DISTRICT ESTIMATES OF HOME DELIVERIES IN GHANA: A SMALL AREA ANALYSIS USING DHS AND CENSUS DATA

Fiifi Amoako Johnson, James J. Brown & Sabu S. Padmadas

Correspondence:
Fiifi Amoako Johnson
Division of Social Statistics
Southampton Statistical Sciences Research Institute
University of Southampton
Highfield Campus
SO17 1BJ
United Kingdom
Email: faj100@soton.ac.uk
Tel:
Fax:

Abstract

National estimates show that home deliveries under the supervision of unskilled attendants are high in Ghana. Spatial inequalities across administrative regions, where health planning and monitoring is focal, are not known. This study aims to derive district-level estimates of home deliveries and assess spatial variations between districts, using data from the 2000 Ghana census and the 2003 DHS. The empirical best linear unbiased prediction (EBLUP) extension of the Fay-Herriot model for small area analysis is used to produce estimates for the 110 districts of Ghana. Most of the districts in the Savannah zone, where maternal mortality is very high, have the highest proportion of home deliveries. Home deliveries range from 73% to 95% in districts of the Northern Region compared to 13% to 22% for districts in the Greater Accra Region. Using five diagnostic procedures, the reliability of the estimates were assessed. The estimates show significant clustering effect in delivery care uptake across districts that warrant policy and programme attention.

1.0 Introduction

The importance of administrative level statistics to aid policy decisions have been well emphasised. However, in sub-Saharan Africa were decentralisation has been herald as one of the main strategies to enhance development; administrative level statistics are almost non-existent. Census from which such statistics could be derived are limited in the amount of information they collect and are also becoming less regular due the high cost involved. Surveys which have become more regular and collect a substantially large amount of information cannot be used to derive administrative level statistics due to the small samples from administrative areas.

In recent times survey statisticians have developed small area estimators that borrow strength from auxiliary information such as census and administrative registers to derive small area statistics (administrative areas) from surveys. Recent developments in small area estimation techniques and computer software for analysing small area data and the availability of relevant demographic and health data provide an opportunity to investigate the derivation of local-level estimates of demographic and health indicators. However, this has not received much attention in demographic research.

This study thus using survey data from the 2003 Ghana Demographic and Health Survey and auxiliary information from the 2000 Ghana Population and Housing Census (GPHC) aims at deriving district estimates of home deliveries in Ghana. Valid measures of error are also used to access the validity of the estimates. This analysis is important in aiding central government allocation of health funds and local government and health practitioner's implementation, monitoring and evaluation of maternal health activities.

Assembling of Data

To derive reliable model-based small area estimates two categories of variables are usually required.

Y - variables: variables (dependent variable) recorded in the survey for which small area/domain estimates are required.

X - variables: variables that are not recorded in the survey but which current or past values of small areas or domain totals and sizes are known.

The *Y* variable for this study is the proportion of home deliveries in a district, recorded in the 2003 GDHS. The *X* variables (auxiliary variables derived from the 2000 GPHS) used for deriving the estimation model is shown in Table 1 below. The auxiliary information is assumed to explain part of the between area variability. That is, if survey data is only available for a limited number of women in a district, the auxiliary information is known for all individuals in the district. The composite indices were derived via principal component analysis. The composite score for access to health services shows that the Greater Accra region has the highest mean score for access to health services, while the Northern region has the lowest mean score, but with a high variability. The composite score for socio-economic development shows that districts within the Greater Accra region have the highest socio-economic development, while those within the three Northern regions (Savannah zone) have the lowest socio-economic development.

In order to link the survey data (GDHS) to the covariates from the census data, there is the need to make a connection between the locations of the sampled PSU districts and its spatial location in the districts. Since the 2000 GPHC was the sample frame for the 2003 GDHS, the spatial co-ordinates of each of the PSUs could be obtained. Using the spatial-coordinates provided by the Ghana Statistical Service, the 412 PSUs sampled in the 2003 GDHS were placed within a specified administrative district and the statistics about these areas were linked to the survey data as auxiliary information. The survey information is restricted to 2,757 women who had a birth in the five years preceding the 2003 GDHS. Only place of delivery of the last child was considered in this study due to the high dependence in uptake of delivery care for successive pregnancies.

