Small Area Estimation and mapping of poverty in Uganda

Dr Marco Pomati

Profile

Background

- **Recent Research Projects:**

Measurement of material deprivation

Malnutrition in Sub-Saharan Africa

Housing affordability and housing deprivation

- **Teaching:**

Quantitative Methods for Social Scientists

Social Policy

Q-Step

A step-change in quantitative social science skills

Funded by the Nuffield Foundation, ESRC and HEFCE

What this talk is about

- Share findings from an empirical example (Small Area Estimation of multidimensional poverty in Uganda)
- Walk you through (some of) the practicalities of doing Small Area Estimation
- Understand the workflow of Small Area estimation to avoid over-reliance on pre-packaged Small Area estimation programs (i.e. what’s under the bonnet?)
- The presentation assumes basic understanding of linear regression modelling

What this talk is <u>not</u> about

- <u>Not a</u>n in-depth class on the mathematics behind Small Area Estimation
- <u>Not </u>a series of simulations with some conclusions about which SAE estimator is better
- <u>Not </u>a class on Bayesian Hierarchical modelling

Outline

- Small Area Estimation in theory

- Small Area Estimation in practice

- Applied example

Small area estimation

- ➢A set of methods aimed at measuring poverty at local level

- ➢Important for:
- Funds-allocation
- Assessment of policies (“No one left behind”)
- Poor areas within richer regions
- Influence of place on individual outcomes

Direct estimates of local poverty

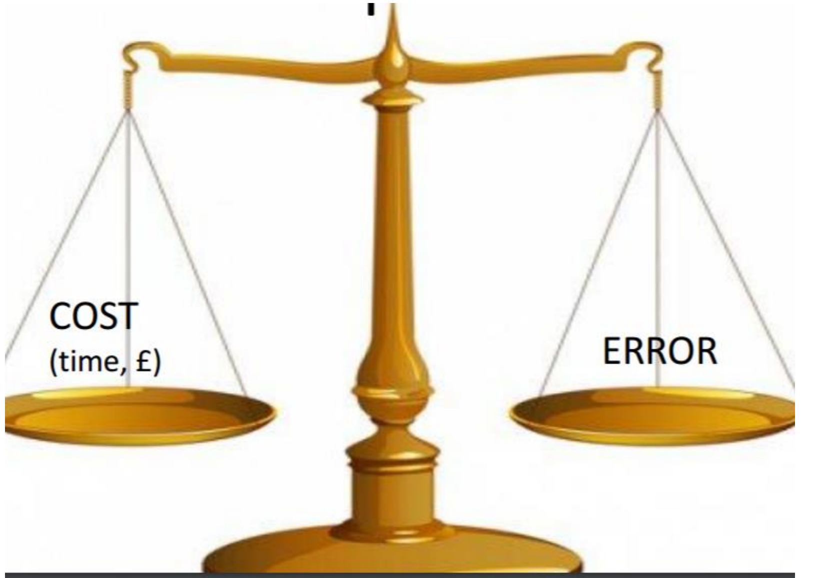
- Margin of error of Proportion $1.96 \text{ x } \sqrt{\frac{\text{Proportion } (1-\text{Proportion})}{\text{number of valid respondents}}}$

- Number of cases needed to achieve a certain Margin of Error for a given proportion:

$$n=\frac{1.96^2 \ x \ Proportion \ x \ (1-Proportion)}{MoE^2}$$

- MoE of small-area estimates too high

- Many small areas often not covered at all!



