University of Nairobi

School of Mathematics

ISSN: 2410-1397

Master's Project in Mathematics

Small area estimation with an application to bivariate spatial modelling of hypertension and diabetes prevalence in Kenya

Research Report in Mathematics, Number 13, 2020

Noel Kanini Joseph

August 2020

Submitted to the School of Mathematics in partial fulfilment for a degree in Master of Science in Biometry

Master's Project in Mathematics **University of Nairobi**
August 2020

Small area estimation with an application to bivariate spatial modelling of hypertension and diabetes prevalence in Kenya

Research Report in Mathematics, Number 13, 2020

Noel Kanini Joseph

School of Mathematics
College of Biological and Physical sciences
Chiromo, off Riverside Drive
30197-00100 Nairobi, Kenya

Master Thesis
Submitted to the School of Mathematics in partial fulfilment for a degree in Master of Science in Biometry

Submitted to: The Graduate School, University of Nairobi, Kenya

Abstract

Background

Comorbidity of hypertension and diabetes leads to significant risks of mortality and other non-communicable diseases (NCDs) such as heart attacks and strokes. Kenya, like many low and middle-income countries (LMICs), faces a rapid increase in NCDs burden. However, sub-national burden profiles to inform health policy at the county level; the current health planning units are implausible due to small sample sizes from the existing NCDs data sources in Kenya. The main objective of this study was to determine the distribution of hypertension and diabetes disease prevalence at county units in Kenya using small area estimation methods.

Methods

Data from a nationally representative Kenya STEPwise survey for NCDs risk factors (STEPs-2015) was used. The survey collected health information (physical and biochemical measurements), risky behaviour and demographic indicators related to NCDs for 4,500 persons aged 18-69 years. Multivariate conditional autoregressive models that account for spatial autocorrelation and dependence between diseases (latent effects) were fit to estimate the county-specific prevalence of hypertension and diabetes. Simple multivariate improper CAR, improper multivariate CAR, proper multivariate CAR and M-model latent effects were explored. A mixed-effects multinomial logistic regression model was fit to identify the macro-risk factors of hypertension and diabetes in Kenya.

Results

The M-model was selected as the best fit based on DIC. Substantial geographical variation in the prevalence of hypertension ranging from 9% in Wajir county and 54% in Nyeri county while diabetes ranged from 0.1% in Narok to 8.1% in Makueni were observed. Overall, 47% (22 counties) and 36% (17 counties) had hypertension and diabetes prevalence estimates above the national burden, 26.4% and 2.7% respectively. Notably, Mombasa, Kiambu, Embu and Nyeri had a substantial burden of both hypertension and diabetes. High cholesterol, central obesity, age, BMI, harmful alcohol intake and high sugar intake were significantly associated with hypertension and diabetes.

Conclusion

The county-specific prevalence estimates provide the first evaluation of hypertension and diabetes burden that policymakers can use to inform interventions aimed at prevention and treatment of NCDs in Kenya. Implementation of comprehensive screening programs and awareness building for NCDs control are crucial in reducing hypertension and diabetes burden in Kenya.

Master Thesis in Mathematics at the University of Nairobi, Kenya.
ISSN 2410-1397: Research Report in Mathematics

DISTRIBUTOR: School of Mathematics, University of Nairobi, Kenya

Declaration and Approval

I, the undersigned declare that this dissertation is my original work and to the best of my knowledge, it has not been submitted in support of an award of a degree in any other university or institution of learning.

Signature Date

Noel Kanini Joseph
Reg No. I56/12351/2018

In my capacity as a supervisor of the candidate's dissertation, I certify that this dissertation has my approval for submission.

Signature Date

Prof. Samuel Mwalili
Associate Professor, JKUAT and
Adjunct professor, Strathmore University
E-mail: samuel.mwalili@gmail.com

Signature Date

Dr. Nelson Owour
School of Mathematics
University of Nairobi,
Box 30197, 00100 Nairobi, Kenya.
E-mail: onyango@uonbi.ac.ke

Dedication

This research project is dedicated to my family and friends for their support, encouragement and contribution towards its completion.

Contents

Figures and Tables

Figures

Tables

Acknowledgments

First, I thank VLIR-OUS Biostatistics Team Project for funding this project. To my supervisors, Prof. Samuel Mwalili and Dr. Nelson Owour, I am truly grateful for your guidance and inspiration throughout this project. I also wish to express my sincere gratitude to my fellow Masters classmates for cheering me on. Finally, I thank my colleagues at KEMRI-Wellcome Trust Research Programme for their support and encouragement.

Noel Kanini Joseph

Nairobi, 2020.

1 Background

1.1 Introduction

This chapter includes; background on disease burden of non-communicable diseases and disease specific (hypertension and diabetes), disease burden within Kenya context, highlights of the policy guidelines in tackling non-communicable diseases, statement of the problem and justification of the study.

1.2 Background

Non-communicable diseases (NCDs), are increasingly a major public health concern causing substantial morbidity and mortality while undermining global progress towards universal health coverage [Habib and Saha, 2010, Martinez-Beneito, 2013]. Globally, NCDs cause about 41 million deaths annually, representing more than two-thirds of all-cause deaths [WHO, 2018], preventable illness and related disability [Richards et al., 2016].The four main classifications of NCDs include; cardiovascular diseases (hypertension, stroke and heart attacks), chronic respiratory diseases (obstructed pulmonary disease and asthma), cancers and metabolic conditions (diabetes and obesity) [Hunter and Reddy, 2013]. Cardiovascular diseases cause the most NCDs deaths of about 17.9 million annual deaths, followed by cancers (9.0 million), respiratory diseases (3.9 million), and diabetes (1.6 million) [WHO, 2013].

NCDs burden is disproportionately distributed across various regions of the world with low- and middle- income countries (LMICs) bearing the highest burden of more than

75% of the global NCDs deaths (WHO, 2016). Sub-Saharan Africa (SSA) faces a rapid increase in NCDs.However, infectious diseases still dominate the overall disease burden causing dual disease burden in the region [Agyepong et al., 2017] and eminent danger of co-infections. Rapid urbanization, unhealthy diet, physical inactivity and increased consumption of tobacco and alcohol products are the main risk factors contributing to increased rates of NCDs that are projected to eclipse infectious diseases by 2030 [Boutayeb, 2006, Mathers and Loncar, 2006, Mufunda et al., 2006]. Hypertension is a common co-morbid of diabetes, leading to significant risks of mortality and other NCDs such as heart attacks and strokes [Torp-Pedersen and Jeppesen, 2011].

Hypertension has progressively become a common public health problem shifting from high-income countries to LMICs in the past decade [Zhou et al., 2017]. The WHO estimates the prevalence of hypertension to be highest in SSA compared to the other regions with 46% of adults aged above 25 years being hypertensive [WHO, 2018]. In 2010, approximately 130.2 million people were hypertensive in Africa with these numbers projected to increase to 216.8 million by 2030 [Adeloye and Basquill, 2014] [Motala and Ramaiya,].Approximately 15.5 (9.8–27.8) adults in Africa were diagnosed with diabetes in 2017, representing a regional prevalence of approximately 6% [Agyemang et al., 2016].SSA was projected to have the highest increase in diabetes of 162.6% by 2040compared to other regions, affecting about 40.7 million people [Agyemang et al., 2016, Ogurtsova et al., 2017]. Additionally, estimates suggest that only 31% of adults living with diabetes are diagnosed, leading to very high undetected diabetes-related complications [Agyemang et al., 2016].

