# Small Area Estimation Check List for QER

Basic information to be provided in the introduction:

* Indicate why the exercise is being carried out.
* The main data sources used.
* The small area approach used.
  + Justify the choice.
* The level at which estimates were obtained (small areas)

## Data description section

* For unit level models and area level models
  + Survey
    - Year(s) of survey
    - Sampling approach and period conducted
    - Number of households sampled
    - Information on temporal and/or spatial deflation applied to the welfare measure.
    - Provide a table illustrating the number of areas at different geographic levels, how many areas are there in total, how many covered by the census, and how many covered by the survey.
      * Number of small areas covered by the survey
        + As a share of those in the country
        + At the area level, sample size percentiles

E.g., 10% of areas have less than 20 households sampled

* + - * + Motivation: In many instances, EB for a given area could rely on very few households. In this instance it may be preferable to not use EB, just for the areas affected.
    - Information on the definition(s) of poverty or wellbeing to be used (i.e., international PPP, the national official poverty rate, etc.)
    - Poverty rate at level of representativeness accompanied by its standard error, confidence interval, or coefficient of variation (CV) (note: the CV measure is less informative as the poverty rate approaches zero).
      * If possible, poverty threshold value
        + Indicate if the threshold is national or differs by geographic level
* For unit level models (only)
  + Mean welfare (e.g., in adult equivalents or per capita) at the level of representativeness accompanied by its standard deviation and its standard error
    - Should correspond to the welfare measure used as the model’s dependent variable
  + Census/or other large household level data used as auxiliary data
    - Year and period conducted
    - Comparability of potential covariates between survey and census to ensure the population used to estimate the model is aligned to the population to which the model parameters will be applied
      * Ideally at the level of survey representativeness or the level of the models’ fit (national, sub-national).
      * Indication of variables that are included/excluded due to comparability
      * Especially in cases with more than 1 or 2-year gap between survey/census, careful attention should be given to comparability of potentially time-varying indicators.
      * Sample design and census weight adjustment (if not a full census)
      * Minimum number of households for the small areas (from the census) to check if there is a need to “merge/join” contiguous areas together and change boundaries.
        + Motivation: This is dependent on the overall estimated noise, and may require teams to join contiguous areas to arrive at a more acceptable noise estimate

## Methods

* Basic description of the method used and its advantages over the alternatives
  + Package or library used for the application
* Highlight the main assumptions of the method

## Model selection

* Approach used for selecting the model(s). For example, stepwise or lasso
  + Number of covariates at the beginning
  + Number of covariates at the end
  + Details on the on the inclusion/exclusion of spatial fixed effects
  + Was cross validation applied? Provide details.
  + Choice of national vs sub-national models and underlying rationale
* Model assumptions and diagnostic checks
  + Model performance across the welfare distribution:
    - Develop model using survey data and simulate welfare on the same survey data and obtain poverty rates at different thresholds.
      * Motivation: Allows teams to see if the model assumptions hold, without introducing potential data comparability issues.
  + Normal Quantile-Quantile Probability Plot (Q-Q plot) for predicted random location effects and household residuals
  + Kdensity plots of random location effects and household residuals against a normal distribution
  + Any other tests for normality that may be applicable. For example, Shapiro-Swilk, or probability-probability plot (P-P plots).
  + Other model diagnostics tests: e.g., Cook’s D, leverage statistics. See [Guidelines to Small Area Estimation for Poverty Mapping](https://openknowledge.worldbank.org/entities/publication/e22cbb4c-653d-508f-9a21-f42fa051bf02).
* Present final model table
  + Coefficients and their standard errors
    - Make sure that variables at different levels (e.g., geographic) are easily discernible.
  + Fit diagnostics:
    - Adjusted R-squared, BIC, AIC, or others
  + Model for heteroskedasticity (unit-level only)
    - Adjusted R-squared
  + Estimated variance of idiosyncratic errors, var epsilon (unit-level only)
  + Estimated variance of random location effect, var eta (unit-level only)
  + Number of observations and degrees of freedom
  + For area-level models, the gamma “shrinkage factor” for each location

## Presenting results

* Comparison of small area estimates aggregated to the survey’s level of representativeness and the direct estimates and their 95% CI
* Comparison of direct vs. synthetic estimates, SE vs RMSE, and CV.
* Special attention to details on the precision of any out-of-sample estimates in an area-level model (i.e., locations for which there was no direct estimate).
* At the small area level present estimated poverty rates and their MSE.
  + Group estimates by the level of representativeness of the survey
  + Highlight which small area estimates are statistically different from the higher-level geographic units to which they belong and for which representative direct survey estimates are available.
    - Motivation: It adds value to the extent that it uncovers areas where predicted poverty rate falls outside the confidence interval of the poverty rate for the larger area to which it belongs, and for which we already have a poverty rate directly from the survey. It provides information on what we have learned about poverty from the exercise.