Strengthening the capacity of Jordan's Department of Statistics in terms of compilation, analysis and reporting of statistical data in line with International and European best practices. *Twinning Reference Number: JO 21 ENI ST 01 22*

Component 2. Small Area Statistics

Objectives component 2

OBJECTIVE OF THE PROJECT*:*

Development of SAE methods to be applied to some SDG indicators on poverty, and their possible extension to other statistical areas within Jordan DoS.

SAE OBJECTIVE OF THIS MISSIONS:

1. The purpose for is to give a theoretical introduction small area estimation with focus on basic smoothing and data manipulation.
2. Application on simulated data - target variable Risk of Poverty (R).
3. The last day we expect dos to make a general description of HEIS surveys with regards to the indicators and the domains of interest. Basic information the size of the theoretical and realized sample size in the planned domains (governorates level) and for unplanned domains (district and sub-district levels).
4. Which relevant data are collected for the estimation needs?

Objectives component 2

ISSUES TO DEAL WITH

- ✓ Definition and theory behind SAE methodology
- ✓ Identification of the informative gaps with respect to the use small area methodology.
- ✓ Clarification of the needs
- ✓ Basic smoothing - Production of direct estimate and their variances,
- ✓ Production of other indirect estimators (synthetic and composite) from designs-based perspective, Assessment – required threshold, etc.
- ✓ Model based small area estimations;
- ✓ Definition and theory behind SAE models on the base of the characteristics of indicators;
- ✓ Quality Assessment.

- The ultimate goal is to provide DoS of the tools needed to deal with the issue of estimating parameters of interest at desired detailed level of disaggregation.

Outline

- Introduction to data disaggregation and Small Area estimation for SDGs Indicators
- Fundamental concepts for direct and indirect estimation
- Direct and Indirect estimators for data disaggregation of SDG indicators (poverty rate)
- Example and application with R

In the next mission:

- Small Area Estimation with Area Level Models
- Small Area Estimation with Unit Level Models
- Possible extension for complex indicator
- Quality Assessment

Data disaggregation and the SDGs

- With the adoption of the **2030 Agenda for Sustainable Development**, Members of United Nation (UN) have pledged to **leave no one behind (LNOB)** and reach the furthest behind first: Need for more disaggregated data than currently available in most countries.

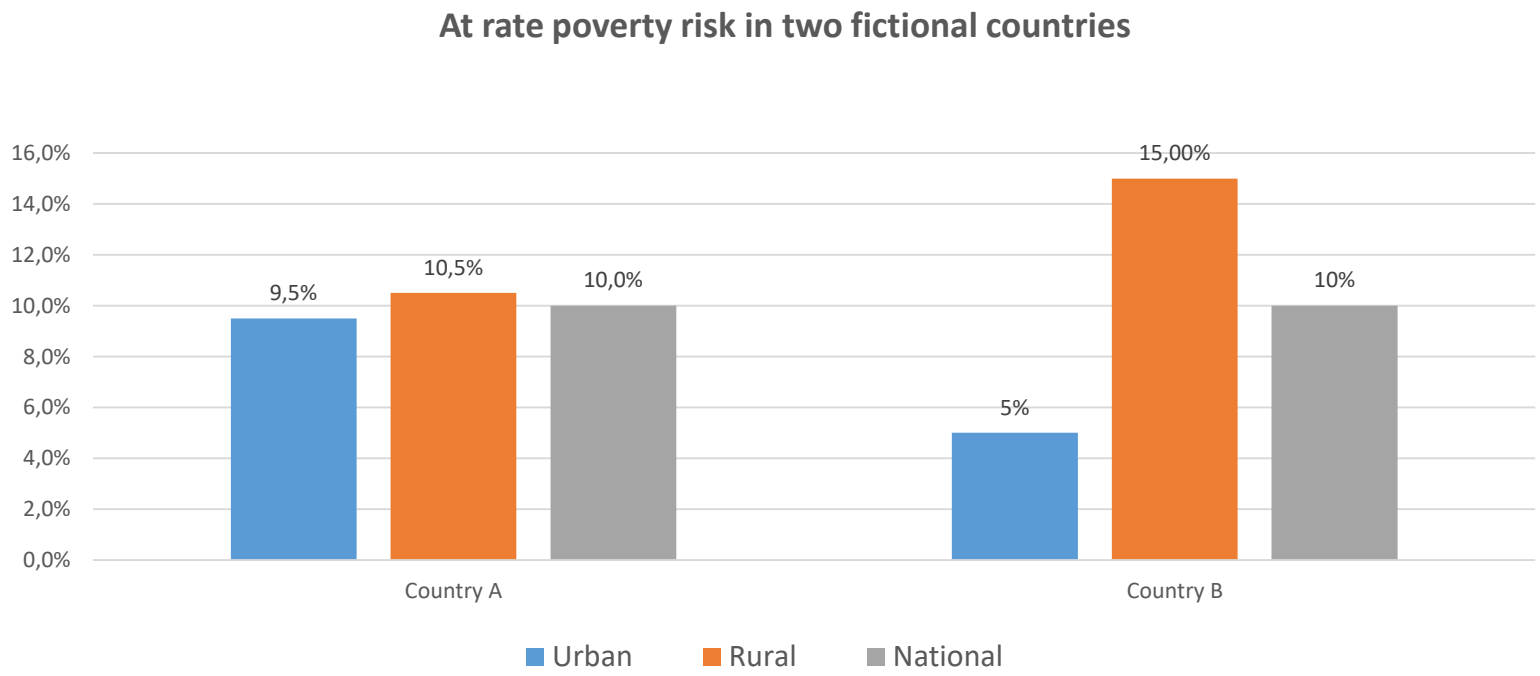
- Some goals, like **eliminate poverty, eradicate hunger** and **preventable child mortality**, are directly concerned with the most vulnerable populations: Need of statistics presented for vulnerable population groups and at granular geographic disaggregation level.

- An overarching principle of data disaggregation is at the core of the SDG Monitoring Framework:

"SDG Indicators should be disaggregated, where relevant, by income, sex, age, race, ethnicity, migratory status, disability and geographic location, or other characteristics in accordance with the Fundamental Principles of Official Statistics."

Data disaggregation and the SDGs

- Data can speak for those left behind.
- The **big picture** does not always portray the **full picture**.
- Need disaggregated data on vulnerable populations to understand: - Who they are; - Where they are; - How many they are.



