Development of a prediction model for hypertensive disorders of pregnancy in a Ghanaian cohort

Shreyanka Shylendrakumar

India

Master of International Health
9th September 2019 – 4th September 2020

Development of a prediction model for hypertensive disorders of pregnancy in a Ghanaian cohort

A thesis submitted in partial fulfilment of the requirement for the degree of

Master of Science in International health by

Shreyanka Shylendrakumar

India

Declaration:

Where other people's work has been used (either from a printed source, internet or any other source) this has been carefully acknowledged and referenced in accordance with departmental requirements.
The thesis **Development of a prediction model for hypertensive disorders of pregnancy in a Ghanaian cohort** is my own work.

Signature:

Master in International Health (MIH)
9th September 2019 – 4th September 2020

Organized by:

KIT/Royal Tropical Institute
Vrije Universiteit Amsterdam
Amsterdam, the Netherlands
September 2020.

In co-operation with:

Vrije Universiteit Amsterdam (VU)
Amsterdam, the Netherlands

Table of Contents

LIST OF TABLES, FIGURES AND ABBREVIATIONS

TABLES

FIGURES

ABBREVIATIONS

ANC	Antenatal care
AI	Artificial intelligence
ANN	Artificial neural networks
aOR	Adjusted odds ratio
AUC	Area under the curve
BMI	Body mass index
BP	Blood pressure
CH	Chronic hypertension
E	Eclampsia
GDP	Gross domestic product
GH	Gestational hypertension
GSS	Ghana statistical service
HDP	Hypertensive disorders of pregnancy
HELPP	Hemolysis, elevated liver enzymes and low platelets
HIC	High-income country
HTN	Hypertension
LMIC	Low-and-middle-income country
MMR	Maternal mortality rate
NICE	National Institute for health and care
NT	Normotensive
OA	Osteoarthritis
OR	Odds ratio
PE	Pre-eclampsia
PIH	Pregnancy induced hypertension
SD	Standard deviation
TRIPOD	Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis
UNICEF	The United Nations International Children's Fund
WHO	World health Organization

Acknowledgements

I gratefully acknowledge the supervisors of this thesis without whom this thesis would not have been possible. My thesis advisor who kept pushing me to work harder, for giving a lot of inputs about research and writing skills and for being a great source of motivation and knowledge. I also thank Maaike Flinkenflögel who supervised this study and always supported and encouraged during times when I felt low, Fernando Maldonado for giving valuable advice and inputs whenever I needed them, other members of KIT faculty and colleagues at KIT.

Abstract

Background

Hypertensive disorders of pregnancy (HDP) are a public health concern worldwide affecting about 10% of women globally. Nearly one-tenth of all maternal deaths in Africa and Asia are due to HDP. In Ghana, HDP is the third leading cause of maternal deaths accounting to 9% of the total deaths, after hemorrhage and abortion. Since the prognosis of HDP is dependent on the severity of the disease process, early recognition and early management of HDP is essential to reduce the mortality and morbidity rates caused by HDP which is facilitated by prediction models. The objective of this study was to develop and internally validate a multivariable logistic regression model that identifies the women at risk for HDP in Ghana by identifying the risk factors.

Methodology

A secondary data analysis nested in a cohort study of 1010 pregnant women from two antenatal clinics in Accra region of Ghana, between July 2012 and March 2014 with outcome of interest being HDP. Information about HDP was available for analysis in 789 women. Univariable analysis, bivariate analysis and multivariable logistic regression were employed on 'Scikit-learn' and 'Statsmodels' libraries of Python to build the prediction model and identify the risk factors of HDP.

Results

The cohort had 88.7% of normotensive women and 11.3% of women with HDP. Sixteen predictors were chosen for analysis and all 16 predictors were used to build the model. Maternal age (OR 1.01), weight (OR 1.01), Body mass index (OR 1.02), systolic blood pressure (BP) 121 – 140 mmHg (OR 1.14), systolic BP >140 mmHg (OR 1.96), diastolic BP 71 - 80 mmHg (OR 1.07), diastolic BP 81 - 90 mmHg (OR 1.21), diastolic BP >90 mmHg (OR 1.72), family history of Hypertension (OR 1.07), history of HDP in previous pregnancy (OR 1.26) showed significant correlation to HDP in bivariate analysis. After multivariable logistic regression, only Hausa ethnic group (aOR 1.12), the Northern ethnic groups of Ghana (aOR 1.10), systolic BP >140 mmHg (aOR 1.42), diastolic BP 81 - 90 mmHg (aOR 1.02), diastolic BP >90 mmHg (aOR 1.42) and multi-fetal pregnancy (aOR 1.27) showed significant correlation to HDP independently. All p-values <0.05. The model showed adequate performance with good efficiency, calibration and discrimination.

Conclusions

A simple prediction model was developed using easily available predictors from primary settings of Ghana and risk factors of HDP in Ghana were identified, which will help reduce mortality and morbidity cases caused by HDP by providing clinical decision-making support in Ghana.

Keywords

Hypertensive disorders of pregnancy, Ghana, prediction model, logistic regression, risk factors.

Word count: 11641

Introduction

Background

My name is Shreyanka Shylendrakumar, a medical doctor from Bangalore, India. I completed my bachelors in 2018 from Rajiv Gandhi University of Health Sciences, Bangalore and have worked in both public and private hospitals. During my bachelors, I found Community and Preventive medicine subject very interesting. This made me participate in many campaigns and awareness programs in and around my city. But somehow, I never took this interest seriously until I saw a young man dying of dengue in one of the hospitals I was working in Bangalore. That day I thought to myself, if this man or his family knew how the disease spreads and how it could be prevented, this man would not have died today. This made me realize the importance of prevention and started focusing on public health which deals more with prevention of diseases rather than curing them and focuses on the whole community and not on a single patient.

Why this thesis?

I was always interested in computer programming languages and used to take online classes on programming such as Python. For my bachelor's thesis, I performed a door-to-door survey in one of the communities in Bangalore to predict the prevalence of Osteoarthritis (OA). To practice the skills that I had learned about programming, I used the data from the survey I did and tried building a prediction model to identify women at risk for OA. Although it was just a practice, I really liked the idea of a prediction model and hence, wanted to do it for my master's thesis and contribute in the field of prevention.

HDP is one of the leading causes of maternal deaths in the world and is responsible for many cases of mortality and morbidity especially in low-and-middle-income countries (LMIC) with low resource settings. About 10% of women in the world are affected by HDP. In Ghana, 14% of all female deaths are pregnancy related and 9% of these deaths are due to HDP therefore being the third leading cause of total maternal deaths after hemorrhage and abortion. It is expected to soon become the leading cause of maternal deaths in Ghana as per few studies. HDP can be prevented by early identification and appropriate management of cases which often does not happen in LMIC due to lack of resources. Prediction models come to use at this point. Prediction models identifies women at risk so that early care is given to the patient thereby preventing the complications. The role of prediction models is rising in clinical settings all around the world, especially in high-income countries but there are a very few models in LMIC such as Ghana even though LMICs such as Africa and Asia report the highest maternal deaths in the world.

Therefore, the topic of this thesis is developing a prediction model that identifies pregnant women who are at risk by identifying the risk factors of HDP in Ghana which helps in early diagnosis and treatment of the women thereby reducing the complications, mortality and morbidity rates caused by HDP

1. Background information on Ghana

Geography

Ghana, officially called the Republic of Ghana, is situated on the west coast of Africa and consists of ten regions. The word 'Ghana' means Warrior King. Geographically, Ghana is almost in the center of planet Earth with a total area of Ghana is 227 540 Km2 [1], [2]. The country has a north to south distance of about 670 km and an east to west distance of about 560 km. It shares borders with Cote d'Ivoire on the west, Burkina Faso on the north and Togo on the east and on the south are the Gulf of Guinea and the Atlantic Ocean. The country is divided into 10 administrative regions. The topography is mostly undulating and has slopes of less than 1%. Although the country has gentle slopes, about 70% of the country is subject to moderate to severe sheet and gully erosion. The highest mountain in Ghana is Mount Afadjato in the Akwapim-Togo Ranges which is 880 meters above sea level [3].

The climate in Ghana is typical tropical climate and is mostly warm and humid. There are mainly two seasons in Ghana, the wet season and the dry season. The wet season begins around March and ends in the month of November. The dry season is between December and March. The mean annual rainfall of the country is 1,187 mm. Mean annual temperatures range from 26.1 C near the coast to 28.9 C in the extreme north and the humidity levels is around 77% [2].

Demography

The country's population, as of early 2020, is around 31,072,940 million, out of which 34% is rural population. The annual population growth rate is 2.19% [4]. The population density in the country is approximately 134 inhabitants/km2 nationwide [1], with a variation of 26 inhabitants/km2 in the Northern Region to 896 inhabitants/km2 in the Greater Accra Region [3]. The Accra region of Ghana is the largest city and also the capital city of Ghana and has a population of 1,963,26479 people thereby becoming the highest populated city in Ghana [1]. According to current projections, it is predicted that Ghana's population will continue to grow the rest of the century and will reach 78.71 million people in 2099. This means that Ghana will have more than double its current population in the next 80 years. In Ghana, around 50.9% is the male population while the rest of 49.1% is female population [2]. Population aged 14 years and younger forms the 38% of the total population in Ghana [5]. Figure [1] shows the population pyramid of Ghana representing the age and sex distribution.

More than 98% in Ghana are Black Africans although the country has multiple ethnic groups. Major ethnic groups in Ghana include Akan (47.5%), Dagbani (17%), Ewe (14%), Ga-Adangbe (7%), Gurma (6%), Guan (4%), Gurunsi (2.5%), and Bissa (1%). Some of the common languages in Ghana are Asante, Ewe, Fante and others but English is considered as an official language in almost all regions of Ghana. In terms of religion, 71% of people in Ghana are Christians and 17% are Muslims [2].

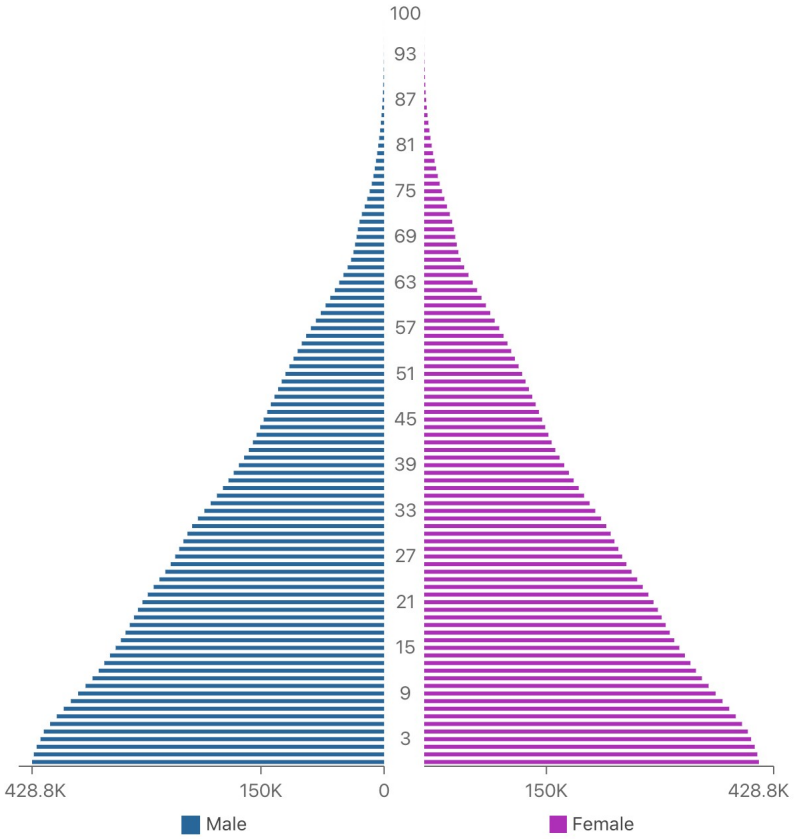

Figure 1: Population pyramid of Ghana, 2020 [2]

Socio-economic status

Ghana recently became a middle-income country [6]. Ghana's economy continued to grow in 2019 and there was a gross domestic product (GDP) growth to 6.7% from 5.4% in 2018 [7]. As of 2019, the country's GDP is worth 190.7 billion US dollars. The country has an economic freedom score of 59.4 and is ranked 11th among the other 47 countries of Sub-Saharan Africa. Ghana is also Africa's second biggest gold producer and second largest cocoa producer along with being rich in diamonds and oil. The country's economy is dominated by agriculture, which employs about 40% of the working population. 6.7% of people in Ghana are unemployed [8]. As of 2018, Ghana's Human Development Index is 0.596 and ranks 142nd position out of the other 189 countries. Around 30% of Ghanans suffer from poverty [9].

Educational status

According to a report from The United Nations International Children's Fund (UNICEF), 2020, nearly 623,500 primary school age children are still not enrolled in a school. Although Ghana has successfully reduced the gender gap when it comes to completing primary school, it is still very high when it comes to secondary school. Adolescent girls in Ghana are usually not able to go to schools due to factors such as poverty and gender inequality [10]. In Ghana, the education system is divided into Primary school, lower secondary school, upper secondary school and upper tertiary education. There is compulsory education till upper secondary school but still the adult literacy rate in Ghana

is only 79% and this figure is unequally distributed between males and females. The adult male literacy rate is 84% and the adult female literacy rate is 75%, as seen in figure [2] which clearly shows gender inequality in the country and nearly 4 million adults in Ghana are illiterate, as of 2018 [11]. The school life expectancy in Ghana is 12 years for males and 11 years for females, as of 2017 [12].

	TOTAL	MALE	FEMALE	
Literacy rate (%)				
15-24 years	92.49	92.76	92.21	(2018)
15 years and older	79.04	83.52	74.47	(2018)
65 years and older	50.93	64.97	39.01	(2018)

Figure 2: Literacy rate in Ghana, 2018 [11]

Marital status

According to Ghana Statistical service (GSS), 2014, it was seen that 42% of women and 38% of men in Ghana are married. 14% of women and 10% of men are living together with a partner and about 48% of men and 33% of women have never been married. The percentage of women divorced, separated or widowed is almost three times more in women than in men, which is 11% and 4% respectively [13].

Health indicators

The life expectancy at birth in Ghana is improving year by year. In 2010, the total life expectancy was 61 years and as of 2016, the life expectancy is 64 years. The life expectancy for males is 63 years and for females, the life expectancy is 65 years. [14], [15]. The fertility rate in Ghana is 3.9 births per woman, as of 2016 [15]. According to GSS, 2014, rural women of Ghana have about 1.7 children more than compared to urban women, that is in rural areas there are 5.1 children per woman compared to 3.4 children per woman in urban areas. There are 121 births per 1,000 women in urban areas whereas 210 births per 1,000 women in rural areas. This low fertility pattern in urban areas is evident in every age group but is most pronounced in women between age 20 and 24. When it comes to receiving antenatal care, (ANC), more than 9 in 10 mothers (97%) receive ANC from a skilled provider and overall, only 3% of mothers receive no ANC during delivery [13]. The maternal mortality ratio in Ghana is 308 (223 - 420) per 100,000 live births [14], [15]. The maternal mortality ratio during pregnancy is 343 (240 - 446) deaths per 100,000 live births, as per GSS, 2017 [16].

