ORIGINAL ARTICLE

Developing and Validating Risk Algorithm for Hypertension in South Africa: Results from a Nationally Representative Cohort (2008–2017)

Handan Wand[1] · Cassandra Vujovich-Dunn[1] · Jayajothi Moodley[2] · Tarylee Reddy[3] · Sarita Naidoo[2]

Abstract

Introduction There is compelling evidence of significant country-level disparities where African countries, particularly South Africa, have the highest hypertension rates in the world.

Aim To develop and validate a simple risk scoring algorithm for hypertension in a large cohort (80,270) of South African men and women.

Methods Multivariable logistic regression models were used to build our hypertension risk scoring algorithm and validated externally and internally using the standard statistical techniques. We also compared our risk scores with the results from the Framingham risk prediction model for hypertension.

Results Six factors were identified as the significant correlates of hypertension: age, education, obesity, smoking, alcohol intake and exercise. A score of ≥ 25 (out of 57) for men and ≥ 35 (out of 75) for women were selected as the optimum cut-points with 82% (43%) and 83% (49%) sensitivity (specificity) for males and females, respectively in the development datasets. We estimated probabilities of developing hypertension using the Framingham risk prediction model, which were higher among those with higher scores for hypertension.

Conclusions Identifying, targeting and prioritising individuals at highest risk of hypertension will have significant impact on preventing severe cardiometabolic diseases by scaling up healthy diet and life-style factors. Our six-item risk scoring algorithm may be included as part of hypertension prevention and treatment programs by targeting older individuals with high body fat measurements who are at highest risk of developing hypertension.

Keywords Prevention · Hypertension · Obesity · Risk prediction · South Africa

✉ Handan Wand
hwand@kirby.unsw.edu.au

Cassandra Vujovich-Dunn
cvujovich-dunn@kirby.unsw.edu.au

Jayajothi Moodley
JMoodley@auruminstitute.org

Tarylee Reddy
Tarylee.Reddy@mrc.ac.za

Sarita Naidoo
S.Naidoo@auruminstitute.org

[1] Biostatistics and Databases Program, Kirby Institute, University of New South Wales (UNSW Sydney), Level 6, Wallace Wurth Building, Kensington, Sydney, NSW 2052, Australia

[2] The Aurum Institute, Johannesburg, South Africa

[3] Biostatistics Unit, South African Medical Research Council, Durban, Kwazulu-Natal, South Africa

1 Introduction

Hypertension is a significant health problem and the leading cause of severe cardiometabolic conditions and premature deaths worldwide [1–3]. Over the past decades, the burden of hypertension has increased substantially with significant geographical disparities where African countries, including South Africa, have been estimated to have the highest prevalence in the region [4–6]. In fact, besides HIV/AIDS epidemic, increasing hypertension rates are considered one of the major health challenges that Black South Africans face [7–9].

Following the United Nation's meeting in 2011, hypertension prevention research has been accelerated across the African nations by developing and implementing new guidelines [10–13]. Despite these efforts, hypertension affects African countries, including South Africa [14, 15]. According to the World Health Organization (WHO),

more than half of the African populations are hypertensive [16, 17]. Based on the most recent guidelines, two-thirds of South Africans are reported as hypertensive [18]. Although these estimates are useful for making global comparisons, they do not provide the most relevant country-specific risk factors for hypertension, which is known to be a proxy for severe cardiometabolic diseases [19, 20]. Therefore, identifying those at increased risk of hypertension and offering appropriate care and treatment before developing its sequelae remains a research priority. This information is particularly crucial due to the asymptomatic nature of the hypertension, which may be undetected until it is too late.

Risk scoring algorithms are known to be one of the most common and effective ways of identifying individuals at increased risk of disease. They have been frequently used in clinical settings and considered alternative diagnostic methods for specific diseases. Several algorithms have been developed and validated in various populations to identify those at high risk for non-communicable diseases, including hypertension [21–24]. The most recent systematic review identified 26 studies that developed risk prediction models exclusively for hypertension using the populations from the United States, Europe, China, Korea, Japan, Iran and India [25]. Although these models mainly included the established risk factors to quantify an individual's risk, they can only be relevant to the populations analyzed and may not apply to South Africans. Given the current status of hypertension and its sequelae in the country, there is an urgent need for a risk-scoring algorithm exclusively for South Africans. Aligned with this urgency, we aimed to develop a risk-scoring algorithm by identifying the minimum information to quantify individuals' risk for hypertension using a set of simple anthropometric, socioeconomic, and lifestyle characteristics without any laboratory measurements which can be impractical outside the clinical settings. Our gender-specific algorithms was validated (internally and external) using the data from more than 80,000 Black South Africans who participated in the five-rounds of the South African National Income Dynamics Studies between 2008 and 2017 [26]. In an additional analysis, using the Framingham risk model, we also predicted the risk of developing hypertension within 1-, 2- and 4-year among those who had normal blood pressure when the surveys were conducted [24].

There is extensive research to report country-level estimates; however, to date, there has not been an attempt to develop and validate a simple risk scoring algorithm that can quantify individuals' risk for hypertension in South Africa. Considering the asymptomatic nature of hypertension and the benefits of early diagnosis and treatment, the results from our study may have significant implications in public health and clinical settings.