Some example of SDGs: Poverty estimation goal of this project

Goal 1. End poverty in all its forms everywhere

Poverty estimation is made through the measurement of income and consumption levels.

A person is considered poor if his/her income level falls below a minimum level that fails to meet his/her basic needs.

This minimum level is known as the ‘poverty line’.

When the income or consumption of an individual or the household he/she belongs to fall below minimum level then they are designated to be Below the Poverty Line

Some example of SDGs: Poverty estimation goal of this project

Goal 1. End poverty in all its forms everywhere

- Indicator 1.1.1 Proportion of the population living below the international poverty line by sex, age, employment status and geographic location (urban/rural)

Goal: By 2030, eradicate extreme poverty for all people everywhere, currently measured as people living on less than $1.25 a day

- Indicator 1.2.1 Proportion of population living below the national poverty line, by sex and age

Goal: By 2030, reduce at least by half the proportion of men, women and children of all ages living in poverty in all its dimensions according to national definitions

Some example of SDGs

Goal 2. End hunger, achieve food security and improved nutrition and promote sustainable agriculture

- Indicator 2.1.2 Prevalence of moderate or severe food insecurity in the population, based on the Food Insecurity Experience Scale (FIES)

Goal: By 2030, end hunger and ensure access by all people, in particular the poor and people in vulnerable situations, including infants, to safe, nutritious and sufficient food all year round

Goal 3. Ensure healthy lives and promote well-being for all at all ages

- Indicator 3.3.1 Number of new HIV infections per 1,000 uninfected population, by sex, age and key populations
- Goal: 3.3 Number of new HIV infections per 1,000 uninfected population, by sex, age and key populations

Some example of SDGs

Goal 4. Ensure inclusive and equitable quality education and promote lifelong learning opportunities for all

- Indicator 4.3.1 Participation rate of youth and adults in formal and non-formal education and training in the previous 12 months, by sex
- Goal: 4.3 By 2030, ensure equal access for all women and men to affordable and quality technical, vocational and tertiary education, including university

Goal 5. Achieve gender equality and empower all women and girls

- Indicator 5.2.1 Proportion of ever-partnered women and girls aged 15 years and older subjected to physical, sexual or psychological violence by a current or former intimate partner in the previous 12 months, by form of violence and by age
- Goal: 5.2 Eliminate all forms of violence against all women and girls in the public and private spheres, including trafficking and sexual and other types of exploitation

Some example of SDGs

Goal 6. Ensure availability and sustainable management of water and sanitation for all

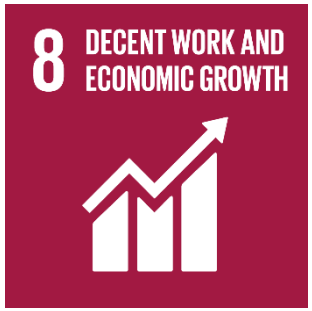
- Indicator 6.3.1 Proportion of domestic and industrial wastewater flows safely treated
- Goal: By 2030, improve water quality by reducing pollution, eliminating dumping and minimizing release of hazardous chemicals and materials, halving the proportion of untreated wastewater and substantially increasing recycling and safe reuse globally

Goal 7. Ensure access to affordable, reliable, sustainable and modern energy for all

- Indicator 7.2.1 Renewable energy share in the total final energy consumption
- Goal: 7.2 By 2030, increase substantially the share of renewable energy in the global energy mix

Some example of SDGs

Others wellbeing dimensions



A complete list here;

https://unstats.un.org/sdgs/indicators/Global%20Indicator%20Framework%20after%202022%20refinement_Eng.pdf

Assessment of data availability

TRADITIONAL

- Surveys: generally carried out to estimate various indicators, but the sample size allow to estimates parameters only at an high level of aggregation
- Census: do not have any sample size problem, but the information are collected less frequently and on more limited number of item respect to surveys

INNOVATIVE

- Administrative records (registers) : innovative sources already in use to produce official statistical information;
- Big data: still experimental ,and still do not in use to produce official statistics.

Using surveys as source of disaggregated data

- About the 30% of indicators in the SDG monitoring framework can be computed using microdata collected with some sort of household survey, and the list increases if we consider also surveys of different kind (e.g. agricultural surveys).

- The use of traditional sampling techniques imposes **limitations** on the production of disaggregated data and reliable estimates for small sub-populations. Techniques that could address some of these issues are far from being mainstreamed in National Statistical Offices.

Using surveys as source of disaggregated data

- **Sample surveys:** designed to provide broad-area estimates with high level of accuracy and precision. E.g. survey sample sizes are typically calculated such that estimates of the parameter of interest are reliable enough at national or regional level, or at certain highly aggregated level.
- In **planned domains**, the estimation of indicators is based on **direct estimators**, i.e. estimators solely based on survey data from the considered domain.
- Direct estimators are conventionally used by NSOs, based on the application of **weights to the sample units** belonging to the estimation domain – also named as design-based estimators

Two main issues:

1. Sampling size often not large enough to guarantee reliable estimates for small domains;
2. Possibility of having non sampled sub-domains.

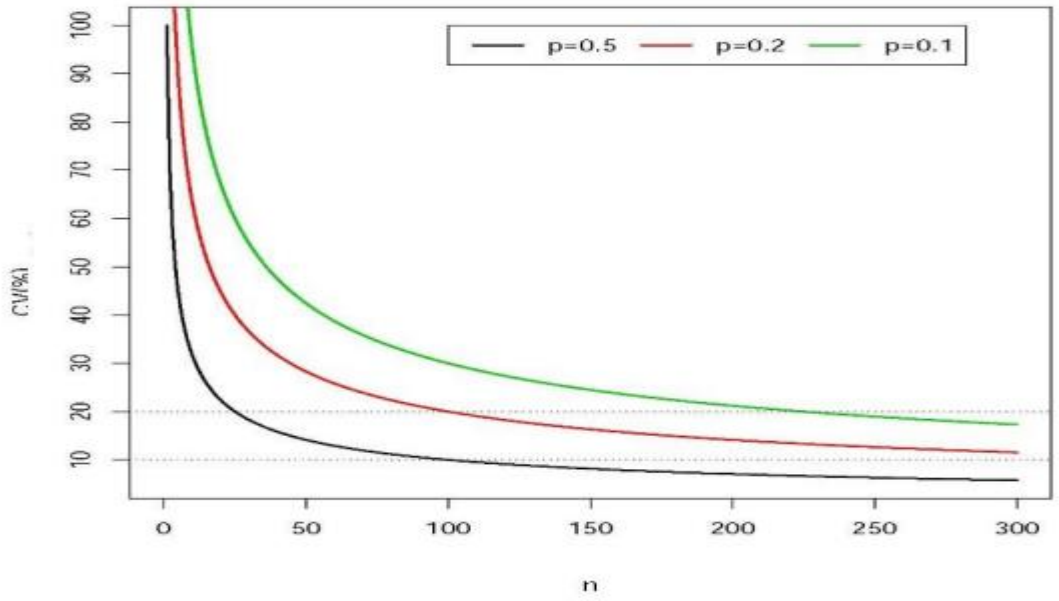
Using surveys as source of disaggregated data