Role of prediction models

Prediction models are gaining much importance in the recent years. They can be used in public health, clinical practice and medical research. They are mainly used to predict the future occurrences of disease which later can be used to target preventive intervention on these target groups [17]. The ultimate goal of prognostic models is timely diagnosis and effective management [18]. The decision-making process in healthcare requires too much thought before arriving to a specific course of action in patient care. Prediction models come to use at such times and aids the health professionals in the decision-making process

[18], [19]. The use of prediction models in clinical settings are increasing day by day as it acts like a guide in the process of decision making [20]. The prediction models also alert who needs immediate medical care and identifies the target groups and estimates a patient's prognosis or a patient's probability of having a particular diagnosis [18], [21].

Clinical prediction models combine multiple predictors and provides information about the relative effects of predictors in the model [22]. Applying a model in clinical settings can help with detection of undiagnosed high-risk cases, which in turn improves the ability to prevent these developing diseases with early interventions [21]. It also increases the accuracy of diagnoses [23].

One of the first and more important decisions while building a model is the choice of an appropriate technique. When the prediction model has only two output variables, presence or absence of a disease for instance, techniques such as Logistic regression, Random forest, etc which performs classification tasks well would be a good choice. The neural networks, on the other hand, does not require such restrictive hypothesis which is why they are becoming an interesting competitor of classical techniques such as Logistic regression [24]. But, in clinical settings, it is seen that the traditional Logistic regression technique performs as well as other machine learning algorithms, particularly when using few strong predictors [25], [26].

2. Problem statement and Justification

Maternal mortality is intolerably high and is a widespread problem worldwide. Although maternal mortality is seen all over the world, a huge disparity can be seen between high income countries (HIC) and low-and-middle-income countries (LMIC), as seen in Figure [3]. According to 2017 World Health Organization (WHO) report, every day, approximately 810 women died from preventable causes related to pregnancy and childbirth and over 94% of these deaths occurred in LMIC with low resource settings. About 295 000 women died during and after pregnancy and childbirth in 2017. The maternal mortality rate (MMR) in LMIC was 462 per 100 000 live births whereas it was only 11 per 100 000 live births in HIC [27]. The lifetime risk of maternal death in HIC is 1 in 5,400 whereas it is 1 in 45 in LMIC [28], as seen in Figure [4].

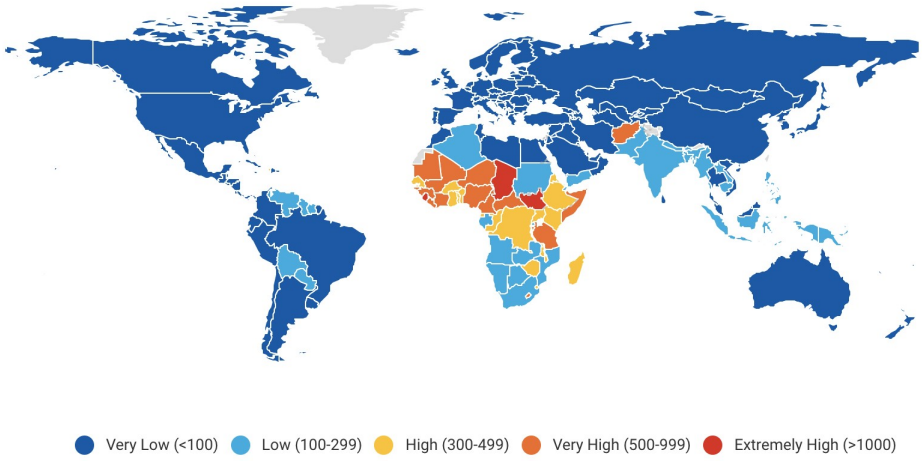

Figure 3: Maternal deaths per 100,000 live births, 2017 [28]

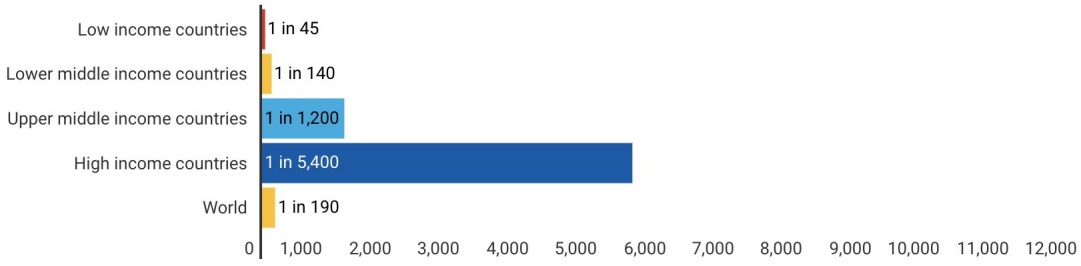

Figure 4: Lifetime risk of maternal death according to income group, 2017 [28]

In LMIC, primarily in Africa and Asia, maternal mortality is still 100-200 times higher than it is in Europe and North America [29]. Sub-Saharan Africa and South Asia reported approximately 86% (254 000) of the total global maternal deaths in 2017. Sub-Saharan Africa alone reported two-thirds (200 000) of maternal deaths with 533 maternal deaths per 100 000 live births a year, accounting to 68% of all maternal deaths per year worldwide, as seen in Figure [5]. It is also seen that the lifetime risk for maternal mortality is highest in African countries and is 1 in 28 in West and Central Africa, 1 is 38 in Sub Saharan Africa, 1 in 58 in Eastern and South Africa [28], as seen in Figure [6].

Maternal deaths occur due to complications during or following pregnancy and childbirth. The causes for maternal deaths can be direct or indirect. The main direct causes responsible for maternal mortality are hemorrhage, infections, Hypertension (HTN) and unsafe abortion. Hemorrhage, accounting for more than one quarter (27%) of the deaths, remains the leading direct cause of maternal mortality and HTN accounts for 14% of all deaths, followed by sepsis, embolism and unsafe abortion which also claims a substantial number of lives [28], as seen in Fig [7].

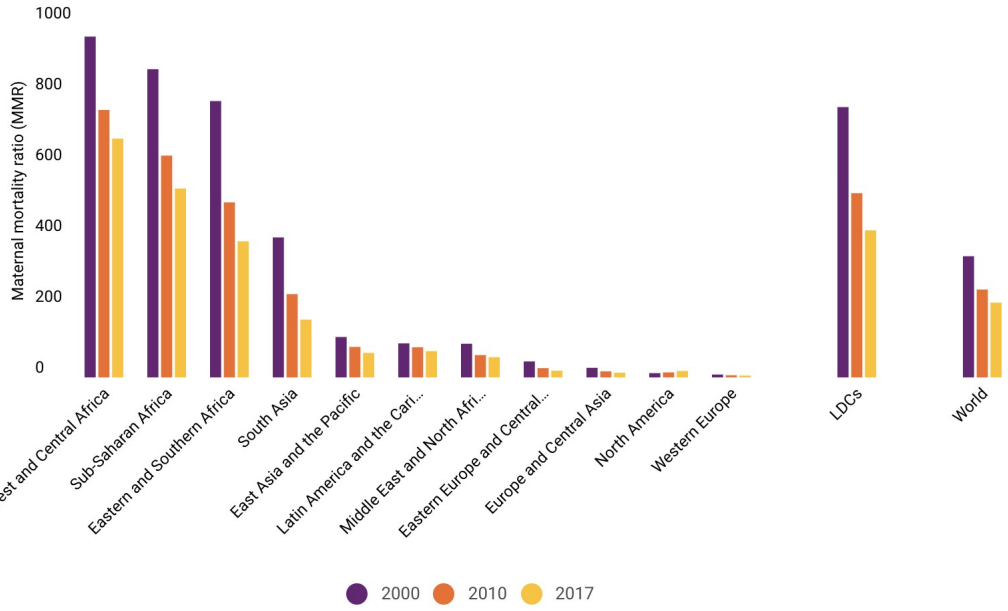

Figure 5: Trend of Maternal deaths per 100,000 live births by region, 2017 [28]

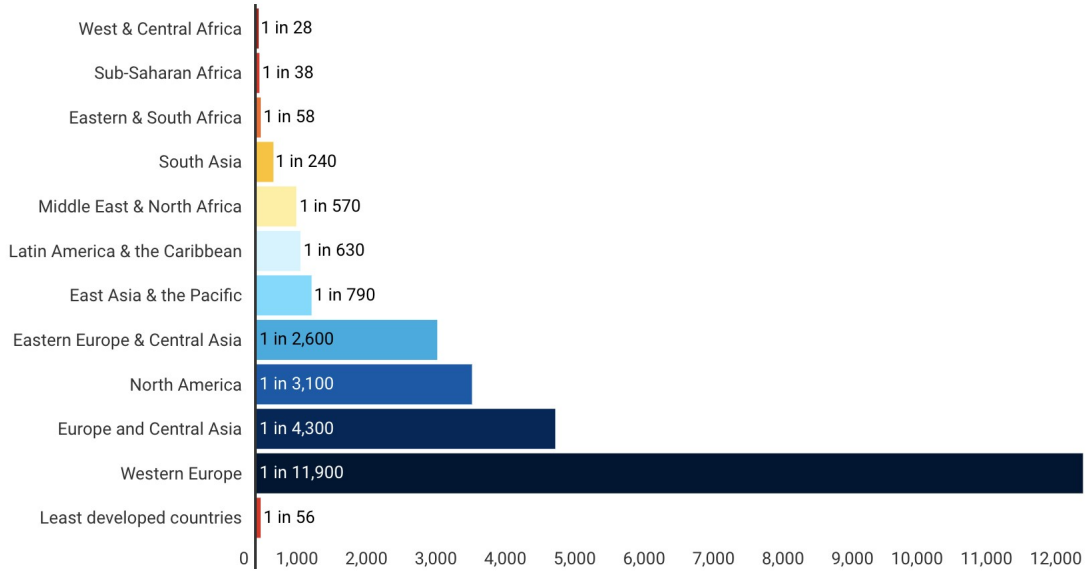

Figure 6: Lifetime risk of maternal death according to region, 2017 [28]

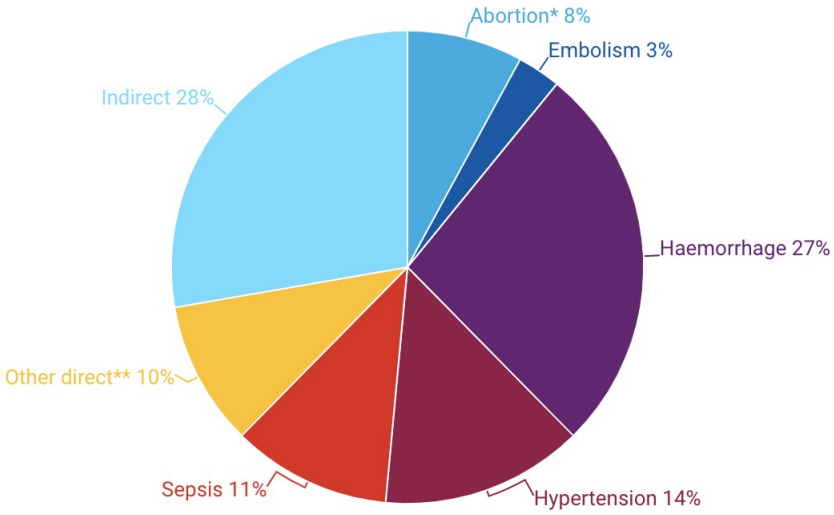

Figure 7: Causes of Maternal Death, 2017 [28]

Hypertensive disorders of pregnancy (HDP) are an important cause of maternal morbidity and mortality and is considered to be one of the most common complications during pregnancy especially in LMIC. It affects about 10% of women all around the world [29], [30]. HDP is seen in about 1 per 2000 deliveries in HIC whereas in LMIC, the incidence of HDP varies from 1/100 to 1/1700 [31]. The International Society for the Study of HTN in Pregnancy recognizes five types of HDP:

1. Chronic HTN (CH) - Systolic blood pressure (BP) 140 mmHg and/or Diastolic BP level of90 mmHg present predating pregnancy or is diagnosed within the first 20 weeks of pregnancy.
2. Gestational HTN (GH) - Systolic BP 140 mmHg and/or Diastolic BP of 90 mmHg after 20 weeks' gestation and normalizes after pregnancy.
3. Preeclampsia-eclampsia (PE/E) – Systolic BP 140 mmHg and/or Diastolic BP of 90 mmHg after 20weeks' gestation which is accompanied by proteinuria or maternal organ dysfunction and with tonic-clonic seizures in cases of eclampsia. HELPP syndrome (Hemolysis, elevated liver enzymes and low platelets) is a serious manifestation of preeclampsia.
4. Pre-eclampsia superimposed on chronic HTN - Systolic BP 140 mmHg and/or Diastolic BP of 90 mmHg prior to preconception or prior to 20 weeks' gestation with signs and symptoms of preeclampsia after 20 weeks' gestation.
5. White coat HTN - Patients with systolic BP 140 mmHg and/or diastolic BP of 90 mmHg only in clinics. [32].

HDP is very commonly seen in LMIC mainly due to poorer health services and delayed identification of cases [29]. Nearly one-tenth of all maternal deaths in Africa and Asia are due to HDP. In West Africa, approximately 8% of the causes of maternal deaths are due to HDP [33]. HDP has also been reported to be the leading cause of maternal mortality in countries of sub-Saharan Africa such as Nigeria and Ghana [34], [35]. In Ghana, HDP is the third leading cause of maternal deaths accounting to 9% of the total deaths, after hemorrhage (23%) and abortion (14%) [36]. Recent studies in tertiary hospitals of Ghana showed that HDPs have overtaken hemorrhage as the leading cause of maternal mortality in the country [35], [37]. HDPs are slowly becoming the leading cause of maternal mortality worldwide with majority of deaths occurring in countries with low resource settings [38]. The complications of HDP are preventable and treatable because they usually occur due to mismanagement. Even though the exact causes for HDP are not fully understood, early recognition and appropriate management could greatly reduce the morbidity and mortality rates caused by HDP [37], [39].

