*South African Statistical Journal*
Vol. 56, No. 1, pp. 1–19
https://doi.org/10.37920/sasj.2022.56.1.1

# Estimating county level overweight prevalence in Kenya using small area methodology

*Noah Cheruiyot Mutai*

Institute of Statistics and Econometrics, Freie Universität Berlin, Berlin, Germany
Department of Mathematics, Statistics and Physical Sciences, Taita Taveta University, Voi, Kenya

Public health surveillance of overweight prevalence is essential to assess the extent of the problem, identify regions and groups most affected and inform policy-making. However, the needed reliable data at disaggregated levels is lacking in Kenya. The Kenya STEPwise Survey for Non-communicable Diseases and Risk Factors (KSSNDRF) was nationally representative. It was used to obtain various indicators of non-communicable diseases and risk factors including overweight. However, due to small sample sizes at lower levels like at the county, overweight prevalence estimates are statistically imprecise (i.e., high variance). Therefore, to increase the effective sample size we combine data from the KSSNDRF and the Kenya Population and Housing Census by model-based small area methods. In particular, we fit an arcsine square-root transformed Fay–Herriot model. To transform back to the original scale, we use a bias-corrected back transformation. For this model, we smooth the design variance using Generalised Variance Functions. We compute the mean squared error estimates using a bootstrap procedure. We found that counties within urban areas — including the major towns like Nairobi, Nakuru, Nyeri and Mombasa — have a higher prevalence of overweight compared to rural counties. Although the paper focuses on overweight prevalence in Kenya, the presented method can also be applied to other indicators in developing countries with similar data sources.

*Keywords:* Direct estimation, Fay-Herriot model, Prevalence mapping, Sample surveys, Transformations.

## 1. Introduction

Globally, the prevalence of overweight and obesity has increased more than three times between 1975 and 2016 (World Health Organization, 2020). Prevalence is the proportion of subjects with a specific characteristic in a population — in this case, the proportion of persons who are overweight. In 2016 the World Health Organization (WHO) estimated that 1.9 billion and 650 million adults were overweight and obese, respectively (World Health Organization, 2020). The WHO defines overweight and obesity as abnormal or excess fat accumulation that present risk to human health. The Body Mass Index (BMI) is the basic and commonly used measure. It is a simple index used to classify overweight (BMI > 25) and obesity (BMI > 30) for adults. The BMI is a ratio of a person's weight in kilograms

*Corresponding author:* Noah Cheruiyot Mutai (ncheruiyot3@gmail.com)

*MSC2020 subject classifications:* 62D05, 62F10.

to the square of the height in meters (kg/m$^2$). According to the World Health Organization (2020), overweight is associated with increased risk for other non-communicable diseases (NCDs) such as type-2 diabetes and hypertension. Worldwide, the prevalence of overweight and obesity is higher for women than men (World Health Organization, 2020). A number of studies on maternal overweight such as Sebire et al. (2001), Kulie et al. (2011), Chowdhury et al. (2016) and Mkuu et al. (2018) have found that maternal overweight can affect both the mother and the unborn child. It can lead to higher rates of miscarriage, still-births and congenital anomalies. During pregnancy, overweight can later affect the health of the mother and child, including increased risk of heart disease, hypertension and diabetes.

Over the last decades, health challenges in low-income and middle-income countries have revolved mainly on communicable diseases and under-nutrition (Pawloski et al., 2012). Sub-Saharan Africa harbour a large proportion of communicable diseases such as Malaria, HIV/AIDS and Tuberculosis (TB). However, due to urbanisation and better incomes, a nutritional transition from health patterns associated with communicable diseases to health patterns associated with over-nutrition has occurred (Pawloski et al., 2012; Jones-Smith et al., 2012; Awuah et al., 2014; Steyn and Mchiza, 2014; Agyemang et al., 2016). Much attention and funding have gone into combating communicable diseases. With emerging NCDs coexisting with communicable diseases, this presents more challenges. For countries in sub-Saharan Africa, overweight and obesity present a tough challenge because persons who grow with under-nutrition, are prone to adding up more weight as they grow up. This is defined by WHO as malnutrition and is characterised by the coexistence of under-nutrition with overweight and obesity within individuals and households for a lifetime (World Health Organization, 2020).

In Kenya, more people move to towns and urban areas in search of jobs. This has the potential of improving their living standards from better income earned. However, a lot of time is spent working causing reduced physical activity. They also have access to high-calorie fast foods within urban settings. Due to this, among other factors, problems of increased body weight is on the rise (Kenya National Bureau of Statistics, 2015). The Ministry of Health Kenya (MOHK) notes that, in addition to existing communicable diseases, this causes a double burden of disease in morbidity, mortality and medical expenses (Kenya National Bureau of Statistics, 2015). NCDs are a major public health concern with significant social and economic effects in terms of health care needs, loss in productivity, and premature death. They are a great setback to attaining the Sustainable Development Goals (SDGs) of the United Nations (UN) (General Assembly, 2015) if appropriate interventions are not implemented. Mkuu et al. (2018), using Kenya Demographic and Health Survey (KDHS) of 2014, found that 20.5% of the Kenyan women are overweight and 9.1% are obese. A study by Muthuri et al. (2014) on Kenyan school-going children established that out of 563 children, aged 9 to 11 years, 3.7% were underweight, 14.4% were overweight, and 6.4% were obese. In a cross-sectional study with 365 women aged 25 to 54 years in Nairobi, Kenya, Mbochi et al. (2012) showed that BMI increased with age, greater socio-economic group, increased expenditure, increased parity and more number of living rooms.

The Kenyan government is committed to improving the overall health of its citizens. To do this, data at disaggregated levels are required. However, this is lacking. Especially to ascertain the extent of the problem and identify the most affected groups and regions. Reliable data will also help to inform policy-making. The government of Kenya has come up with some policies and strategic plans such as the Kenya Health Policy (KHP), 2014–2030. The KHP outlines how Kenya seeks to improve the public health status in line with the Kenyan Constitution, Vision 2030 and SDGs (Kenya National Bureau of Statistics, 2015). Specifically, this policy was developed to respond to local and global development efforts to attain MDGs. It also targets NCDs, social determinants of health and the management of emerging and re-emerging health threats. Another strategy is the Kenya National Strategy for the Prevention and Control of Non-communicable Diseases, 2015–2020. The main objective is to reduce the preventable burden, avoidable death, sickness, risk factors and cost due to NCDs. To fulfil these goals, the Kenya National Bureau of Statistics (KNBS) carried out the inaugural survey on NCDs in 2015. This was a national cross-sectional household survey. It was designed to estimate indicators on risk factors for NCDs for persons aged 18 to 69 years at the national level. According to the survey, the common and important risk factors for NCDs are daily smoking, overweight or obesity, elevated blood pressure, low physical activity and a minimum of 400g of fruit and vegetables per day. Additionally, 28% of those sampled were either overweight or obese. Women (38%) were either overweight or obese as compared to 18% of men (Ministry of Health Kenya, 2014).

The KSSNDRF 2015 was a national survey. It was designed to provide reliable (design-based) estimates at the national level only. The design-based estimators (they rely only on the survey data) are approximately designed unbiased and consistent. However, direct estimators generally have large variances and estimates are unreliable when the sample sizes are small — for example at the county level in Kenya. In contrast, model-based small area methods produce more reliable estimates in terms of smaller MSE and coefficient of variation (CV) (Tzavidis et al., 2018). This is because they combine survey and census/administrative data through a model and therefore increase the effective sample size. For more overviews on small area estimation (SAE), we refer the reader to (Rao and Molina, 2015; Pfeffermann, 2013).

