages, 2010.
32. Price AJ, Crampin AC, Amberbir A, Kayuni-Chihana N, Musicha C, Tafatatha T, et al. 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 & Endocrinology. 2018 Mar; 6 (3):208–22.
33. Guwatudde, D., Mutungi, G., Wesonga, R., Kajjura, R., Kasule, H., Muwonge, J., Ssenono, V. and Bahendeka, S.K., 2015. The epidemiology of hypertension in Uganda: findings from the national non-communicable diseases risk factor survey. *PloS one*, *10*(9).
34. Abebe, S.M., Berhane, Y., Worku, A. and Getachew, A., 2015. Prevalence and associated factors of hypertension: A cross-sectional community based study in Northwest Ethiopia. *PloS one*, *10*(4).
35. Bjertness, M.B., Htet, A.S., Meyer, H.E., Htike, M.M.T., Zaw, K.K., Oo, W.M., Latt, T.S., Sherpa, L.Y. and Bjertness, E., 2016. Prevalence and determinants of hypertension in Myanmar-a nationwide cross-sectional study. *BMC Public Health*, *16*(1), p.590
36. Demisse, A.G., Greffie, E.S., Abebe, S.M., Bulti, A.B., Alemu, S., Abebe, B. and Mesfin, N., 2017. High burden of hypertension across the age groups among residents of Gondar city in Ethiopia: a population based cross sectional study. *BMC Public Health*, *17*(1), p.647.
37. Williams, B., Mancia, G., Spiering, W., Agabiti Rosei, E., Azizi, M., Burnier, M., Clement, D.L., Coca, A., De Simone, G., Dominiczak, A. and Kahan, T., 2018. 2018 ESC/ESH Guidelines for the management of arterial hypertension: The Task Force for the management of arterial hypertension of the European Society of Cardiology (ESC) and the European Society of Hypertension (ESH). *European heart journal*, *39*(33), pp.3021-3104.
38. Wang, B., Liu, Y., Sun, X., Yin, Z., Li, H., Ren, Y., Zhao, Y., Zhang, R., Zhang, M. and Hu, D., 2020. Prediction model and assessment of probability of incident hypertension: the Rural Chinese Cohort Study. *Journal of Human Hypertension*, pp.1-11.

Figures

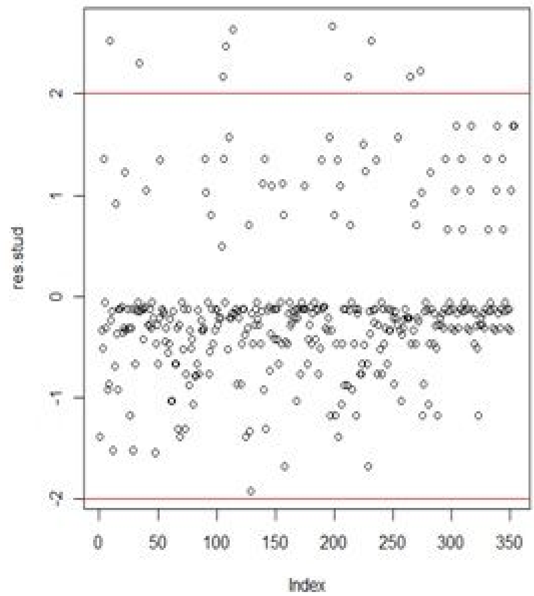

Figure 1

Studentized residual

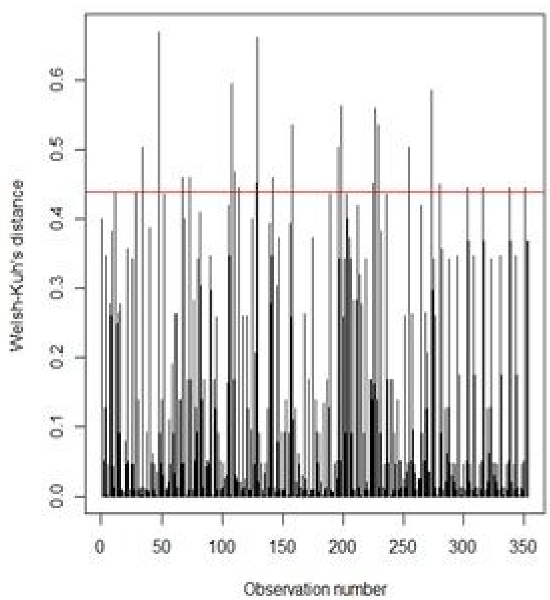

Figure 2

Welsh-Kuh’sdistance

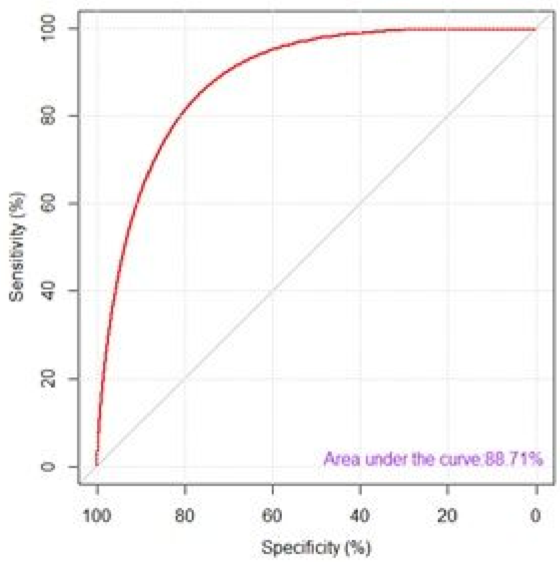

Figure 3

Area under

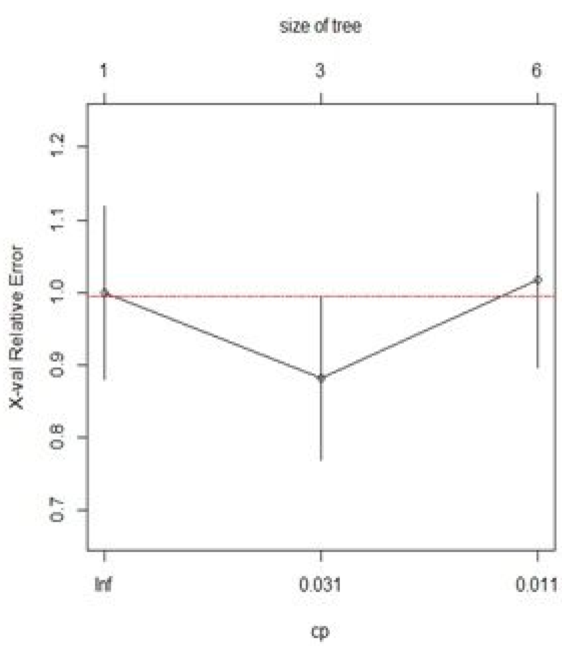

Figure 4

Complexity parameter