Early recognition of these cases can be made easier with prediction models that identifies women with a risk of getting HDP based on one or more maternal clinical characteristics, uterine artery doppler and biomarkers [40],[41], [42]. Many such prediction models have been developed and are in use to identify women at high risk for HTN during pregnancy, especially in HIC. LMIC cannot implement these models due to differences in the availability and cost of diagnostic tools causing a lack of such prediction models in low resource settings and hence attribute to such an increased number of deaths [43]. Since, Ghana reports such a high number of HDP cases and deaths and is rapidly going toward becoming the leading cause of maternal deaths, a prediction model in Ghana will be helpful in early identification of women and providing them with appropriate care.

There is a paucity of predictive models in low resource settings such as Ghana. To the best of knowledge, there are only two models in Ghana developed by Antwi on GH using only maternal clinical characteristics [44] and another improved prediction model of GH by including placental growth factor and pregnancy associated plasma protein-a in Ghana [45], but there are no models on HDP in general in Ghana. There are also prediction models based only on maternal clinical characteristics but no models that considers the effect of socio-demographic and socio-economic factors on HDP which might also be interesting to understand. There are also very few studies conducted in Ghana on the risk factors associated with HDP including the socio-demographic and socio-economic factors. Moreover, there are inconsistent findings for every risk factor in various studies. Therefore, the aim of this study was to develop a multivariable prediction model using easily available information on socio-demographic, socio-economic, obstetric and maternal clinical factors from ANC clinics which identifies the risk factors of HDP in Ghana. The aim was also to develop such a prediction model that can also be used in countries such as Ghana with low resource settings without using expensive procedures. Furthermore, this study also offers recommendations for future research and Ministry of health (MOH), Ghana.

Objectives

General Objective

The objective of this study is to develop a prediction model which identifies women with risk of HDP by identifying the risk factors based on easily available information from primary care settings of Ghana.

Specific Objectives

1. To develop a prediction model that identifies women with risk for HDP using information available from primary care settings of Ghana.
2. To explore the risk factors responsible for the development of HDP in women of Ghana, Africa.
3. To validate the model internally to find out the efficiency and the predictive performance of the model.
4. To give recommendations for future research and MOH, Ghana.

3. Methods

Study design and population

This is a secondary data analysis nested in a cohort study among pregnant women in the Greater Accra Metropolis in Ghana between July 2012 and March 2014. This data set was used to address the objectives mentioned above. Reporting of the methods section complies with the Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis (TRIPOD) recommendations for structure and content except the addition of subsections of 'Search strategy' which was added to explain the search term used for the literature review that was necessary for this study and 'Study approach' which was added to have an overall strategy, that I chose, to integrate the different components of the study, which enables me to effectively address all the components of this study [46].

The study population was categorized into two groups: normotensive pregnancies and pregnancies with any type of HDP. The pregnancies with HDP can be either GH, defined in this study as a systolic BP level of 140 mmHg and/or diastolic BP level of 90 mmHg on two separate occasions without proteinuria after 20 weeks of gestation and normotensive pre-pregnancy level, a group of severe hypertensive disorders with PE/E, defined as a systolic BP level of 140 mmHg and/or diastolic BP level of 90 mmHg on two separate occasions with proteinuria (300 mg/24 h or ++ on a dipstick) after 20 weeks of gestation. With the occurrence of tonic-clonic seizures in a pregnant or recently delivered woman, women were considered to have eclampsia. The fourth group was CH where the women had their first elevated BP before 20 weeks of gestation [47]. If there was only one BP measurement available, then only the women who had both systolic BP and diastolic BP elevated were considered hypertensive.

The inclusion criteria were pregnant women above 18 years of age for legal and ethical issues and pregnant women with less than 17 weeks pregnancy. The exclusion criteria were women with previously established chronic HTN irrespective of parity, women with spontaneous abortions before 24 weeks and death of the participating woman [48].

Search strategy

For the sections of background and discussion, a literature search was performed on Google Scholar and PubMed. Articles were screened based on the topic of interest which was HDP. All searches were done in English language and only English articles were included. Articles older than 2000 were excluded. WHO, UNICEF, FAO, UNESCO and World Bank websites were used to collect more relevant demographical and epidemiological information on HDP and Ghana. A search on specific risk factor was done when required, to get more detailed information about every risk factor. The snowball technique was used to collect information missed during the searches. Table [1] describes the search terms used for the study.

("Hypertensive disorders during pregnancy" OR "Hypertension during pregnancy" OR "Pregnancy induced hypertension" OR "Chronic hypertension" OR "Gestational hypertension" OR "Pre-eclampsia") AND ("Prediction models" OR "Prognostic models" OR "logistic regression" OR "multivariable logistic regression" OR "deep neural networks" OR "Artificial intelligence" OR"machine learning") AND ("Ghana") AND ("Africa" OR "Sub-saharan Africa" OR "low income country" OR "low income countries" OR "low-and-middle-income countries" OR "low resource settings") AND ("maternal mortality" OR "maternal deaths") AND ("background information" OR "geography" OR "demography" OR "socio-economic status" OR "health indicators" OR "educational ststus") AND "("risk factors" OR "area of residence" OR "rural and urban" OR "ethnic group" OR "ethnicity" OR "maternal age" OR "age" OR "weight" OR "BMI" OR "Body mass index" OR "blood pressure" OR "systolic BP" OR "diastolic BP" OR "family history of HTN" OR "history of HDP" OR "history of HTN" OR "multi-fetal pregnancy" OR "twin pregnancy") AND ("calibration") AND ("discrimination").

Table 1: Search strategy

Sources of data

The participant demographic information and obstetric history was obtained by trained interviewers using well-structured questionnaires during their ANC visits to the two participating hospitals in the Greater Accra Metropolis of Ghana: Ridge Regional Hospital's out-patient clinic and Maamobi General Hospital. Information about the obstetric outcomes were collected subsequently from the women's medical records in the labor ward. Follow-up calls were made to obtain information on women who did not return to the clinic during pregnancy or after childbirth. Physician records and hospital's maternity registers were reviewed to determine the clinical outcome. The participants were enrolled into the study after a written informed consent and were followed up during ANC visits until they delivered. Ethical approval for the study was granted by the Ethical Review Committee of the Ghana Health Service, 2015 [48].

Participants

The data was collected by interviewers who were trained about the data collection. Information on maternal age, sociodemographic factors, socio-economic factors, family history of HTN (confirmed diagnosis of HTN in parents or siblings), history of HDP in previous pregnancy, parity, systolic and diastolic BP readings were obtained at booking, that is during the participants first ANC visit. All the other predictors were either obtained during the first ANC visit or during the subsequent visits. A mercury sphygmomanometer was used to measure the BP of the participating women. Two readings were taken at an interval of five minutes and the average was used to represent the woman's BP. A stadiometer was used to measure the height in centimeters, and a bathroom scale to measure the weight in kilogram were also obtained during the first ANC visit or subsequent visits [44].

The total number of enrolled participants were 1010. There was no outcome information in 179 participants, and hence were removed. Thirty-nine participants had spontaneous abortions and there were three maternal deaths. All these were removed and following this step the number of participants considered for the analysis was 789.

Outcome

The outcome of interest was HDP. The risk factors of HDP in Ghana were explored by building a prediction model that differentiates between women with a risk for HDP with those with no risk. HDP was defined as any participant with GH, PE/E/HELPP or CH [48]. To identify the outcome, the participants were followed over time using telephone interviews, death certificates reviews, physician records, hospital maternity registers and informant interviews.

The outcome variables included socio-demographic and socioeconomic factors which included area of residence, level of education, economic activity, marital status, religion and ethnicity of the woman and obstetric factors and maternal characteristics which included age, height, weight, BMI, systolic BP, diastolic BP, history of HDP during previous pregnancy, family history of HTN, parity and multi-fetal pregnancy.

Predictors

The data set was prepared and cleaned initially to train the model. All the predictors from the data which might have a correlation with HDP were shortlisted based on National Institute for health and care (NICE) guidelines for HTN in pregnancy, 2019 [49] and clinical relevance. Predictors that were collected after 17 weeks were ignored as this was a prognostic study. There were initially 512 variables, out of which 16 potential candidate predictors were considered for analysis. Out of the 16 predictors, four were continuous variables and the remaining 12 were categorical variables. All the categorical variables were encoded into one-hot numeric array using one-hot encoding which converted categorical variables into a numerical form that could be provided to machine learning algorithms for better prediction [50], [51]. One-hot encoding transforms a single variable to a binary variable which indicates the presence or absence of the outcome [52]. Both the shortlisted categorical and continuous variables were used for further modeling.

The predictors were categorized into two groups. The first group was sociodemographic and socioeconomic factors which included area of residence, level of education, economic activity, marital status, religion and ethnicity of the participating women. The second group included obstetric factors and maternal characteristics such as age, height, weight, BMI, systolic BP, diastolic BP, history of hypertensive disorder during previous pregnancy, family history of HTN, parity and multi-fetal pregnancy. Height and weight of the women were also considered in the analysis along with BMI to find the relation of HDP with both height and weight independently and also their combined influence by observing the correlation between HDP and BMI.

Socio-demographic and socio-economic factors were included in the study to observe the correlation of these variables with HDP. Knowing about the sociodemographic factors such as area of residence will tell us about the rural and urban differences and factors such as education of the woman will tell us about the awareness level of the pregnant women about such issues. This study also aims to add correlation between ethnic groups of Ghana and HDP which shows us the importance of ethnicity in such research studies. Other factors such as educational level and economic activity were also considered as important socioeconomic indicators associated with HDP and hence were added in the study.

Study approach

In order to meet the objectives of the thesis, this study was divided into 3 phases:

The first phase was data preparation. In this phase, the predictors that are related to HDP were chosen based on NICE guidelines and clinical relevance. There were 16 predictors chosen after this step and these predictors were divided into two groups. Group 1, Socio-demographic and socio-economic factors and Group 2, Obstetric factors and maternal characteristics. The group 1 had area of residence, level of education, economic activity, marital status, religion and ethnicity as the risk factors. The group 2 consisted of age, height, weight, BMI, systolic BP, diastolic BP, history of HDP during pregnancy, family history of HTN, parity and multi-fetal pregnancy. This phase also included univariable and bivariate analysis of the chosen predictors to understand the study participants and to understand the relation of various factors with HDP.

The second phase was model development. A multivariable logistic regression prediction model was developed which identified women with risk for HDP by identifying the risk factors of HDP in this cohort (Objective 1 and Objective 2).

The third phase was model performance. After the model was developed in phase 1, it was tested for its performance by internal validation. The procedures included to test for its performance were efficiency, discrimination and calibration. This step was done to evaluate the feasibility and effectiveness of the model in the clinical settings (Objective 3). Based on the results of the two phases, recommendations were provided for future research (Objective 4). The study approach is shown graphically in Figure [8].

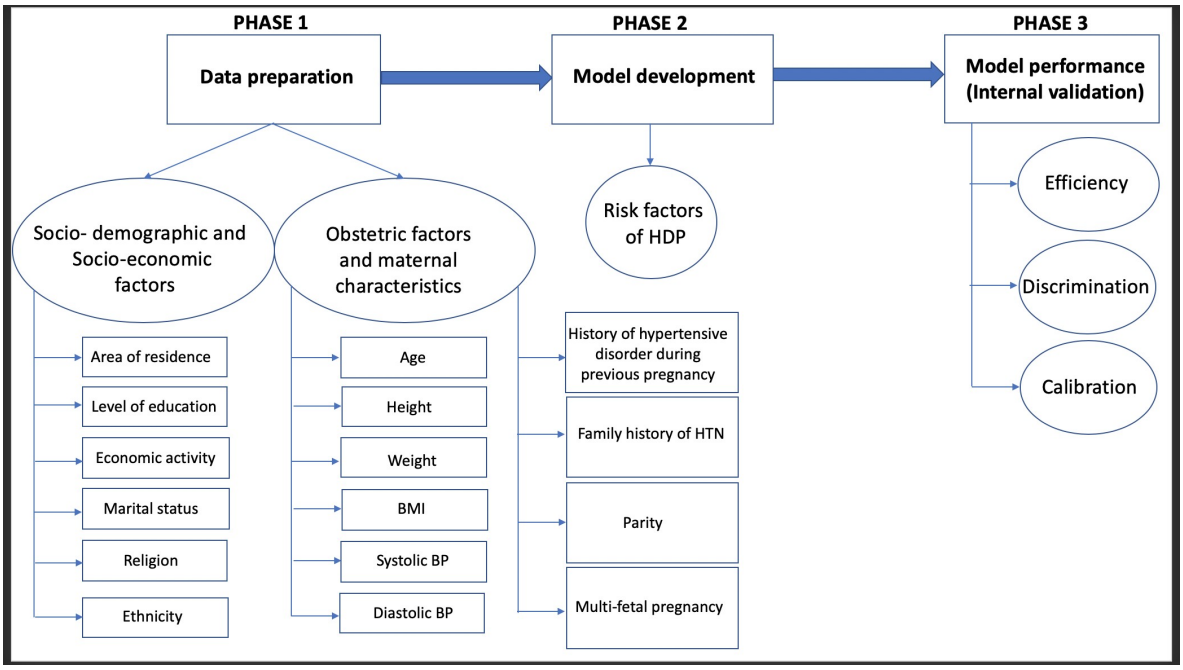

Figure 8: Study approach

Missing data

The missing values of continuous variables in the data set were imputed by multiple imputation using Scikit-learn [53]. Missing values were imputed 10 times and Rubin's rule [54] was applied and the results over the 10 imputed data sets were considered. The variables included in the multiple imputation were height, weight and BMI. The missing values of categorical variables were replaced by the most frequent value. The categorical variables that had missing values were systolic BP, diastolic BP and family history of HTN. Any predictor that was initially shortlisted based on NICE guidelines as an important predictor for HDP, but had more than 50% of missing values, were ignored later and were not included in the modeling process, resulting in the exclusion of diabetes mellitus predictor, which was initially shortlisted based on NICE guidelines and clinical relevance. Also, to ensure reliability of data, participants with no outcome information were removed. In addition, participants with spontaneous abortions and death of the participant were excluded from the analysis.

Sample size

The sample size was estimated based on the incidence of HDP in the population of Ghana and on the principle of 10 outcome events per variable [55]. According to the Ghana Maternal Health Survey, 2007, 9% of all maternal deaths were due to HDP [56]. By using an estimated incidence of HDP as 10% in the study population and for 10 predictors, it was aimed to enroll 1000 women but, in the end, enrolled 1010 women.

Statistical analysis methods

The analysis was done in the above mentioned 3 phases. The first, data preparation, in which the predictors were chosen, and univariable and bivariate analysis were carried out on these chosen predictors. The second, model development where in a multivariable logistic regression model was developed which also identified the risk factors of HDP and the third, model performance, where the model was tested for its performance by internal validation. To go about this, a dichotomous variable was defined, either 0 or 1, 0 being no HDP (Normotensive) and 1 being HDP (either PIH, CH, PE/E/HELPP). Descriptive statistics for variables were presented with frequencies and percentage for categorical variables and means and standard deviations (SD) for continuous variables, as seen in the Figure [9]. Logistic regression technique was chosen to predict the outcome using the selected candidate predictors.