2 Methods

2.1 Study Population

We used data from the five-rounds of cross-sectional South African National Income Dynamics Study (SA-NIDS) surveys conducted in 2008- 2017 by the South African Labour and Development Research Unit based at the School of Economics (University of Cape Town). Details of the surveys were described elsewhere (SA-NIDS) [26]. Briefly, study populations were selected from households using a two-stage random cluster sampling design from 52, district councils across the country where clusters were household units. The current study only included Black South African men and women aged 15 years or older whom constituted approximately 85% of all respondents. Participants provided written informed consent; parent or legal guardians provided consent for those younger than 18 years of age.

2.2 Characteristics

We considered the following characteristics: age, gender, smoking status, frequent alcohol use, exercise per week and education. Participants' weight (kg) and height (cm) were measured twice by a field personal; we used the mean of these two measurements to calculate the body mass index (BMI) (kg/m^2). The World Health Organisation classification was used to categorize participants as underweight: BMI: <24.9 kg/m^2, overweight: BMI: 25–29.9 kg/m^2 and obese: BMI: $\geq$ 30 kg/m^2.

2.3 Primary Outcome Measurement

All the anthropometric measures were collected by the trained study team workers. After a 5-minute rest period, the diastolic and systolic blood pressures were measured twice using an automated monitoring device with an adjustable cuff (Omron M7 BP) by trained study personnel. Quality controllers checked the data for completeness and verification. We restricted the analysis to non-missing data only. After averaging the two measurements of systolic and diastolic pressure measurements, the study population was classified as hypertensive according to the most recent definition suggested by the American College of Cardiology/American Heart Association task force in 2017: a systolic blood pressure $\geq$ 130 mm Hg and/or diastolic blood pressure $\geq$80 mm Hg [27]. Individuals with a pre-existing condition based on a medical history and/or medication were also considered as hypertensive.

2.4 Statistical Analysis

2.4.1 Deriving Hypertension Risk Algorithm: Split-Sample Method

Using the random number generation function in Stata 16.0, which allocated 67% of the randomly selected population who participated in one of the first three rounds of surveys conducted between 2008 and 2014, and allocated to the development data set (n = 11,401 men and n = 16,773 women); 33% of the same population (n = 5744 men and n = 8349 women) was used to validate the final model internally. Data from the most recent two rounds of survey participants were used to validate the risk models externally (n = 15,541 and n = 22,462 women). Our models were finalized using the backwards-selection method with p < 0.05. The Hosmer–Lemeshow goodness-of-fit test was used to assess the statistical appropriateness of the models. The hypertensive risk scoring algorithm included the weighted rounded regression coefficients (i.e. logarithms of the odds ratio, which were multiplied by 10) for each risk factor. This approach was commonly used in various risk scoring algorithms and have been shown to be statistically robust compared to single point approaches [28].

Gender-specific hypertension risk scores were calculated for each participant by adding up the final weighted scores for each risk factor. Using the data-driven techniques such as "quintiles", each participant were classified as being at "low", "mild", "moderate", "high", and "severe" risk for hypertension. We also identified the "optimum" cut-point with high discriminatory power. For this analysis, standard statistical measures including sensitivity, specificity, Area under the curve (AUC), likelihood ratio positive (LR+) and likelihood ratio negative (LR-) were calculated for development as well as internal and external validation datasets where 70% ≤ AUC < 80% was considered to be at statistically acceptable while AUC ≥ 80% was regarded as excellent discriminative power [29].

2.4.2 Framingham Risk Prediction Model for Hypertension

In an additional analysis, we used the Framingham risk prediction model to predict the risk of developing hypertension based on the following characteristics: (1) age (20–79 years old), (2) gender, (3) diastolic blood pressure (<80 mm Hg), (4) systolic blood pressure (<130 mm Hg), (5) current smoker, (6) Age by diastolic blood pressure which were all available in our data sources except parental hypertension [23]. This analysis excluded a total of 41,838 individuals who were classified as hypertensive and/or reported to have high blood pressure (diastolic ≥ 80 mmHg or systolic ≥ 130 mmHg), heart diseases as well as those reported to receive medication/treatment for these conditions (Supplementary

Table S2). Finally, we presented age x BMI specific predicted probabilities using a data visualization technique (Fig. 1a, b: "*heat maps*").

3 Results

80,270 people who participated in one of the five rounds of national surveys from 2008 to 2017 were analyzed. The overall median age of the study population was 32 years old (Interquartile-range (IQR): 22–49), in all time points. Gender distribution also remained similar over time, with 60% women and 40% men. Overall hypertension prevalence for women ranged from 44% (2017) to 56% (2008) and for men 48% (2017) to 54% (2008). In our development datasets, 32% of the men and 42% of the women were 40 years or older (Table 1a, b). The vast majority of the study population had some levels of education (89% for men and 85% for women), one third of the men and women reported their household incomes as being < 3000 ZAR (South African Rand). Based on BMI, 44% of the men and 64% of the women were overweight/obese. Similarly, high waist circumference was also more prevalent in women compared to men (68% and 45% respectively); while 38% of the men and 3% of the women reported being a regular smokers. Frequent alcohol use was more common in men (35% and 10% for men and women respectively); men were also more likely to report exercise (at least once a week) compared to men (42% vs 16% for men and women, respectively). These characteristics were remarkable similar in all validation (i.e. internal and external) datasets (Supplementary Table S1a and S1b).