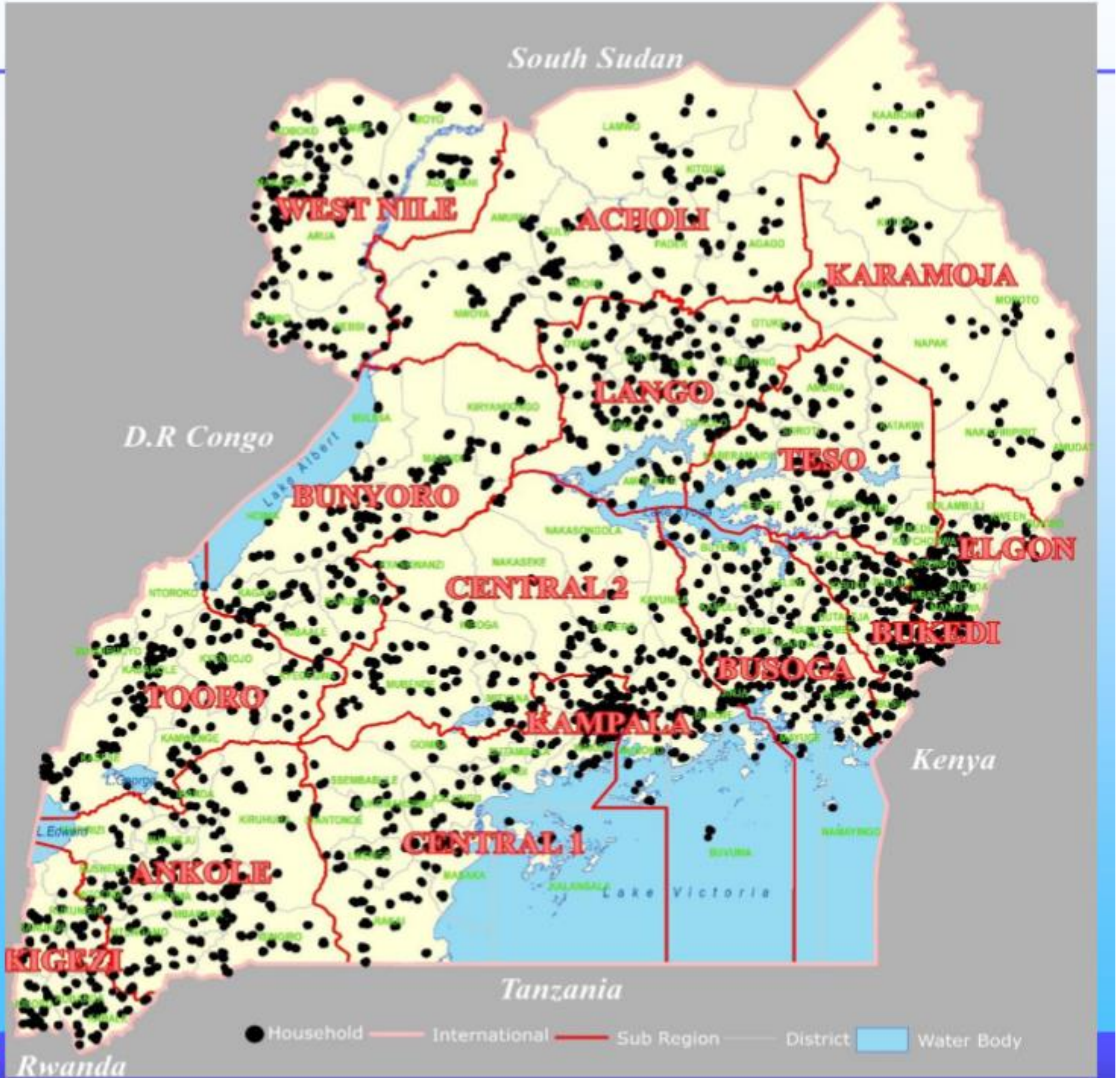
Example: Uganda

➢ Example: Uganda administrative boundaries example

- Region > sub-region > district > county > sub-county > parish > enumeration area

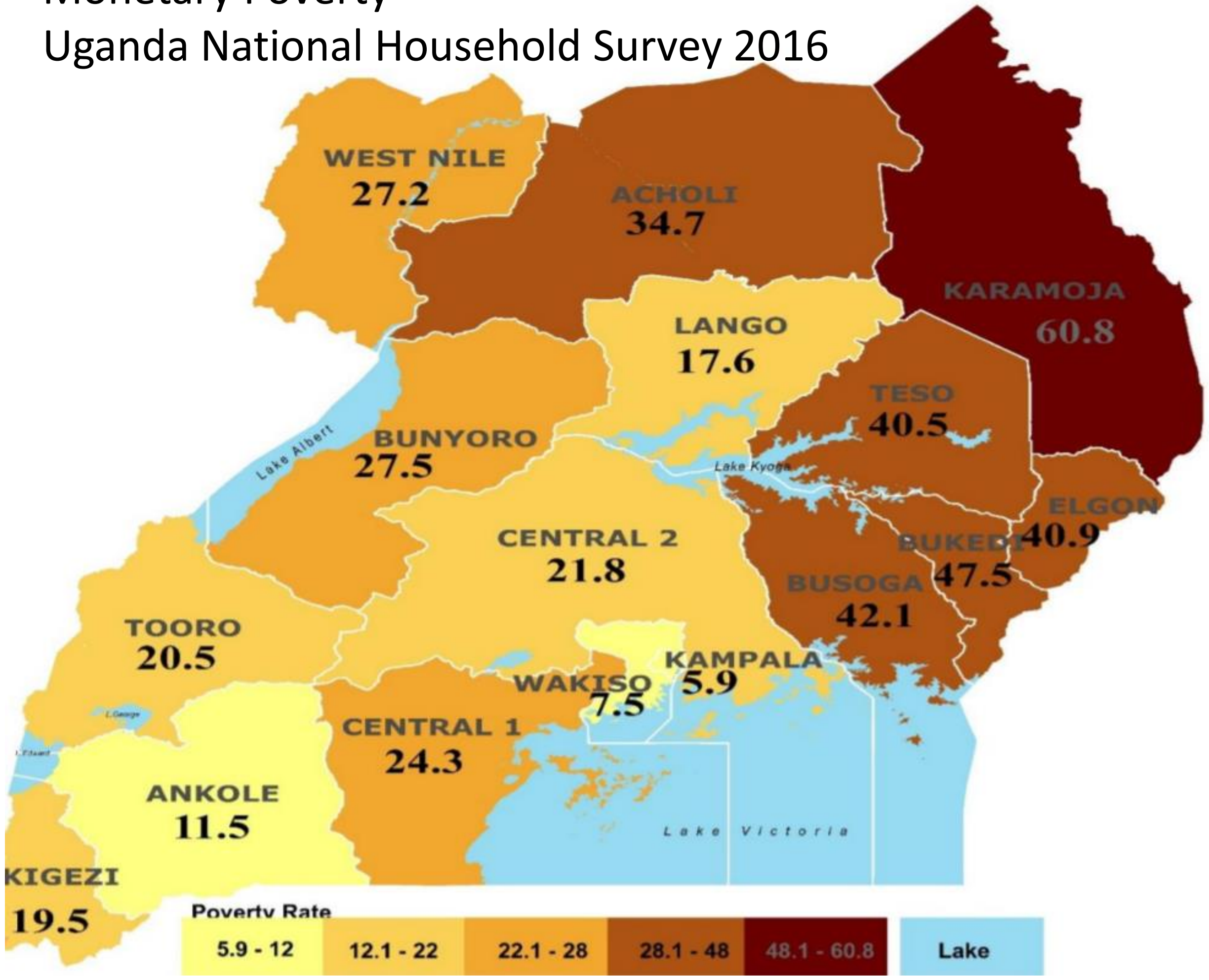
- Can we produce sub-county level estimates of poverty using the National Household Survey?