For this study, therefore, we rely on SAE to better estimate the prevalence of overweight at the county level. To the best of our knowledge, this is the first study to use SAE and estimate the prevalence of overweight in Kenya. Our main data source is KSSNDRF 2015. The prevalence estimates of overweight obtained from survey data only are called direct estimates hereafter in this paper. Initial analysis shows the coefficient of variation for the direct estimates reaches high values given the small sample sizes at the county level. We use an area level model proposed by Fay and Herriot (1979). Since the proportion of persons who are overweight in a particular county must lie between [0,1], we transform the dependent variable with the arcsine square root transformation. This non-linear transformation has been previously applied by Valencia et al. (2016) to estimate poverty in Chile, Schmid et al. (2017) to estimate literacy in Senegal and by Hadam et al. (2020) to estimate regional unemployment in Germany. The estimates obtained are on a transformed scale. To make valid inferences we need to transform back to the original scale. Since bias is introduced due to transformation we use a bias-corrected back transformation. This is similar to the one used by Hadam et al. (2020). To assess the accuracy of our estimates we compute the (MSE) based on a parametric bootstrap that incorporates the additional uncertainty due to the bias-correction.

The rest of this paper is organised as follows. We describe the KSSNDRF 2015 and the Kenya Population and Housing Census (KPHC) 2009 in Section 2. In Section 3, we outline the small area methodology applied in this paper. In particular, the Fay–Herriot model, hereafter called the FH model, transformation, back transformation and MSE estimation. In Section 4, we present the results of the application to estimate the prevalence of overweight in Kenya including model selection.

Lastly, in Section 5, we give the concluding remarks, possibilities for further research and limitation of this study.

## 2. Data sources: survey and census data

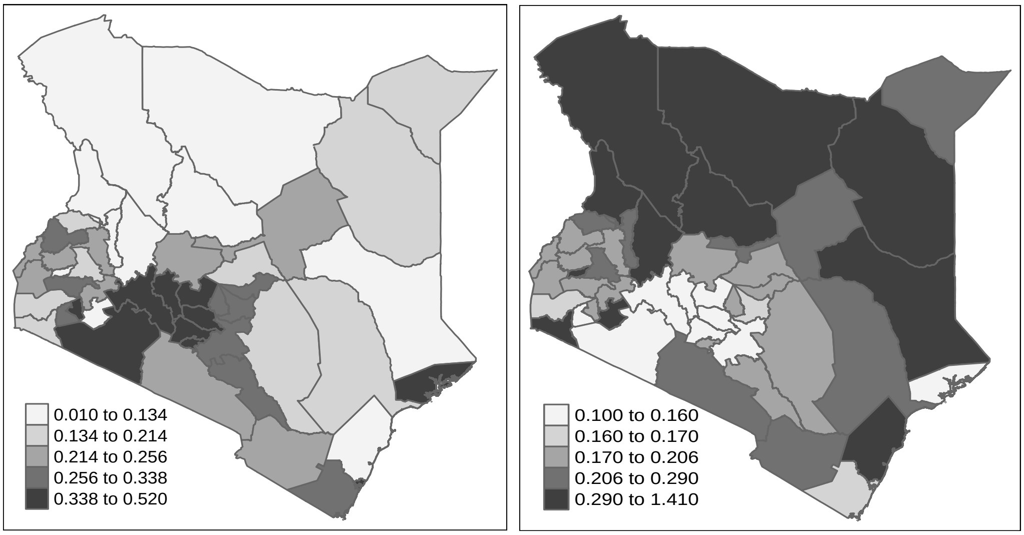
In this section, we describe the data sources used in this paper. We used the KSSNDRF 2015 and the KPHC 2009. The two datasets were provided by the KNBS under the Kenya National Data Archive (KeNADA) as public use files. Since the survey and census data were collected at different years, we assume the functional relation between overweight and auxiliary data remains constant.

### 2.1 Kenya STEPwise Survey for Non-communicable Diseases Risk Factors (KSSNDRF) 2015

The KSSNDRF 2015 adopted the WHO STEPwise approach to Surveillance (STEPS). This approach is a simple, flexible and standardised method for collecting, analysing and disseminating data in countries that are members of WHO. Until 2016, 122 WHO member countries had completed data collection on STEPS surveys (Riley et al., 2016). The WHO uses a tool called the STEPS Instrument to collect and measure NCDs risk factors. The tool covers three different NCDs risk factor assessments, i.e., (i) a questionnaire (ii) physical measurements and (iii) biochemical measurements. The questionnaire gathers data on socio-demographic information, aspects of an individual's medical history related to the main NCDs, and risk behaviours. Physical measurements assess overweight and obesity and increased blood pressure while the biochemical measurements include blood and urine sampling to measure raised blood glucose, cholesterol and lipids (World Health Organization, 2005).

The STEPS Instrument allows each country to adapt and expand on the main variables and risk factors. Kenya adopted the STEPS approach in a sequential process consisting of three steps of information gathering. First, data on demographic and behaviour were collected. Demographic data included questions on age, sex, marital status, education and occupation. It also included questions on housing and social amenities. Questions on behaviour included tobacco use, alcohol consumption, diet, physical activity, history of blood pressure and diabetes, history of cardiovascular diseases, injury and oral health. The second step involved physical measurements on blood pressure, heart rate, height, weight, waist and hip circumference. This is to assess overweight and obesity. The last step collected data on biochemical measurements on blood glucose and blood lipids (Kenya National Bureau of Statistics, 2015).

The KSSNDRF 2015 was a national cross-sectional household survey designed to estimate indicators on risk factors for NCDs for persons aged 18 to 69 years. A sample size of 6 000 individuals was designed to give reliable estimates on the national level by sex (male and female) and residence (rural and urban). The survey used the fifth National Sample Surveys and Evaluation Programme (NASSEP V) maintained by the KNBS. The NASSEP V is a sample frame used for household surveys in Kenya and contains 5 360 clusters split into four sub-samples. The KSSNDRF 2015 adopted a three-stage cluster sampling design which involves the selection of clusters, households and eligible individuals. First, 200 clusters (100 urban and 100 rural) were selected from one subsample of the NASSEP V sample frame. Secondly, a uniform sample of 30 households from the listed households in each cluster was selected. The last step involved randomly selecting one individual from all eligible household members using a programmed Kish selection method of sampling (Kish, 1949). iPAQ



(a) (b)

**Figure 1.** Maps showing: (a) Direct point estimates of overweight prevalence and (b) the corresponding coefficient of variation based on KSSNDRF only.

personal computer and personal digital assistants (PDAs) were used at this stage. Each interviewer was provided with an iPAQ together with its accessories and an extra battery. The PDAs automatically saved the data in their internal memory and also in a Secure Digital Card (SD card) (Kenya National Bureau of Statistics, 2015).

Currently, the KNBS officially reports the prevalence of overweight only on a national level where the survey is reliable. Apart from the KSSNDRF 2015, the other survey that collects and reports data on health related characteristics is the Kenya Demographic and Health Survey (KDHS). Kenya has conducted the KDHS in 1989, 1993, 1998, 2003, and 2008–09. Up to 2014, the previous KDHS has collected data on health characteristics in Kenya except for data on NCDs. The KDHS 2014 was also the first national survey to provide estimates for demographic and health indicators at the county level. However, the KDHS 2014 collected BMI for only women aged 15 to 49 years old. We selected the KSSNDRF 2015 since it collected BMI data on men and women — unrestricted to a particular age group.

Figure 1 presents direct estimates of overweight prevalence and coefficient of variation based on KSSNDRF 2015. Kenya has 47 counties which is the second administrative level after the national. For this survey, all 47 counties were sampled. A total of 4 500 individuals were successfully interviewed at the primary stage sampling giving a response rate of 95%. We had access to a total sample size of 4 288, of which 4 014 are complete cases.