3.1 Development and Internal/External Validations of the Risk Assessment Tool

In gender-specific multivariable models, seven factors were identified as the significant correlates of hypertension in both genders (Table 1a, b): (1) older age; (2) lack of education; (3) being married or cohabitating; (4) overweight/obesity; (5) high-waist circumference; (6) smoking cigarette; (7) frequent alcohol intake; (8) lack of exercise. Living in an urban area was only significantly associated with increased odds of hypertension in men. These risk factors were also significantly associated with increased odds of hypertension in internal and temporal (external) validation datasets in both genders. Results from all the multivariable models were remarkably similar in internal and external validation datasets (Supplementary Table S1a (for men) and S1b (for women)). Non-significant p-values for the Hosmer-Lemeshow test ranged from 0.329 to 0.735 across the development and all validation models, which were interpreted as statistically valid models. Finally, Supplementary Figures S1 presents our 6-item risk scoring tool for men and women
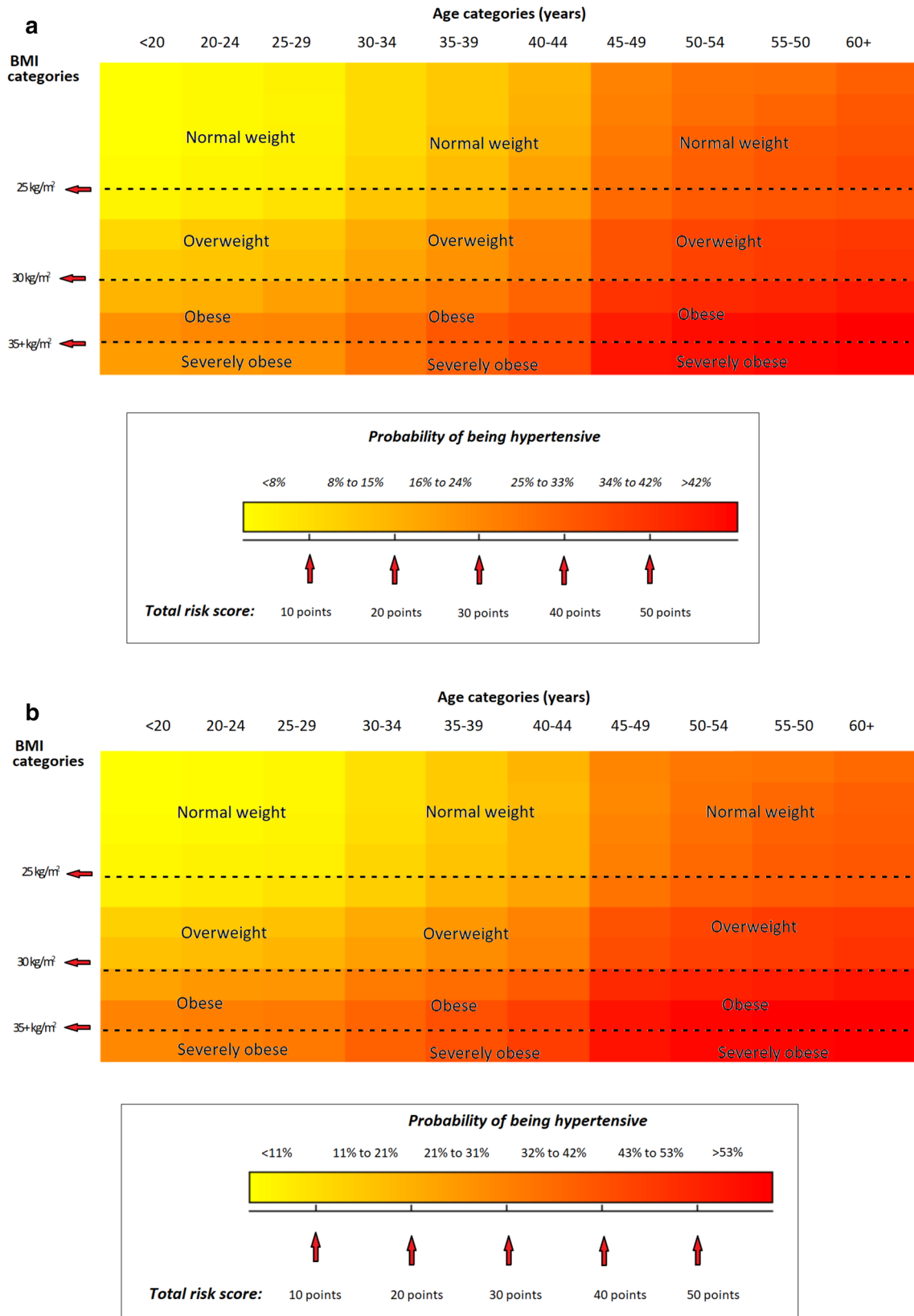
Fig. 1 a Mean hypertension risk scores by the deciles of age and BMI categories: men. **b** Mean hypertension risk scores by the deciles of age and BMI categories: women

Table 1 Gender-specific multivariable logistic regression models for hypertension:

	Development model for men: 2008–2014 (n= 11,401)					Development model for women: 2008–2014 (n=16,773)				
	%	Adjusted OR (95% CI)	P-value	β 10⫶	Score	%	Adjusted OR (95% CI)	P-value	β 10⫶	Score
Age										
<25	38%	1				29%	1			
25–29	12%	1.88 (1.66, 2.13)	< 0.001	6.2	6	12%	1.53 (1.37, 1.72)	< 0.001	4.3	4
30–34	10%	2.26 91.96, 2.60)	< 0.001	8.1	8	9%	2.23 (1.98, 2.53)	< 0.001	8.0	8
35–39	8%	2.80 (2.38, 3.27)	< 0.001	10.2	10	8%	3.13 (2.75. 3.55)	< 0.001	11.4	11
40+	32%	4.51 (4.08, 4.97)	< 0.001	15.1	15	42%	7.63 (7.00, 8.30)	< 0.001	20.3	20
Education										
Some education	89%	1			0	85%	1			0
No education	11%	2.50 (2.57, 2.88)	< 0.001	9.1	9	15%	3.41 (3.08, 3.78)	< 0.001	12.2	12
Marital status										
Single/not cohabiting	69%	1			0	72%	1			0
Married	31%	2.25 (2.10, 2.43)	< 0.001	8.1	8	28%	1.94 (1.80, 2.10)	< 0.001	6.6	7
BMI										
<25§ kg/m²	66%	1			0	36%	1			0
25–29 kg/m²	18%	2.07 (1.87, 2.29)	< 0.001	7.3	7	25%	1.77 (1.63, 1.91)	< 0.001	5.7	6
30+ kg/m²	16%	2.98 (2.60, 3.43)	< 0.001	10.9	11	39%	3.45 (3.20, 3.73)	< 0.001	12.4	12
Waist circumference										
Normal	55%	1			0	32%	1			0
High§§	45%	2.43 (2.24, 2.63)	< 0.001	8.9	9	68%	1.61 (2.44, 2.80)	< 0.001	9.5	10
Smoke regularly										
No	72%	1			0	97%	1			0
Yes	28%	1.31 (1.20, 1.43)	< 0.001	2.7	3	3%	1.58 (1.29, 1.94)	< 0.001	4.6	5
Alcohol intake										
Never	65%	1			0	90%	1			0
<3+ days/week	30%	1.37 (1.26, 1.50)	< 0.001	3.2	3	8%	1.20 (1.06, 1.34)	0.003	1.8	2
3+ days/week	5%	1.73 (1.43, 2.08)	< 0.001	5.5	6	2%	1.62 (1.12, 2.35)	0.010	4.9	5
Frequency of exercise/week										
Never	58%	1			0	84%	1			0
Less than once	42%	1.58 (1.46, 1.70)	< 0.001	4.5	5	16%	1.51 (1.38, 1.65)	< 0.001	4.1	4

§Included individuals with BMI <18.5 kg/m² (<4%)

§§>94 cm for men and >80 cm for women

separately. These figures also included risk assessment scales (i.e. predicted probabilities of hypertension) for the mean scores.

3.2 Impact of the Presence of Multiple Risk Factors

Table 2 presents the odds ratios from the gender-specific logistic regression models across the quintiles of the subject-specific risk scores in development and all validation datasets. We observed strong linear trends between the increasing quintiles of the hypertensive risk scores and the odds of being hypertensive in development and all validation datasets (P_{trend} < 0.001). For example, men in the highest risk score category (i.e. 5th quintile) were 8 times more likely

to be at increased odds of being hypertensive compared to those who had the lowest risk category (i.e. 1st quintile) in development and all validation datasets.

3.3 Gender-Specific Optimum Cut-Points

For men, a total score of 25 or more, which fell into the third quintiles of the total risk score (i.e. at least moderate risk), was associated with hypertension prevalence of 60% and had Area under the curve (AUC) of 75% (95%CI: 72%, 78%). Therefore score ≥ 25 was identified as the optimum cut-point with 82% sensitivity and 43% specificity, indicating a statistically acceptable discriminative power (Table 3a). For women, a score of ≥ 35, which also fell into the third quintiles of the total risk score (i.e. moderate-to-severe risk score), was associated with hypertension prevalence of 64%

Table 2 Gender-specific odds ratios and 95% confidence intervals for hypertension during across the risk score categories in the development dataset

a) Men:

	Development model		Validation model #1 (internal)		Validation model #2 (external)	
	Odds Ratio (95% CI)	p-value[‡]	Odds Ratio (95% CI)	p-value[‡]	Odds Ratio (95% CI)	p-value[‡]
Risk score categories [†]						
Low risk	1		1		1	
Mild risk	1.85 (1.71, 2.00)	<0.001	1.83 (1.64, 2.04)	<0.001	1.95 (1.78, 2.14)	<0.001
Moderate risk	3.05 (2.81, 3.31)	<0.001	3.41 (3.03, 3.83)	<0.001	3.47 (3.15, 3.82)	<0.001
High risk	5.06 (4.61, 5.55)	<0.001	4.86 (4.27, 5.54)	<0.001	5.52 (4.96, 6.15)	<0.001
Severe risk	8.00 (7.05, 8.86)	<0.001	8.73 (7.40, 10.32)	<0.001	8.38 (7.30, 9.63)	<0.001

b) Women:

	Development model		Validation model #1 (internal)		Validation model #2 (external)	
	Odds Ratio (95% CI)	p-value[‡]	Odds Ratio (95% CI)	p-value[‡]	Odds Ratio (95% CI)	p-value[‡]
Risk score categories [†]						
Low risk	1		1		1	
Mild risk	1.68 (1.53, 1.84)	<0.001	1.72 (1.51, 1.96)	<0.001	1.75 (1.56, 1.97)	<0.001
Moderate risk	3.37 (3.10, 3.68)	<0.001	3.34 (2.05, 3.78)	<0.001	3.56 (3.20, 3.97)	<0.001
High risk	6.74 (6.16, 7.39)	<0.001	6.41 (5.63, 7.28)	<0.001	7.20 (6.44, 8.03)	<0.001
Severe risk	12.51 (11.41, 13.71)	<0.001	12.07 (10.61, 13.74)	<0.001	12.50 (11.19, 14.0)	<0.001