Table1: Auxiliary Information

Variables	Categorization	Descriptive statistics
Demographic indicators		
Age category (*i,j*)	Age 15 – 19 years	
	Age 20-34 years	
	Age 35+ years	
Population structure (*j*)	Continuous	Min 39.0%
Proportion of women of reproductive age		Max 57.7%
in a district		Mean 45.2%
		SD 0.032
Total fertility rate *(j)*	Continuous	Min 2.18
		Max 9.78
		Mean 4.44
		SD 1.16
Access to health facilities (j)		
Index of access to health services	Continuous	Min -4.629
A composite index of access to		Max 1.300
traditional health facilities, hospital and		Mean 1.000
clinics derived using principal		SD 0.000
component analysis		
Socio-economic development (j)		
Index of socio-economic development	Continuous	Min -1.747
A composite index of level of literacy,		Max 3.063
employment, educational status,		Mean 1.000
urbanization and proportions employed		SD 0.000
in the different sectors of the economy		
Administrative region (*j*)		Number of districts
	Western[3]	11
	Central	12
	Greater Accra	5
	Volta	12
	Eastern	15
	Ashanti	18
	Brong Ahafo	13
	Northern	13
	Upper East	6
	Upper West	5

Domain Classification

At the time of conducting the 2000 GPHC and 2003 GDHS, Ghana was demarcated into 110 metropolitan, municipal and district assemblies (3 metropolitan authorities, 4 municipal assemblies and 103 district assemblies). Since both census and survey data available for this analysis was based on the 110 districts of Ghana, the analysis is focused on the 110 districts for which data is available. The focus of this analysis is at the district level because it is the level of governance where decision-making and grass roots participation is being encouraged in the decentralization process. Also, all districts were sampled in the 2003 GDHS.

The women in the survey were further categorised into three age-groupings to reflect their reproductive stages. This is aimed at aiding target specific populations according to their needs. Teenagers (15-19 years) are considered to be in their early reproductive stage, women age 20-34 years are assumed to be in their prime reproductive stage and women aged 35 years or above are assumed to be in the latter stage of their reproductive life. Table 2 below illustrates the age-district domain cross-classification. In all 330 age-district domains were derived for this study. Out of the 330 domains derived, 227 (68.0%) have sample sizes less than 10, 90 (27.3%) have sample sizes of between 10 and 30 and only 13 (3.9%) have sample sizes greater than 30.

Table 2: Age-District Domain Classification

	District (*j*)			
Age (*i*)	1	2	…..	110
1 (15-19 years)	(1,1)	(1,2)	…..	(1,110)
2 (20-34 years)	(2,1)	(2,2)	…..	(2,110)
3 (35-49 years)	(3,1)	(3,2)	…..	(3,110)

Methodology

Using the ratio estimator technique which uses only area specific data to derive estimates, it was identified that the ratio estimator technique is not able to produce reliable estimates at the domain level (high Relative Standard Errors). Even at the regional level, the relative standard errors were not encouraging. For some domains, the direct estimate is zero because no sample cases were observed for that domain. In some domains also the direct estimates were zero due to that fact that there were small samples from this domain rather to the non-existence of eligible women in these domains. This is a clear indication that there is the need to use statistical methods that produce local level statistics with adequate precision.

The sample data for this study consists of direct (DHS) estimates of home deliveries and associated age-district covariates from the Ghana Population and Housing Census. If i index age-group and j index district and N_{ij} is the number of women in age group i and district j, with M_{ij} the number of women in age group i and district j that had a home delivery. Then the proportion of women that had home deliveries is given by $z_{ij} = M_{ij} / N_{ij}$. Since, the sample data for the analysis consist of the direct estimate of the proportion of home deliveries, $\hat{N}_{ij}$ is the direct estimate of the number of women in age age-group i and district j. $\hat{M}_{ij}$ is the direct estimate of the number of home deliveries. Then the direct estimate of the proportion of women of who delivered at home is $\hat{z}_{ij} = \hat{M}_{ij} / \hat{N}_{ij}$.