At the design stage:

- Adopting sampling designs that guarantee an observed set of sampling units for every sub-population for which disaggregated data must be produced.
- Potentially optimal strategy, although it translates in an increased sample size and greater survey costs and complexity.

CV % of the sample proportion estimates as a function of the sample size $\boldsymbol{n}$, for each value of the true proportion.

thresholds



Using surveys as source of disaggregated data

- We cannot realistically plan all possible disaggregation domains at the design stage – data users and policy makers will always ask for more disaggregated data.
- Moreover bigger is the sample, bigger can be the respondent burden, the missing values and as consequence the non-sampling errors

At the inferential stage:

- adopting indirect estimation methods that cope with the little information available for “small areas” by borrowing strength from other data sources.

FOCUS OF THE TWINNING PROJECT!

Small Area Estimation

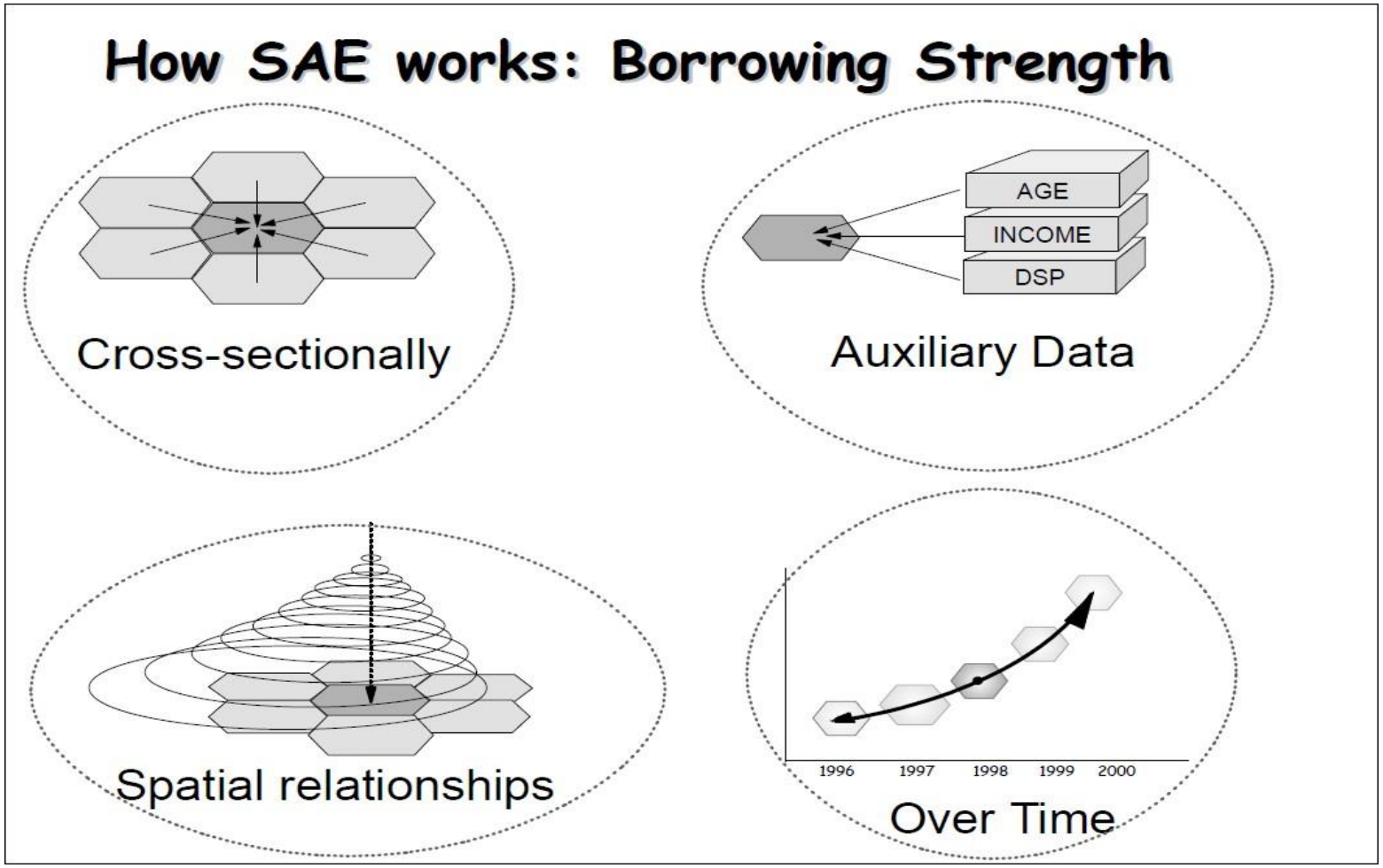
SAE methods

- Concerned with the development of statistical procedures to produce efficient (precise) estimates of the parameter of interest in small areas, i.e. **domains with small or even zero sample size**.

- **Combine survey data with an additional data source**, generally through an implicit or explicit model that allows to borrowing strength from other sources. The additional data source can be a census, administrative data or alternative data sources such as mobile phone or satellite data.

- When direct estimator **does not reach a user-specified level of precision** other estimators can be developed to reach this level of precision and to provide also predictions for domains where no sample information is available.