Distribution of 2016/2017 UNHS sampled households

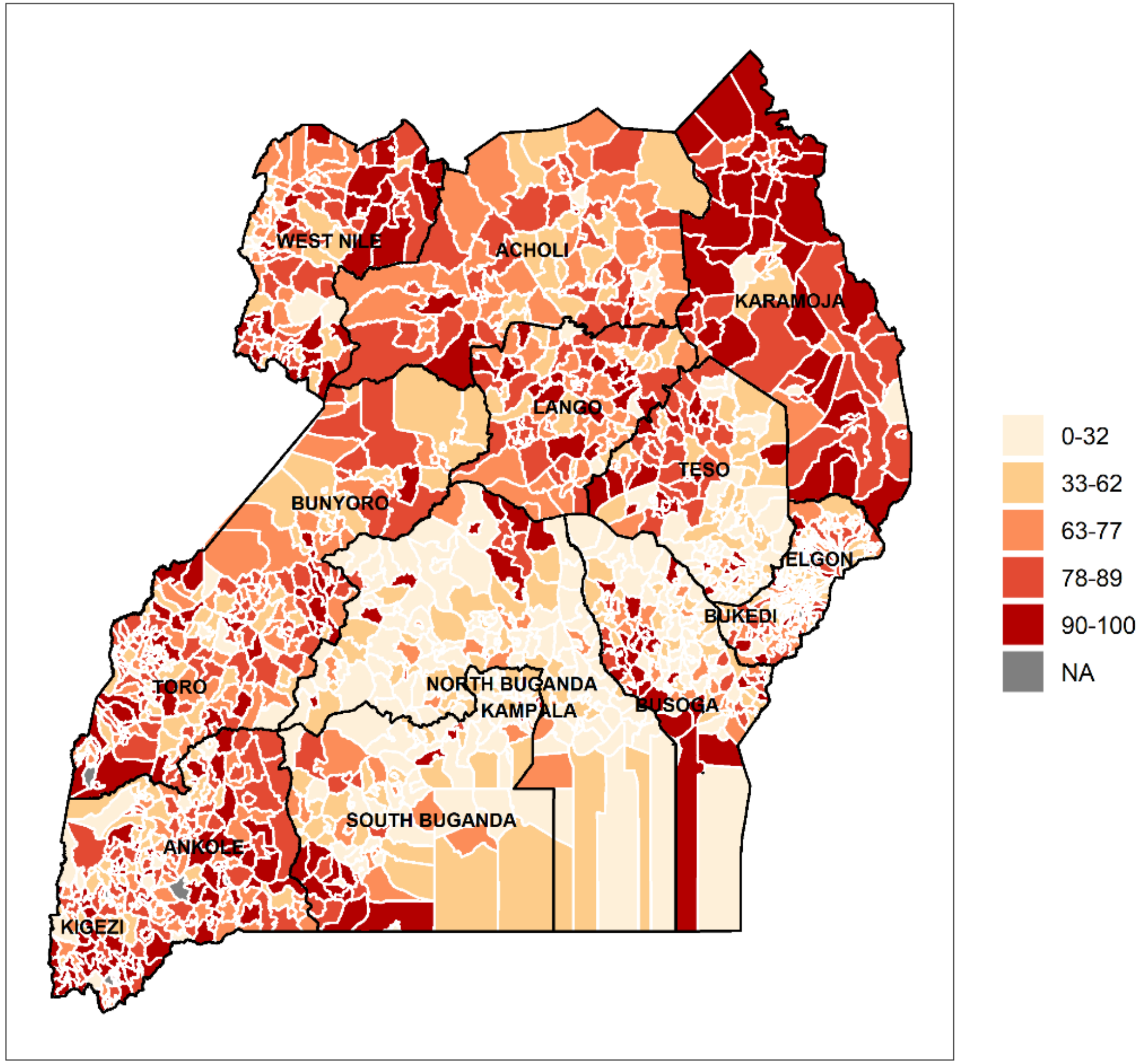


Poverty Estimates

Monetary Poverty
Uganda National Household Survey 2016



% of households experiencing wall deprivation
(Subcounty Census Estimates)



Small Area Estimation: Basic intuition

- Survey data: More questions, more frequent, fewer cases

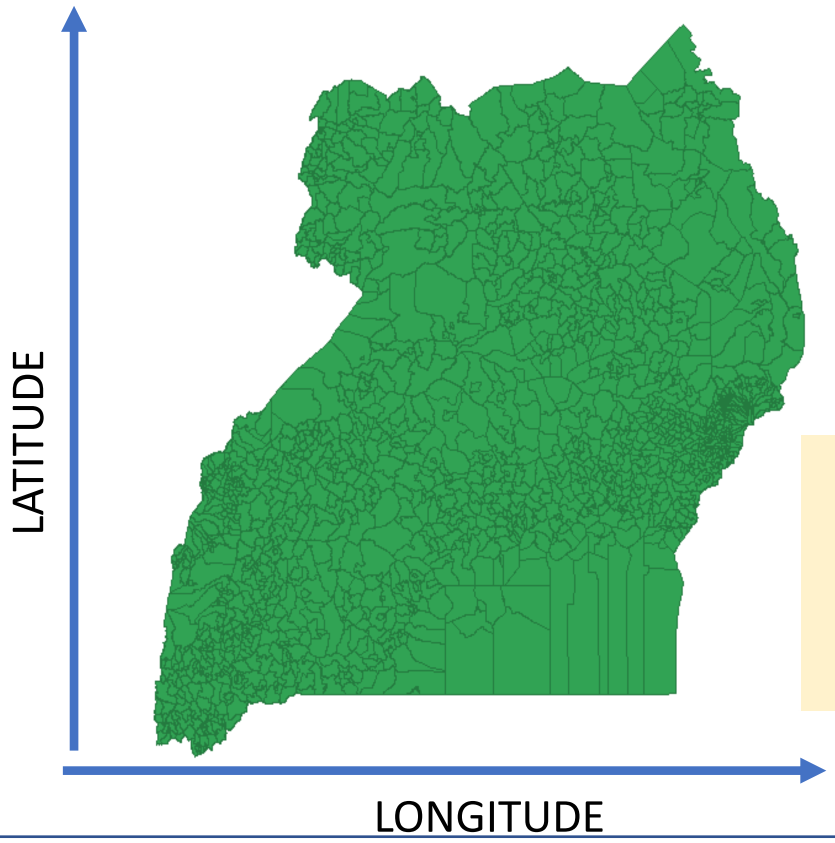
- Census data: More cases, fewer questions, less frequent

SAE: Combine strengths to obtain small area estimates

The Third Dataset

- Your small area Shapefile/s. Just as important as survey and Census

SHAPEFILE



ID	District Code	Subcounty name
21	318	KAABONG EAST

See IPUMS repository https://international.ipums.org/international/gis.shtml

Indirect estimators

- Statistical Modelling

e.g. Multilevel regression modelling (Frequentist or Bayesian)