One of the recent breakthroughs in small area estimation is the situation where the variables of interest are categorical and the small area measures of interest are proportions. In this study the aim is to predict the true area proportions and obtain valid measures of error. A convenient way to model the proportions allowing for the inclusion of covariates is to use a generalized linear mixed model. Considering a *logit* link function, the generalized linear mixed model of interest can be defined as:

$$Logit(\pi_{ij}) = x_{ij}'\beta \qquad 1$$

Where z_{ij} is a vector of known age-district attributes. Within the multilevel modelling framework equation 1 can be expressed as:

$$Logit(\pi_{ij}) = x_{ij}'\beta + \mu_{0j} \qquad 2$$

Where μ_{0j} is the random effect, assumed to be normally distributed with mean zero and variance ϕ. Estimates of the model parameters are obtained using an iterative procedure that combines maximum quasi-likelihood (MPQL) estimation of β and μ_{0j} with residual/restricted maximum likelihood (REML) estimation of ϕ (Saei and Chambers 2003). The random area effect accounts for between area variations beyond that explained by the covariates

Two main modelling estimators have been proposed in the literature – (1) the Battese-Fuller estimator and (2) The Fay-Herriot estimator. The Battese-Fuller estimator assumes that only individual level covariates are available. In the case where area-level data is available the Fay-Herriot estimator could be applied. The estimation procedure adopted for this study is the Fay-Herriot Model. The main reason for choosing a Fay-Herriot Model type approach is that the auxiliary information available for this study is at the area level.

If the expected value of $\hat{z}_{ij}$ is given by $E(\hat{z}_{ij} \mid z_{ij}) = p_{ij}$ and the variance of $\hat{z}_{ij}$ $\text{var}(\hat{z}_{ij} \mid z_{ij}) = p_{ij}(1 - p_{ij})/n_{ij}$, then equation 2 above can be used to specify how the characteristics of an area j influences the value of p_{ij}. This model can be fitted to the sample data using standard regression software that accounts for the area effects (equation 2). An estimate $\hat{\beta}$ of β is then estimated from model 2 above. An estimator of the proportion of women in age group i in district j is then given by the EBLUP estimator $\hat{p}_{ij}$ = $antilogit(x'\hat{\beta} + \hat{\mu}_{0j})$. The problem associated with the EBLUP estimator is the complexity of estimating the mean squared error. Practically, the variance component of $\hat{p}_{ij}$ constitute a fixed and a random part which are not estimated independently, thus a covariance structure between the fixed part and the random part of the model is needed to estimate the variance of $\hat{p}_{ij}$. The limitation of the software package available for this analysis is that it does not estimate the covariance structure. A compromise approach, which uses a Taylor series approximation approach to estimate the variance of $\hat{p}_{ij}$ is equation 3 below

$$\operatorname{var}(\hat{p}_{ij}) \approx ((\hat{p}_{ij}(1-\hat{p}_{ij}))^2[\hat{\sigma}_u^2 + x'_{ij}\hat{V}(\hat{\beta})x_{ij}] \qquad 3$$

Once $\hat{p}_{ij}$ has been estimated, for illustration purposes, the proportion home deliveries in district *j* (weighted average) was estimated from equation 4 below.

$$\hat{P}_j = \sum \frac{\hat{p}_{ij} N_{ij}}{N_j} \qquad 4$$

Where N_j is the proportion of the population of women aged *i* in district *j*.

Results

Model Formulation and Performance

A null model (Model 0) controlling for the area effects was initially fitted to determine the basic partitioning of the variability at the district level. Five modelling process were then used to check for the explanatory power of the covariates. Model 1 controlled for the demographic characteristics. The composite score for access to health services was then added to Model 1 to derive Model 2. Model 3 accounted for the score for socio-economic development of the districts in addition to the covariates in Model 2. A dummy indicating the administrative region of the district was added to Model 3 to derive Model 4. Finally, all possible interactions were investigated. The significant interactions were added to Model 4 to derive Model 5. Where a variable was identified not to be significant except age, it was dropped from the model to avoid introduction of errors in the final model. Below is an illustration of the model building process.

Model 0: $X_0 = 1$

Model 1: X_1 = Demographic indicators

Model 2: $X_1 = 1, X_2$ = Access to health services

Model 3: $X_1 = 1, X_2, X_3$ = Socio-economic characteristics

Model 4: $X_1 = 1, X_2, X_3, X_4$ = Administrative region

Model 5: $X_1 = 1, X_2, X_3, X_4, X_5$ = Significant interactions

The performance of the model was checked using the values of $\hat{\sigma}^2_\mu$ as an indication of the explanatory power of the synthetic part of the model. Tables 3 shows that models 1 – 5 are in general better than model 0 in terms of explaining spatial variability, with models 4 and 5 showing the best results. Model 5 was preferred to model 4 due to the inclusion of interaction effects between covariates. The percentage change in $\hat{\sigma}^2_\mu$ shown in Table 3 reveals that region of residence is very important in uptake of delivery care.