Small Area Estimation: how to overcome the lack of information



Context for using SAE

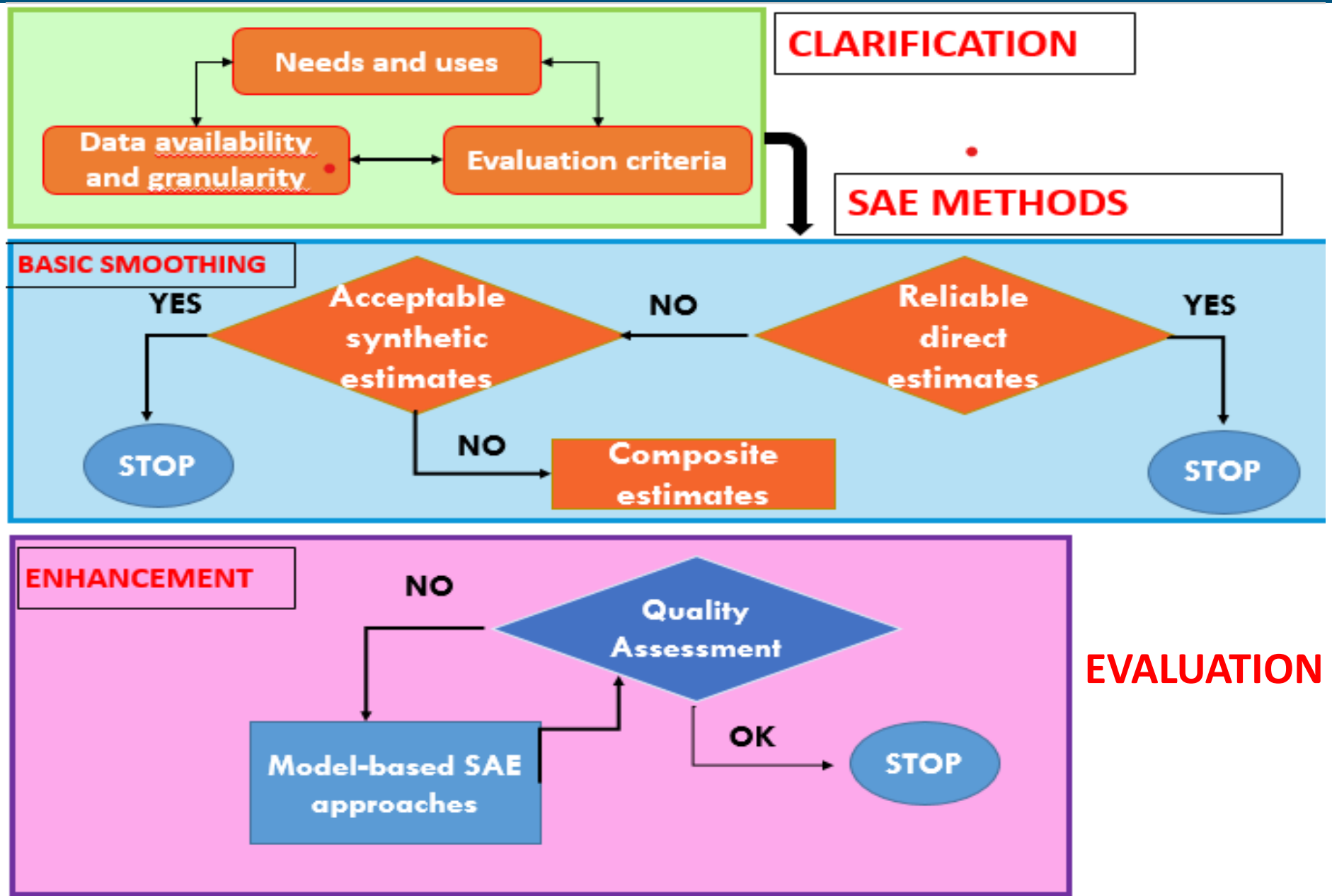
- **SAE** can **not be regarded** as a solution for all data disaggregation problems.

- Indirect estimation methods (including SAE) can be adopted when the **following conditions** coexist:

  - The considered (SDG) indicator is computed using **survey data**;
  - There is **not enough sampling information** to produce estimates at the desired disaggregation level using only survey data
  - **Suitable auxiliary information** is available to be combined with survey data.

The production of **Small Area Estimates is based on various steps** often summarized in flowchart that can support the production process.

Steps to compute SAS the process flow



From direct estimation to SAE: the process flow

Step 1: Clarification

- ✓ Identification and prioritization of needs and uses of SAE
- ✓ Assessment of data availability: survey and auxiliary data
- ✓ Choice of evaluation criteria (e.g. Thresholds CV)

- When developing an SAE implementation plan it is important to carefully examine the characteristics of the variables of interest for which disaggregated estimates need to be produced

- User needs can be identified by looking at the **purpose** of the estimation process:
  - *What are the key policies or funding decisions?*
  - *What are the questions that need to be answered?*

- Important to identify **which indicator** can actually measure the information of interest:
  - *What are you trying to measure?*
  - *What type of indicator is the indicator of interest?*

"The definition of the target parameter to be estimated is of vital importance. The target parameter needs to be well-supported by the available data. An increasing complexity of the indicators of interest simultaneously increases the granularity of the data that are needed for the estimation "

Target parameters: the process flow

Parameter of interest: target phenomenon for which aggregate and disaggregated estimates need to be produced – expressed through indicators for the purpose of this training.

Indicators can have different functional forms that needs to be considered when choosing the appropriate SAE approach. **Examples:**

Type	Description	Example
Total	The sum of values	Number of people with ownership or secure rights over agricultural land
Mean/Average	The sum of values divided by the number of values	Average agriculture income, average labour productivity, average crop yield
Proportion	Fraction of the population with specific characteristics	Proportion of individuals living below poverty line
Rate	Ratio between two quantities in different units	Mortality rate attributed to unsafe water

- The Proportion of individuals living below poverty line is linear if the target population for each small area of interest is known and the threshold is known and fixed

- Indicator 1.1.1 Proportion of the population living below the international poverty line by sex, age, employment status and geographic location (urban/rural)

- Indicator 1.2.1 Proportion of population living below the national poverty line, by sex and

Data disaggregation: What is the relevant dimension of disaggregation?