Kenya, like many other LMICs, is experiencing an increase in NCDs burden. The STEPs survey 2015, showed that NCDs accounted for 50% of hospital deaths and more than

half of hospital admissions in Kenya. The age-standardized prevalence for hypertension, pre-diabetes and diabetes was 24.5%, 3.1% and 2.4%, respectively in 2015 [Mohamed et al., 2018a]. Hypertension is the leading risk factor causing 20% of all cardiovascular disease deaths in the country, while diabetes caused a lower proportion of deaths (1%) [WHO, 2013].

To strengthen global efforts in reducing NCDs burden, the World Health Organisation (WHO) endorsed the WHO Global Action Plan for the Prevention and Control of NCDs 2013-2020 [WHO, 2015]. This global declaration was translated into local policy as the 'Kenya National Strategy for the Prevention and control of NCDs 2015-2020' [MoH, 2015] strategic plan and within other health strategic plans such as the Kenya Health Sector Strategic and Investment Plan (KHSSP 2014-2018) [MoH, 2014] and the Kenya Essential Package of Health Services (KEPHS). These strategies reinforce the WHO global NCDs targets with the overall aim to halt and reverse the increasing NCDs burden in Kenya and specifically to reduce premature deaths from NCDs by 25% in 2025. The main components of the strategic plans constitute national prevention and management programs to mitigate NCDs in Kenya. Various programs have been implemented including; population-based and NCDs management programs.Raising awareness on avoidable NCDs risk factors, passing laws related to tobacco and alcohol products consumption and school health promotion initiatives constitute the health education initiatives while management programs focus on the delivery of care through comprehensive NCDs risk assessment, treatment, and prevention of secondary disease progression [MoH, 2015]. However, these programs are implemented at a national level with few evidence-based implementations at county units; the current health planning units [MoH, 2014].

County-specific prevalence estimates of hypertension and diabetes in Kenya can serve as a tool for prioritizing and optimizing the reduction of NCDs in the country. Small area estimation technique implementation on Bayesian hierarchical models is proposed to describe variations in geographical disease risks and identify areas of elevated risk for small sub-populations (small areas) where a sample is insufficient or no sample is available for the sub-population to be able to make accurate estimates. This study examines the associated risk factors of hypertension and diabetes using multinominal logistic regression and explores county level variations in their prevalence using a bivariate spatial modelling approach.

1.3 Statement of the problem

Most of the estimates of the NCDs burden in Kenya are done at the national level or selected county units due to limited data samples at lower units of administration. Even though counties and its lower units have become the current health planning units in Kenya, there are no published studies on NCDs at sub-national units. A better understanding of the NCDs estimates at the sub-national level, particularly to high-burden areas, is critical for efficient targeting with limited resources. Therefore, in this study, we incorporate disease mapping to inform the spatial distribution of NCDs at lower units of administration(counties) to delineate local disease burden and the associated risk factors. The disease model quantifies the prevalence of hypertension and diabetes. Findings from this study describe the local distribution of hypertension and diabetes disease burden useful in informing policy implementation towards the disease-specific and overall reduction of NCDs burden in Kenya.

1.4 Study objectives

1.4.1 Overall objective

To determine the distribution of hypertension and diabetes disease prevalence at county units in Kenya.

1.4.2 Specific objectives

- To identify the best fit of the bivariate spatial model's latent effects for hypertension and diabetes data.
- To quantify the distribution of hypertension and diabetes prevalence at the county level in Kenya.
- To identify macro-risk factors of both hypertension and diabetes diseases in Kenya.

1.5 Justification of the study

Analytical approaches that account for dependence among diseases while capturing spatial clustering and variation are widely used in epidemiological methods to study the distribution diseases [Jin et al., 2005]. Understanding geographical patterns of diseases and identifying risks corresponding to several diseases is essential for strategic planning and implementing of health programs that often seek to use limited resources through efficient prioritization. This study provides a health metric using spatial modeling to describe the geographical patterns for hypertension and diabetes in Kenya. Although a few studies [Chege, 2016, Mohamed et al., 2018b] have reported the prevalence and risk factors of hypertension and diabetes in Kenya using nationally representative survey

data sets, they did not provide regional variations of the prevalence in the country. This study will borrow strength from extensive literature on spatial modelling of diseases [Best et al., 2005, Held et al., 2005, Jin et al., 2007, MacNab, 2010] to estimate county-specific prevalence, which are useful in designing and implementation of interventions for regions identified as high-risk zones.

2 Literature Review

2.1 Introduction

This chapter provides a review of the literature on disease mapping methods relevant to the implementation of bivariate spatial models for diseases.

2.2 Disease mapping and small area estimation techniques

Disease mapping of disease is essential in understanding the spatial variation of diseases and their correlates. Therefore, disease maps are an integral part of public health epidemiology that inform health interventions, including prevention and control programs [Lawson and Lee, 2017]. Various statistical approaches to spatial analysis have been developed and are mainly classified as cluster detection or disease mapping. Cluster detection adopts a hypothesis testing framework to differentiate a typical disease rate from clustering (hot or cold spots) across a study region while disease mapping approach uses model-based techniques to produce smoothed disease rates estimates suitable for mapping [Gangnon and Clayton, 2003].

Small area estimation (SAE) techniques tackle the problem of unstable estimate by allowing a mechanism that borrows strength across neighbouring regions to improve direct estimates based on insufficient small sample sizes across regions of interest [Saei and Chambers, 2003, Torabi and Rao, 2008]. In addition, SAE improves estimation in non-sampled regions arising from survey design. SAE techniques utilize the spatial

setting and assumes a positive spatial correlation between observations, mainly borrowing more information from neighbouring regions than from those farther away hence smoothing local estimates towards local neighbouring values. This technique is derived from a conceptual framework proposed by [Clayton and Kaldor, 1987], who defined an empirical Bayesian approach built from a poisson regression with random intercepts that included spatial correlation. A hierarchical approach is adopted and a positive spatial correlation is introduced across the estimated regional rates using a conditionally autoregressive random effects distribution that is assigned to the region-specific intercepts [Besag, 1974]. Generalized linear mixed models are specified to estimate disease rates as they can accommodate spatial correlation and effects of disease covariates.

2.3 Conditional autoregressive models

Besag 1974, provides a fundamental understanding of the conditional autoregressive (CAR) models. In this section, CAR model for a univariate spatially random variable is described here. Consider ϕ_i, observed in n regions and define $\phi = (\phi_1 \cdots \phi_n)^T$; a vector with p components that follow a multivariate Gaussian distribution with mean zero and variance-covariance B^{-1}. Under the Markov random field (MRF) specification, the n full conditional distribution is defined as;

$$p(\phi_i|\phi_j, j \neq i, \tau_i^{-1}) = N(\alpha\Sigma_{i\sim j} b_{ij}\phi_j \tau_i^{-1}), i, j = 1, \cdots, n \quad \textbf{(1)}$$

Where $i \sim j$ indicates that region j is a neighbour of region i. The joint distribution is defined as below considering the Hammersley-clifford and Brook's lemma theorems of the full conditional distributions in 1;

$$\phi \sim N(0, [D_\tau(I - \alpha B)]^{-1} \tag{2}$$

Where B is an nxn matrix with $b_i i = 0$, $D_\tau = Diag(\tau_i)$ is an nxn diagonal matrix and α is a smoothing parameter measuring spatial correlation (dependence). From the CAR specification in 2), different CAR model structures can be chosen by defining α, D and B.