Phase 1 : After choosing the predictors based on NICE guidelines, univariable analysis and bivariate analysis were performed to examine the relationship between every predictor and the outcome both individually and in relationship to one another was carried out which calculated the p-values and crude odds ratios (OR) and their 95% confidence intervals (CI) for every category of the predictor to understand the correlation better. Odds ratio >1 and p-values <0.05 were considered to indicate statistical significance [57]. A table with characteristics of the participants who were normotensive (NT) and participants who had HDP was created using the selected 16 predictors to know about the various risk factors of HDP in Ghana.

Phase 2: A multivariable logistic regression model to identify the risk factors using all the 16 predictors was built. This was done without following backward stepwise predictor selection based on p-value obtained from bivariate analysis, because this process might reject important predictors owing to nuances in the data set or confounding by other predictors [17]. Adjusted odds ratio (aOR) and their 95% CI were also reported along with the p-values which were calculated to control the confounding bias and to adjust for the confounders. The beta coefficients and standard errors of the predictors were also reported. Predictors with aOR >1, with a positive beta-coefficient and with p-values <0.05 were considered to indicate statistical significance [57].

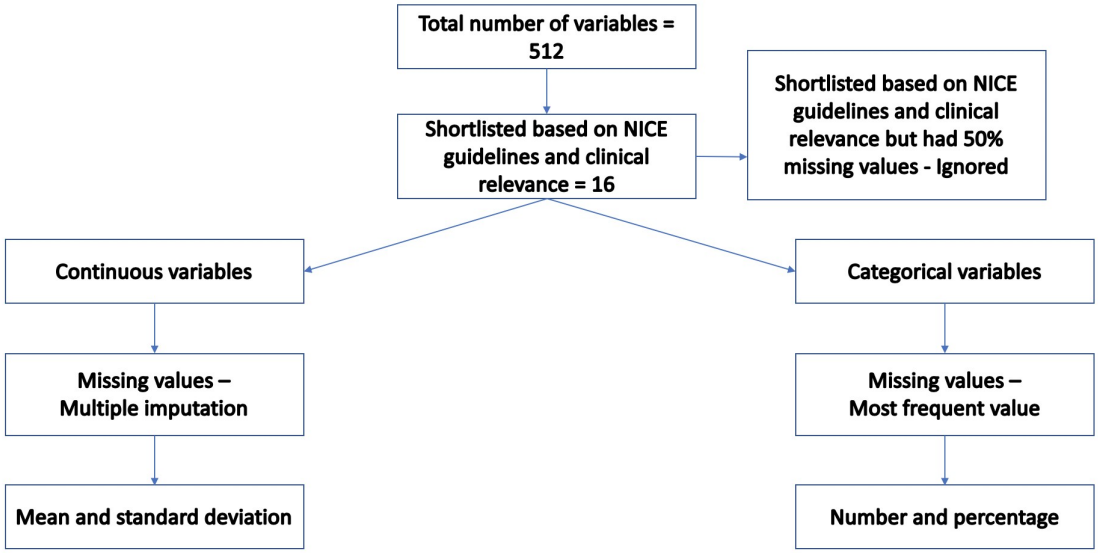

Figure 9: Statistical analysis methods

Phase 3: The predictive performance of the model was assessed by measuring the efficiency, discrimination and the calibration of the model. Efficiency tells us how well the model is able to predict the outcome of the test data set. The efficiency of the model in our study was calculated using split-sample validation procedure which needs to be done manually and through an in-built scikit-learn function called stratified k-fold cross-validation. Both are resampling techniques used to evaluate the model on a limited data set. These techniques are used to reduce the possibility of high bias if we have a small dataset, since we would miss information from the data that we have not used for training. The basic principle behind both the techniques are the same; giving all observations from the original data set, a chance to appear in the training and the test set. In the manual split-sample validation procedure, the data set was randomly split into two data sets. 70% of the data set was used for training the model and the remaining 30% was used for testing the model. The splitting was done randomly for 100 times with replacement of the test sample from the original sample each time after splitting the original data set into training set and test sets to exclude the chance of bias. The final efficiency was the average of all the 100 efficiencies obtained. As this procedure does not sometimes use all available

data for model development and since there Is no specific splitting method (60 and 40, 70 and 30, 80 and 20, or 90 and 10) it can be inefficient at times. So, the efficiency of the prediction model was also compared by using cross-validation procedure which is an in-built functionality in Scikit-learn library of Python. In stratified k-fold cross-validation from scikit-learn, the entire dataset is split into k-folds (k being 10 in this case). The model is then trained using k-1 folds (k-1 being 9 in this case) and later the model's performance is tested using the remaining fold (1 in this case). This is done for 10 times and the average of the efficiencies over 10 iterations is noted down. Statistical data analysis was performed using 'Scikit-learn' library and also 'Statsmodels' library of Python as all functionalities that the study demanded were not available in a single library [53], [58].

Discrimination is the ability of the model to differentiate between women who develop HDP (true positive) and those who do not (false positive). This was done by using receiver operating characteristic (ROC) curve statistic (c-statistic). The c-statistic or the area under the receiver operating characteristic curve (AUC) ranges from a value of 0.5 to 1.0. A value of 0.5 represents no discrimination and a value of 1.0 represents perfect discrimination [59].

Calibration tells us how closely the predicted HDP risk agrees with the observed HDP risk. A calibration curve was plotted between predicted risk and actual risk and the slope and the intercept of the curve were calculated.

4. Results

Phase 1: Data preparation

Study participants and participant characteristics

The total number of enrolled participants was 1010. There was no outcome information in 179 participants, and hence were removed. 39 participants had spontaneous abortions and there were three maternal deaths. All these were removed and following this step the number of participants considered for the analysis was 789. Among the 789 participants, 700 women (88.7%) were normotensive and 89 women (11.3%) had HDP with 59 women (7.5%) having PIH, 14 women (1.8%) having PE/E/HELPP and 16 women (2.0%) having CH, as seen in Figure [10].

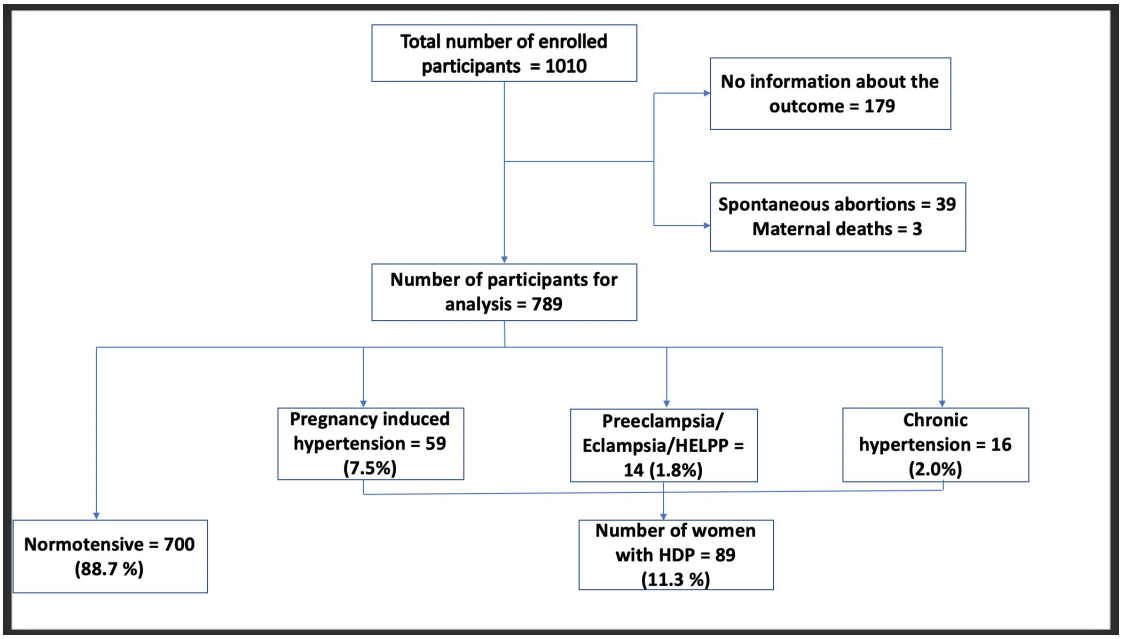

Figure 10: Data preparation

The characteristics of the study participants were understood by performing univariable analysis on the cohort. Most of the participating women in this cohort were from Accra metropolitan area (80.95%, n=618), the next majority of the women (16.25%, n=145) were from other urban areas followed by peri urban and rural areas (2.8%, n=26). When it came to educational status of the participating women, majority of the women were educated and studied up to lower secondary school (47.25%, n=373), 224 women (29.2%) had studied till upper secondary school, 2.25% of women (n=18) had studied till upper tertiary education and 11.0\% of women (n=92) had an education level of primary school. A small percentage of women (10.25%, n=82) had no formal education at all. Almost 3/4th of the women in this cohort (73.25%, n=593) were employed in an informal sector, from the remaining 1/4th of the cohort, 14.1% of the women (n=99) were formally

employed and 12.7% (n=97) of were not employed in any both formal or an informal sector and hence, were economically inactive.

Among the study participants, more than half of them were married (64.55%, n=492), 18.6% of women (n=157) were engaged or living together with their partners and 16.85% (n=14) of women were either single or widowed. The participants enrolled in the study were predominantly Christians (71.6%, n=570 women) and the remaining 28.4% (n=219) belonged to the religion of Islam. When coming to ethnic groups of the participating women, most of the women belonged to the three major ethnic groups of Ghana, with 284 women (32.55%) from Akan, 22.3% (n=161) women from Ewe, 21.95% (n=150) women from the ethnic group of Hausa. A small number of women belonged to other ethnic groups (7.2%, n=73), some women from the Northern: Mole, Dagbon, Gonja groups (9.05%, n=65) and the remaining 6.9% (n=56) were from the Ga ethnic group of Ghana.

It was found that women with HDP were slightly older than NT women (NT 28.0 4.9 years; HDP 29.7 5.6). There was no major difference in mean height between women who developed HDP and those who did not develop HDP (NT 161.16 cm ± 6.62; HDP 160.57 cm ± 11.69). The mean weight differed between women who were NT and women with HDP. It was seen that women with HDP weighed more than the women without HDP (NT 64.41 kg ± 12.42; HDP 74.89 ± 13.63). The mean BMI also differed between women with and without HDP, women with HDP having a greater BMI (NT 25.17 kg/m2 ± 4.70; HDP 28.92 kg/m2 ± 5.43).

More than half of the women had systolic BP between 100 and 120 mmHG (62.99%, n=497) and 1/4th of the women had systolic BP <100 mmHg (25.10%, n=198). 10.77% (n=85) of women had systolic BP between 120 and 140mmHg and 1.14% (n=9) women had systolic BP >140 mmHg. Seven of nine women with systolic BP >140 mmHg were having HDP. Most of the women had diastolic BP of 61 - 70 mmHg (35.11%, n=277) and <60 mmHG (32.19%, n=254). 23.83% (n=188) women had diastolic BP between 71 and 80 mmHg, 6.21% (n=49) between 81 and 90 mmHg and 2.66% (n=21) women had >90mmHg which also included 13 women with HDP.

Out of 789 women, 21.17% of women (n=167) were nulliparous, 1.52% of women (n=12) were primiparous and the remaining 77.31% (n=610) had 2 or more pregnancies in the past. Only 4.56% (n=36) had a family history of HTN in parents or siblings among which only 9 women were having HDP, 3.04% (n=24) reported to have a history of a HDP during previous pregnancies, out of which 8 women were having HDP in the current pregnancy as well and 0.76% (n=6) had multi-fetal pregnancy out of which only 2 women were with HDP. Table [2] describes the study participants and participant characteristics of socio-demographic and socio-economic factors and Table [3] describes the study participants and participant characteristics of obstetric factors and maternal characteristics.

	No HDP (n=700)	HDP (n=89)	Total (n=789)
	n (%)	n (%)	n (%)
Area of residence			
Accra metropolitan area	543 (77.6)	75 (84.3)	618 (80.95)
Other urban area	133 (19.0)	12 (13.5)	145 (16.25)
Peri-urban and rural area	24 (3.4)	2 (2.2)	26 (2.8)
Level of education			
Lower secondary school	331 (47.3)	42 (47.2)	373 (47.25)
Upper secondary school	197 (28.1)	27 (30.3)	224 (29.2)
Upper tertiary education	16 (2.3)	2 (2.2)	18 (2.25)
Primary school	83 (11.9)	9 (10.1)	92 (11.0)
No formal education	73 (10.4)	9 (10.1)	82 (10.25)
Economic activity			
Informal sector employment	530 (75.7)	63 (70.8)	593 (73.25)
Formally employed	87 (12.4)	12 (15.8)	99 (14.1)
Not economically active	83 (11.9)	14 (13.5)	97 (12.7)
Marital status			
Engaged/Living together	142 (20.3)	15 (16.9)	157 (18.6)
Married	432 (61.7)	60 (67.4)	492 (64.55)
Single, widowed	126 (18.0)	14 (15.7)	140 (16.85)
Religion			
Christianity	507 (72.4)	63 (70.8)	570 (71.6)
Islam	193 (27.6)	26 (29.2)	219 (28.4)
Ethnicity			
Akan	259 (37.0)	25 (28.1)	284 (32.55)
Ewe	139 (19.9)	22 (24.7)	161 (22.3)
Hausa	127 (18.1)	23 (25.8)	150 (21.95)
Other	69 (9.9)	4 (4.5)	73 (7.2)
Northern: Mole, Dagbon, Gonja	56 (8.0)	9 (10.1)	65 (9.05)
Ga	50 (7.1)	6 (6.7)	56 (6.9)

Table 2: Study participants and participant characteristics – Socio-demographic and socio-economic factors

HDP: Hypertensive disorders of pregnancy, n: number

	No HDP (n=700)	HDP (n=89)	Total (n=789)
	Mean +/- SD	Mean +/- SD	
Maternal age (years)	**28.0 +/- 4.9**	**29.7 +/- 5.6**	
Height (cm)	**161.16 +/- 6.62**	**160.57 +/- 11.69**	
Weight (kg)	**65.41 +/- 12.42**	**74.89 +/- 13.63**	
BMI (kg/m2)	**25.17 +/- 4.70**	**28.92 +/- 5.43**	
	n (%)	n (%)	n (%)
Systolic BP (mmHg)			
100 - 120 mmHg	445 (63.57)	52 (58.43)	497 (62.99)
121– 140 mmHg	65 (9.29)	20 (22.47)	85 (10.77)
<100 mmHg	188 (26.86)	10 (11.24)	198 (25.10)
>140 mmHg	2 (0.28)	7 (7.86)	9 (1.14)
Diastolic BP (mmHg)			
61 – 70 mmHg	256 (36.57)	21 (23.60)	277 (35.11)
71 - 80 mmHg	161 (23.01)	27 (30.33)	188 (23.83)
81 – 90 mmHg	36 (5.14)	13 (14.61)	49 (6.21)
<60 mmHg	239 (34.14)	15 (16.85)	254 (32.19)
>90 mmHg	8 (1.14)	13 (14.61)	21 (2.66)
Number of pregnancies			
0	147 (21.00)	20 (22.47)	167 (21.17)
1	11 (1.57)	1 (1.12)	12 (1.52)
2	220 (31.43)	21 (23.60)	241 (30.54)
3	160 (22.86)	22 (24.72)	182 (23.07)
>=4	162 (23.14)	25 (28.09)	187 (23.70)
Family history of HTN			
No	673 (96.1)	80 (89.9)	753 (95.44)
Yes	27 (3.9)	9 (10.1)	36 (4.56)
History of HDP during previous pregnancy			
No	684 (97.7)	81 (91.0)	765 (96.96)
Yes	16 (2.3)	8 (9.0)	24 (3.04)
Multi-fetal pregnancy			
No	696 (99.43)	87 (97.75)	783 (99.24)
Yes	4 (0.57)	2 (2.25)	6 (0.76)

Table 3: Study participants and participant characteristics – Obstetric factors and maternal characteristics

BP: Blood pressure, HDP: Hypertensive disorders of pregnancy, HTN: Hypertension, n: number, SD: Standard deviation

Predictors with missing data

Six predictors out of the chosen 16 predictors had missing data, which were height (1 missing value, 0.13%), weight (1 missing value, 0.13%), BMI (5 missing values, 0.63%), Systolic BP (1 missing value, 0.13%), Diastolic BP (1 missing value, 0.13%) and Family history of HTN (23 missing values, 2.92%). Table [4] shows the total number of observations and their corresponding missing values along with the percentage of missing values of these predictors.