Table 3: Values and Percentage Change in $\hat{\sigma}^2_\mu$

Model	$\hat{\sigma}^2_\mu$	% Change in $\hat{\sigma}^2_\mu$
0	1.254 (0.208) *	-
1	1.234 (0.204) *	1.6
2	1.074 (0.182) *	13.0
3	1.074 (0.182) *	13.0[1]
4	0.386 (0.086) *	64.1
5	0.346 (0.080)*	10.4

*Significant *p<0.01

Estimation Model

The estimated coefficients of the estimation model are shown in equation 6 below, with their standard errors shown in parenthesis. The estimated coefficients were used to derive domain estimates. The reliability of the estimates is investigated in section 6.3 below before they are discussed in section 6.4. The estimated coefficients show that maternal age does not predict choice of delivery care. There was a significant interaction between age and access to health services. Access to health services influences young women's uptake of delivery care. The high TFR in most regions of Northern Ghana and low access to health services contributes to the high proportion of home deliveries in the area.

[1] Not significant/no significant interaction

Y_{ij} = 1.21 – 0.13(0.17)*15-19 years* – 0.11(0.10)*20-34 years* – 0.14(0.09) *TFRj* + 0.02(0.11) *IAHS* - 0.020(0.34) *Central* – 2.40 (0.45) *Greater Accra* -0.49(0.34) *Volta* – 0.51(0.32) *Eastern* – 1.07(0.31) *Ashanti* - 0.72(0.31) *Brong Ahafo* – 2.15(1.96) *Northern* + 0.77(0.38) *Upper East* + 0.70(0.41) *Upper West* + 0.31(0.16) *15-19 years *IAHS* + 0.75(0.40) *Northern *TFR* + 0.93(0.50) *Upper West*IAHS*

(μ_{0j})=0.346 (0.080) ***

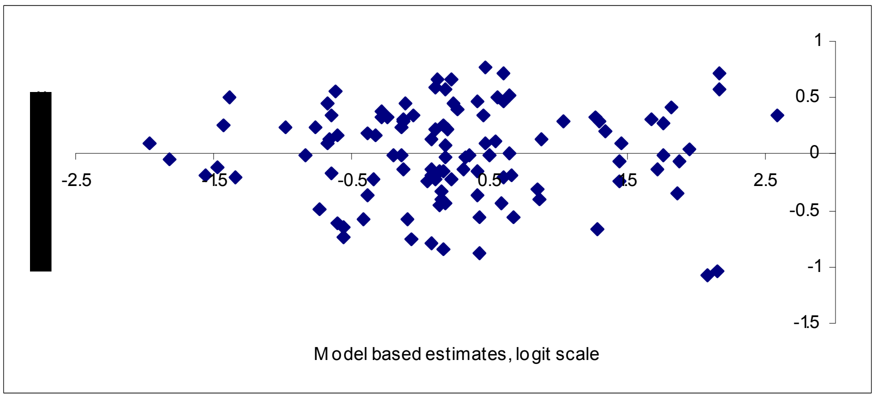
Validation of the Domain Estimates

Three diagnostic procedures discussed in Heady et al (2003) for validating domain estimates are used to validate the reliability of the estimates generated from the estimation model. The diagnostic methods used are: (1) the model diagnostics, (2) the goodness of fit diagnostic, and (3) the stability diagnostics.

Model Diagnostics

As with all statistical models, a number of assumptions are made about small area models. Using the logit link, the district level residuals were assumed to have a normal distribution with mean zero and variance ϕ. The **model diagnostics** are used to verify that the model assumptions are satisfied. If the model assumptions are satisfied the relationship between the area level residuals and the model estimates is expected not to be significantly different from the regression line *y=0*. Figure 1 below shows the plot of the area level residuals against the model base estimates (logit scale) with the relationship shown below the plot. The equations confirm that the regression line is not significantly different to the line *y=0* as required.

Figure 1: Model Diagnostic Plots



y = 0.001 (0.041) – 0.001 (0.043) x

Goodness of fit diagnostics

The **goodness of fit diagnostics** tests if the model-based estimates are close to the direct estimates. Basically, this diagnostics using the Wald goodness of fit statistics test whether there are significant differences between the expected values of the direct estimates and the model based estimates. This diagnostics is carried out by first computing the differences between the model-based and direct estimates. The differences are then squared and then inversely weighted by their variances and summed over all areas. This is then compared to a χ^2 distribution with degrees of freedom equal to the number of domains in the population. This provides a parametric significance test of bias of model-based estimates relative to their precision. In application to the estimates derived from the estimation model, the goodness of fit statistics estimated was 286.55 on 330 degrees of freedom corresponding to a p-value of 0.960. Indicating that there is no significant evidence to reject the χ^2 distribution, hence there is no significant difference between the direct estimates and the model based estimates.