Cut-points: [†]1st quintile (score: <10 points); 2nd quintile (score: 10-19); 3rd quintile (score 30-35); 4th quintile (score: 36-49); 5th quintile (score: 50+ points)

Cut-points: [†]1st quintile (score: <10 points); 2nd quintile (score: 10-19); 3rd quintile (score 20-35); 4th quintile (score: 36-49); 5th quintile (score: 50+ points)

had AUC of 83% (95% CI: 79%, 86%), and was identified as the optimum cut-point with 83% sensitivity and 49% specificity which were considered to be statistically acceptable discriminatory power.

3.4 Hypertension Risk Assessment Tool: Age ⌐ BMI-Specific Probabilities of Hypertension

Age-specific mean hypertension risk scores were presented across the deciles of the BMI categories for in heat-maps (Fig. 1a for men and 1b for women). As expected, there was a strong positive correlation between the increasing levels of hypertension risk score across the increasing deciles of age and BMI categories. In gender-specific analysis, every 5-point increase in the risk score resulted in a 4% and 5% increase in the predicted probabilities of being hypertensive for men and women, respectively. The highest predicted hypertension probabilities were observed in BMI categories ≥30 kg/m^2 regardless of age and gender. For example, individuals younger than 30 years of age with normal weight had a mean hypertension risk score ranging from 0 to 10 in both genders. The probability of these individuals being hypertensive was < 8% for men and < 10% for women (Fig. 1b). These probabilities more than doubled among those who were obese or severely obese in both genders (24% to 40%

for men and 21% to 42% for women). Individuals with scores of ≥ 25 (men) and ≥ 35 (women) were considered to be at "moderate-to-severe risk", and they were more than 3 (for men) and 5-times (for women) more likely to be hypertensive (data not shown).

3.5 Framingham Risk Prediction Model for Developing Hypertension

The risk of developing hypertension within 1, 2 and 4 years was presented using the Framingham hypertension risk prediction model across the categories of the quintiles of the risk scoring algorithm developed in the current study (Supplementary Table S2). There was a positive correlation between the two scoring tools. For example, individuals in the increasing risk score categories based on our scoring algorithm were also estimated to be at higher risk of developing hypertension based on the Framingham risk model.

4 Discussion

The current study identified the most influential risk factors and developed a 6-item risk scoring algorithm for hypertension in a nationally representative of Black South African

Table 3 Performance of the risk scoring algorithm for different cut-points.

For men

Score	Development model AUC (95% CI): 75% (95% CI: 72%, 78%)		
Median (IQR)	20 (7–36) points		
Cut points	Sensitivity (specificity)	Correctly classified	LR+ (LR−)
≥ 10 points	100% (0%)	66%	1 (–)
≥ 15 points	95% (22%)	68%	1.38 (0.25)
≥ 20 points	88% (31%)	70%	1.42 (0.43)
≥ 25 points	82% (43%)	78%	1.62 (0.50)
≥ 30 points	61% (68%)	83%	1.93 (0.57)
	Internal validation model AUC (95% CI): % (95% CI: %, %)		
Score cut points	Sensitivity (specificity)	Correctly classified	LR +/LR−
≥ 10 points	100% (0%)	62%	1 (–)
≥ 15 points	91% (19%)	64%	1.28 (0.22)
≥ 20 points	85% (28%)	68%	1.32 (0.49)
≥ 25 points	80% (39%)	75%	1.58 (0.48)
≥ 30 points	58% (71%)	79%	1.81 (0.61)

For women

Score	Development model AUC (95% CI): 83% (95% CI: 79%, 86%)		
Median (IQR)	34 (16-48) points		
Score cut-points	Sensitivity (specificity)	Correctly classified	LR + (LR−)
≥ 10 points	100% (0%)	50%	1 (–)
≥ 20 points	95% (22%)	58%	1.20 (0.25)
≥ 25 points	94% (23%)	63%	1.40 (0.30)
≥ 30 points	89% (36%)	66%	1.52 (0.50)
≥ 35 points	83% (49%)	71%	2.01 (0.40)
≥ 40 points	75% (60%)	75%	2.37 (0.52)
	Internal validation model AUC (95% CI): % (95% CI: %, %)		
Score cut-points	Sensitivity/specificity	Correctly classified	LR +/LR−
≥ 10 points	100% (0%)	46%	1 (–)
≥ 20 points	91% (25%)	54%	1.18 (0.22)
≥ 25 points	93% (27%)	59%	1.37 (0.28)
≥ 30 points	83% (46%)	61%	1.42 (0.45)
≥ 35 points	80% (47%)	68%	1.81 (0.41)
≥ 40 points	73% (63%)	71%	2.10 (0.50)

AUC area under the curve; *LR* likelihood ratio

men and women. Besides traditional risk factors such as age, body composition measurements (BMI or waist circumference), smoking, alcohol use and exercise, our algorithm also included socioeconomic condition indicators such as education and marital status. Most of the risk factors used in our risk prediction models have been well recognized and identified as significant predictors of hypertension in other populations and ethnic groups, including Europeans, North Americans, and Asian populations [21, 25]. However, their weighted and clustering impacts on hypertension were not

explored for South Africans. Most importantly, we quantified individuals' risk of hypertension with high precision (>80%) and robustness without using laboratory measurements which can be impractical outside the clinical settings. For example, previous risk prediction models included blood glucose, high-density lipoprotein cholesterol and triglyceride levels as independent predictors of developing hypertension [30]. However, their discriminative powers were not superior to our risk prediction model, which only included simple anthropometric measures (AUC <70%).

