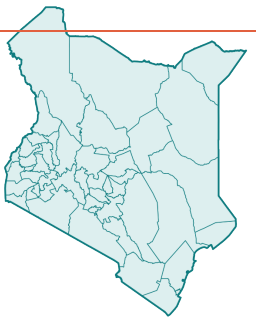
Adding Covariates in sae4health

Worked example: average years of education as a covariate for MCV1

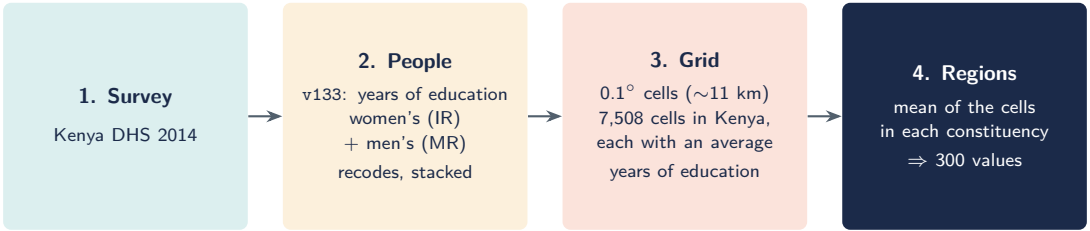
The sae4health team

CEMA workshop

Small Area Estimation for Improving Health in LMICs



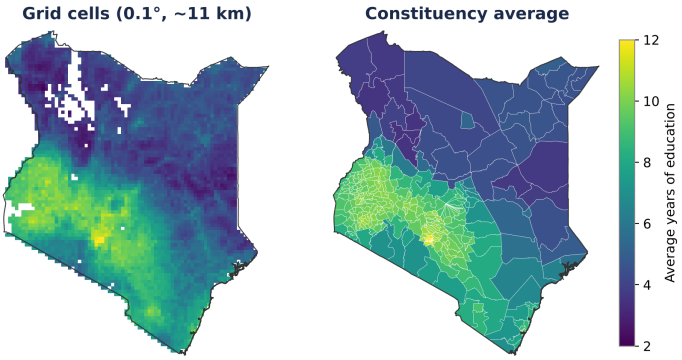
How the covariate is built



The app needs **one number per region**. Step 4 turns a fine-grained surface into exactly that.

Covariate data prepared by Ameer Dharamshi.

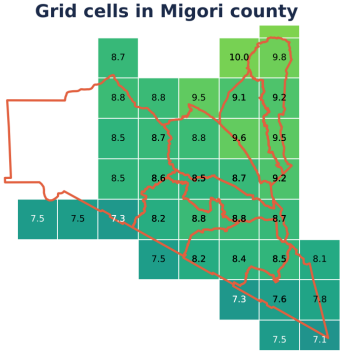
From grid to regions

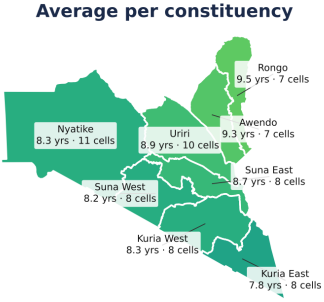


- Highest in **Nairobi** and the **central highlands**, lowest in the arid **north** and **north-east**.
- Averaging keeps the big pattern, but loses detail inside large regions.

White cells: water or no value.

Averaging, up close: Migori county





10
median grid cells per constituency

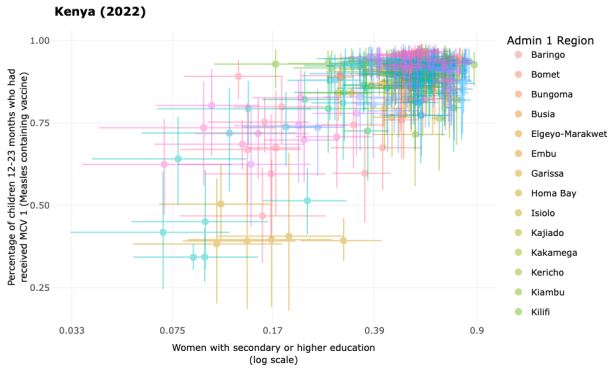
1.5 yrs
median within-area spread

Range: 1–327 cells.

5 constituencies have only one cell.

Simple (unweighted) mean of the grid cells in each constituency.

Why it might help for MCV1



- Caregiver education is a commonly reported predictor of childhood vaccination.
- The covariate covers **every constituency**, including those with few or no sampled clusters.
- Education is lowest in the north and north-east, where MCV1 coverage is also lowest.

> The covariate averages **women and men**, not only mothers.

Covariates in the app

Adding a covariate in the app

1. Add it

After the sparsity check, open **Advanced Options**. Pick the Admin level and **download the template** CSV of regions.

2. Fill it

Add the average years of education column: one value per region, names exactly as in the template, no blanks. Upload, then **Apply**.

3. Refit

Fit the same model **without** and **with** the covariate. Area-level and unit-level models both take covariates. Direct estimates don't change.

Demo: MCV1 (`CH_VACC_C_MSL`), Kenya DHS 2022, constituency level.

What to look at after adding the covariate

Model components

- **Covariate effect:** is it positive and clearly away from zero? It's on the log-odds scale, per extra year of average schooling.
- **Spatial random effect:** the total variance should **shrink**, because education now explains part of the differences between areas.

Cautions

- A covariate needs a value for **every** region.
- Region averages lose detail within regions.
- Choose plausible drivers, not near-copies of the outcome.

Visuals: before vs after

- **Prevalence Map**: does the pattern sharpen (north-east lower, centre higher)?
- CV and 95% interval-width maps: precision should improve most where data are thin.
- **Grey and hatched areas**: their estimates now lean on the covariate, not only on neighbours. This is where covariates matter most.
- **Scatter Plot**, direct vs model: closer to the diagonal, less blanket shrinkage.
- **Interval Comparison** and **Ridge Plot**: narrower intervals where the covariate is informative.