	Number of observations	Missing (%)
Maternal age	789	0 (0)
Area of residence	789	0 (0)
Level of education	789	0 (0)
Economic activity	789	0 (0)
Marital status	789	0 (0)
Religion	789	0 (0)
Ethnicity	789	0 (0)
Height (cm)	788	1 (0.13)
Weight (kg)	788	1 (0.13)
BMI (kg/m2)	784	5 (0.63)
Systolic BP (mmHg)	788	1 (0.13)
Diastolic BP (mmHg)	788	1 (0.13)
Number of pregnancies	789	0 (0)
Family history of HTN	766	23 (2.92)
History of hypertensive disorder during previous pregnancy	789	0 (0)
Multi-fetal pregnancy	789	0 (0)

Table 4: The total number of observations with their corresponding missing values (with percentage missing)

BP: Blood pressure, HTN: Hypertension

Bivariate analysis

Bivariate analysis was performed to assess the risk factors of HDP dependently. It was found that eight of the 16 variables showed a statistically significant relationship (OR >1 and p value <0.05) with HDP in the bivariate analysis. None of the predictors from group 1, which included socio-demographic and socio-economic characteristics, showed any strong association between the two groups other than the ethnic group of Hausa which showed some statistical significance (OR 1.07, 95% CI 1.00 to 1.14, p-value = 0.04). From group 2 of obstetric factors and maternal characteristics, maternal age (OR 1.01, 95% CI 1.00 to 1.01, p-value = 0.00), weight (OR 1.01, 95% 1.00 to 1.01, p-value = 0.00), BMI (OR 1.02, 95% CI 1.01 to 1.02, p-value = 0.00), systolic BP 121 - 140 mmHg (OR 1.14, 95% 1.06 to 1.22, p-value = 0.00), systolic BP >140 mmHg (OR 1.96, 95% 1.60 to 2.39, p-value = 0.00), diastolic BP 71 - 80 mmHg (OR 1.07, 95% CI 1.01 to 1.13, p-value = 0.02), diastolic BP 81 - 90 mmHg (OR 1.21, 95% CI 1.10 to 1.32, p-value = 0.00), diastolic BP >90 mmHg (OR 1.72, 95% CI 1.51 to 1.97, p-value = 0.00), family history of HTN (OR 1.07, 95% CI 0.96 to 1.18, p-value = 0.01) and history of HDP in previous pregnancy (OR 1.26, 95% CI 1.11 to 1.43, p-value = 0.00) showed statistical significance to HDP.

Table [5] and Table [6] describes the bivariate analysis of the cohort for the two groups separately. In conclusion, the bivariate analysis showed seven statistically significant predictors. Women with an increased age (Mean 29.7kg, SD 5.6), increased weight (Mean 74.89, SD 13.63), BMI (Mean 28.92 kg/m2, SD 5.43), systolic BP >120 mmHg, diastolic BP >70 mmHg, having a family history of HTN and having a history of HDP in previous pregnancy increased the odds of getting HDP than in women without these characteristics.

	NT (%)	HDP (%)	OR (95% CI)	P-value
Area of residence				
Accra metropolitan area	543 (77.6)	75 (84.3)	Reference	-
Other urban area	133 (19.0)	12 (13.5)	0.96 (0.91 to 1.02)	0.19
Peri-urban and rural area	24 (3.4)	2 (2.2)	0.96 (0.84 to 1.08)	0.48
Level of education				
Lower secondary school	331 (47.3)	42 (47.2)	Reference	-
Upper secondary school	197 (28.1)	27 (30.3)	1.01 (0.96 to 1.10)	0.94
Upper tertiary education	16 (2.3)	2 (2.2)	1.00 (0.86 to 1.16)	0.69
Primary school	83 (11.9)	9 (10.1)	0.99 (0.92 to 1.06)	0.77
No formal education	73 (10.4)	9 (10.1)	0.99 (0.92 to 1.08)	0.98
Economic activity				
Informal sector employment	530 (75.7)	63 (70.8)	Reference	-
Formally employed	87 (12.4)	12 (15.8)	1.02 (0.95 to 1.09)	0.27
Not economically active	83 (11.9)	14 (13.5)	1.04 (0.97 to 1.11)	0.66
Marital status				
Engaged/Living together	142 (20.3)	15 (16.9)	Reference	-
Married	432 (61.7)	60 (67.4)	1.03 (0.97 to 1.09)	0.36
Single, widowed	126 (18.0)	14 (15.7)	1.00 (0.93 to 1.08)	0.90
Religion				
Christianity	507 (72.4)	63 (70.8)	Reference	-
Islam	193 (27.6)	26 (29.2)	1.01 (0.96 to 1.06)	0.75
Ethnicity				
Akan	259 (37.0)	25 (28.1)	Reference	-
Ewe	139 (19.9)	22 (24.7)	1.05 (0.99 to 1.12)	0.12
Hausa	127 (18.1)	23 (25.8)	1.07 (1.00 to 1.14)	0.04*
Other	69 (9.9)	4 (4.5)	0.97 (0.89 to 1.05)	0.42
Northern: Mole, Dagbon, Gonja	56 (8.0)	9 (10.1)	1.05 (0.97 to 1.15)	0.25
Ga	50 (7.1)	6 (6.7)	1.02 (0.93 to 1.12)	0.68

Table 5: Bivariate analysis - Socio-demographic and socio-economic factors

HDP: Hypertensive disorders of pregnancy, n: number, NT: Normotensive, OR: Odds ratio, CI: Confidence interval
*p-value: statistically significant

	NT (%)	HDP (%)	OR (95% CI)	P-value
	Mean +/- SD	Mean +/- SD		
Maternal age (years)	**28.0 +/- 4.9**	**29.7 +/- 5.6**	**1.01 (1.00 to 1.01)**	**0.00***
Height (cm)	**161.16 +/- 6.62**	**160.57 +/- 11.69**	**0.99 (0.99 to 1.00)**	**0.48**
Weight (kg)	**65.41 +/- 12.42**	**74.89 +/- 13.63**	**1.01 (1.00 to 1.01)**	**0.00***
BMI (kg/m2)	**25.17 +/- 4.70**	**28.92 +/- 5.43**	**1.02 (1.01 to 1.02)**	**0.00***
	n (%)	n (%)	OR (95% CI)	P-value
Systolic BP (mmHg)				
100 - 120 mmHg	445 (63.57)	52 (58.43)	Reference	-
121– 140 mmHg	65 (9.29)	20 (22.47)	1.14 (1.06 to 1.22)	0.00*
<100 mmHg	188 (26.86)	10 (11.24)	0.95 (0.90 to 0.99)	0.04*
>140 mmHg	2 (0.28)	7 (7.86)	1.96 (1.60 to 2.39)	0.00*
Diastolic BP (mmHg)				
61 – 70 mmHg	256 (36.57)	21 (23.60)	Reference	
71 - 80 mmHg	161 (23.00)	27 (30.34)	1.07 (1.01 to 1.13)	0.02*
81 – 90 mmHg	36 (5.14)	13 (14.61)	1.21 (1.10 to 1.32)	0.00*
<60 mmHg	239 (34.14)	15 (16.85)	0.98 (0.93 to 1.04)	0.52
>90 mmHg	8 (1.14)	13 (14.61)	1.72 (1.51 to 1.97)	0.00*
Number of pregnancies				
0	147 (21)	20 (22.47)	Reference	-
1	11 (1.57)	1 (1.12)	0.96 (0.80 to 1.16)	0.70
2	220 (31.43)	21 (23.60)	0.97 (0.91 to 1.03)	0.31
3	160 (22.86)	22 (24.72)	1.00 (0.94 to 1.07)	0.97
>=4	162 (23.14)	25 (28.09)	1.01 (0.95 to 1.01)	0.68
Family history of HTN				
No	673 (96.1)	80 (89.9)	Reference	-
Yes	27 (3.9)	9 (10.1)	1.07 (0.96 to 1.18)	0.01*
History of HDP during previous pregnancy				
No	684 (97.7)	81 (91.0)	Reference	-
Yes	16 (2.3)	8 (9.0)	1.26 (1.11 to 1.43)	0.00*
Multi-fetal pregnancy				
No	696 (99.43)	87 (97.75)	Reference	-
Yes	4 (0.57)	2 (2.25)	1.25 (0.97 to 1.62)	0.09

Table 6: Bivariate analysis - Obstetric factors and maternal characteristics

BP: Blood pressure, HDP: Hypertensive disorders of pregnancy, HTN: Hypertension, n: number, NT: Normotensive, OR: Odds ratio, CI: Confidence interval, SD: Standard deviation, *p-value: statistically significant

Phase 2: Model development

Multivariable Logistic regression model

All 16 predictors were included in the multivariable logistic regression model, regardless of their statistical significance in bivariate testing. This method was chosen because the predictors that were not statistically significantly in bivariate testing will sometimes have a statistical significance independently in the logistic regression model. The multivariable logistic regression gave few more additional statistically significant predictors but also removed a few predictors which showed statistical significance in bivariate analysis when adjusted for confounders.

In the multivariable analysis, only four individual predictors showed statistical significance as opposed to seven in bivariate testing, they were ethnicity, systolic BP, diastolic BP and multi-fetal pregnancy. The categories of individual predictors that showed significant relation to HDP in multivariable analysis were Hausa ethnic group (aOR 1.12, 95% CI 1.02 to 1.23, beta = +0.11, p-value = 0.02), Northern : Mole, Dagbon, Gonga ethnic groups of Ghana (aOR 1.10, 95% CI 1.00 to 1.19, beta = + 0.09, p-value = 0.04), systolic BP >140 mmHg (aOR 1.42, 95% CI 1.13 to 1.79, beta = +0.35, p-value = 0.00), diastolic BP of 81 to 90 mmHg (aOR 1.02, 95% CI 1.02 to 1.24, beta = +0.12, p-value = 0.02), diastolic BP >90 mmHg (aOR 1.42, 95% CI 1.20 to 1.66, beta = +0.35, p-value = 0.00) and multi-fetal pregnancy (aOR 1.27, 95% CI 1.00 to 1.61, beta = +0.24, p-value = 0.05).

The predictors that showed statistical significance in the multivariable analysis but did not in bivariate analysis were the Northern: Mole, Dagbon, Gonga ethnic groups of Ghana and multi-fetal pregnancy whereas there were a few predictors that were unable to hold up as an individual risk factor of HDP when it was adjusted for confounders in multivariable analysis. These were age (aOR 1.00, 95% CI 0.99 to 1.01, p-value = 0.38), weight (aOR 1.02, 95% CI 1.00 to 1.03, p-value = 0.09), BMI (aOR 0.97, 95% CI 0.93 to 1.02, p-value = 0.25), family history of HTN (aOR 1.07, 95% CI 0.96 to 1.18, p-value = 0.22) and a history of HDP in previous pregnancy (aOR 1.09, 95% CI 0.97 to 1.25, p-value = 0.14).

It was also interesting to see that being economically inactive (aOR 1.06, 95% CI 1.00 to 1.13, beta = +0.06, p-value = 0.07) and weight (aOR 1.02, 95% CI 1.00 to 1.03, beta = +0.02, p-value = 0.09) also showed significance in terms of adjusted odds ratio and beta-coefficients but the p-values were not significant. in multivariable analysis, it was also seen that living in urban areas reduced the odds of getting HDP in this cohort (aOR 0.94, 95% 0.89 to 1.00, beta = -0.06, p-value = 0.04).

Table [7] and Table [8] describes the multivariable analysis of the cohort for the two groups separately. In conclusion, it was found from the multivariable analysis that, women from Hausa and Northern : Mole, Dagbon, Gonga ethnic groups of Ghana, women with systolic BP >140 mmHg and a diastolic BP >80 mmHg and women with multi-fetal pregnancy had an increased odds of getting HDP than in women without these characteristics.