Table 1 shows summary statistics of sample sizes, direct estimates and the corresponding CVs of overweight over counties. The minimum and maximum CV are 10% and 141% respectively. Currently, there is no internationally accepted cutoff point for CVs to report official statistics. Further, Kenya hasn't set a threshold based on CVs for reporting official statistics. Therefore, we follow the guidelines of other statistics offices; for instance, the Office for National Statistics (ONS) in the UK

**Table 1.** Summary statistics of sample sizes, overweight point estimates and respective coefficient of variation over the 47 counties in Kenya.

| | **Min.** | **Q1** | **Median** | **Mean** | **Q3** | **Max.** |
|---|---|---|---|---|---|---|
| Sample size | 53 | 75 | 84 | 85 | 95 | 152 |
| Direct estimates | 0.0090 | 0.1601 | 0.2378 | 0.2447 | 0.3224 | 0.5199 |
| CV | 0.1003 | 0.1611 | 0.1918 | 0.2409 | 0.2818 | 1.4121 |

uses a CV of 20% as a threshold for publishing official results. Based on this, 23 domains out of 47 have CVs greater than this threshold.

### 2.2 The Kenya Population and Housing Census 2009

The first comprehensive census in Kenya was done in 1948. The next was in 1962 with 8.6 million people. The census helped in setting up political and administrative structures. After independence in 1969, a third census was conducted with 10.6 million people. Since then, Kenya has continuously conducted a census after every 10 years, i.e., 1969, 1979, and so on, the most recent being 2019. The meta-data for 2019 has not been released for public use. The KNBS under the Statistics Act 2006 of Kenyan law is the main government agency responsible for collecting, analysing and disseminating census and other statistical data.

Census is a large statistical undertaking and requires huge finances, planning and personnel. The implementation of the 2009 KPHC for the Republic of Kenya (RoK) used an estimated 8.4 billion Kenyan shillings (appr. 9 million US dollars) (Government of Kenya, 2010). This huge investment is justified as it is a key exercise for the government of Kenya and interested stakeholders. The statistical information is required for monitoring the implementation of various development objectives and global initiatives, e.g., the United Nations Millennium Development Goals (UNMDGs). It serves as a basis for adequate policy planning. Data on fertility and mortality are important in dispatching services related to births and deaths. The country's growth rate can also be accessed. To provide social amenities to different age groups, the census provides data on the composition of a country's population by age. Minority and age groups who require special amenities are identified through this data. To determine tax relief, data on the dependency ratio is needed. Persons are born in different places and move from one place to another. Data on migration is important to understand migration trends and required interventions.

For this study, we had access to the 2009 KPHC. This was the 5th census after independence. It was done from the night of 24 and 25 to 31 August 2009. The main objective was to provide key information on the demographic, social and economic characteristics of the population and housing. These include size and composition of the population, fertility, mortality and migration rates, levels of education, size of labour force, etc. In this census, data was captured through scanning technology with technical assistance provided by the United States Census Bureau (USCB) (Government of Kenya, 2010). This census was based on old administrative areas, i.e., villages, sub-locations, locations, divisions, districts and provinces. There were 46 legal districts in Kenya excluding Nairobi, the capital city, which constituted the 47th district. These districts were converted to the current 47 counties without change of boundaries after 2010 (Government of Kenya, 2013).

Therefore, we can link the survey and census data. The variables in the census serve as potential covariates in the small area model introduced in Section 3 for predicting overweight in Kenya. At this point, we state that the response variable, overweight, is unreported in the census.

## 3. Small area estimation methodology

In this section, we outline the SAE method. First, in Section 3.1 we describe the standard FH model. Since the FH model does not guarantee that the prevalence of overweight lies in the interval $[0, 1]$, we describe an arcsine square root transformed FH model in Section 3.2. For the standard FH and arcsine square root transformed FH models, we explain the estimation of regression parameters, sampling variances, random effects and the MSE.

### 3.1 The Fay–Herriot model

We assume that a finite population of size $N$ which is divided into $m$ disjoint areas of sizes $N_1, N_2, ..., N_m$, where $i = 1, 2, ..., m$ is the $i$th small area. A sample of size $n$ is taken from this population using a complex sampling design with sample sizes $n_1, n_2, ..., n_m$ for each area $i$. Further, we assume the response variable $y_{ij}$ of individual $j$ in area $i$ has been measured without error in the survey. In this paper, we are interested in estimating the mean prevalence of overweight in Kenya with reduced uncertainty by incorporating extra covariates from census data. We consider the area level FH model (Fay and Herriot, 1979), where the direct estimator of the population mean is

$$\hat{\bar{y}}_i^{\text{dir}} = \frac{1}{n_i} \sum_{j=1}^{n_i} w_{ij}\, y_{ij}, \quad i = 1, 2, ..., m, \tag{1}$$

where $\hat{\bar{y}}_i^{\text{dir}}$ is the direct mean estimator for area $i$ and $w_{ij}$ are sampling weights. The weights compensate for unequal probabilities of sampling and unit non-response. This is the Horvitz–Thompson (HT) estimator of Horvitz and Thompson (1952) for estimating population means and totals. A big advantage of the FH model is the ability to take into account the sampling design using the HT estimator (Särndal et al., 2003). The first stage of the FH model is a function of the direct estimator in (1) above and the sampling errors as

$$\hat{\bar{y}}_i^{\text{dir}} = \theta_i + \varepsilon_i, \tag{2}$$

where $\theta_i$ is the population mean and $\varepsilon_i$ is the sampling error assumed to be normally distributed and independent, i.e., $\varepsilon_i \sim N\left(0, \sigma^2_{\varepsilon_i}\right)$. In theory, the sampling errors $\varepsilon_i$ for the FH model are assumed known. However, in practice these need to be estimated. In Section 3.2, we outline how we estimated the sampling variance for this application. In the second stage, $\theta_i$ is linked to available area level covariates

$$\theta_i = x_i^{tr}\beta + v_i, \tag{3}$$

where $x_i$ are area level auxiliary variables, $\beta$ is vector of regression parameters and $v_i$ are area level random effects. The random effects are assumed to be independently normally distributed i.e. $v_i \sim N\left(0, \sigma^2_v\right)$. Combining the sampling model in (2) and linking model in (3), we obtain the area level model given by

$$\hat{\bar{y}}_i^{\text{dir}} = x_i^{tr}\beta + v_i + \varepsilon_i, \tag{4}$$

which is a linear mixed model with $E[v_i] = E[\varepsilon_i] = 0$. According to Rao and Molina (2015), $\hat{\beta}$ can be estimated as the best linear unbiased estimator (BLUE) of $\beta$ and the random effect $\hat{v}_i$ as the empirical best linear unbiased predictor (EBLUP) of $v_i$ (Henderson, 1975). The variance $\sigma_v^2$ can be estimated by the Maximum Likelihood Method (ML) or the Residual Maximum Likelihood Method (REML) (Hartley and Rao, 1967; Patterson and Thompson, 1971; Datta and Lahiri, 2000). Under this combined model, the EBLUP is obtained as

$$\hat{\bar{y}}_i^{\text{FH}} = x_i^{tr}\hat{\beta} + \hat{v}_i = \hat{\gamma}_i \hat{\bar{y}}_i^{\text{dir}} + (1 - \hat{\gamma}_i)\, x_i^{tr}\hat{\beta},$$

where $\hat{\gamma}_i$ is the shrinkage factor for area $i$ given by $\hat{\gamma}_i = (\sigma_v^2)/(\sigma_v^2 + \sigma_{\hat{\varepsilon}_i}^2)$.This EBLUP is a weighted combination of the direct estimator ($\hat{\bar{y}}_i^{\text{dir}}$) and the synthetic estimator ($x_i^{tr}\hat{\beta}$). In practical applications, many small areas have zero sample sizes and the direct estimator is unavailable, therefore we depend on the synthetic estimator (Rao and Molina, 2015). According to Pfeffermann (2013) and Rao and Molina (2015), when small area estimates are produced, they should be accompanied by a valid measure of precision. The mean squared error (MSE) is still the standard measure of uncertainty in official small area statistics. We therefore determine the accuracy of our EBLUP by calculating the MSE. As stated by Rao and Molina (2015), this MSE can be obtained based on the method used to estimate the variance of the random effect. Following Prasad and Rao (1990) and Datta and Lahiri (2000), an analytical MSE is obtained if the method chosen is ML or REML.

