Small area estimation of health insurance coverage for Kenyan counties

Noah Cheruiyot Mutai[ID]

Abstract Health insurance is important in disease management, access to quality health care and attaining Universal Health Care. National and regional data on health insurance coverage needed for policy making is mostly obtained from household surveys; however, estimates at lower administrative units like at the county level in Kenya are highly variable due to small sample sizes. Small area estimation combines survey and census data using a model to increases the effective sample size and therefore provides more precise estimates. In this study we estimate the health insurance coverage for Kenyan counties using a binary M-quantile small area model for women ($n = 14{,}730$) and men ($n = 12{,}007$) aged 15 to 49 years old. This has the advantage that we avoid specifying the distribution of the random effects and distributional robustness is automatically achieved. The response variable is derived from the Kenya Demographic and Health Survey 2014 and auxiliary data from the Kenya Population and Housing Census 2009. We estimate the mean squared error using an analytical approach based on Taylor series linearization. The national direct health insurance coverage estimates are 18% and 21% for women and men respectively. With the current health insurance schemes, coverage remains low across the 47 counties. These county-level estimates are helpful in formulating decentralized policies and funding models.

Keywords Binary M-quantile · Direct estimation · Health insurance coverage · Universal Health Care · Taylor series linearization

✉ Noah Cheruiyot Mutai
Institute of Statistics and Econometrics, Freie Universität Berlin, Berlin, Germany
E-Mail: ncheruiyot3@gmail.com

Department of Mathematics, Statistics and Physical Science, Taita Taveta University, Voi, Kenya

🌀 Springer

1 Introduction

Health insurance reduces extreme health costs and out-of-pocket spending by pooling resources. It is an important component towards the attainment of Universal Health Care (UHC) (Dye et al. 2015). The goal of UHC was set by the World Health Organization (WHO) member states in 2005 (World Health Organization 2005). The goal is to assist member countries to achieve UHC through health system financing. UHC has been defined as the provision of the needed quality health services to the whole population with less cost (World Health Organization 2013). In 2015, the General Assembly adopted the 2030 agenda that includes 17 Sustainable Development Goals (SDGs). The SDG goal 3.8 seeks: To attain UHC, with financial risk security, access to quality vital health care services and inexpensive key medicines and vaccines for everyone (United Nations General Assembly 2015).

Countries in Sub-Saharan Africa face many health challenges. These include; low investment in health care, slow economic growth, extensive out-of-pocket expenditure and reduced access to health services (Sambo et al. 2013). To achieve the health related SDG's and UHC, the regional committee for Africa suggested strategies including; more investment, efficient use of health resources and expand coverage. The objective is to foster efficient and sustainable health financing and achieve these goals. Over the years these countries have prioritized investments towards achieving UHC (Lagomarsino et al. 2012; Cotlear and Rosemberg 2018). This also follows the "Abuja Declaration" of World Health Organization et al. (2010) that set a minimum of 15% of the total government expenditure.

To mention a few countries; Ghana's National Health Insurance Scheme (NHIS) has been in existence since 2003. The goal is to guarantee fairness and access to health care services by reducing financial barriers to access at the point of use (Kusi et al. 2015). By 2014, over 10.5 million Ghanaians (an estimated 40% of the population) were covered by the NHIS, with inpatient and outpatient visits to health facilities increasing from 0.5 to around 3 per capita between 2005 and 2014 (Wang et al. 2017). A study by Aikins et al. (2021) established that the scheme will likely achieve UHC if protected from political interference and improved accountability. In 2017 Zambia developed the National Health Strategic Plan 2017–2021. The plan outlined strategies to establish a social health insurance scheme (Government of Zambia 2018). This was passed into the National Health Insurance Act 2018 whose goal is to provide reliable health system financing and universal access to health services. Under the NHI, all eligible citizens contribute to the pool of resources in addition to external funding. Households classified as poor by measuring absolute poverty based on monthly consumption expenditures are exempted from contributing (Government of Zambia 2018).

Kenya is a lower-middle-income country with a population of 47.5 million, 12.2 million households and an average household size of 3.9. In Kenya, 75.1% are below 35 years and 32.73 million (68.9%) live in rural areas (Kenya National Bureau of Statistics 2019). Approximately 83% of Kenyans do not have financial protection against extreme health care costs. Around 1.5 million become poor due to health care costs (Ministry of Health Kenya 2014; Okungu et al. 2017). As outlined in Vision 2030 Kenya seeks to achieve UHC by 2030. Towards this goal, several strategies

have been implemented. To start with, the government piloted UHC in four out of 47 counties in Kenya (Isiolo, Kisumu, Machakos, and Nyeri). These counties were selected because they have a high incidence of both communicable and non-communicable diseases, maternal mortality, and road traffic injuries. Results from this pilot showed great success, however, not sustainable by the government capitation (Ministry of Health Kenya 2018). Secondly, the government abolished charges in public hospitals and health care. It also introduced free maternity services in all health care facilities (Maina and Kirigia 2015). Further, it expanded the National Health Insurance Fund (NHIF) package from an inpatient-only package to outpatient services (Mwaura et al. 2015).

Some studies on health insurance in Kenya include; Kazungu and Barasa (2017), where they examined the levels, inequalities (where households are categorized into five socio-economic quintiles) and factors associated with health insurance coverage in Kenya. They analyzed data from the Kenya Demographic and Health Survey (KDHS) 2009 and 2014. Results show health insurance coverage in Kenya remains low and show high inequalities. Otieno et al. (2019) carried out a study to determine the prevalence of health insurance and associated factors among households in urban slum settings in Nairobi, Kenya. They used cross-sectional data of adults aged 18 years or older from randomly selected households in Viwandani slums (Nairobi, Kenya). The study was conducted between June and July 2018. Their findings show that the prevalence of health insurance in the sample was 43%. Health insurance coverage in Viwandani slums in Nairobi, Kenya, is low.

The KDHS 2014 was a national survey. It was designed to provide reliable direct estimates at the national level only and county estimates for some selected indicators (health insurance not included). The direct estimators (they rely only on the survey data) are approximately designed unbiased and consistent (Pfeffermann 2013). However, direct estimators generally have large variances and estimates are unreliable when the sample sizes are small (Rao and Molina 2015) – for example at the county level in Kenya. In contrast, model-based small area methods produce more reliable estimates in terms of smaller mean squared error and coefficient of variation (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 theory on small area estimation, we refer the reader to Ghosh and Rao (1994); Pfeffermann (2002, 2013); Jiang and Lahiri (2006); Rao and Molina (2015); Pratesi (2016); Tzavidis et al. (2018) and Morales et al. (2021).

For this study, therefore, we rely on SAE to better estimate the proportion of persons with health insurance for Kenyan counties. The health insurance status of a person is binary. Some approaches have been proposed to estimating binary variables in the small area context. To mention a few Bayesian approaches; Hierarchical Bayes of Malec et al. (1997); Nandram et al. (1999); Liu et al. (2007) and Empirical Bayes of MacGibbon and Tomberlin (1987); Farrell et al. (1997) and Ghosh et al. (1998). For frequentist approaches Jiang and Lahiri (2001) proposed empirical best predictor (EBP) for a binary response. Chambers et al. (2016) outlines the use of a binary logistic generalized linear mixed model (GLMM) in SAE. However, they note that GLMM based on maximum likelihood is influenced by outliers. M-quantile SAE model provides a robust alternative to GLMM's.

The rest of this paper is organized as follows. We describe the KDHS 2014 and the Kenya Population and Housing Census (KPHC) 2009 in Sect. 2. In Sect. 3, we outline the statistical methodology applied in this paper. In particular, the direct estimation and the binary M-quantile small area model estimation, hereafter called the MQ model, are examined by means of the point estimation and the mean squared error. In Sect. 4, we present the results of the application to estimate health insurance coverage for Kenyan counties. Lastly, in Sect. 5, we give the concluding remarks based on the findings, some possibilities for further research as well limitations.

2 Data description

In this section, we describe the data sources used in this paper. We had access to KDHS 2014 and KPHC 2009. The links to the data sources are provided at the end of this paper. We assume that the functional relation between having health insurance and auxiliary data remains constant between the survey and census time.