2.4 Multivariate conditional autoregressive models

Analogous to the univariate case presented above, the joint distribution for multivariate CAR (MCAR) models are derived from full conditional distribution following MRF assumption as developed by [Mardia, 1988]. The conditional distributions are specified as;

$$p(v_i | v_{j \neq i}, \neq i, \Gamma_i^{-1}) = N(R_i \Sigma_{i \sim j} B_{ij} v_j \Gamma_i^{-1}), i, j = 1, \cdots, n \tag{3}$$

Where $v_i = (\phi_{i1} \cdots \phi_{ip})^T$ is a p dimensional vector and R_i, B_{ij} and Γ_i are pxp matrices. Hence, the joint distribution is defined as below;

$$v \sim N(0, \Gamma(I - B_R)^{-1}) \tag{4}$$

where $v = (v_1, v_2, \cdots, v_n)^T$ is an $npxnp$ diagonal entries of Γ_i while $(B_R)_{ij} = R_i B_{ij}$ and $(B_R)_{ii} = 0$. To obtain various MCAR model structures, we define Γ and B_R matrices. To obtain a proper joint distribution, $\Gamma(I - B_R)$ is a positive definite and symmetric matrix. Therefore, $R_i = \alpha I_{pxp}$ for $i = 1, \cdots, n$ where α is the smoothing parameter and $\Gamma = D \bigotimes \Lambda$ Therefore, equation 4 becomes;

$$v \sim N(0, [(D(I - \alpha B)) \bigotimes \Lambda]^{-1}) \tag{5}$$

Where Λ is a pxp positive definite and symmetric matrix and the matrices D is an nxn diagonal matrix and B is a variance-covariance matrix. Hence, in the specification in 5 the precision matrix is a Kronecker product of the univariate CAR and Λ.

2.5 Model Parameter estimation using Integrated Nested Laplace Approximation

Estimation of county level prevalence involves prediction of random effects using generalized linear mixed models (GLMM) that lead to the prediction of the random effects that represent the prevalence. In this section, Bayesian inference of GLMM using

Integrated Nested Laplace Approximation (INLA) approach in latent GMRF models [Martino and Rue, 2009] a computational alternative to MCMC is explained.

Three stages are involved during the specification of the latent GMRF model in INLA. In the first stage, a distribution for the observed y that are assumed to be conditionally independent given a latent parameter η and an additional hyperparameter θ_1 are found.

$$\pi(y/\eta, \theta_i) = \prod_j \pi(y_j|\eta_j, \theta_1) \tag{6}$$

Where the latent parameter η are part of a larger latent random field x, that is modelled as a GMRF with a precision matrix Q depending on an additional hyperparameter θ_2. This constitutes the second stage of the hierarchial model.

$$\pi(x/\theta_2) \propto exp[-\frac{1}{2}(x-\mu)^T Q(x-\mu)] \tag{7}$$

The final stage of the model includes the prior distribution for the hyperparameters $\theta = (\theta_1, \theta_2)$. The INLA approach provides a fast Bayesian inference using accurate approximations to the marginal posterior density for the hyperparameters; $\pi(\theta|y)$ and the posterior marginal densities for the latent variables; $\pi(x_i|y); i = 0, \cdots, n-1$. The approximated posterior marginal are then used to compute summary statistics of interest such as mean values and variances.

2.6 Model choice

Several models are fitted to find the best model that ensures goodness of fit while adjusting or penalizing for model complexity. Various statistical selection criteria are widely used such as; Akaike's information criterion (AIC), Bayesian information criterion (BIC) and Deviance information criterion (DIC). The DIC, is generally preferred in model selection where hierarchical (complex) models are used [Zhu and Carlin, 2000]. A model with the least DIC is regarded a better fit.

3 Methods

3.1 Introduction

This chapter describes the data set, variables of interest and the statistical analysis used. Statistical analysis will provide details on model derivations, specifications, parameter estimation and selection as well as county-specific prevalence estimation.

3.2 Contextual characteristics

Kenya is located on Africa's east coast. It shares borders with Somalia to the east, Ethiopia to the north, South Sudan to the northwest while Uganda lies to the west and Tanzania to the south. As of 2013, Kenya adopted a devolved system of governance with 47 county governments [KPMG, 2013]. These 47 counties are the current health districts overseeing health planning and offering health services in the country with major policy directives from the national government. According to the Kenya population and housing census conducted in 2019, Kenya has a total population of 47.6 Million with approximately 50.4 % (24.0 million) of the population being female. Approximately 61 % of the population is above 15 years of age while four percent is above 65 years. A considerable proportion of the population 69 % lives in rural areas [KNBS, 2019].

3.3 Research design and data source

This is a retrospective quantitative cross-sectional study, that used secondary data from a nationally representative Kenya STEPwise survey for non-communicable diseases risk

factors (STEPs-2015). The survey collected information on health (physical and biochemical measurements), risky behaviour and demographic indicators related to NCDs for persons aged 18-69 years. A three-stage cluster sampling design was adopted for the survey to allow computation of national estimates by gender and urban or rural residence. Two hundred clusters (100 rural and 100 urban) were selected in stage one, then using a uniform selection 30 households were selected in stage two. Lastly, one adult aged 18-69 years were randomly selected from each household. A total of 4,500 individuals were sampled for the study. The fifth National sample surveys and evaluation programme (NASSEP V) master sample frame from the Kenya National Bureau of Statistics (KNBS) was used to aid sample selection in the Steps-2015 survey. A modular expanded STEPS approach was used during data collection with the following steps; demographic and behavioural information (step 1), physical measurements of the respondents (step 2) and biochemical measures (step 3). Extensive details of the sample design, methodology and questionnaires used are provided in the NCDs formal report produced by the Ministry of Health [MOH, 2015].

3.4 Ethical Considerations

Steps-2015 data collection procedures and questionnaires were reviewed and approved by the Kenya Medical Research Institute ethics review board (SSC NO: 2607). Additional participants' consent the current study data was not required since the dataset is publicly available and de-identifiable. The steps-2015 dataset was downloaded from the Kenya Bureau of Statistics official website.

3.5 Key variables of the study

The outcome variables of this study were hypertension and diabetes. In addition, bio-physical, behavioural and socio-demographic risk factors associated with hypertension and/or diabetes identified from a literature review were used to identify macro-risk factors of hypertension and diabetes in Kenya [Chege, 2016, Mohamed et al., 2018b, Wekesah et al., 2018]. The bio-physical risk factors included were body mass index, central obesity and cholesterol level while the behavioural factors considered were physical activity, alcohol use, tobacco use, sugar, salt and bad fat consumption and healthy diet (fruits and vegetable intake). Age of the participant, gender and socio-economic status derived from a wealth index that is computed from assets in the household using principal component analysis were the socio-demographic factors included in the analysis. Table 1 below shows the description of the outcome variables and risk factors as used in the study.

Table 1. Description of Key Variables

Determinants	**Definition**
Outcome variables	
Hypertension	Having a systolic blood pressure (SBP) $\geq$ 140 mmHg and/or a diastolic blood pressure (DBP) $\geq$90 mmHg and/or self-report of previous diagnosis of hypertension by a health care provider and/or if currently taking anti-hypertensives in the previous 2 weeks

Pre-diabetes and diabetes	Pre-diabetes was defined as impaired fasting blood glucose level (6.1 mmol/l to < 7 mmol/l) while diabetes was defined as impaired fasting blood glucose level ≥ 7 mmol/l and/or a self-report of previous diagnosis of diabetes by a health care professional or currently receiving treatment for diabetes
Biophysical risk factors	
Body mass index (BMI)	Weight divided by height squared
Central obesity	Waist circumference ≥ 94 cm for men and ≥ 80 cm for women
High cholesterol	Total cholesterol ≥5.0 mmol/L or are currently on medication for raised cholesterol
Low HDL-cholesterol	HDL-C < 1 mmol/l for men and < 1.3 mmol/l for women
Behavioural risk factors	
Physical activity	Insufficient physical activity was defined as self-reports of less than 150 min of moderate intensive activity or less than 75 min of vigorous intensive physical activity per week, including walking, running and cycling
Harmful use of alcohol	Consumption of more than 1 standard drink (which is the amount of alcohol found in a small beer, one glass of wine, or one tot of spirits) per day for females and more than 2 standard drinks for males