	OR (95% CI)	aOR (95% CI)	β-coefficient	Standard error	P-value
Area of residence					
Accra metropolitan area	Reference	Reference	-	-	-
Other urban area	0.96 (0.91 to 1.02)	0.94 (0.89 to 1.00)	-0.06	0.03	0.04*
Peri-urban and rural area	0.96 (0.84 to 1.08)	0.93 (0.82 to 1.05)	-0.08	0.06	0.22
Level of education					
Lower secondary school	Reference	Reference	-	-	-
Upper secondary school	1.01 (0.96 to 1.10)	0.98 (0.93 to 1.04)	-0.02	0.03	0.54
Upper tertiary education	1.00 (0.86 to 1.16)	0.96 (0.82 to 1.11)	-0.05	0.08	0.56
Primary school	0.99 (0.92 to 1.06)	0.97 (0.90 to 1.04)	-0.03	0.04	0.35
No formal education	0.99 (0.92 to 1.08)	1.01 (0.93 to 1.09)	+0.01	0.04	0.84
Economic activity					
Informal sector employment	Reference	Reference	-	-	-
Formally employed	1.02 (0.95 to 1.09)	1.02 (0.95 to 1.10)	+0.02	0.04	0.59
Not economically active	1.04 (0.97 to 1.11)	1.06 (1.00 to 1.13)	+0.06	0.03	0.07
Marital status					
Engaged/Living together	Reference	Reference	-	-	-
Married	1.03 (0.97 to 1.09)	0.97 (0.92 to 1.03)	-0.03	0.03	0.34
Single, widowed	1.00 (0.93 to 1.08)	0.98 (0.92 to 1.06)	-0.02	0.04	0.64
Religion					
Christianity	Reference	Reference	-	-	-
Islam	1.01 (0.96 to 1.06)	0.95 (0.88 to 1.03)	-0.05	0.04	0.21
Ethnicity					
Akan	Reference	Reference	-	-	-
Ewe	1.05 (0.99 to 1.12)	1.05 (0.99 to 1.11)	+0.05	0.03	0.13
Hausa	1.07 (1.00 to 1.14)	1.12 (1.02 to 1.23)	+0.11	0.05	0.02*
Other	0.97 (0.89 to 1.05)	0.98 (0.90 to 1.06)	-0.02	0.04	0.57
Northern: Mole, Dagbon, Gonja	1.05 (0.97 to 1.15)	1.10 (1.00 to 1.19)	+0.09	0.05	0.04*
Ga	1.02 (0.93 to 1.12)	0.99 (0.91 to 1.08)	-0.01	0.04	0.89

Table 7: Multivariable analysis - Socio-demographic and socio-economic factors

aOR: Adjusted Odds ratio, HDP: Hypertensive disorders of pregnancy, OR: Odds ratio, CI: Confidence interval
*p-value: statistically significant

	OR (95% CI)	aOR((95% CI)	β-coefficient	Standard error	P-value
Maternal age (years)	**1.01 (1.00 to 1.01)**	**1.00 (0.99 to 1.01)**	**+0.02**	**0.00**	**0.38**
Height (cm)	**0.99 (0.99 to 1.00)**	**0.99 (0.97 to 1.00)**	**-0.01**	**0.01**	**0.11**
Weight (kg)	**1.01 (1.00 to 1.01)**	**1.02 (1.00 to 1.03)**	**+0.02**	**0.01**	**0.09**
BMI (kg/m2)	**1.02 (1.01 to 1.02)**	**0.97 (0.93 to 1.02)**	**-0.03**	**0.02**	**0.25**
	OR (95% CI)	**aOR (95% CI)**			
Systolic BP (mmHg)					
100 - 120 mmHg	Reference	Reference	-	-	-
121– 140 mmHg	1.14 (1.06 to 1.22)	1.05 (0.97 to 1.13)	+0.05	0.04	0.24
<100 mmHg	0.95 (0.90 to 0.99)	0.99 (0.93 to 1.04)	-0.01	0.03	0.63
>140 mmHg	1.96 (1.60 to 2.39)	1.42 (1.13 to 1.79)	+0.35	0.12	0.00*
Diastolic BP (mmHg)					
61 – 70 mmHg	Reference	Reference	-	-	-
71 - 80 mmHg	1.07 (1.01 to 1.13)	1.04 (0.99 to 1.10)	+0.04	0.03	0.15
81 – 90 mmHg	1.21 (1.10 to 1.32)	1.12 (1.02 to 1.24)	+0.12	0.05	0.02*
<60 mmHg	0.98 (0.93 to 1.04)	1.00 (0.95 to 1.06)	+0.00	0.03	0.87
>90 mmHg	1.72 (1.51 to 1.97)	1.42 (1.20 to 1.66)	+0.35	0.08	0.00*
Number of pregnancies					
0	Reference	Reference	-	-	-
1	0.96 (0.80 to 1.16)	0.97 (0.82 to 1.16)	-0.03	0.09	0.77
2	0.97 (0.91 to 1.03)	0.95 (0.89 to 1.01)	-0.05	0.03	0.10
3	1.00 (0.94 to 1.07)	0.98 (0.92 to 1.05)	-0.02	0.03	0.63
>=4	1.01 (0.95 to 1.01)	0.95 (0.88 to 1.02)	-0.06	0.04	0.14
Family history of HTN					
No	Reference	Reference	-	-	-
Yes	1.07 (0.96 to 1.18)	1.07 (0.96 to 1.18)	+0.07	0.05	0.22
History of HDP during previous pregnancy					
No	Reference	Reference	-	-	-
Yes	1.26 (1.11 to 1.43)	1.09 (0.97 to 1.25)	+0.09	0.06	0.14
Multi-fetal pregnancy					
No	Reference	Reference	-	-	-
Yes	1.25 (0.97 to 1.62)	1.27 (1.00 to 1.61)	+0.24	0.12	0.05*

Table 8: Multivariable analysis - Obstetric factors and maternal characteristics

aOR: Adjusted Odds ratio, BP: Blood pressure, HDP: Hypertensive disorders of pregnancy, HTN: Hypertension, OR: Odds ratio, CI: Confidence interval
*p-value: statistically significant

Phase 3: Model performance

After the data was prepared for analysis in phase 1, a model was developed in phase 2 and was tested for its performance in phase 3. The performance of the model was good in this study population. The model performance was assessed using three methods: the efficiency, the calibration and the discriminating ability of the model.

The efficiency of the model was calculated using both split-sample validation procedure and cross-validation procedure and the efficiencies from both the procedures were very similar. The final prediction model had an efficiency of 88.68% after split sample validation procedure and an accuracy of 89.03% after cross-validation procedure. Overall, the efficiency of the model was good.

Figure [11] shows the area under the ROC curve of this cohort. ROC plots the curve of probability and is a measure of prediction accuracy. On the X-axis is the false positive rate and on the Y-axis is the true positive rate at various threshold points. The c-statistic or the AOC of the curve is usually between 0.5 to 1.0 where 0.5 means no discriminative ability and 1 means perfect discrimination. The c-statistic of our model was 0.72 (95% CI 0.66 to 0.81). Our model showed fair discriminative performance.

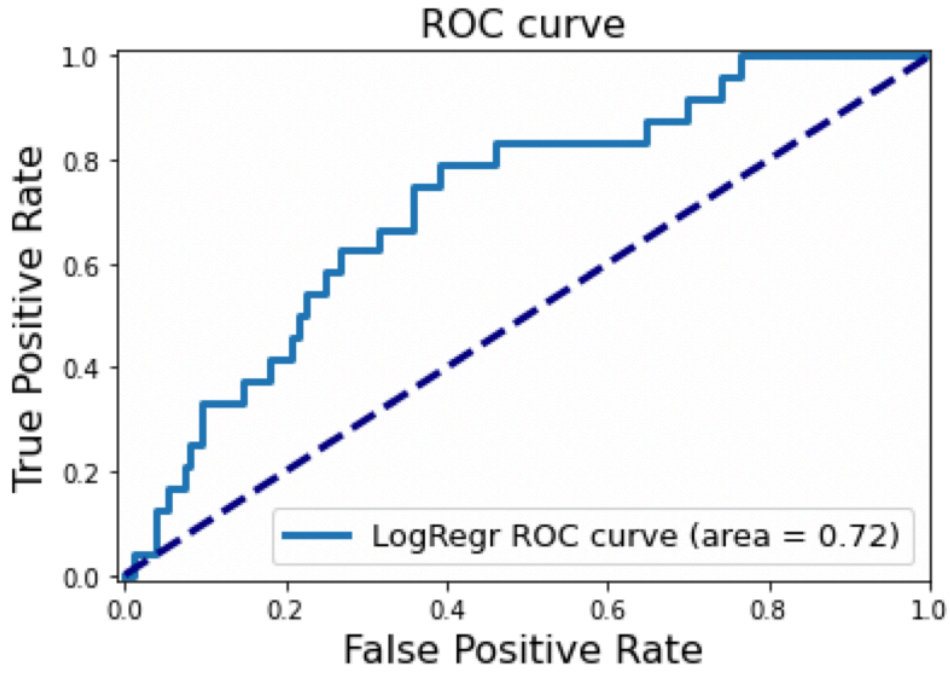

Figure 11: ROC curve of the prediction model

ROC: Receiver operating curve

Figure [12] shows the calibration plot for the cohort. The dotted black line at 45 degrees represents the perfect agreement between predicted risk on the x-axis and the actual risk on the y-axis. The dotted blue line represents the agreement between the predicted probability and the actual probability risks of the pregnant women for HDP in this cohort. The slope of the calibration curve was 1.25 (95% CI 0.70 to 1.80). The intercept of the curve was +0.84 and the standard error of the slope was 0.28. The calibration plot shows a reasonable fit as the slope was close to 1 and as the slope is well within the confidence interval. A positive intercept indicates that the incidence of HDP was a little higher in the test data than in the training data (13.92% in test set and 11.3% in training set).

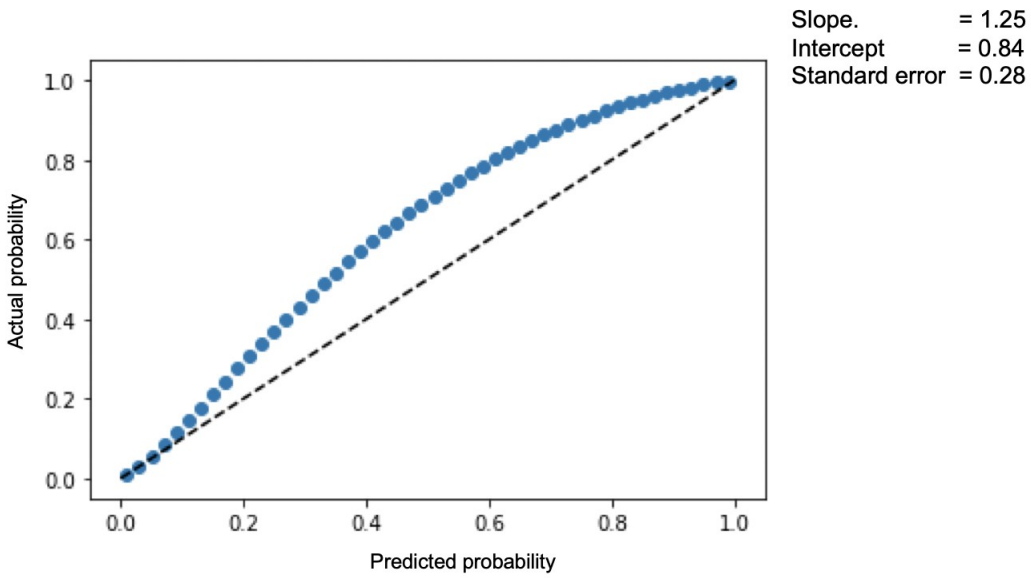

Figure 12: Calibration plot of the cohort

5. Discussion

A simple prediction model was developed and internally validated for HDP in a cohort of pregnant women attending ANC clinics in two hospitals of Ghana between July 2012 and March 2014 using the chosen predictors which were reduced from 512 to 16, based on NICE guidelines and clinical relevance. From the bivariate analysis of our study, it was found that Hausa ethnic group of Ghana, maternal age, weight, BMI, systolic BP >120 mmHg, diastolic BP >70 mmHg, family history of HTN and history of HDP in previous pregnancy were found to have statistically significant relation with HDP. However, in the multivariable logistic regression model, it was found that Hausa ethnic group, the Northern: Mole, Dagbon, Gonga ethnic groups of Ghana, systolic BP >140 mmHg, diastolic BP >80 mmHg and multi-fetal pregnancy were the independent risk factors of HDP in Ghana. As discussed above, we performed the analysis by retaining all the predictors to train the model and did not select the predictors based on p-values as obtained from the bivariate analysis. This is because, as seen in our results, there are a few predictors that turned out to be statistically significant in multivariable analysis even though they were not indicated by the bivariate analysis and vice versa. The predictive performance of the model was assessed by measuring the efficiency, calibration and discrimination. The efficiency of the model using split-sample method was 88.6% and 89.03% using 10-fold cross validation. The calibration and the discrimination of the model were also good.

The initial idea was to develop the model using deep learning algorithms such as Artificial neural networks (ANN) but since the data size was small and ANN needs large data sets, Logistic regression was used. However, there are many studies that report the efficiency of logistic regression and deep learning algorithms in clinical settings and these studies have concluded that both the methods have a similar performance in clinical settings [25], [76], [26], [77]. However, since neural networks are now being extensively used in clinical settings from diagnosis, classification to prediction [78], it would have been interesting to understand how well and how differently it works.

The model also works well for pregnant women in a study setting such as Ghana where expensive procedures are not available and are not affordable because our model uses only socio-demographic, socio-economic, obstetric and maternal clinical factors which are easily available from ANC clinics. Our findings also support that expensive procedures are not required to predict the risk of HDP. Similar to a previous model which was done on predicting GH in Ghana with slightly bigger data set. A study by Antwi et al in 2017, developed a prediction model for GH in Greater Accra region of Ghana using only maternal clinical characteristics and performed external validation of the model. The cohort had a bigger data set with 2529 women for development cohort and 647 women in the external validation cohort. The c-statistic of their model in the development cohort was 0.70 which became 0.68 after external validation. Their final model included six clinical predictors, namely history of GH in previous pregnancies, family history of HTN, Diastolic BP, Height, weight and parity [44].

Another prediction developed by Nijdam et al, in the Netherlands in 2010, was also using simple, easily available clinical variables to identify nulliparous women with risk for hypertension before 36 weeks of gestation. Three main predictors were found to have a strong correlation with HDP, which were weight, systolic and diastolic BP. It was also found that use of these three simple clinical variables obtained easily from ANC visits can accurately identify women with risk for HDP [60].

Data Interpretation and Comparison with Literature

The incidence of HDP in the present study is comparable with findings from other studies done in Ghana and Ghana statistical service [48], [44], [56]. The socio-demographic and socio-economic factors such as area of residence, level of education, employment status, marital status, religion and ethnicity are very particular to this data set. However, there were no major differences found in these factors between women with and without HDP as the outcome except for Hausa ethnic group that was statistically significant in both bivariate and multivariable analysis and the Northern : Mole, Dagbon, Gonga ethnic groups of Ghana that showed significance only in the multivariable analysis. To the best of knowledge, there was no study that reported the significance of ethnicity with HDP in Ghana. In this study, age was found to be associated with HDP in bivariate testing but disappeared after adjusting for confounders. The women with HDP in our study were slightly older compared to women without HDP which was in line with a few other studies which have reported that there is an increased risk of HDP with increased maternal age [61], [62], [63]. However, this was not found in this study after multivariable analysis.