2.1 The Kenya Demographic and Health Survey

The Demographic and Health Survey (DHS) collects, analyzes and disseminates data on population, health, HIV and nutrition in over 90 countries (Croft et al. 2018). In this study, we had access to the Kenya Demographic and Health Survey (KDHS) done in 2014. The KDHS has been conducted in Kenya after every 5 years i.e. 1989, 1993, 1998, 2003, 2008–2009 and 2014. The 2014 KDHS collected several data on household characteristics, education and employment, and health-related indicators such as HIV and child health survival (Kenya National Bureau of Statistics 2015).

The 2014 KDHS sample was drawn from a sample master called the Fifth National Sample Survey and Evaluation Program (NASSEP V). The Kenya National Bureau of Statistics (KNBS) currently uses this framework to conduct household surveys in Kenya. It includes 5,360 clusters derived from the 2009 Kenya Population and Housing Census (Kenya National Bureau of Statistics 2015). The framework has a total of 96,251 enumeration areas (EA's). The KDHS 2014 sought to create representative estimates for the majority of survey variables at the national level, for individual urban and rural regions, for regional (formerly provincial) levels, and selected indicators at the county level. To meet these objectives, the sample was designed to comprise 40,300 households from 1,612 clusters spread across the country, with 995 clusters in rural areas and 617 clusters in urban areas. Samples were selected separately in each sampling stratum using a two-stage sample approach. In the first stage, the 1,612 EA's were chosen with equal probability from the NASSEP V frame. The properties from the listing operations served as the sampling frame for the second round of selection, which included selecting 25 households from each cluster (Kenya National Bureau of Statistics 2015).

Three main questionnaires were used in the KDHS; (i) A household questionnaire, (ii) A questionnaire for women aged 15 to 49, and (iii) A questionnaire for men aged 15 to 54. They were based on model questionnaires designed for the DHS program, as well as questionnaires used in earlier KDHS surveys and Kenya's

Table 1 Summary of sample sizes for women and men over counties in the Kenya Demographic and Health Survey 2014

Questionnaire	Min.	Q1.	Median	Mean	Q3.	Max.
Women	236.0	275.5	310.0	313.5	342.0	460.0
Men	118.0	227.0	250.0	255.5	287.0	370.0

current information needs. During the questionnaire development process, input was sought from relevant stakeholders and data users. Producing county-level estimates necessitated gathering data from a large number of families within each county, resulting in a significant rise in sample size from around 10,000 homes in the 2008–09 KDHS to 40,300 households in 2014 (Kenya National Bureau of Statistics 2015).

A total of 39,679 households were selected in the sample, of which 36,812 were found occupied at the time of the fieldwork. Of these, 36,430 households were successfully interviewed, yielding an overall household response rate of 99%. The shortfall of households occupied was primarily due to structures that were found to be vacant or destroyed and households that were absent for an extended time. Among the households selected using the full questionnaires, a total of 15,317 women were identified as eligible for the full women's questionnaire, of whom 14,741 were interviewed, generating a response rate of 96% (Kenya National Bureau of Statistics 2015). A total of 14,217 men were identified as eligible in these households, of whom 12,819 were successfully interviewed, generating a response rate of 90%. For this application, we use only complete cases for our variable of interest giving a total sample of 12,007 men and 14,730 women. County-specific samples sizes for women and men are summarized in Table 1. The women sample sizes are higher than for men because most indicators in the DHS (fertility, maternal mortality rate, infant mortality rate and neonatal mortality rate) relates to children and women.

2.1.1 Socio-demographic characteristics

Table 2 are socio-demographic characteristics of respondents in the survey. It included women (aged 15 to 49 years) and men (aged 15 to 54 years.) For comparison purposes, we selected men aged 15 to 49 years. The majority of the respondents are between ages 15 to 34 years. Kenya is composed majorly of a youthful population. According to the 2019 Kenyan census the median age is approximately 20 years. The survey also inquired whether respondents lived in urban or rural areas. Most women (63%) live in rural areas while most men (61%) live in urban areas. The KDHS 2014 was planned to give representative estimates for most of the survey indicators at the national levels. Other characteristics include education level (no education, primary, secondary or higher), wealth index (poorer, poorest, middle, richer and richest) and marital status (never married, married, widowed, separated, divorced). The majority of women and men are either never married or married.

Table 2 Socio-demographic characteristics for women and men in the Kenya Demographic and Health Survey 2014

Demographic characteristics	Women		Men	
Age	Frequency	Percent	Frequency	Percent
15–19	2859	19.4	2811	23.4
20–24	2537	17.2	1981	16.5
25–29	2859	19.4	1940	16.2
30–34	2104	14.3	1701	14.2
35–39	1876	12.7	1484	12.4
40–44	1367	9.3	1197	10.0
45–49	1128	7.7	893	7.4
Residence				
Rural	9262	62.9	4644	38.7
Urban	5468	37.1	7363	61.3
Region				
Central	1509	10.2	1246	10.4
Coast	1840	12.5	1503	12.5
Eastern	2494	16.9	2142	17.8
Nairobi	460	3.1	370	3.1
North Eastern	779	5.3	591	4.9
Nyanza	2010	13.6	1542	12.8
Rift Valley	4252	28.9	3483	29.0
Western	1386	9.4	1130	9.4
Education Level				
No education	1980	13.4	4124	34.3
Primary	7398	50.2	4570	38.1
Secondary	4103	27.9	1980	16.5
Higher	1249	8.5	1333	11.1
Wealth index				
Poorer	2864	19.4	2442	20.3
Poorest	3399	23.1	2503	20.8
Middle	2841	19.3	2465	20.5
Richer	2839	19.3	2578	21.5
Richest	2787	18.9	2019	16.8
Marital status				
Never married	9009	61.2	5742	47.8
Married	4053	27.5	5624	46.8
Widowed	580	3.9	56	0.5
Seperated	750	5.1	421	3.5
Divorced	338	2.3	164	1.4

2.1.2 Direct estimation and type of insurance per wealth quantile

The 2014 KDHS asked respondents if they were covered by any health insurance and, if yes, what type. We first estimated health insurance coverage for the whole country (using KDHS only). We used the R package emdi (Kreutzmann et al. 2019)

Table 3 Estimated proportions (direct estimates) with health insurance, mean squared error and coefficient of variation for women and men at the national level in Kenya using the Kenya Demographic and Health Survey 2014

Gender	Percent	Mean squared error	Coefficient of variation
Women	0.18	0.000016	0.021
Men	0.21	0.000027	0.024

Table 4 Descriptive statistics of the coefficients of variation for the distribution of health insurance coverage at the county level in Kenya for the direct estimates

Gender	Min.	Q1.	Median	Mean	Q3.	Max
Women	0.073	0.120	0.157	0.186	0.215	0.634
Men	0.078	0.126	0.150	0.188	0.221	0.670

for direct estimation. Table 3 shows the percentage of women and men age 15–49 covered by health insurance at the national level together with the mean squared error and the coefficient of variation. A small percentage of Kenyans aged 15–49 (18% of women and 21% of men) have health insurance. The mean squared error values are very low, 0.000016 and 0.000027 for women and men respectively. Also, the coefficient of variation is 2.1% and 2.4% for women and men groups. Therefore the estimates are reliable (as expected at the design level of the survey).

In this paper, we are interested in estimating health insurance coverage at the county level. Table 4 below is a summary of the coefficient of variations(CV's) for the direct estimates. The CV's are quite high reaching values of 63% and 67% for women and men, respectively. Model-based SAE methods that borrow strength from other counties of interest are required to increase the accuracy of the estimation.