Behavioural risk factors	
Tobacco use	Self-reported current use of smoked tobacco or smokeless tobacco products
High sugar intake	Self-reports of far too much or too much consumption of sugar in a day
Bad fat intake	Self-reported use of saturated fats e.g. lard, margarine, butter and vegetable fat for cooking
High salt consumption	Self-report of far too much or too much consumption of actual salt and in processed foods, adding salt when cooking and/or to cooked food
Insufficient fruit and vegetable intake	Self-reported consumption of less than 5 servings/day of fruit and vegetables
Socio-demographic variables	
Age	Categorized as 18-29, 30-44, 45-59 and 60-69 age groups
Gender	Male or Female
Socio-economic status	Measured using a household asset and amenities index that assessed household ownership of various assets and amenities

3.6 Bivariate spatial model specification of hypertension and diabetes

The number of observed cases for hypertension and diabetes were aggregated at county units. Considering y_{ik} to represent the number of observed cases of k disease at the i county in which $i = 1, \cdots, I$ are the regions of interest and $k = 1, \cdots, K$ is used to index the number of diseases measured at region i.y_{ik} were modelled as a binomial outcome;

$$y_{ik} \sim B(n_{ik}, p_{ik}) \tag{8}$$

where n_{ik} and p_{ik} represent the sample size and the probability of disease of the ith county and k disease, respectively. For the binomial outcome a logit link function was used as shown below;

$$\log(\frac{p_{ik}}{1-p_{ik}}) = \alpha_k + \theta_{ik} + e_{ik} \tag{9}$$

where α_k is the intercept of k disease, θ_{ik} is the term that models multivariate latent effects and e_{ik} is the error term.

The latent effects (θ_{ik}) consist of variability between k diseases and variability corresponding to the independent spatial correlation for each k disease. A matrix, Θ was used to represent the latent effects with entries of θ_{ik} through a conditional autoregressive (CAR) process which are computationally plausible using multivariate

Gaussian vector that is normally distributed with mean zero and a highly structured precision matrix Σ as shown below;

$$vec(\Theta) \sim N(0, \Sigma^{-1}) \tag{10}$$

Different functions representing various structures and complexities were implemented to model the latent effects, θ_{ik}.

3.6.1 Latent effects specification and prevalence estimation

Five latent effects structures were explored to model θ_{ik} namely; simple multivariate improper CAR (simple IMCAR), simple multivariate proper CAR (simple PMCAR), improper multivariate CAR (IMCAR), Proper multivariate CAR (PMCAR) and M-model. The details of the specifications are as discussed below.

Simple multivariate improper CAR

To model the variability between and within hypertension and diabetes prevalence an improper CAR distribution was used to fit the spatial structure (within variability) and a diagonal covariance matrix to fit between diseases variability. Therefore, Θ is modelled as;

$$vec(\Theta) \sim N(0, \Lambda^{-1} \bigotimes (D - W)^{-1}) \quad \textbf{(11)}$$

where $D = diag(n_1, \cdots, n_I)$ is a diagonal matrix with the number of neighbours for region i and W is an adjacency matrix with entries one if regions i and j are neighbours and zero if else representing with simple independent spatial patterns. The between diseases precision matrix Λ is a K x K diagonal matrix with entries of marginal precision of the kth disease τ_k. Hence the model has K hyperparameters, $(\tau_k)_{K=1}^{K}$ equal to the number of diseases being modelled. A uniform improper prior distribution was assigned to the standard deviation for computational purposes as follows;

$$\sigma_k = \frac{1}{\sqrt{\tau_k}} \sim U_n(0, +\infty); k = 1, \cdots, K \quad \textbf{(12)}$$

Simple multivariate proper CAR

A proper CAR distribution was used to model within disease variability and a common spatial autocorrelation parameter α was introduced. Θ was modelled as;

$$vec(\Theta) \sim N(0, \Lambda^{-1} \bigotimes (D - \alpha.W)^{-1}) \quad \textbf{(13)}$$

Matrices D, W and Λ were defined similarly as above. However the model has $K+1$ hyperparameters that is α the spatial autocorrelation parameter and the $(\tau_k)_{K=1}^{K}$ the precision parameters. As in the previous latent effect specification, an improper uniform prior distribution is assigned to σ_k as well as on α as follows;

$$\sigma_k = \frac{1}{\sqrt{\tau_k}} \sim U_n(0,+\infty); k = 1,\cdots,Kand \tag{14}$$

$$\alpha \sim U_n(\alpha_m in, \alpha_m ax) \tag{15}$$

Improper Multivariate CAR

A dense precision matrix was used to model the between variability of hypertension and diabetes. Therefore, α^{-1} had marginal precision τ_k as diagonal entries and the off diagonals with entries of ρ_{ij} representing correlations between the pair of the diseases. Hence the set of hyperparameters consist of K precision parameters and $K * \frac{K+1}{2}$ correlation parameters. Similar to the simple improper multivariate CAR specification, Θ was modelled as:

$$vec(\Theta) \sim N(0, \Lambda^{-1} \bigotimes (D-W)^{-1}) \tag{16}$$

Unlike in simple improper multivariate CAR, a joint prior distribution was assigned to the between variability matrix Λ and not on each hyperparameter. A wishart prior distribution was set on Λ as follows:

$$\Lambda^{-1} \sim Wishart_k(r, R^{-1}) \tag{17}$$

where r is the degrees of freedom specified as $(2k+1)$ and R is a kxk identity matrix.

For computational plausibility the vector of hyperparameters consists of precision hyperamaters plus the correlations in the lower triangular matrix of Λ.

Proper Multivariate CAR

This specification corresponds to the simple proper multivariate CAR latent effects where Θ was modelled as;

$$vec(\Theta) \sim N(0, \Lambda^{-1} \bigotimes (D - \alpha.W)^{-1}) \tag{18}$$

However a joint Wishart prior distribution is specified for Λ as discussed above and a uniform prior distribution is assigned to the spatial autocorrelation parameter as below;

$$\alpha \sim U_n(\alpha_m in.\alpha_m ax) \tag{19}$$

The vector of hyperparameters in this alternative comprise of α the spatial autocorrelation parameter and the $K, (\tau_k)_{K=1}^{K}$ the precision parameters. Therefore the total number of hyperparaameters are;

$$K * \frac{K+1}{2} \tag{20}$$

Similar to the improper specification above, the vector of hyperparameters consists of precision hypeparamaters plus the correlations in the lower triangular matrix of Λ.