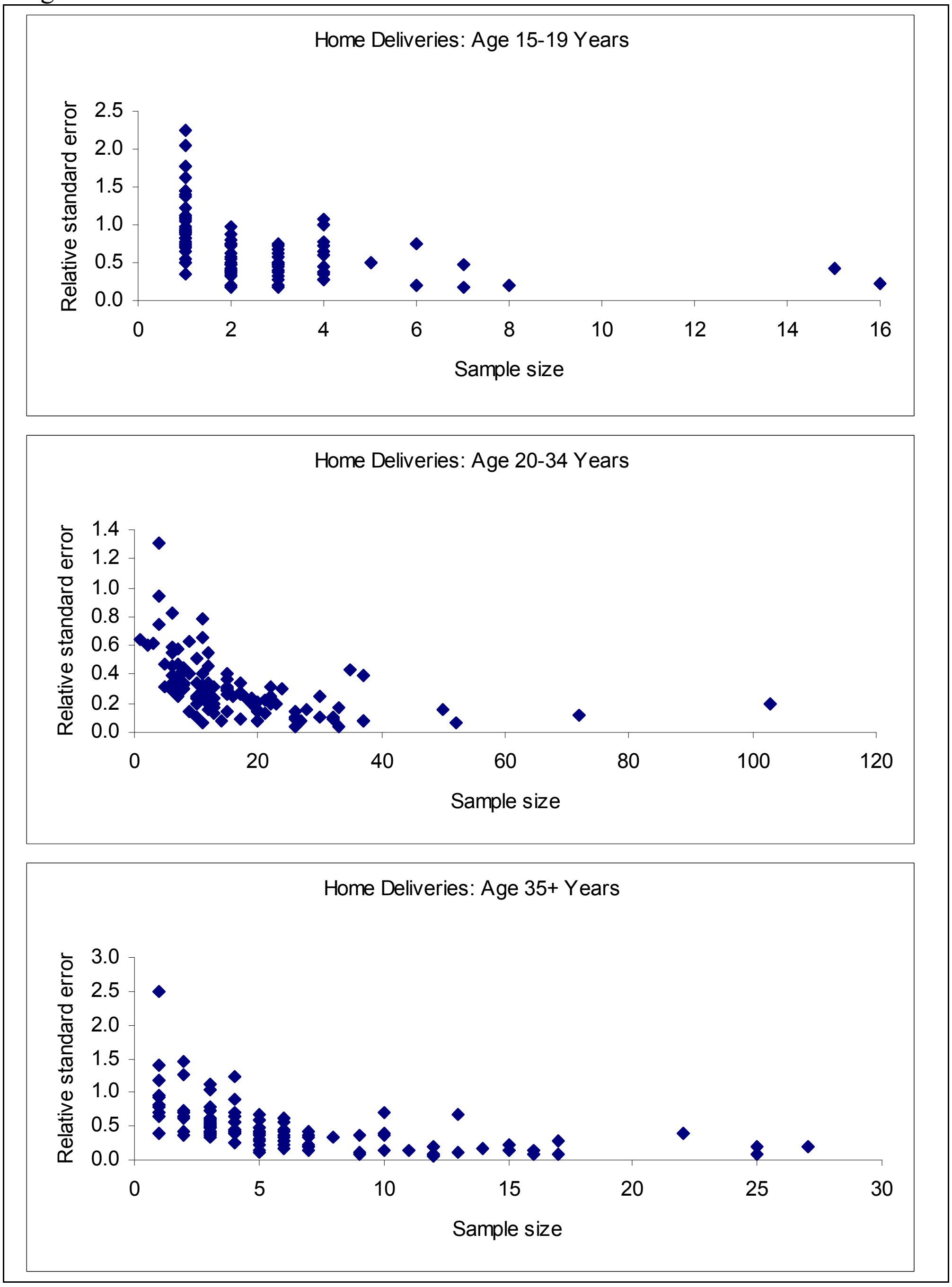
Relative Standard Error Diagnostics

The **relative standard error** (RSE) is a measure of an estimate's reliability. The RSE is obtained by dividing the standard error of the estimate $(SE(\hat{P}))$ by the estimate $\hat{P}$ and expressed as a percentage.