Table 5 shows the type of health insurance for each wealth quantile (built based on household asset data (Kenya National Bureau of Statistics 2015)) category for women and men. Among those covered, the national insurance scheme is the most common type for both genders. Employer-based insurance is the next most common

Table 5 Percentages for each type of health insurance per socio-economic quintiles in Kenya from Kenya Demographic and Health Survey 2014

Gender	Type	National	Employer	Mutual	Private	Prepayment	Other	None
Women	Wealth quintile							
	Poorest	1.54	0.24	0.15	0.12	0.00	0.00	97.96
	Poorer	4.78	0.95	0.11	0.35	0.00	0.11	93.71
	Middle	10.57	1.14	0.21	0.39	0.00	0.14	87.55
	Richer	18.41	2.28	0.57	0.39	0.00	0.21	78.14
	Richest	26.25	6.13	0.83	2.56	0.00	0.25	63.97
Men	Poorest	2.74	0.07	0.00	0.20	0.00	0.07	96.92
	Poorer	7.35	0.80	0.07	0.22	0.00	0.22	91.34
	Middle	11.54	1.04	0.30	0.52	0.00	0.15	86.45
	Richer	20.05	2.96	0.08	1.52	0.00	0.72	74.68
	Richest	30.53	10.96	0.10	4.20	0.00	0.61	53.59

Table 6 Summary of population sizes in Kenya Population and Housing Census 2009 at the county level in Kenya

	Min.	1st Quartile	Mean	Median	3rd Quartile	Max.
Census	2,205	10,676	18,586	15,408	20,572	98,289

type of insurance. This is because employers are obliged by law to provide insurance to their employees. The trend in insurance coverage varies per wealth quintile, with the richer and richest most covered across all insurance types. Ilinca et al. (2019) also found out that there are significant levels of inequality in access to health services in Kenya across the wealth quintile. Mwenda et al. (2021) established that poor households pay more for health care especially for outpatient services. This is because poor households cannot pay or do not have health insurance hence more out of pocket spending. In the study they also noted that the rich also spent more on outpatient care owing to their financial abilities.

2.2 The Kenya Population and Housing Census

For model-based SAE we need supplemental data collected from all areas. We had access to the Kenya Population and Housing Census (KPHC) 2009 in this case. Kenya has consistently conducted a census every ten years, i.e. 1969, 1979, and so on, with the most recent being in 2019. Under Kenyan legislation, the KNBS is the primary government body in charge of collecting, processing, and disseminating census and other statistical data. Statistics are needed to track the progress of numerous development goals and worldwide initiatives, such as the SDG's. The main goal of the KPHC 2009 was to offer essential information on the population's demographic, social, and economic features, as well as housing. These include population size and composition, fertility, mortality and migration rates, levels of education, labour force size, and so on. The data for this census was taken using scanning technology, with technical help from the United States Census Bureau (USCB) (Government of Kenya 2010). This census was conducted based on old administrative areas, such as villages, sub-locations, locations, divisions, districts, and provinces. Kenya had 46 legal districts, minus Nairobi, the capital city, which was the 47th district. After 2010, these districts were changed to the present 47 counties with no changes in borders (Government of Kenya 2013). As a result, we can connect the survey and census data. The census data serve as potential covariates in the small area model described in Sect. 3.4. Table 6 is a summary of population sizes at the county level in Kenya. The census is the 10% sample, i.e. every 10th household of the whole data set is released by the KNBS (Government of Kenya 2010).

3 Statistical Methodology

In this section, we outline the methodology applied in this paper. To begin with we describe the direct estimation using the survey data only. We then introduce M-quantile regression differentiating it from standard mean regression. Next, we

give the general small area estimation setting. Thereafter, we discuss the M-quantile small area model together with the point and mean squared error estimation.

3.1 Direct estimation

The Horvitz-Thompson (HT) estimator of Horvitz and Thompson (1952) is used to estimate the population proportion $\bar{Y}_i$ for area i, $i = 1, 2, ..., m$, where m is the total number of areas of the whole population, from a complex sampling design. Using this estimator, the direct estimator of the target proportion for area i based on sample data is defined as $\widehat{\bar{y}}_i^{\text{dir}} = \frac{1}{N_i} \sum_{j=1}^{n_i} w_{ij} y_{ij}$, $i = 1, 2, ..., m$, where $\widehat{\bar{y}}_i^{\text{dir}}$ is the direct proportion estimator for area i, N_i is the population size in area i, y_{ij} is the response of individual/household j in area i and w_{ij} are sampling weights – inverse of first order inclusion probabilities. The weights compensate for unequal probabilities of sampling and unit non-response. The HT-estimator for population proportions is design-unbiased (Särndal et al. 2003). However, the variance reaches high values for areas with small sample sizes. For KDHS 2014, all counties were sampled, although sample sizes in some counties are not sufficient to provide reliable direct estimates as seen from the high values of coefficient of variation (beyond 20% using guidelines set by the UK Office for National Statistics [ONS]).

3.2 Small area estimation

In SAE we assume the following idealized setting: There is a finite population U 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 ith 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. The sampled and non-sampled units will be denoted by s and r respectively. Let y_{ij} be the response variable of interest of individual/household j in area i and has been observed for sampled units only; $\mathbf{x}_{ij}$ denote a $p \times 1$ vector of unit level covariates with intercept. In general it is assumed that the values of $\mathbf{x}_{ij}$ are known for all units in the population, as are the values $\mathbf{z}_i$ of a $q \times 1$ vector of area level covariates. We are interested in using sample values of y_{ij} and the population values of $\mathbf{x}_{ij}$ and $\mathbf{z}_i$ to estimate the small area i proportion of health insurance coverage given by $\bar{Y}_i = N_i^{-1} \sum_{j \in U_i} y_{ij}$.

3.3 M-quantile regression

The standard linear regression summaries the average relationship between a continuous response y_i given explanatory variables x_i i.e. $E[y_i|x_i]$ where $i = 1, 2, ..., n$ is the number of observations. This does not give a complete picture of the conditional distribution of the response variable given the explanatory variables, and we might be interested in other parts of this distribution for example the 10th percentile. In the same manner, a relationship between the response and the explanatory variables can also be established using the conditional median function instead. The quantile $q \in (0,1)$ is that y which splits the data into proportions q below and $(1-q)$ above such that $F\left(y_q\right) = q$ and $y_q = F^{-1}(q)$. The median has $q = 0.5$. Whereas the

mean regression minimizes the squared error, a regression model based on the median (or median regression) minimizes the least absolute deviation (LAD). Median regression is also more robust to outliers than mean regression and no parametric assumption is required. To generalize the mean and median regression, we discuss the expectile and quantile regression. Expectile regression (Newey and Powell 1987) generalizes the mean regression to estimate the expectiles while quantile regression generalizes the median regression to estimate other parts of the conditional distribution (quantiles) of y given x (Koenker and Bassett Jr 1978; Koenker and Hallock 2001). M-quantile regression (Breckling and Chambers 1988) estimates the conditional distribution lying between the quantiles and expectiles. It is an extension of M-estimation of Huber (1992). The M-quantile of order q of a continuous random variable y with distribution function $F(y)$ is the value Q_q that satisfies

$$\int \psi_q \left(\frac{y - Q_q}{\sigma_q} \right) dF(y) = 0, \tag{1}$$

where $\psi_q(t) = 2\psi(t)\{qI(t > 0) + (1 - q)I(t \leqslant 0)\}$, ψ is an influence function defined by the user and σ_q is an appropriate scale measure for the random variable $Y - Q_q$. According to Chambers and Tzavidis (2006) when the response variable y is binary, there is no obvious definition of a quantile function as in the continuous case in 1. But, given that the influence function ψ is continuous and monotone nondecreasing, the M-quantiles of a binary variable exist and are unique. In that case we are interested in predicting $P(y = 1) = p$ which means that 1 becomes

$$pq\psi \left(\frac{1 - Q_q}{\sigma_q} \right) = (1 - p)(1 - q)\psi \left(\frac{Q_q}{\sigma_q} \right). \tag{2}$$

Since y is binary, following Chambers and Tzavidis (2006), we impose a linear logistic function as

$$Q_q \left(\mathbf{x}_j; \psi \right) = \exp \left(\mathbf{x}_j^{\mathrm{T}} \boldsymbol{\beta}_q \right) \left\{ 1 + \exp \left(\mathbf{x}_j^{\mathrm{T}} \boldsymbol{\beta}_q \right) \right\}^{-1}, \tag{3}$$

where $\boldsymbol{\beta}_q$ are regression coefficients estimated using a robust maximum likelihood estimating equations following (Cantoni and Ronchetti 2001).

3.4 M-quantile small area model

Small area estimation mostly uses random-area effects to characterize between area variations beyond that explained by auxiliary variables in the model (Rao and Molina 2015). However, mixed effect models depend on distributional assumptions (for example, the assumption of normally distributed residuals). Further, it requires the specification of the random part of the model. An alternative approach to mixed effect modeling is the use of M-quantile models in SAE. The M-quantile model for SAE was proposed by (Chambers and Tzavidis 2006). They model the between-area heterogeneity using M-quantile coefficients. In this case, the population model is specified and fitted at unit level without specifying any small area geography.