### 3.2 The arcsine square root transformed FH model

The prevalence of overweight is a proportion and must lie on the interval of $[0, 1]$. However, the FH model outlined above can give estimates outside this range. Secondly, the FH model is a linear mixed model(LMM) in which some assumptions are made. Specifically, normality, linearity and homoscedasticity of error variance. To meet the condition of $[0, 1]$ interval and assumptions for the LMM, we therefore transform the vector of direct estimators. As in Schmid et al. (2017) and Hadam et al. (2020) we use the arcsine square root transformation. This transformation also stabilises the variance (Carter and Rolph, 1974; Efron and Morris, 1975). In this case, we transform only the response variable, vector of direct estimators. Both-sides transformation for linearity can make the error term heteroscedastic (Carroll and Ruppert, 1988). Since we have chosen our model a prior, we assume it fits the data adequately. We first define the function $g(z) = \sin^{-1}(\sqrt{z})$. Equation (4) above becomes

$$\sin^{-1}\left(\sqrt{\hat{\bar{y}}_i^{\text{dir}}}\right) = x_i^{tr}\beta + v_i + \varepsilon_i, \tag{5}$$

where $\varepsilon_i \sim N\left(0, \tilde{\sigma}_{\varepsilon_i}^2\right)$ and $v_i \sim N\left(0, \sigma_v^2\right)$. As mentioned in Section 3.1, in theory the sampling variances are assumed known, but estimated in practice (Rao and Molina, 2015). In this study we estimate the sampling variance directly from the sample data. The sampling variances can be unstable especially for small sample sizes (Bell, 2008; Hawala and Lahiri, 2010). Generalised Variance Functions (GVF) have been used to smooth the sampling variance (Maples et al., 2009; Hawala and Lahiri, 2010; Pratesi, 2016; Hawala and Lahiri, 2018). In this paper we adopt a similar approach as used in Pratesi (2016). The variance smoothing model is given by $\hat{p}_i(1 - \hat{p}_i)/\text{var}(p_i) = \beta n_i + e_i$, where $e_i \sim D(0, \tau^2)$, $\hat{p}_i$ is the prevalence in area $i$, $n_i$ is the sample size in area $i$ and $\beta$ is a linear regression coefficient.

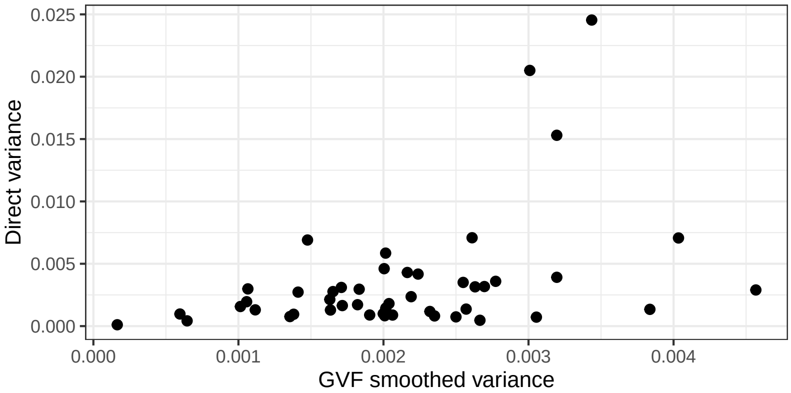


**Figure 2.** A scatter plot of direct and GVF smoothed sampling variance.

For the arcsine square root transformed model, we estimate the sampling variance as in Jiang et al. (2001), Schmid et al. (2017) and Hadam et al. (2020) by $\tilde{\sigma}^2_{\varepsilon_i} = (4n_i^*)^{-1}$, where $n_i^*$ is the effective sample size in area $i$. The effective sample size is estimated by $n_i^* = n/(\mathrm{def}\, f)$, where $n$ is sample size and def $f$ is the design effect. The design effect is the ratio of the variance of the direct estimator under simple random sampling to the variance under the complex sampling design of the survey (Särndal et al., 2003). Figure 2 show a plot of the direct variances against GVF smoothed variances.The graph show that the direct variances follow a similar pattern as the GVF smoothed variances. However, the later show a smooth behavior. From (5) the parameters $\beta$ and the random effects $v_i$ are estimated as in the standard FH model described in Section 3.1. The arcsine square root transformed FH model is obtained by replacing these parameters with their respective estimates, yielding

$$\hat{\bar{y}}_i^{\text{FH, trans}} = \hat{\gamma}_i \left( \sin^{-1} \sqrt{\hat{\bar{y}}_i^{\text{dir}}} \right) + (1 - \hat{\gamma}_i)\, x_i^{tr} \hat{\beta}. \tag{6}$$

With (6) we obtain the prevalence of overweight on a transformed scale. To obtain estimates on the original scale, we transform them back to the original scale. Mathematically, one would use $z = \sin^2(z)$. This kind of back transformation is naive as it introduces bias due to the non-linear transformation. Some studies that have used this transformation include Valencia et al. (2016) and Schmid et al. (2017). In this paper, we adopt a bias-corrected back transformation as proposed in Hadam et al. (2020). In their paper, if for the transformed FH-model the assumptions are fulfilled, the transformed FH estimator is normally distributed with $a \sim N(a,\, b)$, where $a = \hat{\bar{y}}_i^{\text{FH, trans}}$ and $b = \hat{\sigma}_v^2 \tilde{\sigma}^2_{\varepsilon_i} / (\hat{\sigma}_v^2 + \tilde{\sigma}^2_{\varepsilon_i})$. The bias-corrected back transformed FH estimator $\hat{\bar{y}}_i^{\text{FH, back}}$ is obtained by computing the expected value of the naive back transformation under the assumed normal distribution. This integral can be solved using numerical integration techniques. Through simulation studies, they show a reduction in bias due to this correction. To estimate the MSE of the bias-corrected back transformed FH estimates, we also adopt a parametric bootstrap method used in Hadam et al. (2020). They present a procedure for estimating confidence intervals and MSE for an arcsine square root transformed bias-corrected FH estimator. Through simulation studies, they show this bootstrap shows a good performance for MSE estimation and confidence intervals.

## 4. Application: estimating the prevalence of overweight in Kenya

In this section, we apply the small area method presented in Section 3.2 to estimate the prevalence of overweight in Kenya. We implement this in the R package emdi Kreutzmann et al. (2019). From this section hereafter, the estimates obtained from this methodology will be referred to as FH_trans or FH estimates or simply model-based estimates.