- The key dimensions for disaggregation include:

- Characteristics of the individual or household such as sex, ages, income, disability, religion, ethnicity, Economic activity and so on
- Spatial dimensions such as metropolitan areas, urban and rural, or districts.
  - **Estimation domains:** sub-populations of the target population, they can be planned or not at the survey design stage.
    - Geographic areas (regions, provinces, municipalities, health service areas)
    - Socio-demographic groups (sex, age, race)
    - Other sub-populations (e.g. the set of firms belonging to an industry subdivision)
    - Cross classifications of the above domains
  - **Small domain/area:** estimation domain where the direct estimator does not reach a pre-specified level of precision due to 1) small sampling size or 2) absence of

Data disaggregation and the SDGs: the process flow

Many of the SDG indicators, like labor, health, poverty and other dimensions of wellbeing, can all be derived from household surveys.

➢**Sample surveys:** designed to provide broad-area estimates with high level of accuracy and precision. E.g. survey sample sizes are typically calculated such that estimates of the parameter of interest are reliable enough at national or regional level, or at certain highly aggregated level.

➢In **planned domains**, the estimation of indicators is based on **direct estimators**, i.e. estimators solely based on survey data from the considered domain.

- The direct estimator cannot be obtained for some areas due to missing survey information on non-sampled areas.
- The direct estimates may be unreliable for unplanned domains.

From direct estimation to SAE: the process flow

Step 2: Basic smoothing

- Production of direct estimates and their variance
- Assessment : CV is below the required threshold. For instance Statistics Canada uses three categories of reliability for the Labour Force Survey: no release restriction for a CV ≤ 16.5% , added warning when 16.5% < CV≤ 33.3% and otherwise, the data is not recommended for release
- Production of other «intermediate» indirect estimators (synthetic and composite) from a design-based perspective (Mean Square Error- MSE)

Advantages and disadvantages of direct estimates

Data requirements

- Sampling weights ωdi for survey units.
- True population size Nd of area d for the basic HT estimator; sample observations of the auxiliary variables related to the variable of interest and correspondent Population total Xd in the area for direct calibrated estimator (GREG).

Pros:

- They make no model assumptions (non-parametric).
- They are exactly or approximately unbiased under the sampling design
- They are consistency: as the sample size increases, the probability that the estimator differs from the true value by more than ε approximates to 0, for every $\varepsilon > 0$.
- Additivity (Benchmarking property): the direct estimate of the total for a larger area covering several areas coincides with the aggregation of the estimates of the totals for the areas within the larger area.

Cons:

- Inefficient for small areas. For an area d with small nd, traditional area-specific direct estimators do not provide adequate precision.
- Direct estimates cannot be calculated for non-sampled domains.

Data disaggregation in unplanned domains: indirect estimators

Fundumentally, there are two types of indirect estimators:

- **Synthetic Estimators:**
  - A reliable direct estimator for a broad area, covering several small areas, is used to derive an indirect estimator for a small area.
  - *Produced under the assumption* (implict or explicit model) *that the small areas have the same characteristics as the broad area*.

  More precise than the direct but can be very biased is the assumption is not hold

- **Composite Estimators:**
  - A linear combination between a direct estimator and a synthetic one *using a design-based approach or by assuming an explicit area or unit-level model*.

  Represents a good compromise in terms of efficiency and bias of the two components. It can computed under a design or model based approch.

Inferential framework

Three main inferential approaches to estimation:

- **Design Based Approach:** The properties of the estimator are assessed with respect to the sampling design. This is the traditional and simplest approach. The estimation only relies on the observed values for the variable of interest and the sampling weights.

- **Model Assisted Approach:** The values of Y are typically defined by assuming a model for the distribution of Y given X, i.e. use of models in order to identify optimal strategies to estimate T_y.

- **Model Based Approach:** design-unbiasedness is no longer a requirement; the required property is instead:

$$E\left(\hat{T}_Y - T_Y \middle| S, X\right) = 0$$

In other words, in the model-based inferential framework, the **estimator is model-unbiased** given the sample S and auxiliary information X

Data availability

Auxiliary information can be available at **different levels of aggregation**:

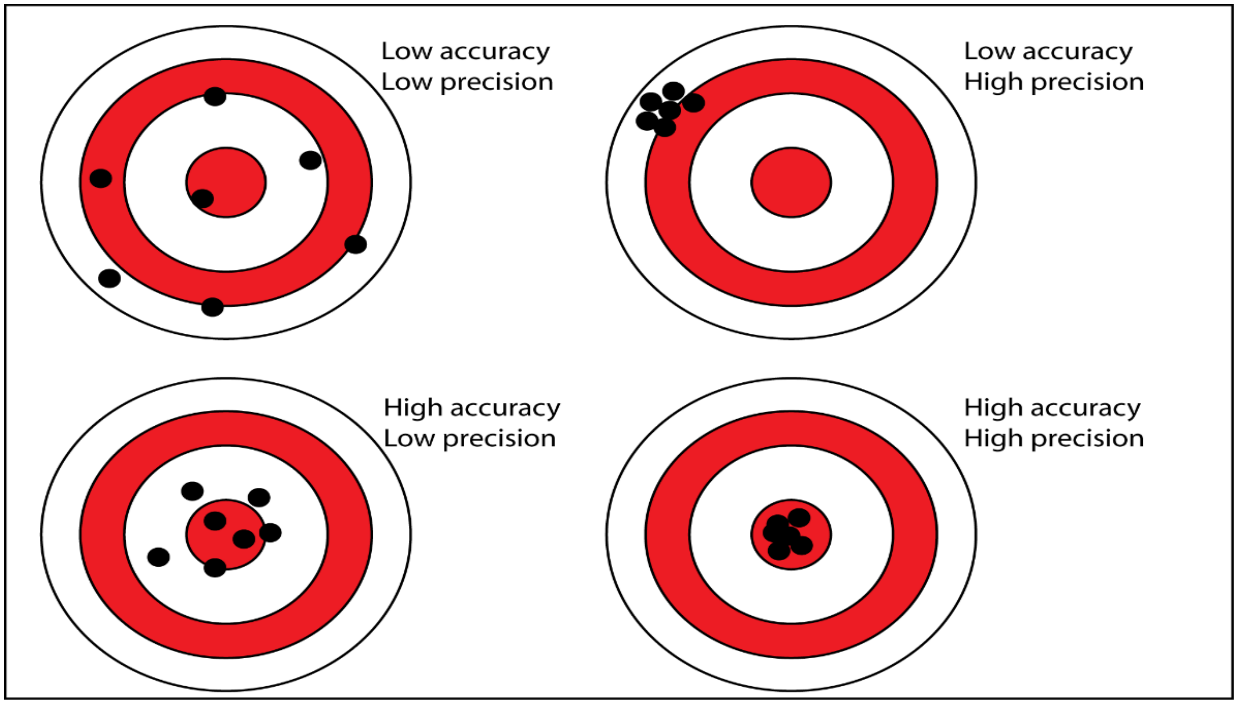
- **Area-level data:** auxiliary data are aggregated at the level of the considered estimation domain (e.g. district, municipality, etc.) and not available at the unit level.