First define q_{ij} such that $y_{ij} = Q_{q_{ij}}(\mathbf{x}_{ij}; \psi)$, i.e. q_{ij} is a random index that varies between 0 and 1. Since the response variable is binary, we specify a linear logistic function, where the population M-quantile model for q_{ij} (and hence y_{ij}) is then defined by

$$Q_{q_{ij}}(\mathbf{x}_{ij}; \psi) = \exp\left(\mathbf{x}_{ij}^{\mathrm{T}} \boldsymbol{\beta}_{q_{ij}}\right) \left\{1 + \exp\left(\mathbf{x}_{ij}^{\mathrm{T}} \boldsymbol{\beta}_{q_{ij}}\right)\right\}^{-1}. \tag{4}$$

Chambers and Tzvidis (2006) called q_{ij} the M-quantile coefficients. The M-quantile coefficient for area i is given by; $\theta_i = E\left[q_{ij} \mid i\right]$, where the expectation is conditional on the distribution of the random indices q_{ij} within area i.

3.4.1 Point estimation

To estimate the population proportion we proceed as follows. We first note that the empirical value $\widehat{q}_{ij}$ of the random index q_{ij} is the solution to $y_{ij} = Q_{\widehat{q}_{ij}}(\mathbf{x}_{ij}; \psi)$ and this value is referred to as the estimated M-quantile coefficient of y_{ij} (Chambers and Tzvidis 2006).

1. Obtain sample observations in area i using a non-informative sampling method (for example a two-stage cluster sampling design).
2. Derive the Estimate $\widehat{\theta}_i$ of the area i specific M-quantile coefficient θ_i as the sample average of the estimated M-quantile coefficients for that area; otherwise it is set $\widehat{\theta}_i = 0.5$.
3. Compute the corresponding M-quantile predictor of the average $\bar{y}_i$ in small area i as $\widehat{\bar{y}}_i^{MQ} = N_i^{-1} \left\{\sum_{j \in s_i} y_{ij} + \sum_{j \in r_i} \widehat{Q}_{\widehat{\theta}_i}(\mathbf{x}_{ij}; \psi)\right\}$. If y is binary, model the regression M-quantiles of order q by 4.

3.4.2 Estimating the mean squared error

In this study, we estimate the mean squared error (MSE) of the proposed estimator using the approach by Chambers et al. (2014) based on the linearisation approach. In essence, the assumption is that the working model used in concluding influences the obtained values from the area under study. Therefore, the MSE of interest is relied upon it and is equal to a conditional prediction variance plus a squared conditional prediction bias. Further, the estimated area level M-quantile coefficient values are assumed as having some slight variations and can be considered as fixed. According

to Chambers et al. (2016), a first order approximation to the conditional prediction variance of $\widehat{\bar{y}}_i^{MQ}$ is then

$$
\text{Var}\left(\widehat{\bar{y}}_i^{MQ} - \bar{y}_i \mid \theta_i\right) = N_i^{-2} \left\{ \text{Var}\left[\sum_{j \in r_i} \widehat{Q}_{\theta_i}\left(\text{x}_j; \psi\right)\right] + \sum_{j \in r_i} \text{Var}\left(y_j\right) \right\}
$$
$$
\approx N_i^{-2} \left\{ \left[\sum_{j \in r_i} Q_{\theta_i}\left(\text{x}_j; \psi\right) \mathbf{x}_j^T\right] \text{Var}\left(\widehat{\boldsymbol{\beta}}_{\theta_i}\right) \left[\sum_{j \in r_i} Q_{\theta_i}\left(\text{x}_j; \psi\right) \mathbf{x}_j^T\right]^T + \sum_{j \in r_i} \text{Var}\left(y_j\right) \right\},
$$

(5)

which can be estimated by

$$
\widehat{\text{Var}}\left(\widehat{\bar{y}}_i^{MQ}\right) =
$$
$$
N_i^{-2} \left\{ \left[\sum_{j \in r_i} \widehat{Q}_{\widehat{\theta}_i}\left(\text{x}_j; \psi\right) \mathbf{x}_j^T\right] \widehat{\text{Var}}\left(\widehat{\boldsymbol{\beta}}_{\widehat{\theta}_i}\right) \left[\sum_{j \in r_i} \widehat{Q}_{\widehat{\theta}_i}\left(\text{x}_j; \psi\right) \mathbf{x}_j^T\right]^T + \sum_{j \in r_i} \widehat{\text{Var}}\left(y_j\right) \right\}.
$$

(6)

In this case $\widehat{\text{Var}}\left(\widehat{\boldsymbol{\beta}}_{\widehat{\theta}_i}\right)$ is a sandwich-type estimator. The $\widehat{\text{Var}}\left(y_j\right)$ can be calculated either by using the sample data from area i, $\widehat{\text{Var}}\left(y_j\right) = \widehat{\bar{y}}_i\left(1 - \widehat{\bar{y}}_i\right)$, or by pooling data from the entire sample, in which case $\widehat{\text{Var}}\left(y_j\right) = \widehat{\bar{y}}(1 - \widehat{\bar{y}})$. According to Chambers et al. (2016) the pooled estimator should lead to more stable prediction variance estimates when area sample sizes are very small and the conditional prediction bias can be approximated using the results of Copas (1988) as

$$
E\left(\widehat{\bar{y}}_i^{MQ} - \bar{y}_i \mid \theta_i\right) \approx
$$
$$
-\frac{1}{2N} \left\{ \frac{\partial}{\partial \boldsymbol{\beta}_{\theta_i}} \Psi\left(\boldsymbol{\beta}_{\theta_i}\right) \right\}^{-1} \left\{ \text{tr}\left[\left\{\frac{\partial}{\partial \boldsymbol{\beta}_{\theta_i} \partial \boldsymbol{\beta}_{\theta_i}^T} \Psi\left(\boldsymbol{\beta}_{\theta_i}\right)\right\} \text{Var}\left(\widehat{\boldsymbol{\beta}}_{\theta_i}\right)\right] \right\} \left\{ \frac{\partial}{\partial \boldsymbol{\beta}_{\theta}} \sum_{j \in r_i} Q_{\theta_i}\left(\text{x}_j; \psi\right) \right\},
$$

(7)

with corresponding plug-in estimator

$$
\widehat{\text{Bias}}\left(\widehat{\bar{y}}_i^{MQ}\right) = -\frac{1}{2N} \left\{ \frac{\partial}{\partial \boldsymbol{\beta}_{\theta_i}} \Psi\left(\boldsymbol{\beta}_{\theta_i}\right) \bigg|_{\boldsymbol{\beta}_i = \widehat{\boldsymbol{\beta}}_{\widehat{\theta}_i}} \right\}^{-1}
$$
$$
\left\{ \text{tr}\left[\left\{\frac{\partial}{\partial \boldsymbol{\beta}_{\theta_i} \partial \boldsymbol{\beta}_{\theta_i}^T} \Psi\left(\boldsymbol{\beta}_{\theta_i}\right) \bigg|_{\boldsymbol{\beta}_{\theta_i} = \widehat{\boldsymbol{\beta}}_{\widehat{\theta}}}\right\} \widehat{\text{Var}}\left(\widehat{\boldsymbol{\beta}}_{\widehat{\theta}_i}\right)\right] \right\} \left\{ \frac{\partial}{\partial \boldsymbol{\beta}_{\theta_i}} \sum_{j \in r_i} Q_{\theta_i}\left(\mathbf{x}_j; \psi\right) \bigg|_{\boldsymbol{\beta}_{\theta_i} = \widehat{\boldsymbol{\beta}}_{\widehat{\theta}_i}} \right\}.
$$

(8)

The estimator of the conditional mean squared error of $\widehat{\overline{y}}_i^{MQ}$ is then

$$\text{Mse}\left(\widehat{\overline{y}}_i^{MQ}\right) = \widehat{\text{Var}}\left(\widehat{\overline{y}}_i^{MQ}\right) + \left\{\widehat{\text{Bias}}\left(\widehat{\overline{y}}_i^{MQ}\right)\right\}^2. \tag{9}$$