### 4.1 Model selection and diagnostics

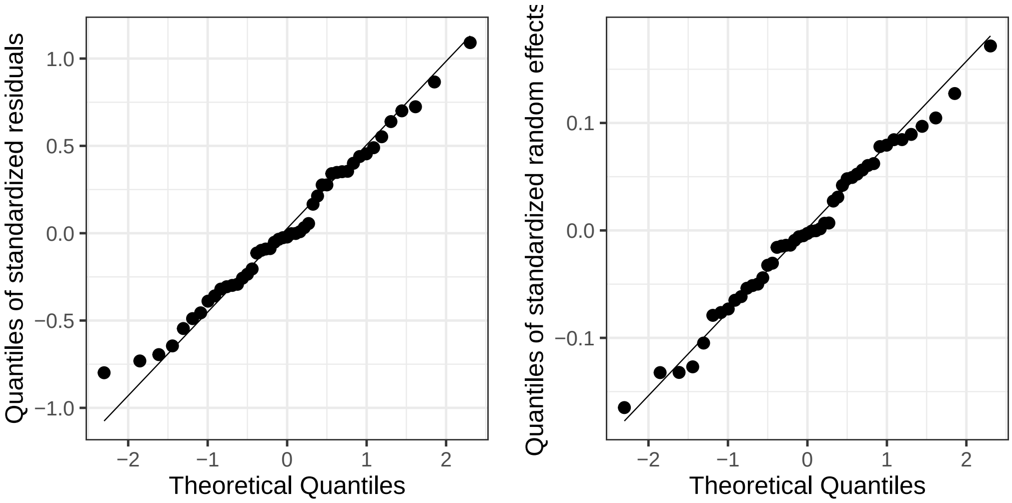
The model in Section 3.2 requires aggregated auxiliary data. For this study, we had access to census data. According to Rao and Molina (2015), a key requirement for the success of small area methods is the availability of good useful auxiliary data from census or administrative records. Some studies such as Schmid et al. (2017) and Hadam et al. (2020) have used auxiliary data from mobile phone data as an alternative. Based on Mkuu et al. (2018), Mbochi et al. (2012), and Asiki et al. (2018), we first selected likely predictors of overweight from census data. We then fit a full model with all the covariates. Lastly, we select predictive covariates using the Akaike Information Criterion (AIC) for the FH model. The STEPwise procedure we used involved forward and backward selection. The final model had an adjusted $R^2$ of 59%. The selected covariates are: age; gender with categories male and female; education with categories no formal schooling, completed primary school, secondary school and above; marital status with categories never married, married, widowed and divorced; employment with categories government employed, self-employed, unemployed, employed by NGO and others.

Table 2 shows the significant predictors of overweight at $\alpha = 0.05$. It also shows the corresponding standard errors, t-values and p-values. We note that age, marital status and employment have positive coefficients, while gender, level of education and household size have negative coefficients. Our results are in agreement with other studies on predictors of overweight. To mention a few, Groenveld-van Dijk (2013) found that gender, age, education, wealth and ethnicity are highly correlated with the prevalence of overweight and obesity in Kenya. A study by Mbochi (2010) showed that age, parity, socioeconomic status and physical activity are all significant predictors of overweight and obesity in Kenya. In a study of Kenyan schoolchildren, Muthuri et al. (2014) found that parent's education level, income, and type of school attended (either private or public) were positively associated with

**Table 2.** Significant predictors of overweight in Kenya and corresponding coefficients, standard errors, t-values and p-values.

| Parameter | Estimate | SE | t-value | p-value |
|---|---|---|---|---|
| Intercept | −0.2763 | 0.6716 | −0.4114 | 0.6808 |
| Age | 0.0372 | 0.0149 | 2.4981 | 0.0125 |
| Female | −2.3412 | 0.8679 | −2.6974 | 0.0070 |
| Married | 1.6591 | 0.5294 | 3.1339 | 0.0017 |
| Secondary | −0.2512 | 0.0967 | −2.5969 | 0.0094 |
| Unemployed | 6.9896 | 2.4552 | 2.8469 | 0.0044 |
| Household size | −0.0508 | 0.0230 | −2.2049 | 0.0275 |

The estimated random effects variance $\hat{\sigma}_v^2 = 0.00817$.



(a)

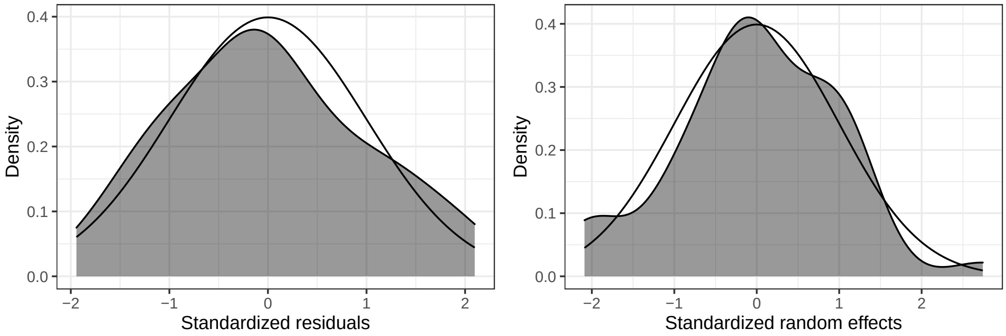
(b)

**Figure 3.** (a) Quantiles of standardised residuals and (b) Standardised random effects for the arcsine square root transformed model with GVF smoothed variance.

overweight/obesity. Mkuu et al. (2018) and Christensen et al. (2008) found urbanisation to be significant in predicting the prevalence of overweight and obesity in Kenya. Since physical inactivity has been shown to significantly predict overweight, Gichu et al. (2018) established that gender, age, education level and wealth index significantly predict physical inactivity.

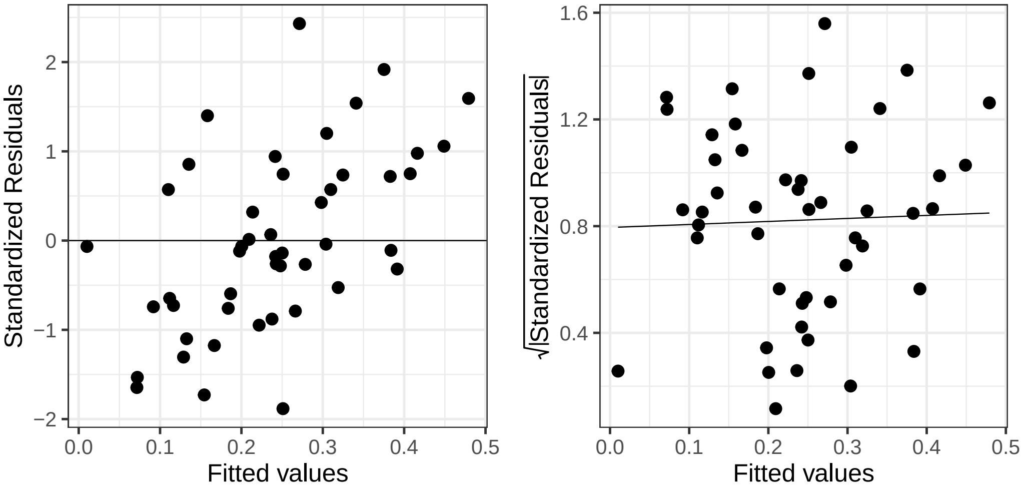
Figure 3 shows QQ-plots of (a) standardised residuals and (b) standardised random effects for assessing normality assumptions in the sampling and linking models. The residuals and random effects lie close to the QQ-line with only a few deviations, especially in the tails. Therefore, it is reasonable to assume the approximate normality of error terms. To further confirm this, we present density plots in Figure 4. Lastly, the Shapiro–Wilk test for standardised residuals equals 0.98552 with a p-value of $0.8213 > 0.05$. Therefore, we fail to reject the null hypothesis and conclude the distribution of the data is not significantly different from the normal distribution.

Figure 5 shows a plot of (a) standardised residuals versus fitted values and (b) a scale-location plot. This is to test for linearity and the homoscedasticity assumption in the model. The residuals are randomly distributed around $y = 0$ with no apparent pattern. They form an approximate horizontal band around the zero line. Therefore, linearity can be assumed. The scale-location plot shows the smooth line is roughly horizontal across the plot. There is also no clear pattern among the residuals. Therefore, the homoscedasticity assumption is met. Further, to test for randomness in this pattern we use the runs test. The p-value is 0.4588 which is greater than $\alpha = 0.05$ and we fail to reject the null hypothesis of nonrandomness. Therefore, we have sufficient evidence to say that the residuals are random.



(a) (b)

**Figure 4.** Density plots for (a) standardised residuals and (b) standardised random effects for the arcsine square root transformed model with GVF smoothed variance.



(a) (b)

**Figure 5.** Scatter plots for (a) standardised residuals verses fitted values and (b) scale-location plot for the transformed Fay–Herriot model.