- Spatial Microsimulation

Involves creation of synthetic micro-population or reweighting using Census area level benchmarks

Borrowing strength (Individual level modelling)

Simple Regression Framework example

$$Y_i = \alpha_0 + \beta_1 X1_i + \beta_2 X2_i + e_i$$

$$Income_i = \alpha_0 + \beta_1 \text{ Secondary Education}_i + \beta_2 \text{ Urban}_i + e_i$$

$$Income_i = 50 + 150_1 \text{ Secondary Education}_i + 3 \text{ Urban}_i + e_i$$

	Income	Has Secondary Education	Lives in Urban area	Survey model predicted Income	Residual
Survey Respondent 1	213	Yes	Yes	50+150+3=203	10
Survey Respondent 2	10	No	No	50	-40

Borrowing strength (individual level modelling)

SURVEY

$\text{Income}_i = 50 + 150_1 \text{ Secondary Education}_i + 3 \text{ Urban}_i$ **Survey data model estimates**

	Income	Has Secondary Education	Lives in Urban area	Survey model predicted Income	Residual
Survey Respondent 1	213	Yes	Yes	50+150+3=203	10
Survey Respondent 2	10	No	No	50	-40

CENSUS

	Income	Has Secondary Education	Lives in Urban area	Survey model predicted Income
Census Respondent 1	Not collected	Yes	Yes	50+150+3=203
Census Respondent 2	Not collected	No	No	50

Poverty SAE models

Most widespread methodologies in poverty SAE:

- Fay-Harriot
- ELL/World Bank
- Empirical Bayes
- Hierarchical Bayes

Borrowing strength- Multilevel modelling

Simple Regression Framework example

$$Income_i = \alpha_0 + \beta_1 X1_i + e_i$$ **Survey data model** **(example with just one independent variable)**

Multilevel Framework

$$Income_{ij} = \alpha_{0j} + \beta_1 X_{1ij} + \varepsilon_{ij}$$

$$\alpha_{0j} = \gamma_{00} + \gamma_{01} W_j + \mu_{0j}$$

$$\beta_{1j} = \gamma_{10} + \gamma_{11} W_j + \mu_{1j}$$

Borrowing strength (individual level modelling)

SURVEY

$$Income_{ij} = \alpha_{0j} + \beta_1 X_{1ij} + \varepsilon_{ij}$$

$$\alpha_{0j} = \gamma_{00} + \gamma_{01} W_j + \mu_{0j}$$

CENSUS

	Income	Has Secondary Education	Lives in Urban area	Survey model predicted Income
Census Respondent 1 in Busaana	Not collected	Yes	Yes	50+150+3+ $\mu_{0BUSAANA}$
Census Respondent 2 in Busaana	Not collected	No	No	50+ $\mu_{0BUSAANA}$

Small Area Estimation in practice

Pre-modelling choices: Which small area?

- Choose your target levels (your “Small Area”).

Uganda 2016 example:

- Region > sub-region > district > county > <u>sub-county</u> > parish > enumeration area

Pre-modelling checks: Survey and Census data Comparison

1) Check

- Check that the distribution of your dependent and independent variables match in sample and Census
- How do they derive household size?
- How is the household head selected?

2) Inform variable selection and possible post-stratification

Do differences between Census and survey reflect genuine differences or sampling error and/or measurement error and/or nonresponse error?

Pre-modelling choices: boundaries

1) Identify a consistent and relevant nesting structure of geographical boundaries. Choose your target level (your “Small Area”)

2) Establish that this is consistent in both survey, Census data and shapefile.

3) Research changes in boundaries between Census and survey

4) Establish a strategy to update Census and Survey boundaries and create updated geography variables and shapefiles.

5) Create a unique ID in your (GIS) shapefile and then match it to your survey and Census data

5) Create a unique ID in your (GIS) shapefile and then match it to your survey and Census data

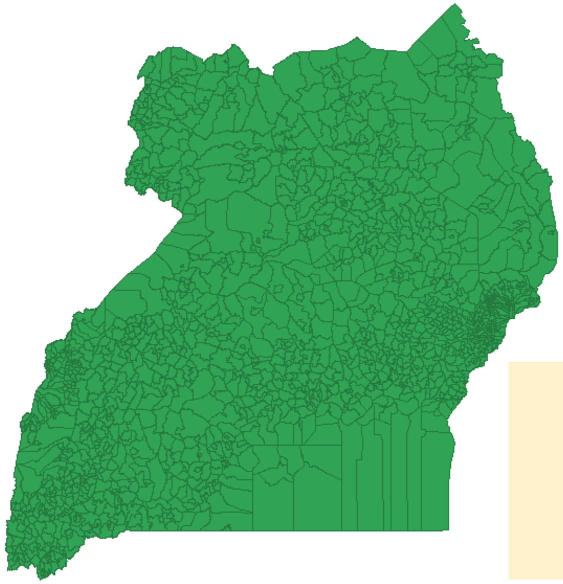
SURVEY

District Code	Subcounty name	Shapefile subcounty ID
318	KAABONG EAST	21

CENSUS

District Code	Subcounty name	Shapefile subcounty ID
318	KAABONG EAST	21

SHAPEFILE

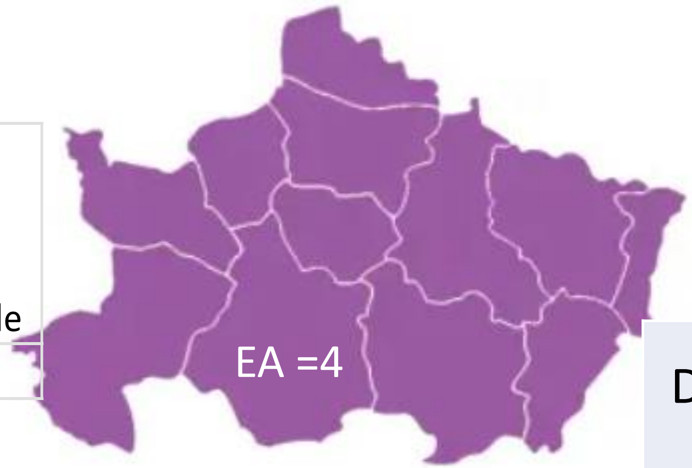


ID	District Code	Subcounty name
21	318	KAABONG EAST

5) If the match is not complete, try to match on the lowest-level area available in shapefiles and datasets (e.g. enumeration areas).