4 Results

In this section we present the results of estimating health insurance coverage in Kenya at the county level. The respondents were asked the question; Are you covered by any health insurance? Therefore the response is a binary variable coded as 0 (No)
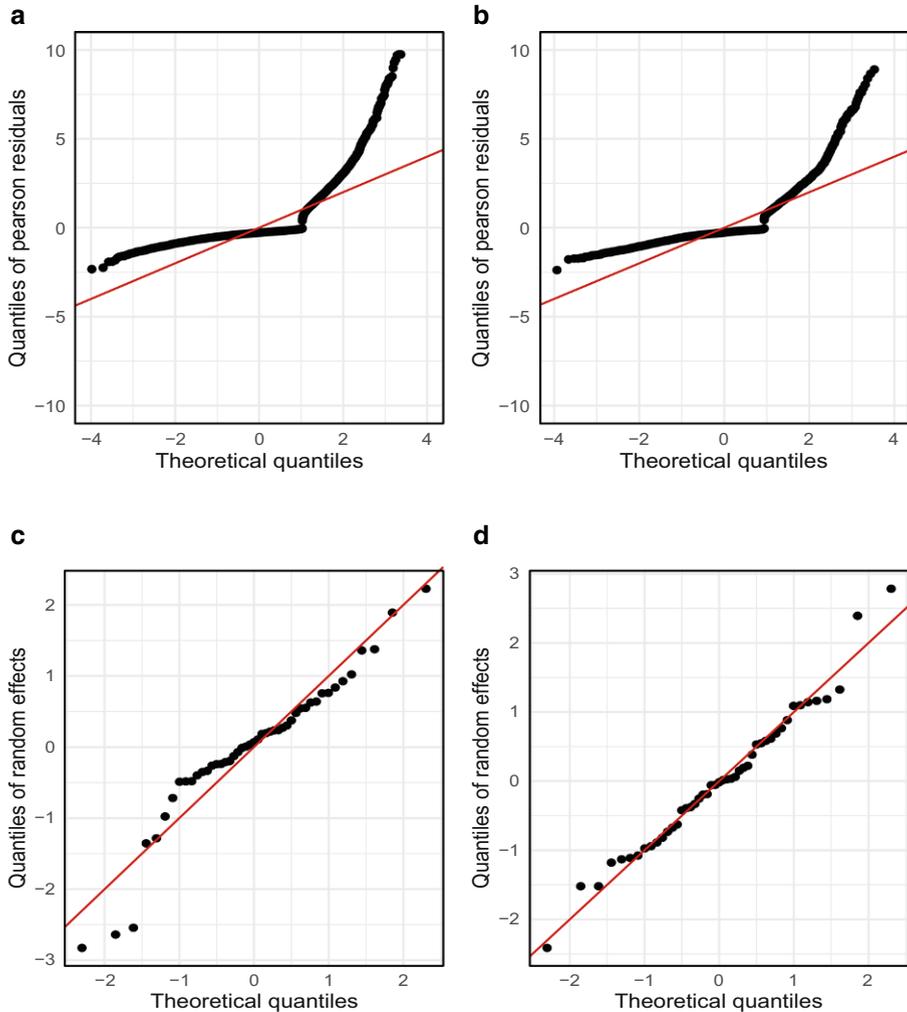
Fig. 1 QQ-plots for Pearson residuals for **a** women **b** men and random effects for **c** women **d** men of health insurance coverage at the county level in Kenya

Table 7 Estimated fixed effects coefficients, variance components for the random effect and likelihood ratio test from fitting a generalized linear mixed model to health insurance data from Kenya

	Women				Men			
	Estimate	Std. Error	z value	Pr(> |z|)	Estimate	Std. Error	z value	Pr(> |z|)
Intercept	−3.87664	0.37817	−10.25100	0.00000	−3.36832	0.29500	−11.41800	0.00000
Age	0.03095	0.00338	9.14900	0.00000	0.03077	0.00408	7.54800	0.00000
RelationToHead=2	−0.42258	0.06806	−6.20900	0.00000	0.47509	0.23885	1.98900	0.04669
RelationToHead=3	−0.57801	0.09281	−6.22800	0.00000	−0.47476	0.09738	−4.87500	0.00000
EmploymentStatus=1	−0.12982	0.05279	−2.45900	0.01393	0.17550	0.10854	1.61700	0.10588
EducationLevel=2	0.96049	0.06132	15.66400	0.00000	1.47206	0.05942	24.77400	0.00000
EducationLevel=3	1.63818	0.07047	23.24700	0.00000	−1.12156	0.26877	−4.17300	0.00003
Residence=1	0.73861	0.05487	13.46000	0.00000	−0.51425	0.05928	−8.67500	0.00000
MaritalStatus=2	−0.41803	0.09746	−4.28900	0.00002	−0.52183	0.09138	−5.71100	0.00000
MaritalStatus=3	−1.09058	0.15780	−6.91100	0.00000	−1.37671	0.54140	−2.54300	0.01099

Table 7 (Continued)

	Women				Men			
	Estimate	Std. Error	z value	Pr(> |z|)	Estimate	Std. Error	z value	Pr(> |z|)
MaritalStatus=4	−0.66043	0.19122	−3.45400	0.00055	−1.28689	0.31528	−4.08200	0.00004
MaritalStatus=5	−0.90623	0.13709	−6.61000	0.00000	−1.32604	0.20376	−6.50800	0.00000
Region=2	−1.08478	0.61113	−1.77500	0.07589	−0.37307	0.41679	−0.89500	0.37073
Region=3	0.60538	0.46117	1.31300	0.18928	0.55371	0.29280	1.89100	0.05861
Region=4	1.47144	0.51299	2.86800	0.00413	0.85915	0.32349	2.65600	0.00791
Region=5	0.54445	0.41713	1.30500	0.19182	0.66483	0.26522	2.50700	0.01219
Region=6	0.24644	0.55106	0.44700	0.65473	0.47969	0.34717	1.38200	0.16706
Region=7	0.98432	0.49054	2.00700	0.04479	0.56304	0.30997	1.81600	0.06930
Region=8	1.09669	0.90439	1.21300	0.22527	0.93010	0.55562	1.67400	0.09413
Variance component	0.6829	LRT = 1087.248	Pr(> χ^2)	0.00000	0.2448	LRT= 1111.015	Pr(> χ^2)	0.00000

and 1 (Yes). We first fitted a binary logistic generalized linear mixed model (GLMM) and show that the assumption of normality of random effects is not met.

4.1 Initial analysis using binary logistic GLMM

To begin with we first fitted a binary logistic GLMM with normally distributed random effects using the function glmer in R package lme4 (Bates et al. 2015). Plots a and b from Fig. 1 represent the QQ plots for Pearson residuals obtained from fitting a logistic GLMM for women and men respectively. Within the same fitted model, the random effects for women and men were obtained are also displayed in plots c and d respectively. The Pearson residuals are not normally distributed. Although the random effects show normality especially for the men data, there is a slight departure from the tails. The Shapiro-Wilk normality test using significance level of 0.05 does not reject the null hypothesis that the random effects are normally distributed for men (p-value $= 0.591$) but it does for women (p-value $= 0.00473$).

Table 7 shows the estimated model parameters, standard errors and corresponding *p-values* for women and men. The fixed effects are age (15–49) years, relationship to household head (1=head, 2=spouse, 3=others), employment status (0=unemployed, 1=employed), education level (1=completed primary, 2=secondary school and above, 3=no formal schooling), residence (0=rural, 1=urban), marital status (1=never married, 2=married, 3=widowed, 4=divorced, 5=seperated), region (1=Coast, 2=North Eastern, 3=Eastern, 4=Central, 5=Rift Valley, 6=Western, 7=Nyanza, 8=Nairobi). The regression coefficient of age has a positive sign (for both women and men), implying age increases the probability of access to health insurance. The table also shows the variance component for the random part of the model. To test whether the variance components are significant to measure unobserved heterogeneity we use the Likelihood Ratio Test (LRT). For women, the test statistic is 1,087.248, with a *p-value* of 0.0000, and for men equals 1,111.015, *p-value* $= 0.0000$. Therefore we reject the null hypothesis of no significance and conclude there is evidence of significant unobserved heterogeneity.