### 4.2 Diagnostics for model-based small area estimates

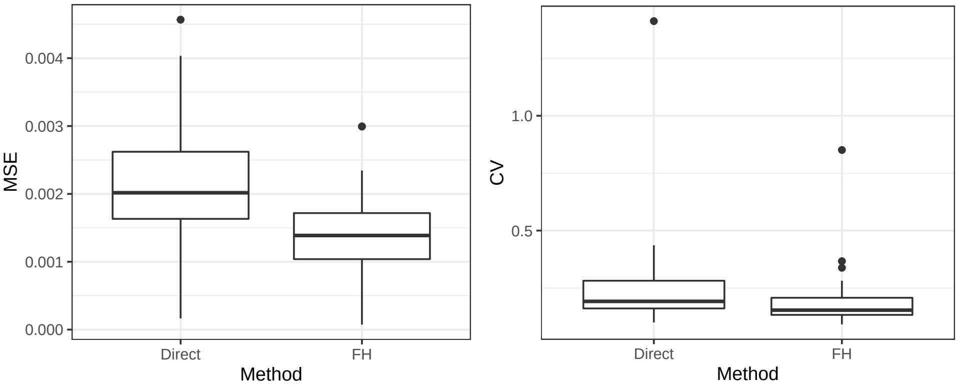
In addition to model diagnostics in Subsection 4.1 we present diagnostics for small area model-based estimates. We assess the reliability and validity of the FH estimates. We follow closely the guidelines of Brown et al. (2001). As noted by Chandra et al. (2018), the FH estimates should (i) be more precise than direct estimates, (ii) be consistent with the unbiased direct estimates, and (iii) give useful results to users. To assess for precision, we use the MSE and CV. Table 3 shows the summary statistics for the point estimates, the MSE and CV.

From Table 3 one sees that the direct estimates range between 0.009088 and 0.51998 while the FH estimates lie between 0.01013 and 0.47918. Therefore, the FH estimates have a smaller range compared to the direct estimates. An important advantage of SAE is the shrinkage of direct estimates towards the regression estimates from additional auxiliary data (Datta and Ghosh, 2012). The summary statistics for the MSE and CV show the accuracy gained in using the small area method outlined in this paper. There is a reduction in MSE and CV for the FH estimates. For instance, the maximum MSE for the direct and FH estimates are 0.00456 and 0.00386, respectively. We also note the FH estimates are shrunk such that the maximum CV reduced significantly from 141% to 65% for the direct and FH estimates, respectively. The gain in accuracy is further shown in the box plots for MSE and CV in Figure 6.

For bias diagnostics as outlined by Brown et al. (2001), we first plot the direct ($x$ axis) and FH estimates ($y$ axis). Figure 7 (a) shows a scatter plot of the fitted regression line and the identity line ($y = x$), i.e., $\beta_0 = 1$ and $\beta_1 = 0$. The regression line is fitted by least squares. As stated by Chandra et al. (2018), if the direct estimates are unbiased, then regressing them on the true values should be linear and correspond to the line $y = x$. Therefore, the scatter plot would be evenly distributed around the identity line. In our case, $\beta_0 = 0.898$ and $\beta_1 = 0.02236$, implying the FH estimates are approximately design unbiased. Further, Brown et al. (2001) provide a test to compare estimators. The test computes the correlation between the regression synthetic part of the model and direct estimates. Using the Brown test, we fail to reject the null hypothesis that the FH estimates are statistically significantly different from the direct estimates at $\alpha = 0.05$. The line graph in Figure 7 (b) is a comparison of direct and FH estimates where the counties are ordered by decreasing the MSE of the direct estimator. We note as the MSE of the direct estimator reduces, the direct estimates approach the FH estimates. This is a gain as, according to Rao and Molina (2015), the main reason for using model-based estimators is the reduction in MSE.

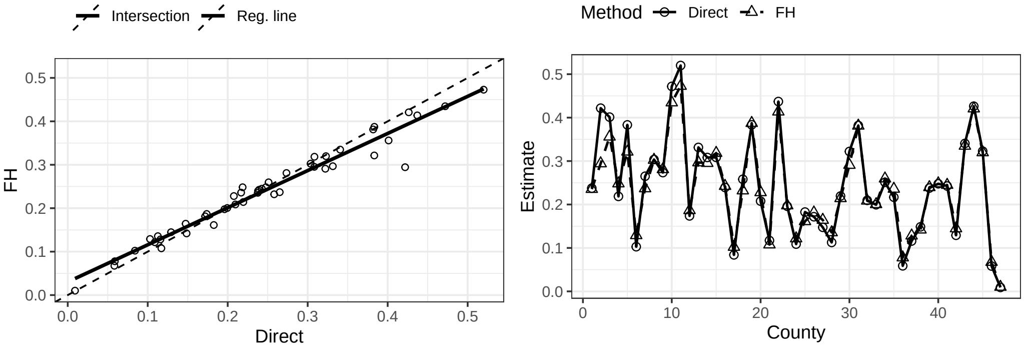
**Table 3.** Summary statistics for the point estimates, mean squared error and coefficient of variation for the county level overweight prevalence in Kenya.

| | | **Min.** | **Q1** | **Median** | **Mean** | **Q3** | **Max.** |
|---|---|---|---|---|---|---|---|
| Point est. | Direct | 0.009 08 | 0.160 18 | 0.237 85 | 0.244 78 | 0.322 45 | 0.519 98 |
| | FH | 0.010 13 | 0.162 47 | 0.242 08 | 0.242 06 | 0.307 30 | 0.479 18 |
| MSE | Direct | 0.000 16 | 0.001 63 | 0.002 01 | 0.002 12 | 0.002 62 | 0.004 56 |
| | FH | 0.000 04 | 0.001 02 | 0.001 33 | 0.001 43 | 0.001 77 | 0.003 86 |
| CV | Direct | 0.100 30 | 0.161 10 | 0.191 80 | 0.240 90 | 0.281 80 | 1.412 10 |
| | FH | 0.091 18 | 0.126 92 | 0.150 30 | 0.182 95 | 0.205 57 | 0.659 55 |



(a) (b)

**Figure 6.** Box plots showing (a) mean squared errors and (b) coefficient of variation for direct and FH estimates of overweight prevalence.



(a) (b)

**Figure 7.** A plot of direct and FH estimates with (a) regression and intersection line ($y = x$) and (b) a line graph showing overweight prevalence point estimates — direct and FH and counties where counties are ordered by decreasing MSE of the direct estimator.

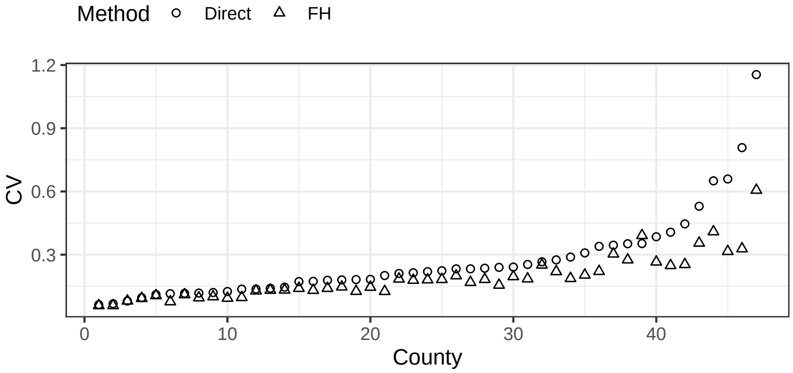
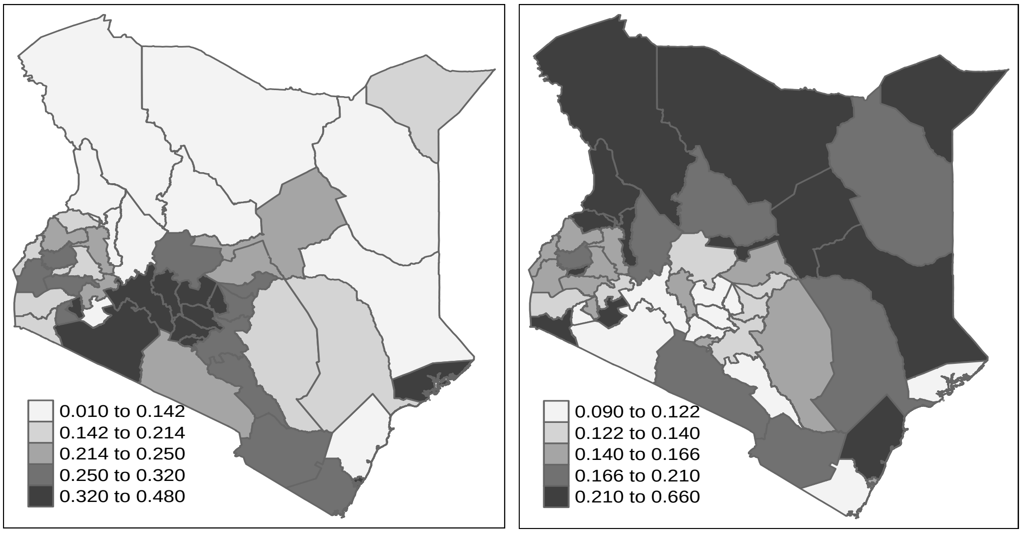