M-Model

A linear combination of K proper CAR spatial effects were considered as defined below:

$$\phi_k \sim N(0, (D - \alpha_k W)^{-1}), k = 1, \cdots, K \tag{21}$$

Where phi_k is a vector of length I and the spatial random effect of each disease k is defined as

$$\theta_{ik} = \phi_1 m_1 + \cdots + \phi_k m_k \cdots, K \tag{22}$$

Such that matrix M with entries m_{ik} that define the loadings of disease specific CAR spatial effects. Here, Θ was modelled as;

$$vec(\Theta) \sim N(0, M^T M \bigotimes I) diag((\Sigma_W)_1 \cdots (\Sigma_W)_k)(M \bigotimes I) \tag{23}$$

Where $(\Sigma_w)_k$ is the variance matrices of the $k$proper CAR and is given by;

$$(\Sigma_w)_k = (D - \alpha_k I)^{-1} \tag{24}$$

A wishart joint prior is set on $M^T M$ as below:

$$M^T M = Wishart_k(K, \tau I) \tag{25}$$

Table 2. Summary of latent effects

Latent effects	Latent effects structure	Number of hyperparameters	Matrix complexity	Joint Prior distribution
Independent improper CAR	$vec(\Theta) \sim N\{0, \Lambda^{-1} \otimes (D - W)^{-1}\}$	K	Sparse matrix	$\sigma_k \sim Un(0, +\infty); k = 1 \ldots K$
Independent proper CAR	$vec(\Theta) \sim N\{0, \Lambda^{-1} \otimes (D - \alpha . W)^{-1}\}$	K+1	Sparse matrix	$\sigma_k \sim Un(0, +\infty); k = 1 \ldots K$ $\alpha \sim Un(\alpha_{min}, \alpha_{max})$
Improper multivariate CAR	$vec(\Theta) \sim N\{0, \Lambda^{-1} \otimes (D - W)^{-1}\}$	K* (K+1)/2	Dense symmetric matrix	$\Lambda^{-1} \sim Wishart_k(r, \boldsymbol{R}^{-1})$
Proper multivariate CAR	$vec(\Theta) \sim N\{0, \Lambda^{-1} \otimes (D - \alpha . W)^{-1}\}$	K* (K+1)/2+1	Dense symmetric matrix	$\Lambda^{-1} \sim Wishart_k(r, \boldsymbol{R}^{-1})$ $\alpha \sim Un(\alpha_{min}, \alpha_{max})$
M-model	$vec(\Theta) \sim N\{0, (\mathbf{M}^T \otimes \mathbf{I}) diag((\Sigma_w)_1 \ldots (\Sigma_w)_k)(\mathbf{M} \otimes \mathbf{I})\}$ Where the variance matrices of the k proper CAR $(\Sigma_w)_k$ is given by; $(\Sigma_w)_k = (D - \alpha_k \mathbf{I})^{-1}$ **Note:** Matrix M consists of entries from linear combinations of the different CAR spatial effects for each disease that is $\theta_{ik} = \phi_1 m_1 + \ldots + \phi_k m_k$ where; $\phi_k \sim N\{0, (D - \alpha_k W)^{-1}\}$	K*K	Dense symmetric matrix	$\mathrm{M}^T \mathrm{M} \sim Wishart_k(K, \tau \mathbf{I})$

With parameters K and τI such that τ is a fixed precision parameter at 0.001. The vector of hyperparameters in this specification therefore, comprise of spatial autocorrelation parameters and columns of matrix **M**. The average mean squared error and correlation coefficient were estimated to evaluate the predictive fit of the selected model.

The latent effects discussed above are summarized in Table 2.

3.6.2 County-specific prevalence estimation

The spatial models described above were fit within a Bayesian framework. Therefore a posterior distribution was obtained and used to compute the mean to describe county-specific prevalence of hypertension and diabetes. Bayes theorem was applied such that the posterior distribution $p(y_i|\theta)$ was given by the equation below:

$$p(y_i|\theta) = \frac{p(\theta)L(\theta|y)}{p(y)} \tag{26}$$

where $p(y_i|\theta)$ is the posterior probability, $p(\theta)$ is the prior probability, $L(\theta|y)$ is the likelihood function and $p(y)$ is a marginal likelihood obtained from the observed values.

The county level prevalence was obtained by summarizing the posterior distribution and obtaining the mean as follows:

$$\bar{\theta}_i = \int \theta p(\theta|y) d\theta, i = 1, \cdots, 47 \tag{27}$$

A correlation coefficient of hypertension and diabetes prevalence was obtained to assess their association.

3.7 Multinomial Logistic Regression

A mixed effects multinomial logistic regression model was fit to identify risk factors associated with both hypertension and hypertension (macro-risk factors). The response variable y was classified into four categories as; normal $(y = 0)$, hypertension only $(y = 1)$, diabetes only $(y = 2)$ and hypertension and diabetes $(y = 3)$. The normal category $(y = 0)$ was used as the reference group in the regression. Due to the hierarchical structure of the Steps-2015 data a random effect to account for clusters

nested within counties was included in the model. The outcome y was modelled as follows:

$$y_{ijk} = \log \frac{P(R_{ij} \leq k)}{1 - P(R_{ij} \leq k)} = \beta_{0j} + \beta_{1j} + \mu_k \quad \textbf{(28)}$$

for an individual in the ith cluster in the jth county being at or below the kth level of the response variable y_{ijk}. μ_k represents the random effect due to clusters nested within counties. The final model was selected based on the least Akaike's information criterion (AIC). To define the final model, we reported adjusted odds ratios with their $95percent$ confidence intervals, variance and it's standard error to infer on the effects of the random effects μ_k and Wald test (p-value <0.05) to inform the overall significance of the models. Multicollinearity test among risk factors identified in the final model was checked using variance inflation factor (VIF) with a cut off 5. The analyses were done using gsem function in STATA v.14 (Stata Statistical Software: Release 14. College Station, TX: StataCorp LP) and Steps-2015 sampling weights were incorporated throughout the analysis

4 Results

4.1 Introduction

Chapter four gives the results of the descriptive analysis, the models selection criteria, county-specific prevalence and the macro-risk factors of hypertension and diabetes.

4.2 Study participants characteristics

The final analysis included 4,485 out of the 4,500 participants. Fifteen study participants that were missing diabetes results and had inconsistent age (age outside 15-69 years range) were dropped from the analysis. Table 3 shows the prevalence of hypertension and diabetes by participants characteristics. The study revealed that the national prevalence for hypertension was 26.4 percent while that of diabetes is 2.7 percent. The prevalence of hypertension and diabetes were found to be high among female and urban residents. However, these differences were not significantly different. Approximately 4.5 percent and 6.6 percent had pre-diabetes and diabetes respectively as a comorbidity of hypertension.