4.2 Binary M-quantile modeling

The diagnostic plots in Sect. 4.1 show that the model assumptions of GLMM are not met. According to Chambers et al. (2016), GLMM's have attractive properties that can be used to model binary response variables. However, using GLMM's in SAE is not straightforward since the estimation of model parameters can be numerically demanding. Apart from computational complexity for using GLMM in small area estimation, if model are not met, invalid conclusions could be obtained. To reduce the adverse effects from deviations from distributional assumptions (provide robust inference), while at the same time borrowing strength from domains, we explore the use of M-quantile small area estimation model. Robust in this case means the estimator is reasonably efficient and unbiased, small deviations from model assumptions do not substantially affect the model performance and large deviations will not totally invalidate the model entirely. Since SAE via M-quantile uses M-quantile coefficients as opposed to random effects in GLMM, we find the
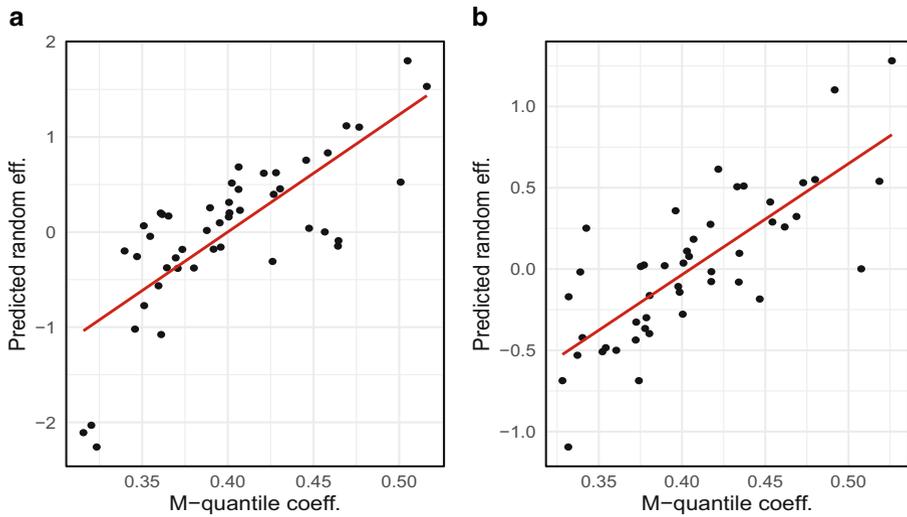
Fig. 2 Scatter plots for the predicted random effects (estimated with GLMM) and the M-quantile coefficients (estimated with the M-quantile model) for **a** women **b** men at the county level in Kenya

correlation between the predicted area effects and area-level M-quantile coefficients. This was suggested by Chambers et al. (2016). The correlation equals 0.769 for women model and 0.77 for men. Fig. 2 visualizes the scatter plots between the predicted random effects from mixed model and M-quantile model.

There is a high correlation between the area level M-quantile coefficients and the predicted area effects from GLMM to capture area variability. However, according to Chambers et al. (2016) M-quantiles provides an alternative SAE method when GLMM assumptions are not met. Table 8 shows quantiles of the point estimates of the proportion of persons with health insurance at the county level in Kenya. On average the mean and median for both women and men for direct and MQ estimates are comparable. A higher proportion of men are covered with health insurance compared to women. Overall, these proportions are quite low despite the efforts put by the government. This finding implies the government should explore other better options to increase coverage.

Fig. 3 shows smooth maps of health insurance coverage for the 47 counties in Kenya using M-quantile estimation. The observed distribution is similar for women and men. Counties with highest coverage for women are Bomet (24%), Embu

Table 8 Distribution of health insurance coverage proportions over counties in Kenya for women and men aged 15–49 years.

| Gender	Estimator	Min.	Q1.	Median	Mean	Q3.	Max.
Women	Direct	0.0127	0.0891	0.1467	0.1518	0.1987	0.3654
	MQ	0.1033	0.1278	0.1478	0.1559	0.1724	0.2403
Men	Direct	0.0291	0.1001	0.1673	0.1731	0.2414	0.4193
	MQ	0.1042	0.1311	0.1548	0.1611	0.1804	0.2715
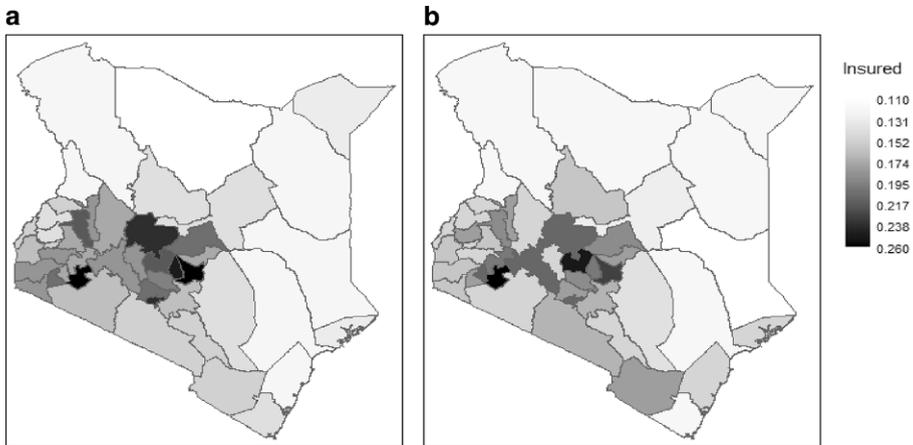
Fig. 3 Maps showing the proportion covered with health insurance for **a** women and **b** men for the 47 counties in Kenya

(23.6%), Kirinyaga (22.8%) and Nairobi (22%). The five least covered counties are West Pokot (10.9%), Turkana (10.8%), Garissa (10.7%) and Marsabit (10.3%). For men the leading counties are; Nairobi (27.2%), Bomet (24.8%), Nyeri (23.4%), Nyamira (22.2%) while the least covered counties are; Mandera (11%), Tana River (10.6%), Garissa (10.6%) and Kwale (10.4%). From the findings, we note that counties neighboring Nairobi have higher coverage rates. Since these counties are close to Nairobi which is the capital city of Kenya, with more employment opportunities, people living here are able to afford health insurance premiums. This is in contrast to counties further away like Turkana and Garissa. These results have been possible with the use of SAE methodology.

4.3 Evaluation of the M-quantile SAE model estimates

We evaluate the model-based results based on three criteria: (i) smaller MSE and CV for MQ compared to direct estimates, where the MSE is the sum of the variance and bias squared of the estimator, while the CV measures the dispersion of the estimates around the mean. (ii) consistency and (iii) usefulness to users. This has been proposed by Brown et al. (2001). The same approach has been used by Chandra et al. (2018) when estimating poverty incidence in the state of Bihar in India.

For MSE, Fig. 4a and b are the box plots of estimated MSEs of the estimated health insurance coverage for women and men. The MSE for both women and men are smaller for MQ compared to direct estimates. For CV's, Table 9 shows quantiles of the coefficient of variation for the estimated health insurance percentages at county level in Kenya. For direct estimates especially for counties with small samples, the CV's reach values greater than 60% for both women and men. For MQ estimates all the CV's were less than 20. Since Kenya has not set a guideline for publishing official statistics, we use other statistical offices like the UK's Office for National
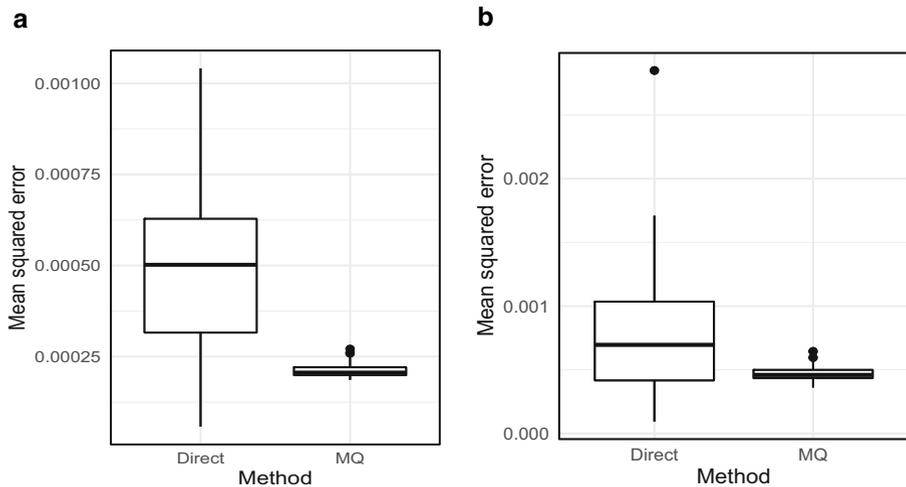
Fig. 4 Box-plots showing the mean squared error for the distribution of health insurance coverage percentages at the county in Kenya for **a** women **b** men

Statistics (ONS). They set a CV of 20% for publishable official statistics. Therefore, our estimates meet this cut-off.