**Figure 8.** A scatter plot showing coefficients of variation for direct and FH estimates and counties ordered by decreasing sample size.

Figure 8 presents a line graph of CVs with counties ordered by decreasing sample size. For all counties, the CV for FH estimates is smaller than those of the direct estimator. As the sample size reduces per county, the difference in CV between the direct and FH estimates increases. This partly explains the need for using FH estimates for areas with small sample sizes.

### 4.3 Distribution of overweight prevalence in Kenya

We found the national overweight prevalence to be 27.98%. The KNBS reported 28% of Kenyans are overweight (Kenya National Bureau of Statistics, 2015). Figure 9, are two maps showing (a) the county level distribution of overweight prevalence in Kenya and (b) corresponding CVs. The CVs indicate the sampling variability as a percentage of FH estimates. The map is essential in identifying regions with high and low overweight prevalence. It shows that the prevalence of overweight varies geographically. Kenya is administratively divided into 47 counties, 290 sub-counties, and 1 450 wards. These administrative units are clearly defined by geographical boundaries. Therefore, for this study, $m = 1, 2, ..., 47$. The counties Nairobi (39%), Kiambu (45%), Nakuru (41%), Murang'a (48%), Nyeri (41%), Mombasa (38%) and Lamu (38%) have relatively high proportions of overweight. There are 15 counties with a prevalence above the national mean of 28%. It is important to note that the counties Nairobi, Nakuru, Mombasa and Nyeri are located in major towns of Kenya. The high prevalence of overweight seen in Murang'a and Lamu could be explained by socio-cultural factors. However, more research should be done to establish the underlying reasons and targeted interventions implemented.



(a) (b)

**Figure 9.** Maps showing the county level distribution of model-based overweight prevalence in Kenya: (a) Point estimates and (b) the corresponding coefficient of variation.

## 5. Conclusion

Establishing the extent of overweight prevalence is important for the public health surveillance of a country. The information acquired is vital to inform policy-making and resource allocation. This study presents a new data source of overweight prevalence at the county level relevant to the Kenya Health Policy (KHP), 2014–2030 and the Kenya Vision 2030. We have combined survey and census data through a model-based SAE methodology to better estimate the prevalence of overweight at the county level. Our model-based prevalence estimates have smaller MSEs and CVs than designed-based estimates. We found that counties within urban areas, including the major towns like Nairobi, Nakuru, Nyeri and Mombasa, have a higher prevalence of overweight compared to rural counties. Although we focus on overweight prevalence in Kenya, the presented method can also be applied to other indicators in developing countries with similar data sources. Health is devolved in Kenya. Therefore, counties with high prevalence should do more research and tailor interventions. We provide overweight prevalence estimates at the county level. It will be interesting to further extend this research to include more disaggregated domains like age, sex, gender and ethnicity. One limitation of this study is the time difference between the survey and census data. Data collected around the same time might yield more accurate results. Despite the limitation, this study has estimated the prevalence of overweight at the county level in Kenya with better precision.

**Acknowledgments.** I am grateful to the Editors and Referees for comments that significantly improved the paper. I also thank Sylvia Harmening, Nora Würz, Natalia Rojas-Perilla and Timo Schmid for providing useful comments that improved this paper. Lastly, I thank the Kenya National Bureau of Statistics for providing data and the Kenyan-German postgraduate programme through which this study is partly funded.

## References

Agyemang, C., Boatemaa, S., Agyemang Frempong, G., and de Graft Aikins, A. (2016). *Obesity in Sub-Saharan Africa.* Springer International Publishing, Cham, 41–53.

Asiki, G., Mohamed, S. F., Wambui, D., Wainana, C., Muthuri, S., Ramsay, M., and Kyobutungi, C. (2018). Sociodemographic and behavioural factors associated with body mass index among men and women in Nairobi slums: AWI-Gen Project. *Global Health Action*, **11**, 1470738.

Awuah, R. B., Anarfi, J. K., Agyemang, C., Ogedegbe, G., and de Graft Aikins, A. (2014). Prevalence, awareness, treatment and control of hypertension in urban poor communities in Accra, Ghana. *Journal of Hypertension*, **32**, 1203–1210.

Bell, W. R. (2008). Examining sensitivity of small area inferences to uncertainty about sampling error variances. *In JSM Proceedings, Survey Research Methods Section.* American Statistical Association, Alexandria, Virginia, 327–333.

Brown, G., Chambers, R., Heady, P., and Heasman, D. (2001). Evaluation of small area estimation methods—an application to unemployment estimates from the UK LFS. *In Proceedings of Statistics Canada Symposium.* Statistics Canada.

Carroll, R. J. and Ruppert, D. (1988). *Transformation and Weighting in Regression.* Chapman & Hall, New York.

Carter, G. M. and Rolph, J. E. (1974). Empirical Bayes methods applied to estimating fire alarm probabilities. *Journal of the American Statistical Association*, **69**, 880–885.

Chandra, H., Aditya, K., and Sud, U. (2018). Localised estimates and spatial mapping of poverty incidence in the state of Bihar in India. An application of small area estimation techniques. *Plos One*, **13**, e0198502.

Chowdhury, M. A. B., Uddin, M. J., Haque, M. R., and Ibrahimou, B. (2016). Hypertension among adults in Bangladesh: Evidence from a national cross-sectional survey. *BMC Cardiovascular Disorders*, **16**.

Christensen, D. L., Eis, J., Hansen, A. W., Larsson, M. W., Mwaniki, D. L., Kilonzo, B., Tetens, I., Boit, M. K., Kaduka, L., Borch-Johnsen, K., and Friis, H. (2008). Obesity and regional fat distribution in Kenyan populations: Impact of ethnicity and urbanization. *Annals of Human Biology*, **35**, 232–249.

Datta, G. and Ghosh, M. (2012). Small area shrinkage estimation. *Statistical Science*, **27**, 95–114.

Datta, G. S. and Lahiri, P. (2000). A unified measure of uncertainty of estimated best linear unbiased predictors in small area estimation problems. *Statistica Sinica*, **10**, 613–627.

Efron, B. and Morris, C. (1975). Data analysis using Stein's estimator and its generalizations. *Journal of the American Statistical Association*, **70**, 311–319.

Fay, R. and Herriot, R. (1979). Estimates of income for small places: An application of James-Stein procedures to census data. *Journal of the American Statistical Association*, **74**, 269–277.

General Assembly (2015). Resolution adopted by the General Assembly on 19 September 2016. Accessed on 15/11/2019.
URL: *https://www.un.org/en/development/desa/population/migration/generalassembly/docs/globalcompact/A_RES_71_1.pdf*

Gichu, M., Asiki, G., Juma, P., Kibachio, J., Kyobutungi, C., and Ogola, E. (2018). Prevalence and predictors of physical inactivity levels among Kenyan adults (18–69 years): An analysis of STEPS survey 2015. *BMC Public Health*, **18**, 1–7.