- **Unit-level data:** auxiliary data available for each unit in each domain. A fundamental requirement of Unit-level SAE approaches is that auxiliary variables share the same definition in the survey and additional considered data sources (e.g. census, admin register, etc.).

Quality Assessment of indirect estimator in general

Key Concept: Reliability of small area estimates

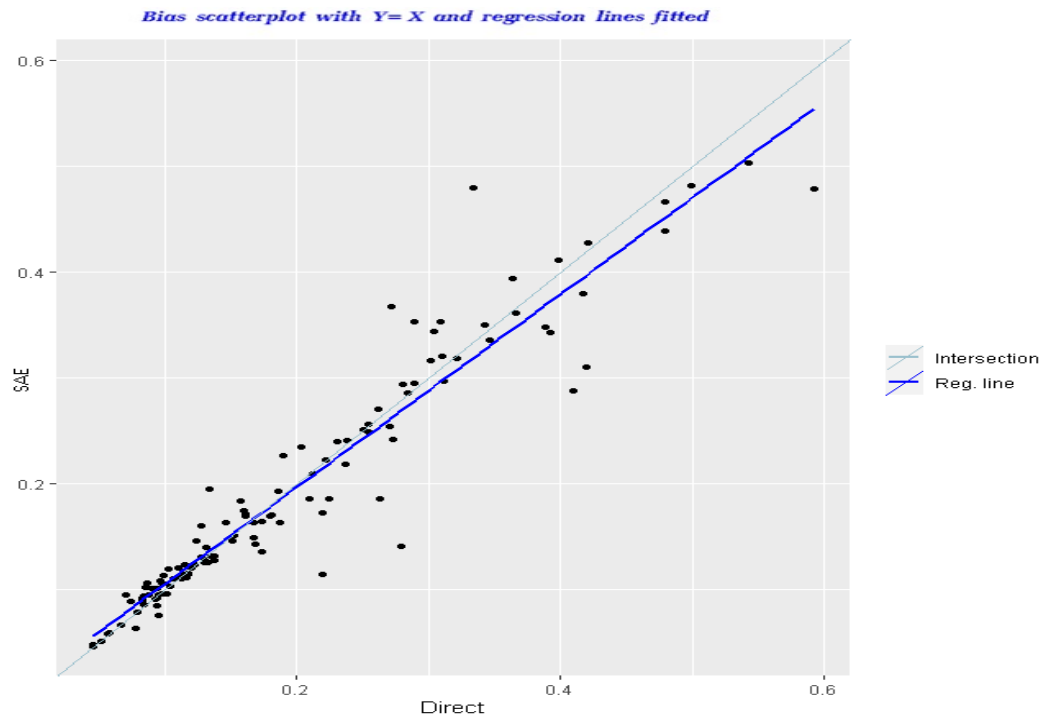
- **Precision** measures the closeness of the estimates to each other (variance or standard error) – the possible range of variation of estimates.
- **Accuracy** measures the closeness of the estimates to the true value of the parameter being estimated (bias).
- Joint measure of precision and accuracy = **Mean Sqare Error**



Quality Assessment of indirect estimator in general

Over shrinkage

• The degree of shrinkage can be empirically observed also in the regression of indirect estimates vs. direct estimates, especially when the difference between the slopes of the fitted line and Y=X line is large



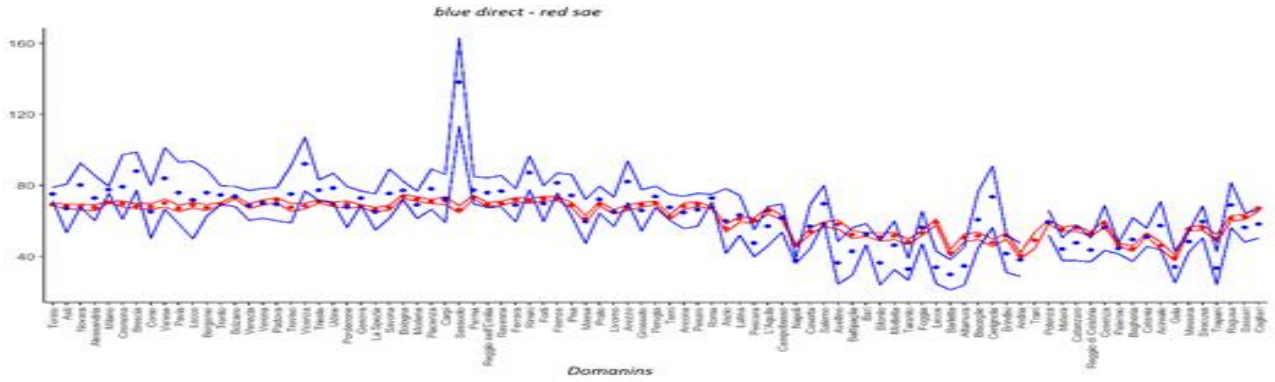
Quality Assessment of indirect estimator in general

Over shrinkage

• A numeric measure of shrinkage based also on the comparison between direct and indirect estimates is the Average Shrinkage Indicator (ASHR) (in percentage):

$$ASHR\left(\hat{\theta}^{INDIRECT}\right) = \frac{1}{D}\sum_{d=1}^{D}\left|\frac{\hat{\theta}_d^{INDIRECT} - \hat{\theta}_d^{DIRECT}}{\hat{\theta}_d^{DIRECT}}\right| \cdot 100$$

The SAE estimates should also fall in the confidence interval of the direct estimates Anyway, if the relative standard error for each small area whose direct estimates fall outside the confidence interval of the indirect estimates is significantly higher than those that do fall within then the smoothing, i.e. the shrinkage, is warranted.