In our scoring algorithm, older age and obesity were identified as the most influential risk factors for hypertension. These two characteristics have been consistently reported as key predictors of hypertension and included in all previous risk prediction models [24, 25, 28, 30]. Despite their substantial positive impacts on the prevalence of hypertension, their individual discriminative powers were lower than their combination effect. For example, AUC was calculated as 62% (men), 59% (women) for age and 60% (men) and 64% (women) for obesity alone compared to the AUC: 75% and 83% for the combination of the risk factors. A similar strong association was observed between high waist circumference and hypertension. In our risk prediction models, BMI and waist circumference measurements were identified as significant correlates of hypertension with similar weighted scores. However, because of their strong correlation with each other, individuals who have one of these risk factors would likely to have the other one (spearman correlation coefficient ~ 90%). In practice, waist-circumference is a simpler measurement than BMI calculation and is also reported to be a better predictor for visceral fat [30]. Therefore, the final self-assessment questionnaire included the waist circumference measurement as an alternative to BMI. To our knowledge, this is the first study to give an option for both measurements.

Our study also highlighted significant associations between hypertension and smoking status, alcohol use, and lack of exercise. These results are also consistent with previous research, where they are considered as traditional risk factors for several cardiometabolic diseases, including CVD, diabetes, stroke and hypertension in other populations. However, despite statistically significant odds ratios associated with these factors, their weighted risk scores were modest. Other characteristics, marital status and education had moderate impacts on hypertension with weighted scores ranging between 8 and 12. Previous research reported a correlation between obesity and marriage in sub-Saharan Africa [22]. This association was primarily explained by a sedentary lifestyle of married couples compared to unmarried [31–33]. Consistent with these reports, lack of exercise was also more common among married couples than single individuals who exercise more. These results were also confirmed in our study. For example, 71% of the married men and 85% of the married women reported that they never exercise, compared to 50% and 77% of unmarried men and women, respectively (data not shown).

Similar to the other developing countries, in South Africa, the burden of hypertension is shifting towards the populations with low socioeconomic conditions which was also supported by our findings [34]. For example, in our study population, individuals with no education were approximately three times more likely to be hypertensive than those who had some levels of education. A possible explanation is the strong bi-directional correlation between the lack of education and high levels of unemployment and low-income rates, which potentially lead to disadvantaged socioeconomic conditions [15]. This is the first study to include socioeconomic indicators in identifying individuals at high risk of hypertension; individuals with these characteristics alone cannot be categorized as hypertensive without the presence of others. Nevertheless, they can be considered potential confounders and adjust the results for their impact. These results collectively indicate the importance of population-specific risk scoring algorithms rather than those developed in other populations. Our findings confirmed that individuals with high scores were also at increased risk of developing hypertension according to the Framingham risk prediction model [18]. Although we cannot conduct direct comparisons, we believe that the Framingham risk scoring model is likely to underestimate individuals' risk of developing hypertension in our study population for several reasons, including the scores assigned to BMI and age-specific diastolic blood pressure categories, which are expected to be much higher in Black South African populations compared to the population used in Framingham study who were primarily white. For example, in our study population, every one-unit increase in BMI leads to a 12% increase in the odds of being hypertensive compared to a 4% increase in the hazard ratio from the Framingham risk model [23].

4.1 Comparison with Existing Risk Prediction Models

We compared our results with 26 risk prediction models developed for hypertension presented in the most recent systematic review, and four additional studies were published later on [24, 25]. Similar to the results from our study, traditional risk factors including age, BMI/waist circumference and smoking were included in almost all hypertension risk prediction models. There were only two studies with physical activity, and another included dietary factors. At the same time, several of them used other laboratory measurements, including lipoprotein, C-reactive protein, cholesterol, uric acid and glucose levels [25]. None of the models included socioeconomic characteristics, and only one of them had an indicator for marital status [22]. A gender-specific risk

prediction model was reported in one study, while the others included "gender" as a risk [25]. While these studies used a diversity of population and ethnicities, none was conducted in South Africa. However, we observed a strong correlation between the predicted probabilities from the Framingham risk scoring algorithm and the risk scores calculated from the current study.

4.2 Clinical Implications

Given the current status of hypertension in South Africa, identifying those at high risk of hypertension would provide an opportunity to offer appropriate care and treatment before the development of severe cardiometabolic conditions. This is particularly crucial given the asymptomatic nature of hypertension; most individuals would not be aware of their status. In this setting, our risk scoring algorithm can serve as an initial step by identifying and alerting those who may need non-routine screening and further clinical assessments and treatment. For example, in our study population, 15,976 (51%) men and 22,993 (50%) women were classified as being hypertensive. However, only 12% of the men (1992/15,976) and 29% of the women (6772/22,993) were aware of their status (data not shown). Most importantly, hypertension is an established risk factor for severe cardiometabolic diseases and premature death. Therefore, it is one of the most commonly used risk factors to predict individuals' risk of developing cardiovascular diseases (CVD), diabetes and stroke. For example, a risk of CVD is reported to be increased by 15.82 for every one-unit increase in the log of systolic blood pressure [23]. Hypertension was also linked to development of diabetes and stroke in various populations globally [18].