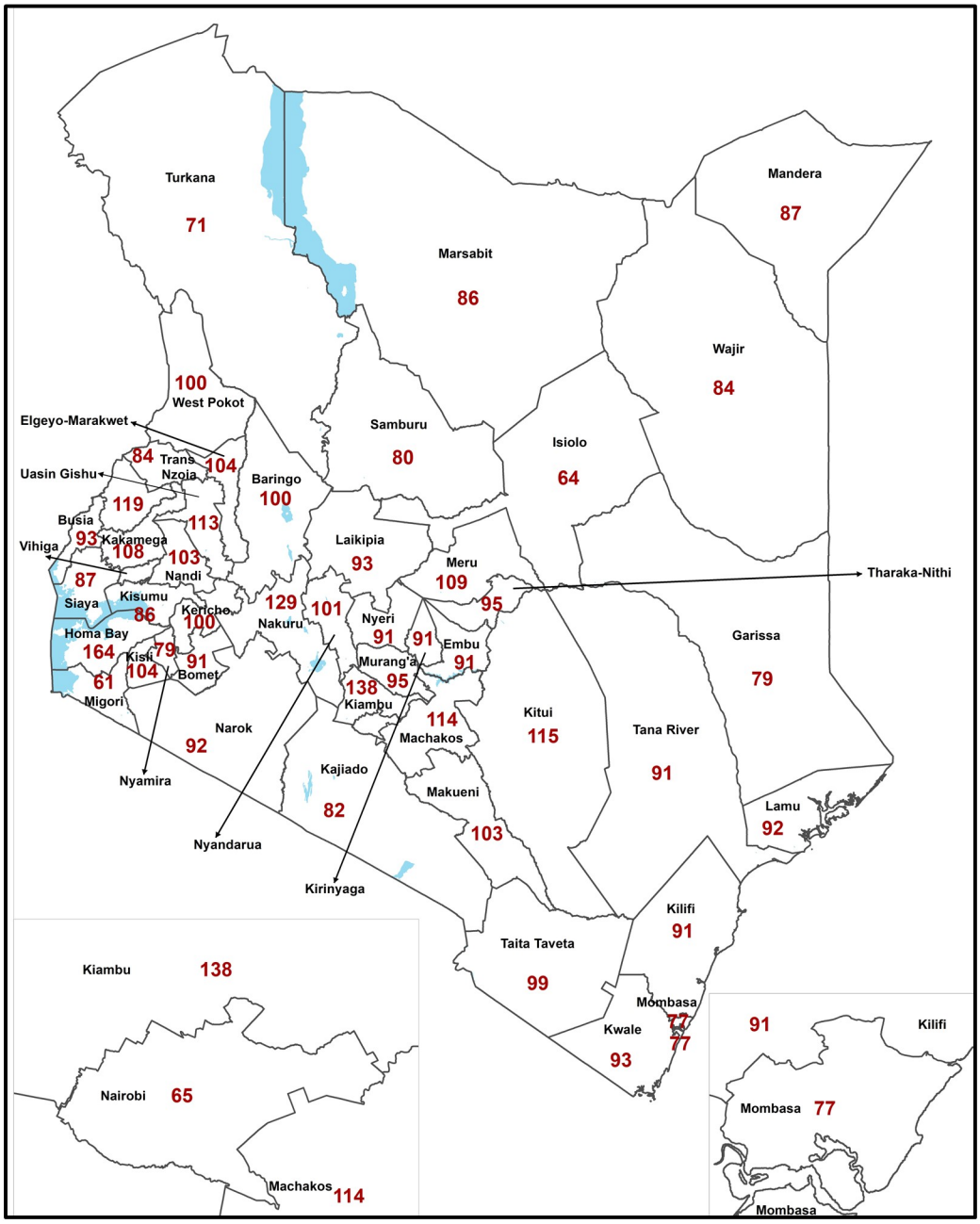
4.3 Observed counts of hypertension and diabetes at county-units

County-wise disaggregation of the study samples as shown in Figure 2 indicate that the sample sizes for each region are small to provide county level crude prevalence. As discussed in the methods, we addressed this issue by using a Bayesian spatial analytical approach.

Table 3. Study population Characteristics

		Hypertension	Diabetes
Characteristics	**Sample size**	**Prevalence (95% CI)**	
Gender			
Male	1,791	26.0(8.1-60.2)	2.4(0.9-14.3)
Female	2,694	26.6(6.9-60.0)	3.1(1.1-16.1)
Place of residence			
Urban	2,189	27.1(3.4-57.5)	3.3(0.8-19.0)
Rural	2,296	25.1(13.4-63.9)	2.1(0.4-12.9)
Pre-diabetes			
No	4,329	48.1(24.2-89.8)	
Yes	156	4.5(0.02-28.0)	
Diabetes			
No	4,333	46.0(20.0-78.5)	
Yes	152	6.6(0.08-34.8)	
Overall	**4,485**	**26.3(3.1-49.5)**	**2.7(0.6-11.3)**

Figure 1. County-specific sample size



4.4 County-specific prevalence of hypertension and diabetes

Five different spatial models described earlier were fit and DIC values of each model computed to identify the best fit model (Table 4). The results revealed that, M-model that included a spatial autocorrelation parameter and as well as correlation between hypertension and diabetes had the lowest DIC value.

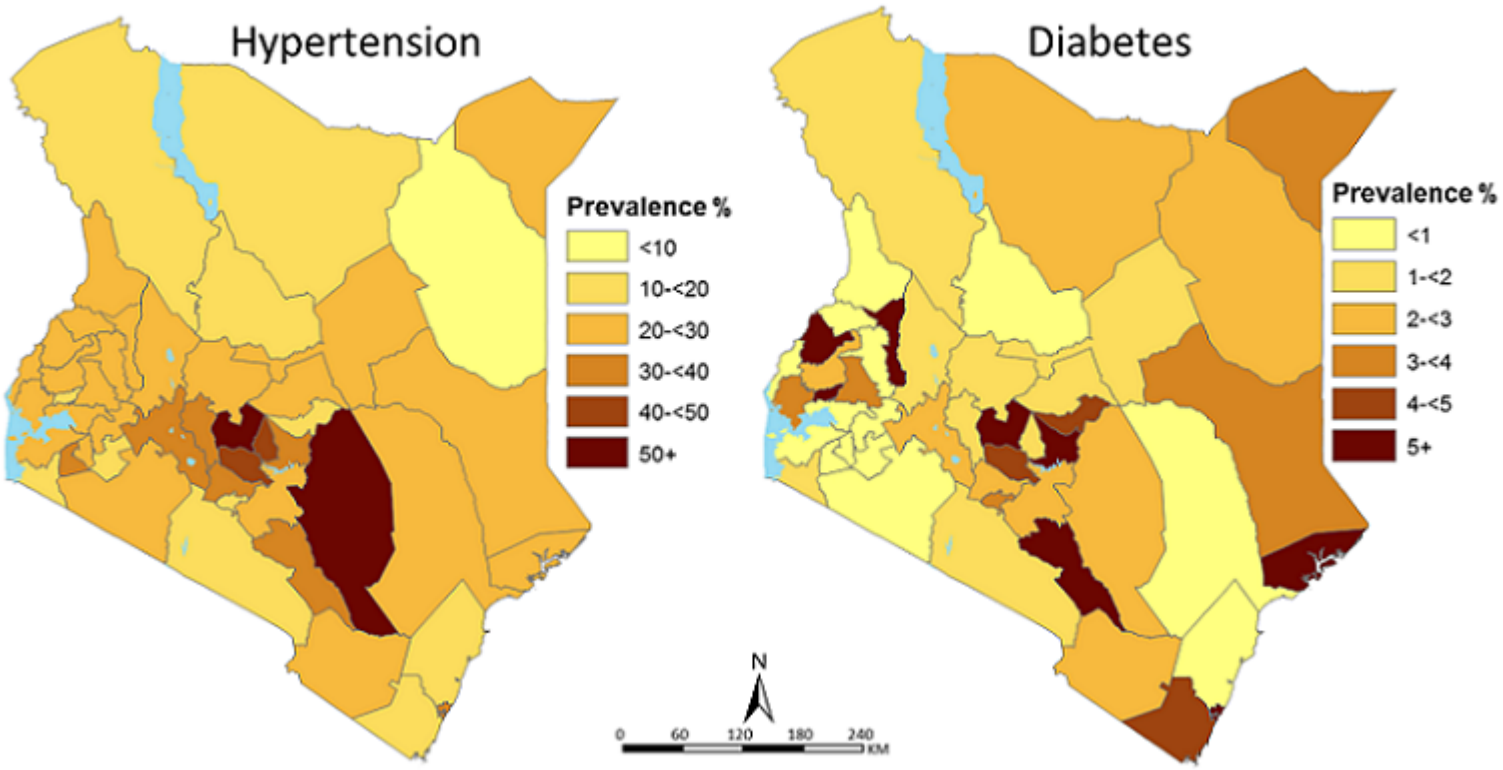
Table 4. Latent effects specification selection

Model	DIC	WAIC
Independent improper CAR	1956.48	1619.53
Independent proper CAR	1695.63	1527.03
Improper multivariate CAR	1264.39	1243.2
Proper multivariate CAR	1263.52	1242.01
M-model	1260.2	1240.65

The M-Model was therefore, used to estimate the county-specific prevalence shown in Figure 2.