From direct estimation to SAE: the process flow

Step 3: Enhancement

- Model-based (SAE) approaches. The model specification will strong depends on:
  - ✓ The characteristics of the indicator;
  - ✓ The disaggregation of the small domains of interest and thus the corresponding sample sizes
  - ✓ The data availability determines which SAE methods could actually be applied.

  Importantly, some SAE methods are applicable to one form of variable, but not to others. For other type of indicators different SAE method might be necessary

Data disaggregation in unplanned domains: indirect estimators

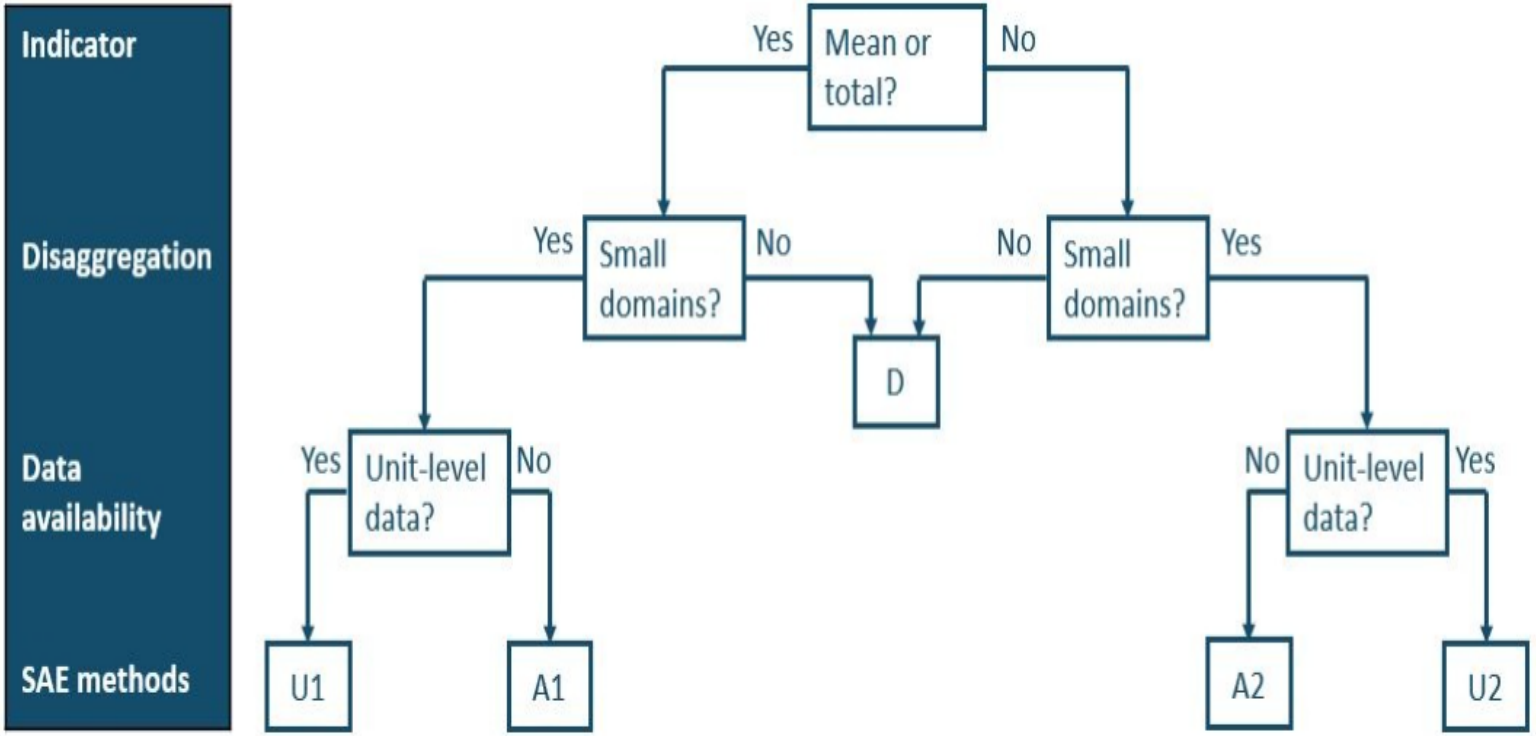
About Model Based Approach

- **Area Level Model:**
  - Area level model are appropriate if only area-level summary data available for the auxiliary and/or response variables. It is possible to take into account the sampling weights into model.

- **Unit Level Model:**
  - Unit level model are generally to prefer if unit-specific information is available. They usually ignore survey weights.

Model Based SAE methods

Initial specification depending on input factors



D: Direct estimation approaches.

A1: The basic area-level model and its extensions for means and totals including, e.g., the spatial-correlation and robust models.

A2: Area-level models for other indicators such as ratios. These can either use transformations or a non-linear model specification.

U1: The basic unit-level model and its extensions for means and totals, including robust models.

U2: Extended unit-level models such as the ELL and the EBP approaches.

Quality assessment of model based estimates

Model diagnostics

- ✓ Proper model specification;
- ✓ the checking of the model assumptions
- ✓ deal with departures from model assumptions

Reliability of small area estimates

- ✓ Precision measures (variance or standard error)
- ✓ Accuracy measures (bias)