SURVEY

District Code	Subcounty name	EA name	EA code
318	NA	KALODEKE	4



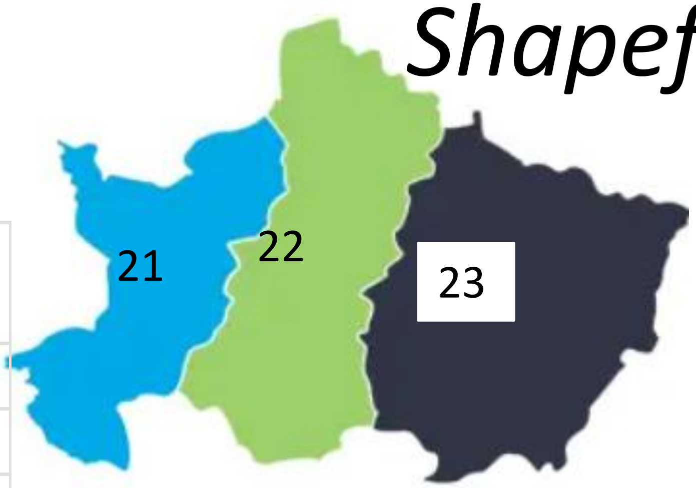
shapefile

CENSUS

District Code	Subcounty name	EA name	Subcounty name
318	NA	KALODEKE	4

5) If matching is not complete…persevere and eventually triangulate, inspect manually etc.…

Shapefile subcounties



SURVEY

District Code	EA name	EA code	Shapefile subcounty ID
318	KALODEKE	4	21
318	SOKODU B	8	22
318	KOBUIN_a	11	?

CENSUS

District Code	EA name		Shapefile subcounty ID
318	KALODEKE		21
318	SOKODU North		22
318	SOKODU South		22
318	KOBUIN		23

Modelling choices: Explore and choose final survey data model

<u>Different coefficients specifications</u>

- Individual, Aggregate and Individual + Aggregate level covariates

<u>Different random error specifications</u>

Random intercepts? Random Slopes? At what levels?

<u>Assess the model fit</u>

- E.g. Individual-level Logistic regression: Pseudo R^2, Sensitivity, Sensibility, AIC, BIC
- E.g. Multilevel Logistic regression: above + WAIC, Loo (Vehtari et al, 2016)

<u>Reproduce direct estimates</u>

Can you survey model reproduce accurate poverty rates at regional level (using survey data)?

Validation

- After you have used your survey model to predict poverty in Census cases and produced small area estimates there are still a few checks you need to carry out

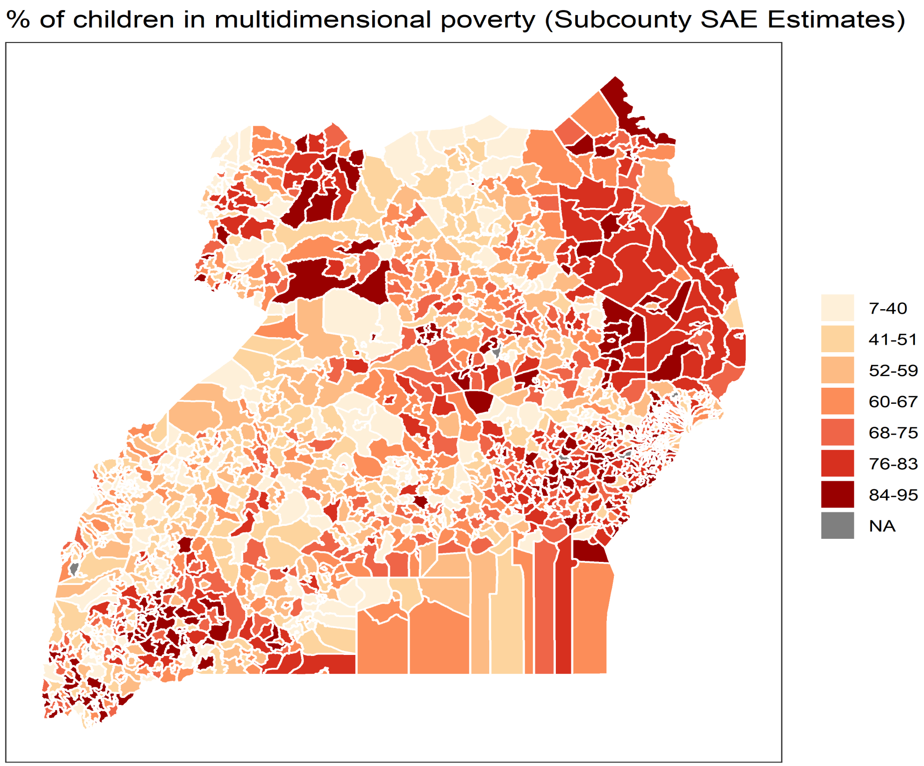
- Can you reproduce direct survey poverty estimates using Census predicted values?

- Variation of your estimates (see Molina and Rao, 2010)

- Leave one variable out of your independent variables set to validate your small area estimates. Ideally use auxiliary data.