There were substantial geographical variation in the prevalence of hypertension ranging from 9 % in Wajir county and 54 % in Nyeri county while diabetes ranged from 0.1 % in Narok to 8.1 % in Makueni. In general, hypertension and diabetes prevalence were comparatively higher in Central, Eastern and parts of western and coastal Kenya. Mombasa, Nakuru, Nyeri, Embu, Kirinyaga, Murang'a, Kiambu, Kitui, Makueni and Kisii had the highest (>30 %) hypertension prevalence. Lamu, Mombasa, Nyeri, Embu, Murang'a, Kwale, Makueni, Vihiga, Elgeyo-marakwet and Bungoma had high diabetes burden (>5 %). Overall, 47 percent (22 counties) and 36 % (17 counties) had hypertension and diabetes prevalence estimates that were above the national burden.The prevalence of hypertension and diabetes estimates was positive and significant 0.326 (p value=0.03).

Figure 2. County-level prevalence of hypertension and diabetes



4.5 Macro-risk factors of hypertension and diabetes

Table 5, shows results from a multivariate multinomial mixed effects regression model. High cholesterol, central obesity, age, BMI, harmful alcohol intake and high sugar intake were significantly associated with both hypertension and diabetes. People with high cholesterol level and central obesity were 2.1 and 3.5 times more likely to have both hypertension and diabetes respectively. Similarly, harmful alcohol intake and high sugar intake increased the odds of having hypertension and diabetes, 1.83[95 % CI; 1.18-2.85] and 1.17[95 % CI; 1.15-2.57], respectively. Older people are more likely to have hypertension and diabetes than younger people with odds ratio; 9.85[95 % CI; 5.04-19.24] and 18.5[95 % CI; 8.88-38.40] for people aged 30-44, 45-59 and 60-69 years compared to those in 18-29 years old, respectively. BMI measures of 25+ was significantly associated to hypertension and diabetes. Insufficient physical inactivity

and bad fat intake increased the odds of having hypertension and diabetes, however these effects were not significant.

Table 5. Adjusted odds ratios of macro-risk factors of hypertension and diabetes

Determinants	Hypertension only	Diabetes only	Hypertension & diabetes
High cholesterol	1.75(1.36-2.25) ***	1.50(0.87-2.59)	2.09(1.32-3.29) ***
Central obesity	1.40(1.14-1.72) ***	1.40(0.89-2.23)	3.53(2.13-5.86) ***
Age			
18-29	Ref	Ref	Ref
30-44	1.55(1.27-1.90) ***	1.44(0.94-2.22)	1.98(0.97-4.02) *
45-59	2.92(2.32-3.67) ***	1.62(0.96-2.74)	9.85(5.04-19.24) ***
60-69	6.38(4.76-8.56) ***	3.30(1.76-6.20) ***	18.5(8.88-38.40) ***
Harmful alcohol intake	1.29(1.06-1.57) **	1.35(0.88-2.08)	1.83(1.18-2.85) ***
High sugar intake	1.13(0.94-1.37)	1.14(0.72-1.74)	1.17(1.15-2.57) ***
BMI			
<18.5	Ref	Ref	Ref
18.5-24.9	1.30(0.98-1.73) *	0.64(0.38-1.07)	1.12(0.45-2.77)
25+	2.15(1.54-3.00) ***	0.83(0.44-1.58)	2.80(1.07-7.34) **
Insufficient physical inactivity	0.95(0.78-1.17)	0.95(0.61-1.50)	1.37(0.89-2.10)
Bad fat intake	1.15(0.97-1.36)	0.64(0.45-0.92)	1.13(0.77-1.67)
Intercept	0.11(0.08-0.15) ***	0.05(0.03-0.09) ***	0.03(0.001-0.007) ***

5 Discussion and Conclusion

5.1 Discussion

In kenya, there is need for county specific health statistics to inform optimal allocation of resources aimed at prevention and control of diseases. In this study, a bivariate spatial modelling approach is used to determine county units prevalences of hypertension and diabetes, which is important in determining similarities and divergence in the patterns of the diseases as well as understanding their association. Several multivariate conditional autoregressive models with varied latent effects were fit to overcome the challenge of small samples at county units using the Steps survey data. The M-model was the best fit with the lowest DIC and the correlation of hypertension and diabetes was found to be positive and significant. These findings are in line with others studies that indicate a correlation between hypertension and diabetes [Lago et al., 2007] Smoothed prevalence estimates producesd using the M-model, indicate substantial spatial variation of the disease burden. Counties in Central, Eastern and parts of Western and Coastal Kenya had relatively high burden. Specifically, Mombasa, Kiambu, Embu and Nyeri had high burden of both hypertension and diabetes. Correspondingly, high burden reflects population distribution and lifestyles that characterize these regions with most of their populance living in urban settings. Overall, 47 % (22 counties) and 36 % (17 counties) had hypertension and diabetes prevalence estimates that were above the national burden, 26.4 % and 2.7 % respectively. Urgent and precision targeting in the identified counties is essential to

counter the rising burden of NCDs in the country. High cholesterol, central obesity, age, harmful alcohol intake, high sugar intake and BMI of 25+ were identified as significant macro-risk factors of both hypertension and diabetes. Similar findings were reported by [Mohamed et al., 2018b, Wekesah et al., 2018] Wekesah and colleagues as well as Mohamed et al indicating the presence of multiple NCDs risk factors in Kenya. Health benefits resulting from the reduction of risk factors through public awareness of healthy lifestyles, particularly, addressing poor dietry habits are cost-effective in tackling hypertension and diabetes burden in Kenya. In addition, advocacy for co-testing and treatment of diabetes and hypertension especially for older age groups (45+ years) could have huge effects in early detection of comorbities and prevention of related health complications.

5.2 Conclusion

This study provides the first evaluation of county-level hypertension and diabetes burden in Kenya, against which future analysis can be monitored. Sub-national NCDs data sources are essential in improving the surveillance of NCDs in the country. An understanding of NCDs profiles matched with appropriate interventions including implementation of comprehensive screening programs and awareness building for NCDs control through mass media campaigns are key in reducing hypertension and diabetes burden in Kenya.

Bibliography

[Adeloye and Basquill, 2014] Adeloye, D. and Basquill, C. (2014). Estimating the prevalence and awareness rates of hypertension in africa: a systematic analysis. *PloS one*, 9(8).

[Agyemang et al., 2016] Agyemang, C., Meeks, K., Beune, E., Owusu-Dabo, E., Mockenhaupt, F. P., Addo, J., de Graft Aikins, A., Bahendeka, S., Danquah, I., and Schulze, M. B. (2016). Obesity and type 2 diabetes in sub-saharan africans–is the burden in today's africa similar to african migrants in europe? the rodam study. *BMC medicine*, 14(1):166.

[Agyepong et al., 2017] Agyepong, I. A., Sewankambo, N., Binagwaho, A., Coll-Seck, A. M., Corrah, T., Ezeh, A., Fekadu, A., Kilonzo, N., Lamptey, P., and Masiye, F. (2017). The path to longer and healthier lives for all africans by 2030: the lancet commission on the future of health in sub-saharan africa. *The Lancet*, 390(10114):2803–2859.

[Besag, 1974] Besag, J. (1974). Spatial interaction and the statistical analysis of lattice systems. *Journal of the Royal Statistical Society: Series B (Methodological)*, 36(2):192–225.

[Best et al., 2005] Best, N., Richardson, S., and Thomson, A. (2005). A comparison of bayesian spatial models for disease mapping. *Statistical methods in medical research*, 14(1):35–59.

[Boutayeb, 2006] Boutayeb, A. (2006). The double burden of communicable and non-communicable diseases in developing countries. *Transactions of the Royal society of Tropical Medicine and Hygiene*, 100(3):191–199.