$$RSE = \frac{SE(\hat{p}_{ij})}{\hat{p}_{ij}} * 100$$

Estimates with large RSEs are considered unreliable. The criterion recommended for publication of estimates of local authority unemployment by the Office of National Statistics (ONS), United Kingdom is that the relative standard error is not greater 20 percent (Curtis 2005). The criterion recommended by ONS was adopted for investigate the reliability of the estimates derived in this study. Figure 2 below showing the RSE diagnostic plots for the estimates reveals that the results hold for all the estimates.

Figure 2: Relative Standard Error Plots



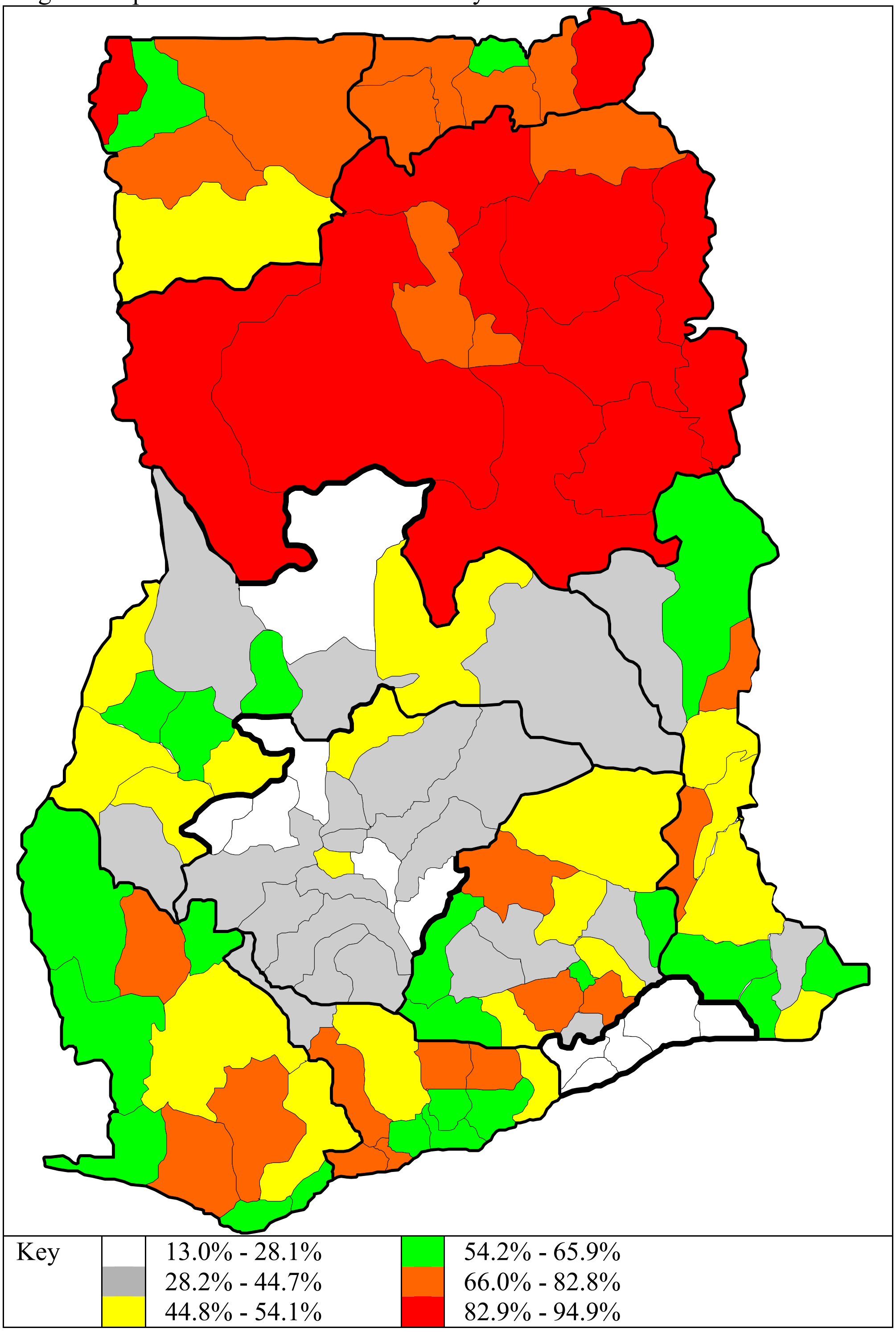
Mapping Spatial Variations in Home Deliveries