Summary of a simple SAE workflow

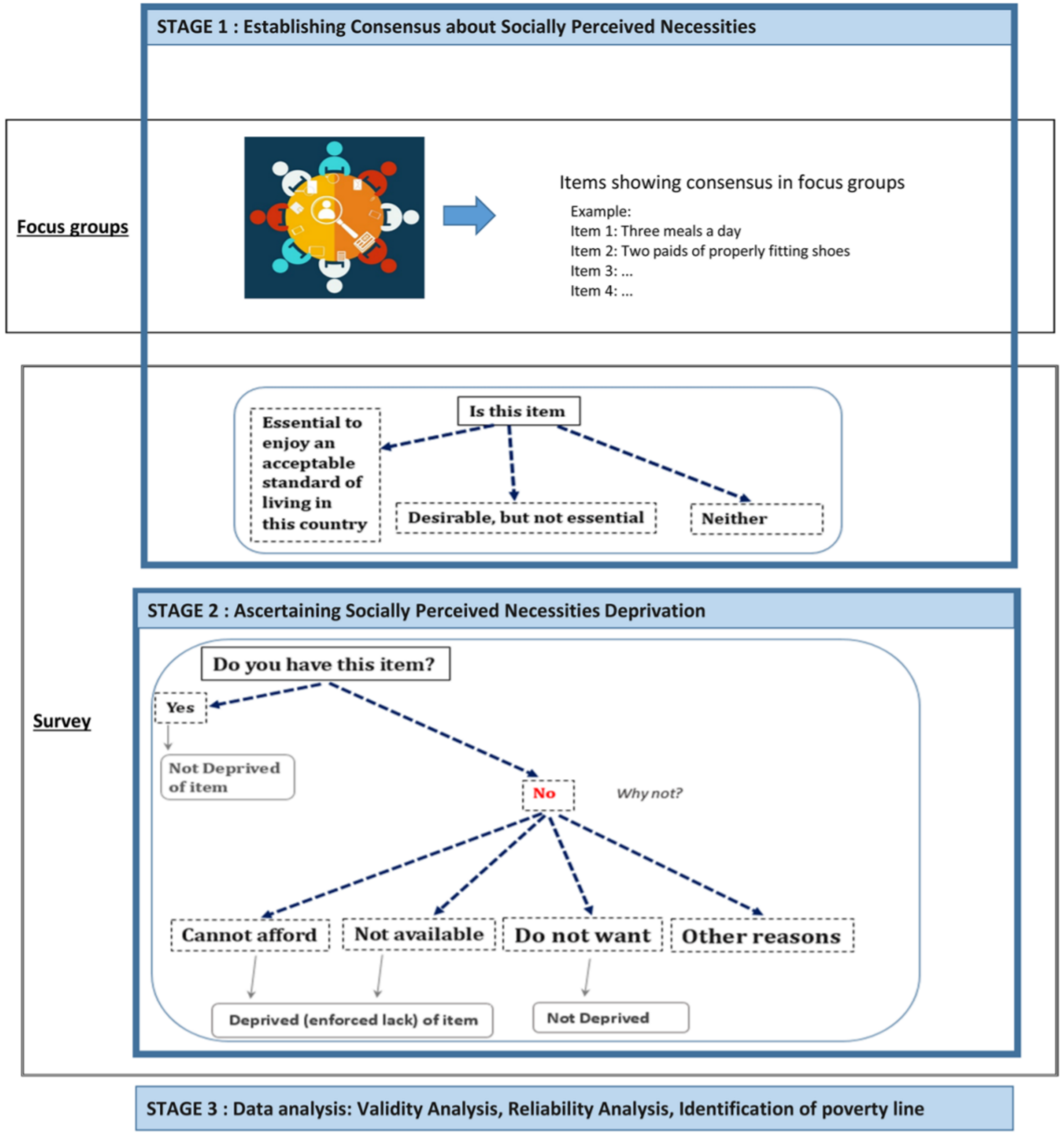
- ➢Pre-modelling choices:
- • Which small area?
- • Survey and Census data Comparison
- • Boundaries
- ➢Modelling choices: Explore, compare and choose final survey data model
- ➢Apply survey model to Census (predict poverty status of Census respondents) and produce Small Area Estimates
- ➢Validate



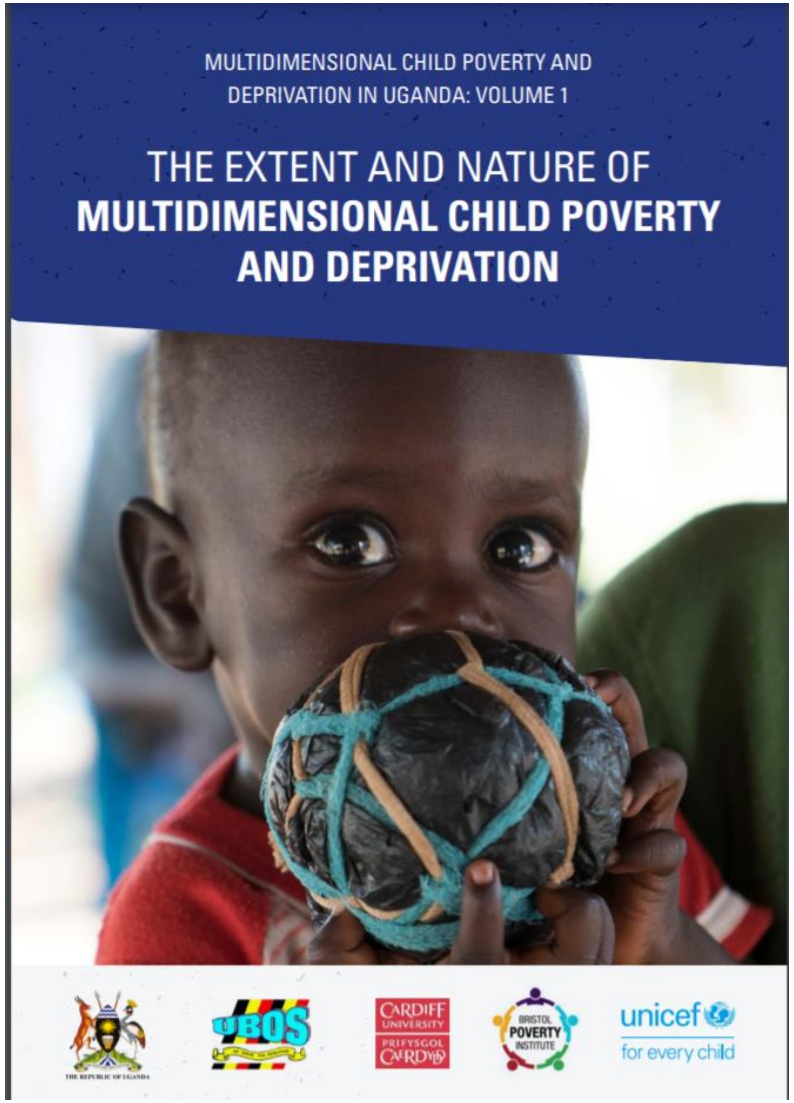
Applied example

Sub-county estimates of multidimensional poverty in Uganda 2016

Multidimensional poverty in Uganda



https://link.springer.com/article/10.1007%2Fs11205-019-02198-6



https://www.unicef.org/uganda/reports/multidimensional-child-poverty-and-deprivation-uganda-report-volume-1