[Chege, 2016] Chege, P. (2016). Multiple cardiovascular disease risk factors in rural kenya: evidence from a health and demographic surveillance system using the who step-wise approach to chronic disease risk factor surveillance. *South African Family Practice*, 58(2):54–61.

[Clayton and Kaldor, 1987] Clayton, D. and Kaldor, J. (1987). Empirical bayes estimates of age-standardized relative risks for use in disease mapping. *Biometrics*, pages 671–681.

[Gangnon and Clayton, 2003] Gangnon, R. E. and Clayton, M. K. (2003). A hierarchical model for spatially clustered disease rates. *Statistics in Medicine*, 22(20):3213–3228.

[Habib and Saha, 2010] Habib, S. H. and Saha, S. (2010). Burden of non-communicable disease: global overview. *Diabetes Metabolic Syndrome: Clinical Research Reviews*, 4(1):41–47.

[Held et al., 2005] Held, L., Natário, I., Fenton, S. E., Rue, H., and Becker, N. (2005). Towards joint disease mapping. *Statistical methods in medical research*, 14(1):61–82.

[Hunter and Reddy, 2013] Hunter, D. J. and Reddy, K. S. (2013). Noncommunicable diseases. *New England Journal of Medicine*, 369(14):1336–1343.

[Jin et al., 2007] Jin, X., Banerjee, S., and Carlin, B. P. (2007). Order[U+2010]free co[U+2010]regionalized areal data models with application to multiple[U+2010]disease mapping. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 69(5):817–838.

[Jin et al., 2005] Jin, X., Carlin, B. P., and Banerjee, S. (2005). Generalized hierarchical multivariate car models for areal data. *Biometrics*, 61(4):950–961.

[KNBS, 2019] KNBS (2019). 2019 kenya population and housing census. volume iii: Distribution of population by age and sex. Report.

[KPMG, 2013] KPMG (2013). Devolution of healthcare services in kenya: Lessons learnt form other countries.

[Lago et al., 2007] Lago, R. M., Singh, P. P., and Nesto, R. W. (2007). Diabetes and hypertension. *Nature clinical practice Endocrinology metabolism*, 3(10):667–667.

[Lawson and Lee, 2017] Lawson, A. and Lee, D. (2017). *Bayesian disease mapping for public health*, volume 36, pages 443–481. Elsevier.

[MacNab, 2010] MacNab, Y. C. (2010). On bayesian shared component disease mapping and ecological regression with errors in covariates. *Statistics in medicine*, 29(11):1239–1249.

[Mardia, 1988] Mardia, K. (1988). Multi-dimensional multivariate gaussian markov random fields with application to image processing. *Journal of Multivariate Analysis*, 24(2):265–284.

[Martinez-Beneito, 2013] Martinez-Beneito, M. A. (2013). A general modelling framework for multivariate disease mapping. *Biometrika*, 100(3):539–553.

[Martino and Rue, 2009] Martino, S. and Rue, H. (2009). Implementing approximate bayesian inference using integrated nested laplace approximation: A manual for the inla program. *Department of Mathematical Sciences, NTNU, Norway*.

[Mathers and Loncar, 2006] Mathers, C. D. and Loncar, D. (2006). Projections of global mortality and burden of disease from 2002 to 2030. *Plos med*, 3(11):e442.

[MoH, 2014] MoH (2014). Kenya health sector strategic and investment plan (khssp 2014-2018). Report.

[MoH, 2015] MoH (2015). Kenya national strategy for the prevention and control of ncds 2015-2020. Report.

[MOH, 2015] MOH (2015). Who. kenya step wise survey for non communicable diseases risk factors 2015 report. Report.

[Mohamed et al., 2018a] Mohamed, S. F., Mutua, M. K., Wamai, R., Wekesah, F., Haregu, T., Juma, P., Nyanjau, L., Kyobutungi, C., and Ogola, E. (2018a). Prevalence, awareness, treatment and control of hypertension and their determinants: results from a national survey in kenya. *BMC Public Health*, 18(3):1–10.

[Mohamed et al., 2018b] Mohamed, S. F., Mwangi, M., Mutua, M. K., Kibachio, J., Hussein, A., Ndegwa, Z., Owondo, S., Asiki, G., and Kyobutungi, C. (2018b). Prevalence and factors associated with pre-diabetes and diabetes mellitus in kenya: results from a national survey. *BMC public health*, 18(3):1215.

[Motala and Ramaiya,] Motala, A. and Ramaiya, K. Diabetes: the hidden pandemic and its impact on sub-saharan africa. In *Diabetes leadership forum*.

[Mufunda et al., 2006] Mufunda, J., Chatora, R., Ndambakuwa, Y., Nyarango, P., Kosia, A., Chifamba, J., Filipe, A., Usman, A., and Sparks, V. (2006). Emerging non-communicable disease epidemic in africa: preventive measures from the who regional office for africa. *Ethn Dis*, 16(2):521–526.

[Ogurtsova et al., 2017] Ogurtsova, K., da Rocha Fernandes, J., Huang, Y., Linnenkamp, U., Guariguata, L., Cho, N. H., Cavan, D., Shaw, J., and Makaroff, L. (2017). Idf diabetes

atlas: Global estimates for the prevalence of diabetes for 2015 and 2040. *Diabetes research and clinical practice*, 128:40–50.

[Richards et al., 2016] Richards, N. C., Gouda, H. N., Durham, J., Rampatige, R., Rodney, A., and Whittaker, M. (2016). Disability, noncommunicable disease and health information. *Bulletin of the World Health Organization*, 94(3):230.

[Saei and Chambers, 2003] Saei, A. and Chambers, R. (2003). Small area estimation: A review of methods based on the application of mixed models.

[Torabi and Rao, 2008] Torabi, M. and Rao, J. (2008). Small area estimation under a two-level model. *Survey Methodology*, 34(1):11.

[Torp-Pedersen and Jeppesen, 2011] Torp-Pedersen, C. and Jeppesen, J. (2011). Diabetes and hypertension and atherosclerotic cardiovascular disease: related or separate entities often found together. *Hypertension*, 57(5):887–888.

[Wekesah et al., 2018] Wekesah, F. M., Nyanjau, L., Kibachio, J., Mutua, M. K., Mohamed, S. F., Grobbee, D. E., Klipstein-Grobusch, K., Ngaruiya, C., Haregu, T. N., and Asiki, G. (2018). Individual and household level factors associated with presence of multiple non-communicable disease risk factors in kenyan adults. *BMC public health*, 18(3):1220.

[WHO, 2013] WHO (2013). Who global action plan for the prevention and control of ncds 2013-2020.

[WHO, 2015] WHO (2015). A global brief on hypertension. 2013. *Geneva, Switzerland: World Health Organization*, pages 7–15.

[WHO, 2018] WHO (2018). Noncommunicable diseases: Key facts. *World Health Organization website. Retrieved from http://www. who. int/news-room/fact-sheets/detail/noncommunicable-diseases*.

[Zhou et al., 2017] Zhou, B., Bentham, J., Di Cesare, M., Bixby, H., Danaei, G., Cowan, M. J., Paciorek, C. J., Singh, G., Hajifathalian, K., and Bennett, J. E. (2017). Worldwide trends in blood pressure from 1975 to 2015: a pooled analysis of 1479 population-based measurement studies with 19· 1 million participants. *The Lancet*, 389(10064):37–55.

[Zhu and Carlin, 2000] Zhu, L. and Carlin, B. P. (2000). Comparing hierarchical models for spatio[U+2010]temporally misaligned data using the deviance information criterion. *Statistics in Medicine*, 19(17[U+2010]18):2265–2278.