Overweight women with a high BMI were at high risk of developing HDP as compared to women who had a normal weight and a normal BMI in agreement with other studies done in Ethiopia, Ghana and the Netherlands [64], [44], [60]. History of HDP during previous deliveries was another important risk factor of this study found in bivariate analysis. This is in line with other studies done in Ghana by Antwi [44] and in Cameroon [65]. Family history of HTN was another predictor found to be important in this bivariate analysis of our study and other studies done on HDP in Ghana [44] and other African countries such as Ethiopia [64] and Cameroon [65] also reported an increased risk of HDP in women with a positive family history in parents or siblings. However, little is known separately about the relation of HDP with histories of the women from the paternal and maternal sides. All these predictors were found significant in bivariate analysis but disappeared after adjusting for confounders other than weight which had a significant aOR but not a significant p-value.

Systolic and diastolic BP measurements at booking were also found to be significantly related to HDP in this study. This was also true in other studies on HDP in Ghana [44] and in the Netherlands who only used 3 variables, systolic BP, diastolic BP and weight of the women to identify women with risk for HDP [60]. Another predictor found significant in multivariable analysis was multi-fetal pregnancy. Studies from other African countries [66], other countries such as Nepal [67], China [68], India [69] and a study done considering the outcomes from 29 LMIC which included six countries from Africa and other countries from Asia, Latin America and the Middle East [70] have also reported that women with multi-fetal pregnancy are at a higher risk of developing HDP compared to women with singleton pregnancies. This is in agreement with our study which also indicates that multi-fetal pregnancy increased the odds of developing HDP. It has been suggested that the increased risk of HDP with women with multi-fetal pregnancies might be due to the large placental mass and increased circulating levels of soluble levels of Soluble fms-like tyrosine kinase-1and other placental growth factors [71].

One of the major risk factors reported by other studies in Africa and a study done in 23 developing countries in Africa, Latin America and Asia, is gestational diabetes [64], [72]. However, this predictor was not considered in our study due to a large number of missing data. It is recommended for future studies to add this predictor to understand the correlation between gestational diabetes and HDP in Ghana.

There were few other predictors that were found significant in multivariable analysis which had significant aOR but not significant p-values. They were being economically inactive and living in urban areas. It was interesting to see that being economically inactive increased the odds of getting HDP in our study population. Unfortunately, there was no study found in Ghana or other LMIC to compare this with. It was also found that living in urban areas reduced the odds of getting HDP in this cohort. Similar to other studies in Ethiopia [64] but a study done in Ghana reported otherwise [73].

Strengths and Limitations

One of the main strengths is the prospective character of the study and taking into consideration the predictors in early pregnancy. This can help minimize the impact of such pregnancy disorders and can help identify women with risk in an early stage thereby reducing the number of mortality and morbidity cases. Secondly, feasibility of the model was of utmost importance and the model's main aim was to identify women at risk for HDP in settings that lack an access to expensive procedures. This was done by considering only the predictors that are easily obtainable to develop the model and ignoring the predictors obtained by expensive or high-tech procedures such as uterine artery doppler and biomarkers. Prognostic models based on such easily available factors can be easy to use, informative, low cost and can have a potential to improve maternal health outcomes in LMIC. The application of models which includes the expensive procedures might become a limitation in low-resource settings due to constraints in the availability of the diagnostic equipment and the affordable capacity of the women to such procedures. Therefore, use of easily obtainable predictors is ideal for prediction models in Ghana for better feasibility. Thirdly, there are a very few models developed in Ghana and other low resource settings [74], hence models like these could be helpful, especially for diseases like HDP which needs early identification and early initiation of care.

No study is without limitations and our study also had several limitations. One of the main limitations of our study was considering the composite of HDP as the outcome which includes GH, CH, PE, E and HELPP. The study does not focus on the risk factors of the different types of HDP separately as the risk factors might not be the same for all types of HDP. Also, for every type of HDP, addition of new risk factors is necessary, for example, kidney disease is a major risk factor for PE and E, the age group and parity might for every type differs [49]. There might be another limitation with respect to predictor selection in this study. The predictors for this study were chosen based on NICE guidelines and clinical relevance; however, it is a possibility that this study may have excluded some major predictors or included predictors differently than other studies done in Ghana.

This study had several other limitations as well. First, the application of the model is probably limited to Accra region of Ghana because there might be differences especially in the socio-demographic and socio-economic factors, hence the efficiency of the model might not be the same without an external validation of the study in Ghana or any other low resource setting or updating the model as per the context of the setting without which the credibility of the model will be unsure. Also, limiting our study to only two hospitals of Ghana may have caused unintended selection bias for which external validation before implementation of the model might be one of the solutions. Second, the sample size of the study was too small, and it looks at a small number of women in Ghana, hence the model can be overfitting. Also, the confounders were not controlled due to lack of sufficient information. Hence, the reproducibility of this model in other settings is

unknown. Third, the risk was not categorized into high, moderate and low. This might become hard in the clinical setting while prioritizing the women so that the care is given to someone who needs it immediately. Fourth, use of only socio-demographic, socio-economic , obstetric and maternal factors and clinical characteristics excluding biomarkers and uterine artery Doppler in this study could also become a limitation at times because a combination of all the three along with the BP measurements of the women is proven to be an effective prediction tool than a model considering only one of these factors [75].

Challenges in evaluating performance of prediction models

Before a prediction model is used in clinical settings, it is very essential to ensure that the model is ready for use in such settings [79]. A major difficulty about prediction models is evaluating the performance or the efficiency of prediction models due to the lack of a good reference standard. Without evaluating the efficiency, it becomes hard to assess the reproducibility of the models. Although the model's performance was tested using discrimination and calibration, there is no gold standard for these methods which can decide how well the model can work. The model's performance will vary with a different data set with different data size and with different predictor selection. Questions like: How helpful will the models be in management and treatment decisions? What is the effect of these models on patient and disease outcomes? can arise. Another biggest challenge about prediction models is the unreliability of the predictions. For instance, the model can identify patients with an actual risk of HDP and it can sometimes identify women without an actual risk for HDP. This can lead to overuse of resources on someone who might have a low risk and therefore depleting the resources for people who really need them. This might lead to negative consequences especially in such low resource settings. Also, the usage of such models depends mainly on a physician's perspective about the patient. A physician always needs to use an additional information to base their decisions and cannot solely rely on the information presented by the prediction model [80]. Therefore, it is important that both overestimation and underestimation of the risk and the irrespective consequences are taken into account when evaluating the performance of a prediction model.

Despite the availability of many such models in clinical settings, not many are used in routine practice. There are various reasons for this. As mentioned above, the first reason being doubtfulness of clinicians on the probabilities provided by these models [81]. Second, the models can be too complex for daily use in clinical settings without computer support [82]. Third, preventive treatment for the outcome may not exist in these settings, so clinicians prefer management of the disease after it develops, hence do not wish to use the model. Fourth, models are usually not validated in other populations making it hard to know their generalizability [83].

However, despite the fact that prediction models have certain challenges and limitations, it surely has the potential of identifying cases who are at an increased risk in early stages which can be helpful to plan the treatment protocol and start subsequent care and monitoring. This will greatly reduce the effects in terms of mortality and morbidity rates, especially in Ghana and other LMIC.

6. Conclusion and Recommendations

Conclusions

This study developed a prediction model to identify women at risk for HDP by identifying the risk factors of HDP in Ghana and can provide clinical decision-making support to improve maternal health outcomes in Ghana and other LMIC. This study highlights the need for prediction models in such low resource settings and encourages the use of easily available predictors from ANC clinics for a cheap and affordable prediction of HDP. As this study does not need the information from expensive procedures such as uterine artery Doppler or biomarkers, it is easier for the physicians to decide the treatment plan immediately without any waiting time for the test results. This also enables the physicians in early detection of women at risk of HDP. The models can also be used by ANC providers after a brief training and an advice to refer the woman to the hospital immediately when a risk is found for appropriate management. This helps women who cannot go to hospitals due to access issues.

Early detection and timely management of HDP and its risk factors at antenatal care becomes very important in improving maternal health. Screening of all pregnant women at ANC to identify potential cases of HDP and initiating timely care is essential. This clinical prediction tool can be an important contribution in such situations as it offers the potential to improve health outcomes of women for a condition that is responsible for a large number of morbidity and mortality in Ghana and other LMIC which can prevent many women from reaching a stage where the effects of HDP are irreversible. This model may be reliable for countries where only predictors collected during ANC are available.

The study also identified certain risk factors of HDP. The risk factors for HDP in Greater Accra region of Ghana after bivariate analysis included Hausa ethnic group, maternal age, weight and BMI of the woman, systolic BP >120 mmHg and diastolic BP >70 mmHg, history of HDP during previous pregnancies and family history of HTN in either father or mother. From multivariable analysis, the independent risk factors were identified which were Hausa ethnic group, the Northern: Mole, Dagbon, Gonga ethnic groups of Ghana, systolic BP >140 mmHg, diastolic BP >80 mmHg and multi-fetal pregnancy. The knowledge of risk factors for HDP gives a defined track for its prevention in the population. It is important for this information to be collected during ANC checkups to identify the women at risk for HDP. Some more extended prospective studies are necessary to confirm the findings of this study.

However, the ultimate value of a prediction model will only be known after it has been implemented and after it is proven to demonstrate its potential to help change the health systems and improve the maternal health outcomes because no matter how good the efficiency of the model is with this data set, all algorithms tend to underperform with an external data set [25]. Therefore, external validation on this study is still required to understand the effectiveness of the model better and to confirm the performance of the model in a new data set. The possible research questions for future studies in regard to external validation could be:

a. Is there a difference in the feasibility of the model with a different data set from Ghana?
b. Is there a difference in the feasibility of the model across various geographical regions with low resource setting?

Another important factor in such studies is the classification of risk. Classification and prediction based on risk becomes important in settings such as Ghana where the physician needs to prioritize the women who needs immediate care and plan accordingly the treatment protocol based on the level of risk so that the resources are kept for moderate and severe risk cases. Hence, it is recommended for the future studies to also include a more detailed classification of the severity of HDP which was not possible to do in this study die to time constraints.

In conclusion, the findings of this study encourage and supports increased research studies and interventions related to areas of education and awareness in the social and community context that could help women with HDP. For LMIC such as Ghana to achieve the Millennium Development Goal 5 with respect to maternal mortality, the country should take immediate measures to reduce the number of HDP cases due to the matter of fact that it is the third major killer of pregnant women. Various strategies must be put in place to address these preventable causes of maternal deaths including better education for women receiving the ANC and increased numbers of trained maternal health personnel who are capable of handling such common causes of maternal deaths in the community level itself.

In the light of above-mentioned points in the discussion and conclusion, following recommendations are made for future research and MOH of Ghana. These recommendations are aimed to be beneficial in Ghana and other low resource settings, as was the aim of this study.

Recommendations

Recommendations for future research

Recommendation no 1: Other techniques to develop the model

The choice of the technique becomes very essential to develop such models that can be helpful in clinical settings because models can be developed using many techniques. The choice of one or other technique is not evident or proven and none of the techniques can be assumed better than the other, hence, the right choice depends on the size and the type of data. Therefore, it is recommended that this model is developed in the future with other techniques such as ANN and observe if the performance of the model improves. Also, use of other statistical measures to evaluate the performance of the model is recommended to get a better clarity of its performance in a clinical setting.

Recommendation no 2: More prospective studies.

Many more prospective studies in Ghana are necessary to confirm the findings of this study. Future research can further explore the risk factors based on the local setting and interventions and can be tailored to specifically address the areas which need attention immediately. Also, trying to replicate this study with a larger sample size is recommended as this study was not large enough to assess the real application of the model and this could be an area of continued research.

Recommendation no 3: Data collection

A systematic interview guide is advised while collecting the data to develop the model because lack of proper case taking will miss out information on a lot of important factors and since, the model is completely based on data obtained during ANC, missing out on a single factor also can be problematic.

Recommendation no 4: External validation to ensure the robustness of the model

External validation of this study is recommended in Ghana or any other low-income settings because a model's effectiveness is unknown until the model is validated with a new data set. Therefore, a vigorous validation of the model with multiple external data sets and updating the model as per the context of the setting should occur before actually using the model in practice. Later, an implementation research has to be done to understand the process of implementation, monitoring and evaluation of the model in Ghana.

Recommendation no 5: Anthropological studies

It would also be interesting and helpful to capture qualitatively the acceptance of models by physicians, ANC workers and the community of Ghana itself. Because it is often observed that very few physicians appreciate such models as the decision cannot be based completely on the prediction given by the model. Hence, by knowing their perspectives and by knowing the strengths and limitations of the model in clinical settings, they can be tried to design better in the future so that many physicians use them.

Recommendations to the MOH, Ghana

Recommendation no 6: Educating women about the risk factors

The health department of the country should conduct many campaigns and awareness programs with the help of medical staff, paramedical staff, community health workers and volunteers in order to educate women about the various risk factors of HDP and the essential lifestyle and pharmacological interventions necessary. Also, counseling mothers about pregnancy related complications must be done. The main target groups should be women with risk factors of HDP to have an early recognition and preparedness for a better management. Community engagement is the key in handling such diseases.

Recommendation no 7: Filling the knowledge gap

The knowledge of prediction models in Ghana is low. The MOH Ghana must promote the uses of the model among women for improved pregnancy outcomes and also among the frontline workers for an early identification and effective management of such cases. For the pregnant women, it can be done through contextual health education at ANC, media channels or through national education programs. Meanwhile, for frontline workers, it can be through training programs and workshops, especially if such models have not been used in the usual care so far.

Recommendation no 8: Involvement of various departments

The involvement of several departments such as policy makers, hospitals, frontline workers, researchers and medical universities will be pivotal in achieving significant improvement in the quality of care for women with HDP. When the clinical efficacy of the model is proven, the policy makers must adopt strategies to deliver the model and implemented in routine care. When the tool is implemented in a clinical setting, other health system outcomes should be also evaluated such as costs, human resources, acceptability of the model by the frontline professionals and the community.

Recommendation no 9: Model presentation

As a final recommendation, this study proposes to consider the presentation of a prediction model in a best possible way that addresses the clinical needs appropriately. There are many paper-based alternatives that are available for an easy presentation of the model such as score charts, monograms, web-based calculators or as apps for mobile phones, so that both the community and the health workers can use it.