Table 2.5 Final child deprivation index components

Relevant age range	*Item*
0–17	A visit to the health facility when ill and all prescribed medication
0–17	Three meals a day
0–17	Two sets of clothing
0–17	Toiletries to be able to wash everyday
6–17	All fees, uniforms of correct size, and equipment
0–17	Own blanket
0–17	Own bed
0–17	Two pairs of properly fitting shoes
11–17	Own room for children over 10 of different sexes
3–17	Books at home for their age
0–17	Some new clothes
6–17	To be able to participate in school trips
6–17	Bus/taxi fare or other transport
6–17	A desk and chair for homework
6–17	Presents for children once a year on special occasions
3–12	Educational toys and games
0–17	To be able to make regular savings for emergencies
0–17	Repair a leaking roof for main living quarters
0–17	Repair or replace any worn out furniture
0–17	Replace broken pots and pans for cooking
0–17	Take children to a medical facility when sick
0–17	Pay school fees for children

Pre-modelling choices

Uganda National Household Survey 2016/17→Uganda Census 2014

➢Pre-modelling choices:

- **Which small area?** Sub-county nationally and parishes for Kampala
- **Survey and Census data Comparison**

- Applied post-stratification to UNHS 2016 (under-representation of rural households in the survey)

- Inconsistent definition of household size and household head across two datasets (variables flagged/excluded)

- Boundaries: Region> sub-region > district > county > sub-county > parish > enumeration area . Several administrative changes, many days spent harmonising shapefile, Census and survey boundaries.

Modelling choices: Explore, compare and choose final survey data model

- Started with basic individual logistic regression nested models
- Candidate set of independent variables draws on previous literature and advise from UBOS
- Led to final set of fixed effects

Logistic regression models predicting individual-level poverty status. Log-odds	
	Model 1
Urban	-0.15 *
clothes deprivation	0.29 **
shoes deprivation	1.33 ***
roof deprivation	0.29 ***
wall deprivation	0.44 ***
Sanitation type (Flush toilet)	
Latrine	1.59 ***
Covered pit latrine	2.55 ***
Covered pit latrine with a slab	3.02 ***
Covered pit latrine without a slab	2.60 ***
Uncovered pit latrine with a slab	3.29 ***
Uncovered pit latrine without a slab	2.13 *
No facility	3.39 ***
Other	4.28 ***
tv deprivation	1.59 ***
Improved water	0.16 *
Number of children	0.27 ***
Overcrowding	0.52 ***
bicycle deprivation	0.74 ***
Household head working in subsistence agriculture	
Household head Illiterate	
N	15646
Nagelkerke R2	0.32
Specificity	0.75
Sensitivity	0.75

Modelling choices: Explore, compare and choose final survey data model

We then moved to multilevel modelling using Hierarchical Bayes Logistic Regression models

- Model 1 without including Sub-region-level intercepts.
- Model 1 + household head working in subsistence agricultural activities + household head working in subsistence agricultural activities (variables previously flagged as problematic)
- Model 1 + sub-region intercepts.
- **Model 1 + random intercepts at the district level**
- Model 1 + random slopes at the district level

The final survey model

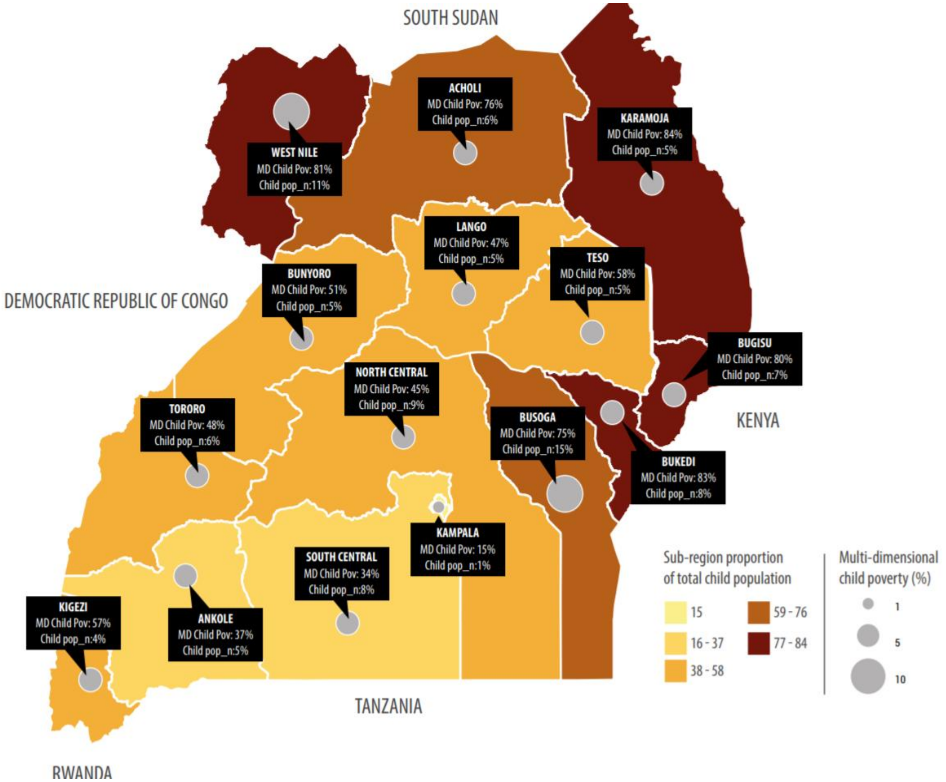
- Model 1 + random intercepts at the district level

- Statistical fit: the WAIC (widely applicable information criterion) statistic of fit and Loo (leave-one-out cross-validation for fitted Bayesian models).
- Capacity to reproduce the subregional point estimates, i.e. whether the model reproduces the observed data (design estimates of multidimensional poverty).