From the previous section, it has been established that the estimates generated from the estimation model are reasonably reliable and representative of the areas to which they belong. The mapping of the spatial variations in home deliveries is shown in Figure 3. The range of the estimated proportions was chosen such that 20 percent of districts are included in the ranges for the districts with the highest and lowest proportions. Forty percent of districts are included in the ranges for the second highest and second lowest proportions. The remaining 40 percent of districts are included in the ranges nearest the average values. The white shading represents districts with low home deliveries, while the orange and red shadings represent districts with high and very high home deliveries. The regional boundaries are marked with thick lines.

The estimates derived from the small area model confirm high inequalities in choice of delivery care in Ghana. The estimates shows that the proportion of home deliveries ranges from a low 13 percent in the Dangbe West District of the Greater Accra region to a high 95 percent in the Savelugu-Nanton district of the Northern region. The lowest estimate of home deliveries in Northern region was recorded in the East Mamprusi district (75 percent), an indication of the high proportion of home deliveries across all the districts of the Northern region of Ghana.

Figure 3 showing the spatial inequalities in home deliveries clearly depict that majority of the districts with high home deliveries are concentrated within the three Northern regions (Savannah Zone). The Greater Accra and Ashanti regions, the two most developed regions in Ghana unsurprisingly have the lowest proportions of home deliveries. In the Greater Accra region, home deliveries ranges from 13 percent in the Dangme West District to 23 percent in the Accra Metropolis. The high proportion of home deliveries in the Accra Metropolis compared to the rest of the Greater Accra region may be due to the growing slum communities in the area mainly due to the growing migrant population. This finding is similar to what pertains in the Ashanti region, where the Kumasi Metropolis has a higher proportion of home deliveries compared to the rest of the region. Localities within metropolitan areas with high home deliveries need to be identified and their needs taken into account in the provision of essential obstetric care. At the Special Session of the UN General Assembly in 1999, it was agreed that all countries should strive to ensure that 80 percent of all deliveries are attended by skilled attendants by 2005 (AbouZahr and Wardlaw 2001). The district-level estimates shows that only five percent of all districts in Ghana as at 2003 have achieved this target.

Figure 3: Spatial variation in Home Delivery



Discussion and Conclusions

As set out, reliable estimates of home deliveries representative of the districts they belong have been derived. The estimates derived from the analysis confirm high inequalities in home deliveries in Ghana. This can be attributed mainly to the varying levels of access to health services and socio-economic developments between areas. The proportion of home deliveries under unskilled attendant in the three Northern regions of Ghana is way above the

2003 GDHS national average of 54 (GSS, MI, NMIMR 2004), the African average of 58 percent and global average 43 percent (WHO 1998).

In sub-Saharan Africa, small area statistics has remained almost non-existent, particularly in the field of demography and health. This study has shown that with the availability of good auxiliary data, survey information such as the DHS could be used to derive reliable estimates to complement census which are becoming less regular in the region. The successful continuation of studies to derive reliable small area statistics in sub-Saharan Africa to aid the decentralisation process and developmental dispensation highly depends on governments, statistical offices and related organisations making available the needed information to statisticians.

References

Curtis, D. (2005) *Area Random Effects Model Based Estimation of Unemployment at Local Authority Level.* Heady, P., P. Clarke, G. Brown, A. D'Amore, B. Mitchell (2000) *Small Area Estimates Derived from Surveys: ONS Central Research and Development Programme.* Statistics in Transition, March 2000. Vol. 4, No. 4, pp. 635-648.

Ghana Statistical Service, Nouguchi Memorial Institute for Medical Research and ORC Macro (2004) *Ghana Demographic and Health Survey 2003.* Calverton, Maryland: GSS,

Saei, A. and R. Chambers (2003) *Small Area Estimation Under Linear and Generalized Linear Mixed Models with Time and Area Effects.* Southampton Statistical Sciences Research Institute, University of Southampton. S^3RI Methodology Working Paper M03/15.

World Health Organisation. (1998). *World Health Day: Safe Motherhood.* Maternal Mortality (WHD 98.1). WHO Geneva.