References

1. Worldometer. Ghana Demographics 2020 (Population, Age, Sex, Trends) Worldometer. 2020. Available from: https://www.worldometers.info/demographics/ghana-demographics/
2. World Population Review. Ghana Population 2020 (Demographics, Maps, Graphs). 2020. Available from: https://worldpopulationreview.com/countries/ghana-population
3. Food and Agriculture Organization of the United Nations. Geography, climate and population of Ghana. 2003.
4. Food and Agriculture Organization of the United Nations. Ghana at a glance — FAO in Ghana. 2020.
5. United Nations Educational, Scientific and Cultural Organization – UNESCO Institute for Statistics. Ghana - Culture. 2020.
6. United Nations Development Programme. About Ghana — UNDP in Ghana. 2020. Available from: https://www.gh.undp.org/content/ghana/en/home/countryinfo.html
7. The World Bank. Ghana Overview – The World Bank in Ghana. 2020. Available from: https://www.worldbank.org/en/country/ghana/overview
8. 2020 Index of Economic Freedom. Ghana Economy: Population, GDP, Inflation, Business, Trade, FDI, Corruption. 2020. Available from: https://www.heritage.org/index/country/ghana
9. United Nations Development Programme — Human development reports. Ghana — Human Development indicators. 2019. Available from: http://hdr.undp.org/en/countries/profiles/GHA
10. The United Nations International Children's Fund. Education — UNICEF Ghana. 2020. Available from: https://www.unicef.org/ghana/education
11. United Nations Educational, Scientific and Cultural Organization. Ghana — UNESCO UIS – Education and Literacy. 2020.
12. Central Intelligence Agency. Field Listing: School life expectancy (primary to tertiary education) — The World Factbook. 2019.
13. Ghana Statistical Service (GSS), Ghana Health Service (GHS), and ICF International. Ghana - Demographic and Health Survey 2014. The DHS program. ICF International. Rockville, Maryland, USA: GSS, GHS, and ICF International. 2015.
14. World Health Organization. Global Health Observatory. GHO - Key Country Indicators — Ghana key indicators. 2016.
15. The World Bank. Life expectancy - Ghana — Data. 2020.
16. Ghana Statistical Service (GSS), Ghana Health Service (GHS), and ICF International. Ghana Maternal Health Survey 2017. Rockville, Maryland, USA: GSS, GHS, and ICF International. 2018.
17. Ewout S. Clinical Prediction Models: A Practical Approach to Development, Validation, and Updating. Research Gate. 2009.
18. Kleinrouweler CE, Cheong-See FM, Collins GS, et al. Prognostic models in obstetrics: Available, but far from applicable. American Journal of Obstetrics and Gynecology. 2016; 214:79–90.E36.
19. Vogenberg FR. Predictive and prognostic models: Implications for healthcare decision-making in a modern recession. American health and drug benefits. 2009; 2(6):218-22.
20. Jaimes F, Farbiarz J, Alvarez D, et al. Comparison between logistic regression and neural networks to predict death in patients with suspected sepsis in the emergency room. Critical care. 2005; 9:R150.
21. Lee YH, Bang H, and Kim DJ. How to establish clinical prediction models. Endocrinology and Metabolism. 2016; 31(1):38-44.
22. Ewout S, Vergouwe Y. Towards better clinical prediction models: seven steps for development and an ABCD for validation. European heart journal. 2014; 35(29):

1925–31.
23. Winters-Miner LA. Seven ways predictive analytics can improve healthcare. Elsevier. 2014. Available from: https://www.elsevier.com/connect/seven-ways-predictive-analytics-can-improve-healthcare
24. Vergara J, Esteban L, Sanz G, et al. Logistic regression versus neural networks for medical data. Ninth international conference Zaragoza-Pau on applied mathematics and statistics: Jaca (Spain). 2006; 245-52.
25. Lynam AL, Dennis JM, Owen KR, et al. Logistic regression has similar performance to optimised machine learning algorithms in a clinical setting: application to the discrimination between type 1 and type 2 diabetes in young adults. Diagnostic and Prognostic Research. 2020; 4:6.
26. Nusinovici S, Tham YC, Chak-Yan MY, et al. Logistic regression was as good as machine learning for predicting major chronic diseases. Journal of Clinical Epidemiology. 2020; 122:56–69.
27. World Health Organisation. Maternal mortality. 2019. Available from: https://www.who.int/news-room/fact-sheets/detail/maternal-mortality
28. The United Nations Children's Fund. Maternal mortality - UNICEF DATA. 2019. Available from: https://data.unicef.org/topic/maternal-health/maternal-mortality/
29. Duley L. The Global Impact of Pre-eclampsia and Eclampsia. Semin Perinatol. 2009; 33:130-137.
30. Steegers EAP, Dadelszen PV, Duvekot JJ, et al. Pre-eclampsia. The Lancet. 2010; 376(9741):631-44.
31. Duley L. Pre-eclampsia, eclampsia, and hypertension. BMJ clinical evidence. 2011; 1402.
32. Tranquilli AL, Gustaaf D, Laura M, et al. The classification, diagnosis and management of the hypertensive disorders of pregnancy: A revised statement from the International Society for the Study of Hypertension in Pregnancy. Pregnancy Hypertension - An International Journal of Women's Cardiovascular Health. 2014;4(2):97– 104.
33. Khan KS, Wojdyla D, Say L, et al. WHO analysis of causes of maternal death: a systematic review. Lancet. 2006; 367(9516):1066–74.
34. Adu-Bonsaffoh K, Obed SA, Seffah JD. Maternal outcomes of hypertensive disorders in pregnancy at Korle Bu Teaching Hospital, Ghana. International Journal of Gynecology & Obstetric. 2014; 127(3):238–42.
35. Lee QY, Odoi AT, Opare-Addo H, Dassah ET. Maternal mortality in Ghana: A hospital-based review. Acta Obstetricia et Gynecologica Scandinavica. 2012; 91(1):87–92.
36. Asamoah BO, Moussa KM, Stafström M, et al. Distribution of causes of maternal mortality among different socio-demographic groups in Ghana; A descriptive study. BMC Public Health. 2011; 11:159.
37. Adu-Bonsaffoh K, Samuel OA, Binlinla G, et al. Maternal deaths attributable to hypertensive disorders in a tertiary hospital in Ghana. International Journal of Gynecology and Obstetrics. 2013; 123(2):110–3.
38. Berhan Y, Endeshaw G. Maternal mortality predictors in women with hypertensive disorders of pregnancy: a retrospective cohort study. Ethiopian journal of health sciences. 2015; 25(1):89–98.
39. Campbell OM, Graham WJ. Strategies for reducing maternal mortality: getting on with what works. Lancet. 2006; 368:1284-99.
40. Akolekar R, Syngelaki A, Sarquis R, et al. Prediction of early, intermediate and late pre-eclampsia from maternal factors, biophysical and biochemical markers at 11 to 13 weeks. Prenat Diagn. 2011; 31(1)66–74.

41. Kuc S, Wortelboer EJ, Rijn BBV, et al. Evaluation of 7 serum biomarkers and uterine artery Doppler ultrasound for first-trimester prediction of preeclampsia: a systematic review. Obstet Gynecol Surv. 2011; 66(4):225–39.
42. Poon LC, Nicolaides KH. First-trimester maternal factors and biomarker screening for preeclampsia. Prenat Diagn. 2014; 34(7):618–27.
43. Kenny LC, Black MA, Lucilla P. Early pregnancy prediction of preeclampsia in nulliparous women, combining clinical risk and biomarkers the screening for pregnancy endpoints (SCOPE) international cohort study. Hypertension. BMC Public Health. 2014; 64(3):644–52.
44. Antwi E, Groenwold RH, Browne JL, et al. Development and validation of a prediction model for gestational hypertension in a Ghanaian cohort. BMJ Open 2017; 7(1): e012670.
45. Antwi E, Klipstein-Grobusch K, Browne JL, et al. Improved prediction of gestational hypertension by inclusion of placental growth factor and pregnancy associated plasma protein-a in a sample of Ghanaian women. Reproductive Health. 2018; 15(1):56.
46. Collins GS, Reitsma JB, Altman DG, et al. Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis (TRIPOD): the TRIPOD Statement. Br J Surg. 2015; 102:148–58.
47. Brown MA, Magee LA, Kenny LC, et al. Hypertensive disorders of pregnancy: ISSHP classification, diagnosis, and management recommendations for international practice. AHA Journals. 2018; 72:24-43.
48. Browne JL, Vissers KM, Antwi E, et al. Perinatal outcomes after hypertensive disorders in pregnancy in a low resource setting. Tropical Medicine & International Health. 2015; 20(12):1778–86.
49. National Institute for health and care Excellence. Overview — Hypertension in pregnancy: diagnosis and management — Guidance — NICE. 2019. Available from: https://www.nice.org.uk/guidance/ng133
50. Brownlee J. How to One Hot Encode Sequence Data in Python. 2017. Available from: https://machinelearningmastery.com/how-to-one-hot-encode-sequence-data-in-python/
51. Vasudev R. What is One Hot Encoding? Why and When do you have to use it? — Hacker Noon. 2017. Available from: https://hackernoon.com/what-is-one-hot-encoding-why-and-when-do-you-have-to-use-it-e3c6186d008f
52. Potdar K, Pardawala ST, Pai CD. A Comparative Study of Categorical Variable Encoding Techniques for Neural Network Classifiers. International Journal of Computer Applications. 2017; 175(4):7–9.
53. Pedregosa F, Varoquaux G, Gramfort A et al. Scikit-learn: Machine Learning in Python. Journal of Machine Learning Research. 2011; 12(85):2830.
54. Rubin DB. Multiple Imputation for Nonresponse in Surveys. Wiley-Interscience. 2004.
55. Harrell FE Jr. Regression modelling strategies with application to linear models, logistic regression, and survival analysis. SpringerVerlag New York. 2001.
56. Ghana Statistical Service (GSS), Ghana Health Service (GHS), and Macro International. Ghana Maternal Health Survey 2007. Calverton, Maryland, USA: GSS, GHS, and Macro International. 2009.
57. Gilliver S, Valveny N. How to interpret and report the results from multivariable analyses. Medical writing. 2016; 25(3).
58. Seabold S, Perktold J. Statsmodels: Econometric and statistical modeling with Python. 9th Python in Science Conference. 2010.
59. Steyerberg E. Clinical prediction models. A practical approach to development, validation and updating 2009. Springer Science +Media. 2009.

60. Nijdam ME, Janssen KJ, Moons KG, et al. Prediction model for hypertension in pregnancy in nulliparous women using information obtained at the first antenatal visit. Journal of Hypertension. 2010; 28(1):119–26.
61. Xiaoli L, Yan R, Yajun L, et al. Relationship between maternal age and hypertensive disorders in pregnancy. National Medical Journal of China. 2015; 95(1):19–22.
62. Gaillard R, Bakker R, Steegers EA, et al. Maternal age during pregnancy is associated with third trimester blood pressure level: The generation R study. American Journal of Hypertension. 2011; 24(9):1046–53.
63. Assis TR, Viana FP, and Rassi S. Study on the major maternal risk factors in hypertensive syndromes. Arq Bras Cardiol. 2008; 91(1):11–7.
64. Kahsay HB, Gashe FE, Ayele WM. Risk factors for hypertensive disorders of pregnancy among mothers in Tigray region, Ethiopia: matched case-control study. BMC Pregnancy and Childbirth. 2018; 18:482.
65. Tebeu PM, Foumane P, Mbu R, et al. Risk factors for hypertensive disorders in pregnancy: A report from the Maroua Regional Hospital, Cameroon. Journal of Reproduction and Infertility. 2011; 12:227–34.
66. Mbouemboue OP, Cellou D, Tamanji MT, et al. A Study on Factors Related to Hypertensive Disorders in Pregnancy in Ngaoundere (Adamawa Region, Cameroon). Clinical Medicine Research. 2016; 5:6.
67. Das S, Das R, Bajracharya R, et al. Incidence and risk factors of preeclampsia in the paropakar maternity and women's hospital, Nepal: A retrospective study. International Journal of Environmental Research and Public Health. 2019; 16.
68. Ye C, Ruan Y, Zou L, et al. The 2011 Survey on Hypertensive Disorders of Pregnancy (HDP) in China: Prevalence, Risk Factors, Complications, Pregnancy and Perinatal Outcomes. PLoS ONE. 2014; 9(6):e100180.
69. Agrawal S. Prevalence and Risk Factors for Symptoms Suggestive of Pre-Eclampsia in Indian Women. Journal of Women's Health, Issues & Care. 2014; 3.
70. Abalos E, Cuesta C, Carroli G, et al. Pre-eclampsia, eclampsia and adverse maternal and perinatal outcomes: a secondary analysis of the World Health Organization Multicountry Survey on Maternal and Newborn Health. BJOG: an international journal of obstetrics and gynaecology. 2014; 121(1):14–24.
71. Bdolah Y, Lam C, Rajakumar A, et al. Twin pregnancy and the risk of preeclampsia: bigger placenta or relative ischemia? American Journal of Obstetrics and Gynecology. 2008; 198(4):428.e1–428.e6.
72. Bilano VL, Ota E, Ganchimeg T, Mori R, and Souza JP. Risk Factors of Pre-Eclampsia/Eclampsia and Its Adverse Outcomes in Low- and-Middle-Income Countries: A WHO Secondary Analysis. PLoS ONE. 2014; 9:e91198.
73. Middendorp DV, Asbroek AT, Bio FW, et al. Rural and urban differences in blood pressure and pregnancy-induced hypertension among pregnant women in Ghana. Globalisation and health. 2013; 9:59.
74. Heestermans T, Payne B, Kayode GA, et al. Prognostic models for adverse pregnancy outcomes in low-income and middle-income countries: a systematic review. BMJ Global Health Journals. 2019; 4(5).
75. Onwudiwe N, Yu CKH, Poon LCY, et al. Prediction of pre-eclampsia by a combination of maternal history, uterine artery Doppler and mean arterial pressure. Ultrasound in Obstetrics and Gynecology. 2008; 32:877–83.
76. Pua YH, Kang H, Thumboo J, et al. Machine learning methods are comparable to logistic regression techniques in predicting severe walking limitation following total knee arthroplasty. Knee Surgery, Sports Traumatology, Arthroscopy 2019:1–10.

77. Christodoulou E, Ma J, Collins GS, et al. A systematic review shows no performance benefit of machine learning over logistic regression for clinical prediction models. Journal of Clinical Epidemiology. 2019.
78. Topol, EJ. High-performance medicine: the convergence of human and artificial intelligence. Nat. Med. 2019; 25:44–56.
79. Kappen TH, Peelen LM. Prediction models: The right tool for the right problem. Curr Opin Anaesthesiol. 2016; 29(6):717-26.
80. Kappen TH, Klei WAV, Wolfswinkel LV, et al. Evaluating the impact of prediction models: lessons learned, challenges, and recommendations. Diagnostic and Prognostic Research. 2018; 2:11.
81. Moons KG, Altman DG, Vergouwe Y, et al. Prognosis and prognostic research: Application and impact of prognostic models in clinical practice. BMJ Journals. 2009; 338: b606.
82. James BC. Making it easy to do it right. Pathways to Quality Health Care. Performance measurement: Accelerating improvement. 2001.
83. Reilly BM, Evans AT. Translating clinical research into clinical practice: Impact of using prediction rules to make decisions. Ann Intern Med. 2006; 144(3):201-9.