4.3 Limitations

The current study has several limitations. Our risk scoring models were developed using serial cross-sectional data. However, based on the current evidence, same risk factors to play key role in predicting incidence of hypertension in prospective design settings which was also confirmed in our data. However, our findings are exclusively unique to the Black South African men and women who have the highest burden of hypertension and related adverse outcomes in the world.

Despite these limitations, this is also the first study to compare the burden of hypertension using the current and previous definitions in South Africa. As expected, the crude prevalence was higher according to the 2017 guidelines compared to the previous definition of hypertension. However, this difference was not as pronounced as for the other populations. For example, in our population, there was a 5% relative increase in hypertension prevalence based on the new guidelines compared to the previous guideline (61%

versus 58%) while this increase was 44% in the US population (46% versus 32%).

5 Conclusion

The score-specific probabilities may be used as a screening tool before the more costly laboratory assessments in local health care and clinical research setting. Our study particularly highlighted significant associations between hypertension and smoking status, alcohol use, and lack of exercise which are all considered as traditional risk factors for several cardiometabolic diseases, including CVD, diabetes, stroke and hypertension in other populations. Identifying, targeting and prioritising individuals at highest risk of hypertension will have significant impact on preventing severe cardiometabolic diseases by scaling up healthy diet and life-style factors. Implementing targeted screening strategies remain a crucial and effective approach to targeting those at high risk of hypertension and have significant implications in public health and clinical settings.

Declarations

References

1. 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 [published correction appears in Lancet. 2013 Apr 13;381(9874):1276] [published correction appears in Lancet. 2013 Feb 23;381(9867):628 (AlMazroa MA, Memish ZA). *Lancet*. 2012;380(9859):2224–2260. https://doi.org/10.1016/S0140-6736(12)61766-8

2. Lloyd-Sherlock P, Beard J, Minicuci N et al. Hypertension among older adults in low- and middle-income countries: prevalence, awareness and control. *Int J Epidemiol* 2014;43:116–28.https://doi.org/10.1093/ije/dyt215pmid. http://www.ncbi.nlm.nih.gov/pubmed/24505082

3. Mendis S, Puska P, Norrving B, World Health Organization, World Heart Federation et al. Global atlas on cardiovascular disease prevention and control/edited by: Shanthi Mendis [et al.]. World Health Organization. 2011. https://apps.who.int/iris/handle/10665/44701

4. Kearney PM, Whelton M, Reynolds K, Muntner P, Whelton PK, He J. Global burden of hypertension: analysis of worldwide data. Lancet. 2005;365(9455):217–23. https://doi.org/10.1016/S0140-6736(05)17741-1.

5. Kruger HS, Venter CS, Vorster HH, Margetts BM. Physical inactivity is the major determinant of obesity in black women in the North West Province, South Africa: the THUSA study. Nutrition. 2002;18:422–7.

6. Leibbrandt M, Woolard I, de Villiers L. Methodology: report on NIDS wave 1. In: Technical paper, vol 1. http://www.nids.uct.ac.za/nids-data/data-access

7. Wand H, Ramjee G. Assessing and evaluating the combined impact of behavioural and biological risk factors for HIV seroconversion in a cohort of South African women. AIDS Care. 2012;24(9):1155–62. https://doi.org/10.1080/09540121.2012.687820.

8. Wand H, Ramjee G. High prevalence of obesity among women who enrolled in HIV prevention trials in KwaZulu-Natal, South Africa: healthy diet and life style messages should be integrated into HIV prevention programs. BMC Public Health. 2013;13:159.

9. Houle B, Gaziano TA, Angotti N, et al. Hypertension incidence among middle-aged and older adults: findings from a 5-year prospective study in rural South Africa, 2010–2015. BMJ Open. 2021;11: e049621. https://doi.org/10.1136/bmjopen-2021-049621.

10. van de Vijver S, Akinyi H, Oti S, et al. Status report on hypertension in Africa—consultative review for the 6th Session of the African Union Conference of Ministers of Health on NCD's. Pan Afr Med J. 2013;16:38. https://doi.org/10.11604/pamj.2013.16.38.3100 (**published 2013 Oct 5**)).

11. Schutte AE, Gona PN, Delles C, et al. The African Prospective study on the Early Detection and Identification of Cardiovascular disease and Hypertension (African-PREDICT): design, recruitment and initial examination. Eur J Prev Cardiol. 2019;26(5):458–70.

12. Osafo C, Raji YR, Burke D, et al. Human heredity and health (H3) in Africa Kidney Disease Research Network: a focus on methods in Sub-Saharan Africa. Clin J Am Soc Nephrol. 2015;10(12):2279–87. https://doi.org/10.2215/CJN.11951214.

13. Puoane T, Steyn K, Bradshaw D, et al. Obesity in South Africa: the South African demographic and health survey. Obes Res. 2002;10(10):1038–48. https://doi.org/10.1038/oby.2002.141.

14. SADHS: Department of Health MRC, OrcMacro. South Africa Demographic and Health Survey 2003. National Department of Health: Pretoria; 2007.

15. Sartorius B, Veerman LJ, Manyema M, Chola L, Hofman K. Determinants of obesity and associated population attributability, South Africa: empirical evidence from a National Panel Survey, 2008–2012. PLoS ONE. 2015;10(6):e0130218. https://doi.org/10.1371/journal.pone.0130218.