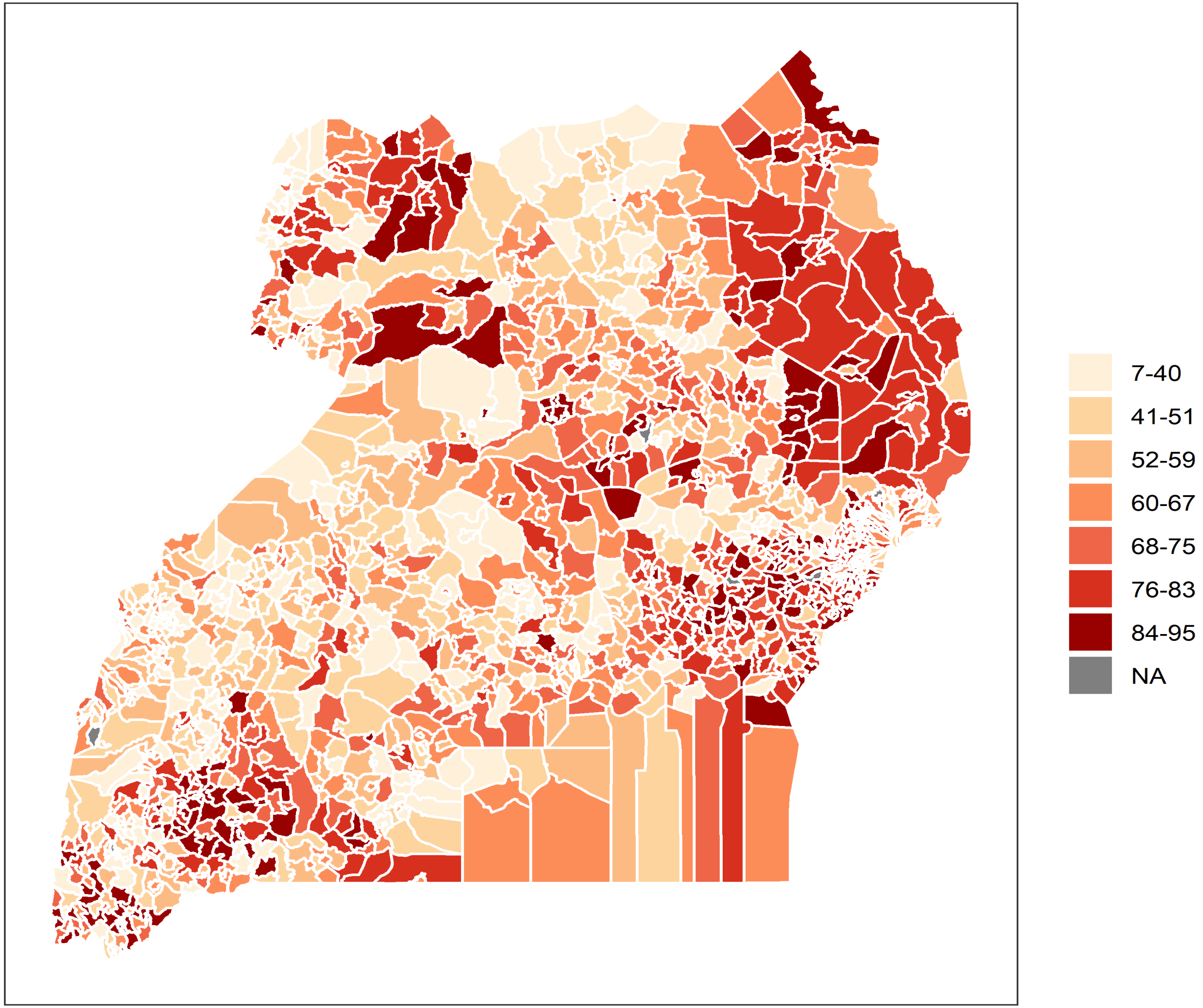
Validation

% Poor (multidimensional poverty)

SubRegion	Survey direct estimate	Survey: Model prediction	Census: Model Prediction
Kampala	9 [6-11]	9	8
Central 1	27 [24-29]	27	27
Central 2	38 [34-41]	38	39
Busoga	59 [56-63]	60	59
Bukedi	74 [69-78]	74	74
Bugishu	63 [59-68]	63	64
Teso	53 [49-57]	53	52
Karamoja	75 [70-80]	75	75
Lango	42 [39-46]	42	42
Acholi	61 [57-65]	61	59
Westnile	70 [66-74]	70	72
Bunyoro	41 [37-45]	41	41
Tooro	41 [37-45]	41	39
Ankole	29 [25-34]	29	31
Kigezi	50 [44-55]	50	52



% of children in multidimensional poverty (Subcounty SAE Estimates)



Findings for Uganda 2016 (forthcoming)

- Areas in the north, particularly in the north east, tend to have very high multidimensional poverty rates (Above 60%).

- However, there are pockets of high poverty in subregions that do not appear to have very high poverty rates at sub-region level

- Areas with the highest levels of need are also those with the lowest levels of infrastructure

Conclusion

- Spend time making sure that shapefiles are compatible across datasets
- Use statistical software that can easily handle several types of datasets at the same time (e.g. R, Python)
- Avoid relying on pre-packaged SAE functions (e.g. try to understand the process yourself)
- Importance of auxiliary information for cross-validation
- Importance of relevant small area

Acknowledgments

- None of this work would have been possible without the contribution of Héctor E. Nájera Catalán, Shailen Nandy, Diego Angemi, David Gordon, Sheila Depio, Gemma Ahaibwe, Ibrahim Kasirye, James Muwonge, Vincent Ssennono, Stephen Baryahirwa, Baylon Twesigye, Sebnem Eroglu-Hawksworth, Eldin Fahmy, Acomo Oloya, Arthur Muteesasira, Sarah Kabaija, Vincent Ssennono, Beatrice Winnie Nyemera and Stephen Baryahirwa and others…