Government of Kenya (2010). *The 2009 Kenya Population and Housing Census, Volume 1C: Population Distribution by Age, Sex, and Administrative Units*. KNBS Nairobi.

Government of Kenya (2013). *Kenya: The Constitution of Kenya*. National Council for Law Reporting.

Groenveld-van Dijk, E. (2013). *The burden of overweight and obesity in Kenya analyses of the known determinants and control*. Master's thesis, Royal Tropical Institute, Amsterdam.

Hadam, S., Würz, N., and Kreutzmann, A.-K. (2020). Estimating regional unemployment with mobile network data for functional urban areas in Germany. Technical report.

Hartley, H. O. and Rao, J. N. K. (1967). Maximum-likelihood estimation for the mixed analysis of variance model. *Biometrika*, **54**, 93–108.

Hawala, S. and Lahiri, P. (2010). Variance modeling in the us Small Area Income and Poverty Estimates program for the American Community Survey. *In JSM Proceedings, Section on Bayesian Statistical Science*. American Statistical Association, Alexandria, Virginia.

Hawala, S. and Lahiri, P. (2018). Variance modeling for domains. *Statistics and Applications*, **16**, 399–409.

Henderson, C. R. (1975). Best linear unbiased estimation and prediction under a selection model. *Biometrics*, **31**, 423–447.

Horvitz, D. and Thompson, D. (1952). A generalization of sampling without replacement from a finite universe. *Journal of the American Statistical Association*, **47**, 663–685.

Jiang, J., Lahiri, P., Wan, S.-M., and Wu, C.-H. (2001). Jackknifing in the Fay-Herriot model with an example. *In Proceedings of the Seminar on Funding Opportunity in Survey Research*. Council of Professional Associations on Federal Statistics, Arlington, Virginia, 75–97.

Jones-Smith, J. C., Gordon-Larsen, P., Siddiqi, A., and Popkin, B. M. (2012). Is the burden of overweight shifting to the poor across the globe? Time trends among women in 39 low-and middle-income countries (1991–2008). *International Journal of Obesity*, **36**, 1114–1120.

Kenya National Bureau of Statistics (2015). Kenya STEPwise survey for non-communicable diseases risk factors 2015 report. Accessed on 15/11/2019.
URL: *https://www.health.go.ke/wp-content/uploads/2016/04/Steps-Report-NCD-2015.pdf*

Kish, L. (1949). A procedure for objective respondent selection within the household. *Journal of the American Statistical Association*, **44**, 380–387.

Kreutzmann, A.-K., Pannier, S., Rojas-Perilla, N., Schmid, T., Templ, M., and Tzavidis, N. (2019). The R package emdi for estimating and mapping regionally disaggregated indicators. *Journal of Statistical Software*, **91**, 1–33.

Kulie, T., Slattengren, A., Redmer, J., Counts, H., Eglash, A., and Schrager, S. (2011). Obesity and women's health: An evidence-based review. *Journal of the American Board of Family Medicine*, **24**, 75–85.

Maples, J., Bell, W., and Huang, E. T. (2009). Small area variance modeling with application to county poverty estimates from the American community survey. *In JSM Proceedings, Section on Survey Research Methods*. American Statistical Association, Alexandria, Virginia, 5056–5067.

Mbochi, R. W. (2010). *Overweight and obesity prevalence and associated socio-economic factors, physical activity and dietary intake among women in Kibera division, Nairobi*. Ph.D. thesis, Doctoral Dissertation, Kenyatta University.

Mbochi, R. W., Kuria, E., Kimiywe, J., Ochola, S., and Steyn, N. P. (2012). Predictors of overweight and obesity in adult women in Nairobi Province, Kenya. *BMC Public Health*, **12**, 1–9.

Ministry of Health Kenya (2014). Kenya Health Policy 2014-2030. Accessed on 20/05/2020.
URL: *http://publications.universalhealth2030.org/uploads/kenya_health_policy_2014_to_2030.pdf*

Mkuu, R. S., Epnere, K., and Chowdhury, M. A. B. (2018). Prevalence and predictors of overweight and obesity among Kenyan women. *Preventing Chronic Disease*, **15**, 170401.

Muthuri, S. K., Wachira, L.-J. M., Onywera, V. O., and Tremblay, M. S. (2014). Correlates of objectively measured overweight/obesity and physical activity in Kenyan school children: Results from ISCOLE-Kenya. *BMC Public Health*, **14**, 1–11.

Patterson, H. D. and Thompson, R. (1971). Recovery of inter-block information when block sizes are unequal. *Biometrika*, **58**, 545–554.

PAWLOSKI, L. R., CURTIN, K. M., GEWA, C., AND ATTAWAY, D. (2012). Maternal-child overweight/obesity and undernutrition in Kenya: A geographic analysis. *Public Health Nutrition*, **15**, 2140–2147.

PFEFFERMANN, D. (2013). New important developments in small area estimation. *Statistical Science*, **28**, 40–68.

PRASAD, N. N. AND RAO, J. N. (1990). The estimation of the mean squared error of small-area estimators. *Journal of the American Statistical Association*, **85**, 163–171.

PRATESI, M. (2016). *Analysis of Poverty Data by Small Area Estimation*. Wiley & Sons, Chichester.

RAO, J. N. K. AND MOLINA, I. (2015). *Small Area Estimation*. Wiley & Sons, Hoboken, New Jersey.

RILEY, L., GUTHOLD, R., COWAN, M., SAVIN, S., BHATTI, L., ARMSTRONG, T., AND BONITA, R. (2016). The world health organization stepwise approach to noncommunicable disease risk-factor surveillance: Methods, challenges, and opportunities. *American Journal of Public Health*, **106**, 74–78.

SÄRNDAL, C.-E., SWENSSON, B., AND WRETMAN, J. (2003). *Model Assisted Survey Sampling*. Springer, New York.

SCHMID, T., BRUCKSCHEN, F., SALVATI, N., AND ZBIRANSKI, T. (2017). Constructing sociodemographic indicators for national statistical institutes by using mobile phone data: estimating literacy rates in Senegal. *Journal of the Royal Statistical Society: Series A*, **180**, 1163–1190.

SEBIRE, N. J., JOLLY, M., HARRIS, J., WADSWORTH, J., JOFFE, M., BEARD, R., REGAN, L., AND ROBINSON, S. (2001). Maternal obesity and pregnancy outcome: A study of 287 213 pregnancies in London. *International Journal of Obesity*, **25**, 1175–1182.

STEYN, N. P. AND MCHIZA, Z. J. (2014). Obesity and the nutrition transition in Sub-Saharan Africa. *Annals of the New York Academy of Sciences*, **1311**, 88–101.

TZAVIDIS, N., ZHANG, L.-C., LUNA, A., SCHMID, T., AND ROJAS-PERILLA, N. (2018). From start to finish: A framework for the production of small area official statistics. *Journal of the Royal Statistical Society: Series A*, **181**, 927–979.

VALENCIA, C. C.-C., ENCINA, J., AND LAHIRI, P. (2016). Poverty mapping for the Chilean comunas. *In* PRATESI, M. (Editor) *Analysis of Poverty Data by Small Area Estimation*, chapter 20. Wiley & Sons, Chichester, 379–403.

WORLD HEALTH ORGANIZATION (2005). WHO STEPS surveillance manual: The WHO STEPwise approach to chronic disease risk factor surveillance. Technical report, World Health Organization, Geneva.

WORLD HEALTH ORGANIZATION (2020). World Health Organization obesity and overweight fact sheet. Accessed on 25/05/2020.
URL: *https://www.who.int/news-room/fact-sheets/detail/obesity-and-overweight*

Manuscript received 2020-09-08, revised 2021-04-22, accepted 2021-11-16.