Fig. 5 are line graphs with counties ordered by increasing sample sizes. To start with as expected the CV's for MQ estimates are smaller than direct estimates for all counties. As the sample sizes per county increase, the CV's for direct estimates reduce, especially for the women sample. By contrast, however, the CVs for direct estimates do not reduce with increasing sample size for men. For this analysis, the samples ranged from 236 to 460 for women and from 118 to 370 for men.

For bias-diagnostics, Fig. 6 is a scatter plot comparing direct estimates of the proportion of persons covered with health insurance and corresponding M-quantile estimates. According to Brown et al. (2001) the plots is based on the idea that, if the model-based estimates are "close" to the model-based SAE values of interest, then unbiased direct estimators should behave like random variables whose expected values correspond to the values of the model-based estimates. That is, the model-based estimates should be unbiased predictors of the direct estimates. To check this, we plot appropriately scaled values of these estimates (x-axis) against similarly scaled direct estimates (y-axis) and then test whether the OLS (ordinary least squares) regression line fitted to these points is significantly different from

Table 9 Quantiles of the coefficient of variation for the estimated health insurance percentages at county level in Kenya

| Gender	Estimator	Min.	Q1.	Median	Mean	Q3.	Max.
Female	Direct	0.073	0.120	0.157	0.186	0.215	0.634
	MQ	0.101	0.130	0.148	0.146	0.160	0.204
Male	Direct	0.078	0.126	0.150	0.188	0.221	0.670
	MQ	0.068	0.083	0.092	0.095	0.103	0.136
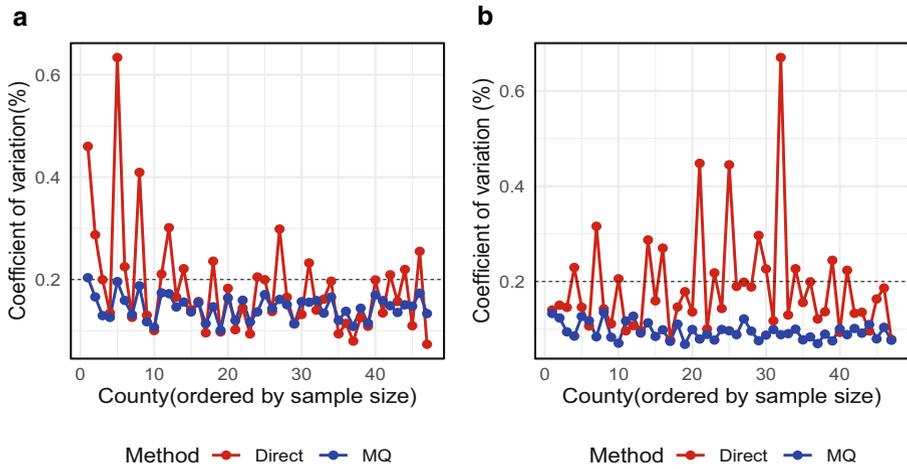
Fig. 5 Line graphs showing coefficient of variation of health insurance coverage with counties ordered by increasing sample sizes for: **a** women **b** men

the identity line. To check homoscedasticity assumption required for OLS fitting, we ran the Goldfeld-Quandt test. Under the null hypothesis, the Goldfeld-Quandt test statistic follows an F distribution with degrees of freedom as specified in the parameter. For men (GQ = 0.39191, df1 = 22, df2 = 21, p-value = 0.983), and women (GQ = 0.44733, df1 = 22, df2 = 21, p-value = 0.9662) we do not reject the null hypothesis and conclude heteroscedasticity is not present. Therefore, the homoscedasticity assumption is satisfied. We note that the M-quantile estimates are generally consistent with the direct estimates. Even though the model-based results depict some bias, the aggregated results in Table 10 are close.
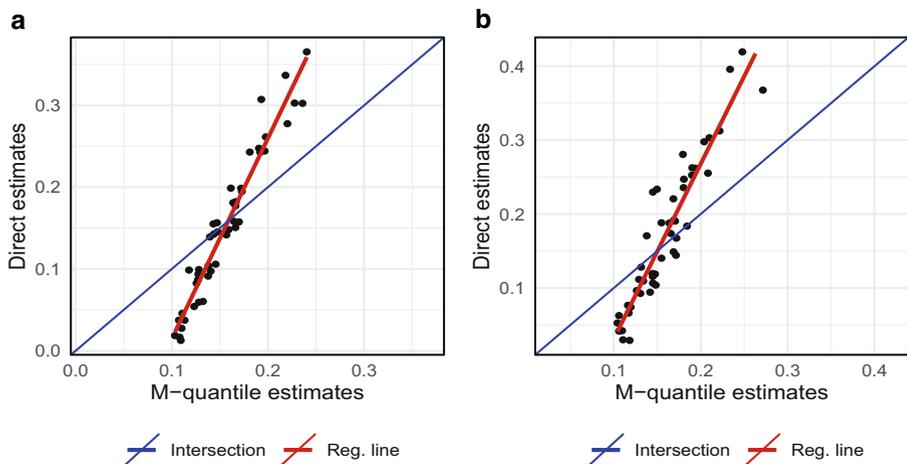
Fig. 6 Scatter plots with regression and intersection line ($y = x$) comparing direct and M-quantile health insurance coverage estimates for **a** women **b** men at the county level in Kenya

Table 10 Aggregated direct and MQ estimates at the national level

Gender	Estimator	Proportion
Women	Direct	0.1546
	MQ	0.1573
Men	Direct	0.1796
	MQ	0.1628

For usefulness to users, we still adopt the criteria proposed by Brown et al. (2001). Accordingly, in SAE applications aggregation or bench marking of small area estimates at higher level is always desirable by the users. In small area applications, National statistical offices involved in generating the small area estimates always expect that the small area estimates are aggregated/bench marked to higher level estimate. At higher level of aggregation, the direct estimates are considered to be reliable and therefore the model-based small area estimates are expected to be near to the direct estimates when they are aggregated. We checked the aggregation of model-based small area estimates at the county level. We computed national level insurance coverage by aggregating the direct estimates and MQ small area estimates, as $\sum_i (n_i.\text{Direct}_i)/\sum_i n_i$ and $\sum_i (n_i.MQ_i)/\sum_i n_i$, respectively. Table 10 shows the aggregated estimates. The MQ estimates aggregate well to national level direct estimate.

5 Concluding remarks

In conclusion, health insurance reduces health care costs by pooling resources. It is an important component towards achieving UHC. Assessing health insurance coverage for policymaking requires reliable data especially at disaggregated levels. In this study, we have combined survey and census data through an M-quantile model to get better estimates when sample sizes are small. This has the advantage that we avoid specifying random effects while providing robust inference against deviations from model assumptions. Findings show model-based estimates have smaller MSE's and CV's than direct estimates. Health insurance coverage remains low overall in Kenyan counties. Among those covered, our findings show inequality in health insurance coverage across the wealth quintile with the highest coverage being the richer and richest, especially for private insurance which requires monthly contributions. Health insurance in Kenya is mostly voluntary except for public and civil servants. The majority of Kenyans also work in the informal sector where health insurance is not compulsory. With the current voluntary health insurance scheme, health insurance coverage remains low. Kenya should establish a mechanism mainly funded by taxation to extend prepaid coverage to its population. Despite financial constraints, Kenya should provide total subsidy to the poor through NHIF. Further, Kenya should give a partial insurance subsidy, through the NHIF to people within the informal sector. Two possible directions for further research are a) to allow for more disaggreaged domains like the sub-county level and b) to incorporate additional predictors of health insurance coverage like geospatial data. A limitation of this study is the time difference between the survey and census data. Data collected

around the same time might yield to more accurate results. Despite the limitation, this study has estimated the health insurance coverage at the county level in Kenya with better precision compared to direct estimates. It has been possible to establish the variation in health insurance coverage between counties, noting that counties neighboring Nairobi have more proportions of persons with health insurance.