16. WHO. A global brief on hypertension: silent killer global public health crisis. World Health Day 2013. http://ish-world.com/downloads/pdf/global_brief_hypertension.pdf. Accessed 28 Dec 2021. (World Health Organization; 2021).

17. WHO. Raised blood pressure. Global Health Observatory Data 2016. http://www.who.int/gho/ncd/risk_factors/blood_pressure_prevalence_text/en/. Accessed Jan 2022.

18. Akpa OM, Made F, Ojo A, Ovbiagele B, Adu D, Motala AA, Mayosi BM, Adebamowo SN, Engel ME, Tayo B, Rotimi C, Salako B, Akinyemi R, Gebregziabher M, Sarfo F, Wahab K, Agongo G, Alberts M, et al. Regional patterns and association between obesity and hypertension in Africa: evidence from the H3Africa CHAIR Study. Hypertension. 2020;75(5):1167–78. https://doi.org/10.1161/HYPERTENSIONAHA.119.14147.

19. Colosia AD, Palencia R, Khan S. Prevalence of hypertension and obesity in patients with type 2 diabetes mellitus in observational studies: a systematic literature review. Diabetes Metab Syndr Obes. 2013;6:327–38. https://doi.org/10.2147/DMSO.S51325Crossref.PubMed.

20. Turner R, Holman R, Matthews D, Bassett P, Coster R, Stratton I, Cull C, Peto R, Frighi V, Kennedy I, et al. Hypertension in Diabetes Study (HDS): I. Prevalence of hypertension in newly presenting type 2 diabetic patients and the association with risk factors for cardiovascular and diabetic complications. J Hypertens. 1993;11:309–17. https://doi.org/10.1097/00004872-199303000-00012.

21. Wang B, Liu Y, Sun X, et al. Prediction model and assessment of probability of incident hypertension: the Rural Chinese Cohort Study. J Hum Hypertens. 2021;35(1):74–84. https://doi.org/10.1038/s41371-020-0314-8.

22. Fava C, Sjogren M, Montagnana M, Danese E, Almgren P, Engstrom G, et al. Prediction of blood pressure changes over time and incidence of hypertension by a genetic risk score in Swedes. Hypertension. 2013;61:319–26. https://doi.org/10.1161/HYPERTENSIONAHA.112.202655[PubMed][GoogleScholar].

23. Kivimäki M, Tabak AG, Batty GD, et al. Incremental predictive value of adding past blood pressure measurements to the Framingham hypertension risk equation: the Whitehall II Study. Hypertension. 2010;55(4):1058–62. https://doi.org/10.1161/HYPERTENSIONAHA.109.144220.

24. Echouffo-Tcheugui JB, Batty GD, Kivimäki M, Kengne AP. Risk models to predict hypertension: a systematic review. PLoS ONE. 2013;8(7): e67370. https://doi.org/10.1371/journal.pone.0067370.

25. Sun D, Liu J, Xiao L, Liu Y, Wang Z, Li C, Jin Y, Zhao Q, Wen S. Recent development of risk-prediction models for incident hypertension: An updated systematic review. PLoS ONE. 2017;12(10): e0187240. https://doi.org/10.1371/journal.pone.0187240.

26. National Income Dynamics Study—, Waves 1, 2, 3, 4, 5. Version 2.0.0. Johannesburg and Cape Town: Version 1.0.0. In: Pretoria: Department of Planning M, and Evaluation. 2018. Cape Town: Southern Africa Labour and Development Research Unit, 2021. https://www.datafirst.uct.ac.za/data-portal/index.php/catalog/NIDS/.

27. Center for Disease Control and Prevention (CDC). Defining adult overweight & obesity. https://www.cdc.gov/obesity/adult/defining.html. Accessed Dec 2021.

28. Schnabel RB, Sullivan LM, Levy D, Pencina MJ, Massaro JM, D'Agostino RB Sr, Newton-Cheh C, Yamamoto JF, Magnani JW, Tadros TM, Kannel WB, Wang TJ, Ellinor PT, Wolf PA, Vasan RS, Benjamin EJ. Development of a risk score for atrial fibrillation (Framingham Heart Study): a community-based cohort study.

29. Bang H, Edwards AM, Bomback AS, Ballantyne CM, Brillon D, Callahan MA, Teutsch SM, Mushlin AI, Kern LM. Development and validation of a patient self-assessment score for diabetes risk. Ann Intern Med. 2009;151(11):775–83. https://doi.org/10.7326/0003-4819-151-11-200912010-00005.

30. Lim NK, Lee JY, Lee JY, Park HY, Cho MC. The role of genetic risk score in predicting the risk of hypertension in the Korean population: Korean Genome and Epidemiology Study. PLoS ONE. 2015;10: e131603.

31. Wilson SE. Marriage, gender and obesity in later life. Econ Hum Biol. 2012;10:431–53.

32. Sobal J, Rauschenbach BS, Frongillo EA Jr. Marital status, fatness and obesity. Soc Sci Med. 1992;35:915–23.

33. Tzotzas T, Vlahavas G, Papadopoulou SK, Kapantais E, Kaklamanou D, Hassapidou M. Marital status and educational level associated to obesity in Greek adults: data from the National Epidemiological Survey. BMC Public Health. 2010;10:732.

34. Monteiro CA, Moura EC, Conde WL, Popkin BM. Socioeconomic status and obesity in adult populations of developing countries: a review. Bull World Health Organ. 2004;82:940–6.