Conflict of interest The author declares no conflict of interest.>

References

Aikins M, Tabong PT-N, Salari P, Tediosi F, Asenso-Boadi FM, Akweongo P (2021) Positioning the national health insurance for financial sustainability and universal health coverage in Ghana: a qualitative study among key stakeholders. PLoS ONE 16(6):e253109

Bates D, Mächler M, Bolker B, Walker S (2015) Fitting linear mixed-effects models using lme4. J Stat Soft 67(1):1–48

Breckling J, Chambers R (1988) M-quantiles. Biometrika 75(4):761–771

Brown G, Chambers R, Heady P, Heasman D (2001) Evaluation of small area estimation methods. An application to unemployment estimates from the UK LFS. Proceedings of statistics Canada symposium

Cantoni E, Ronchetti E (2001) Robust inference for generalized linear models. J Am Stat Assoc 96(455):1022–1030

Chambers R, Tzavidis N (2006) M-quantile models for small area estimation. Biometrika 93(2):255–268

Chambers R, Chandra H, Salvati N, Tzavidis N (2014) Outlier robust small area estimation. J R Stat Soc Series B Stat Methodol 76(1):47–69

Chambers R, Salvati N, Tzavidis N (2016) Semiparametric small area estimation for binary outcomes with application to unemployment estimation for local authorities in the UK. J R Stat Soc Ser A Stat Soc 179(2):453–479

Chandra H, Aditya K, 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(6):e198502

Copas JB (1988) Binary regression models for contaminated data. J R Stat Soc Series B Stat Methodol 50(2):225–253

Cotlear D, Rosemberg N (2018) Going universal in Africa: how 46 African countries reformed user fees and implemented health care priorities. World Bank

Croft TN, Marshall AM, Allen CK, Arnold F, Assaf S, Balian S et al (2018) Guide to DHS statistics

Dye C, Boerma T, Evans D, Harries A, Lienhardt C, McManus J, Pang T, Terry R, Zachariah R (2015) Research for universal health coverage: world health report 2013. Sci Transl Med 199(5):199ed13

Farrell PJ, MacGibbon B, Tomberlin TJ (1997) Empirical Bayes small-area estimation using logistic regression models and summary statistics. J Bus Econ Stat 15(1):101–108

Ghosh M, Rao JN (1994) Small area estimation: an appraisal. Stat Sci 9(1):55–76

Ghosh M, Natarajan K, Stroud T, Carlin BP (1998) Generalized linear models for small-area estimation. J Am Stat Assoc 93(441):273–282

Government of Kenya (2010) The 2009 Kenya population and housing census: volume 1C: population distribution by age, sex, and administrative units. Technical report. Kenya National Bureau of Statistics

Government of Kenya (2013) The constitution of Kenya: 2010. Chief Registrar of the Judiciary

Government of Zambia (2018) National health insurance act 2018. https://www.moh.gov.zm/?page_id=5620. Accessed 20 Jan 2022

Horvitz D, Thompson D (1952) A generalization of sampling without replacement from a finite universe. J Am Stat Assoc 47(260):663–685

Huber PJ (1992) Robust estimation of a location parameter. Breakthroughs in statistics. Springer, pp 492–518

Ilinca S, Di Giorgio L, Salari P, Chuma J (2019) Socio-economic inequality and inequity in use of health care services in Kenya: evidence from the fourth Kenya household health expenditure and utilization survey. Int J Equity Health 18(1):1–13

Jiang J, Lahiri P (2001) Empirical best prediction for small area inference with binary data. Ann Inst Stat Math 53(2):217–243

Jiang J, Lahiri P (2006) Mixed model prediction and small area estimation. Test 15(1):1–96

Kazungu JS, Barasa EW (2017) Examining levels, distribution and correlates of health insurance coverage in Kenya. Trop Med Int Health 22(9):1175–1185

Kenya National Bureau of Statistics (2015) Kenya demographic and health survey 2014. Technical report. Kenya National Bureau of Statistics

Kenya National Bureau of Statistics (2019) The 2019 Kenya population and housing census: population by county and sub-county. Technical report. Kenya National Bureau of Statistics

Koenker R, Bassett G Jr (1978) Regression quantiles. Econometrica. https://doi.org/10.2307/1913643

Koenker R, Hallock KF (2001) Quantile regression. J Econ Perspect 15(4):143–156

Kreutzmann A-K, Pannier S, Rojas-Perilla N, Schmid T, Templ M, Tzavidis N (2019) The R package emdi for estimating and mapping regionally disaggregated indicators. J Stat Soft 91(7):1–33

Kusi A, Enemark U, Hansen KS, Asante FA (2015) Refusal to enrol in Ghana's national health insurance scheme: is affordability the problem? Int J Equity Health 14(1):1–14

Lagomarsino G, Garabrant A, Adyas A, Muga R, Otoo N (2012) Moving towards universal health coverage: health insurance reforms in nine developing countries in Africa and Asia. Lancet 380(9845):933–943

Liu B, Lahiri P, Kalton G (2007) Hierarchical Bayes modeling of survey-weighted small area proportions. Proceedings of the American Statistical Association. Survey Research Section, pp 3181–3186

MacGibbon B, Tomberlin TJ (1987) Small area estimates of proportions via empirical Bayes techniques, pp 341–346

Maina T, Kirigia D (2015) Annual evaluation of the abolition of user fees at primary healthcare facilities in Kenya. Technical report. Futures Group, Health Policy Project, Washington, DC

Malec D, Sedransk J, Moriarity CL, LeClere FB (1997) Small area inference for binary variables in the national health interview survey. J Am Stat Assoc 92(439):815–826

Ministry of Health, Kenya (2014) 2013 Kenya household health expenditure and utilisation survey. Technical report. Government of Kenya

Ministry of Health, Kenya (2018) CS health launches UHC pilot registration. https://www.health.go.ke/cs-health-launches-uhc-pilot-registration-machakos-kenya-november-10-2018/. Accessed 10 Oct 2021

Morales D, Esteban MD, Pérez A, Hobza T (2021) A course on small area estimation and mixed models: methods, theory and applications in R. Springer Nature

Mwaura RN, Barasa E, Ramana GN, Coarasa J, Rogo K (2015) The path to universal health coverage in Kenya: the path to universal health coverage in Kenya : repositioning the role of the national hospital insurance fund. https://openknowledge.worldbank.org/handle/10986/23485. Accessed 15 Jan 2022

Mwenda N, Nduati R, Kosgey M, Kerich G (2021) What drives outpatient care costs in Kenya?an analysis with generalized estimatingequations. Front Public Health 8:648465

Nandram B, Sedransk J, Pickle L (1999) Bayesian analysis of mortality rates for US health service areas. Sankhyā 61(Series B, Pt 1):145–165

Newey WK, Powell JL (1987) Asymmetric least squares estimation and testing. Econometrica. https://doi.org/10.2307/1911031

Okungu V, Chuma J, McIntyre D (2017) The cost of free health care for all Kenyans: assessing the financial sustainability of contributory and non-contributory financing mechanisms. Int J Equity Health 16(1):1–13

Otieno PO, Wambiya EOA, Mohamed SF, Donfouet HPP, Mutua MK (2019) Prevalence and factors associated with health insurance coverage in resource-poor urban settings in Nairobi, Kenya: a cross-sectional study. BMJ Open 9(12):e31543

Pfeffermann D (2002) Small area estimation-new developments and directions. Int Stat Rev 70(1):125–143

Pfeffermann D (2013) New important developments in small area estimation. Stat Sci 28(1):40–68

Pratesi M (2016) Analysis of poverty data by small area estimation. John Wiley & Sons

Rao JNK, Molina I (2015) Small area estimation. Wiley Online Library

Sambo LG, Kirigia JM, Orem JN (2013) Health financing in the African region: 2000–2009 data analysis. Int Arch Med 6(1):1–17

Särndal C-E, Swensson B, Wretman J (2003) Model assisted survey sampling. Springer Science & Business Media

Tzavidis N, Zhang L-C, Luna A, Schmid T, Rojas-Perilla N (2018) From start to finish: a framework for the production of small area official statistics. J R Stat Soc Ser A Stat Soc 181(4):927–979

United Nations General Assembly (2015) Transforming our world: The 2030 agenda for sustainable development. Technical report. United Nations

Wang H, Otoo N, Dsane-Selby L (2017) Ghana national health insurance scheme: improving financial sustainability based on expenditure review. The World Bank

World Health Organization (2005) Sustainable health financing, universal coverage and social health insurance. Technical report. World Health Organization

World Health Organization (2013) Research for universal health coverage: world health report 2013. Technical report. World Health Organization

World Health Organization et al (2010) The Abuja declaration: ten years on. Technical report. World Health Organization

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.