Evaluation of set small area estimates

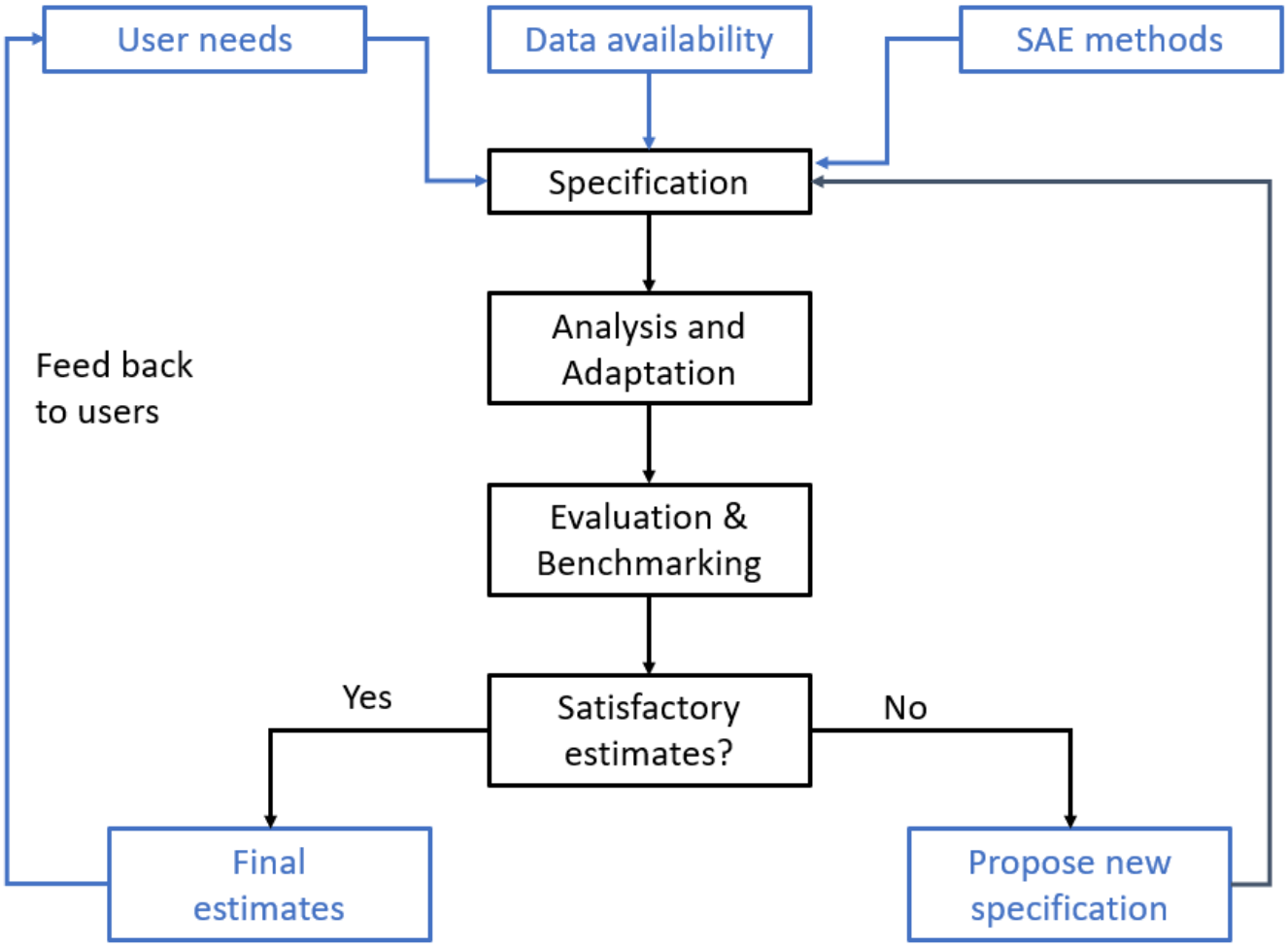
- ✓ Simulation studies can be conducted to evaluate the SAE methods;
- ✓ Comparison with direct estimates and benchmarking properties;
- ✓ Thematic analysis by users/experts are also important.

Main advantages of model-based approaches

Adopting a model-based approach has the following main advantages:

- Model diagnostics can be used to find suitable models that fit the data well.
- Area-specific measures of precision can be associated with each small area estimate, solving the problem of instability that affect synthetic and composite estimators.
- Various types of models can be used: linear mixed models as well as non-linear mixed models
- Complex data structures, such as spatial dependence and time series structures, can also be handled
- Methodological developments for random effects models can be utilized to achieve accurate small area estimates.

SAE production: summary



Communicating SAE methods and results

Fruitful communication and dissemination of small area estimation results requires knowledge about the target audience (e.g. policy makers, researchers, users in statistical offices);

The important aspects that should be communicated with users includes:

- ✓ Methods applied;
- ✓ Input data and models used;
- ✓ Assumptions and validation of followed process;
- ✓ Quality assessment of the estimates, including a possible comparison of the results over the time;
- ✓ Indication on how to interpret the results and ways to (or not to) use the results.

Challenges in using SAE for official data production- Cit: UNSTATS wiki page/ The SAE Toolkit

Overall: “The main challenges for NSIs when producing small area estimates is the ability to master the complexities of the required statistical theory (e.g. the assessment of the estimation error is recognized as a complex problem in the small area estimation context), the availability of relevant administrative data and the capacity to overcome internal and external barriers for the acceptance of model based estimates as trustworthy official statistics outputs.“

In particular:

- Lack of interest and support from the top management, hence lack of resource;
- Lack of dedicated resources for SAE research and implementation. Compared to other output of a National Statistical Office, SAE is usually a relatively minor one. While one household survey can produce a large number of indicators, great efforts are necessary for deriving just one indicator for small domains;
- Lack of in-house technical capacity;
- Lack of proper input data;
- Reluctance about the use of model-based estimates;
- Difficulties in communicating the technical aspects to users

Some International SAE application on SDG's (GOAL 1)

Goal 1. End poverty in all its forms everywhere

Case studies

Poverty mapping is one of most common applications in small area estimation. Many examples are available for the proportion of population living below the international or national poverty line (indicators 1.1.1 and 1.2.1).

World Bank applications

The World Bank proposed a poverty mapping process that was conducted in several countries. Based on surveys and additional data sources, various poverty and inequality estimates such as the Foster-Greer-Thorbecke poverty estimates and the Gini coefficient were derived.

The report More than a pretty picture - Using poverty maps to design better policies and interventions published in 2007 shows case studies for the countries **Albania, Bolivia, Bulgaria, Cambodia, Yunnan Province (China), Ecuador, Indonesia, Mexico, Morocco, Sri Lanka, Thailand** and **Vietnam** that describe all poverty mapping steps and also lessons learned. Hence this can be a good starting point for a new poverty mapping study.

In 2005, the World Bank provided technical assistance to the **Philippine** national statistical system to leverage on small area estimation techniques to produce municipality- and city-level poverty statistics. The Philippine Statistics Authority conducts the Family Income and Expenditure Survey (FIES), which is the main source of official poverty statistics in the country, every three years. The small area estimation technique used in the Philippines is based on the ELL method. It entails regressing (log) per capita household income from the FIES with auxiliary information from the FIES, the Labor Force Survey, and the Census of Population and Housing. The model regressors include survey-obtainable variables such as educational attainment of the household head and other household characteristics, and census-derivable information like average family size in a village, and other village-level information. Since small area poverty statistics became available in 2005, numerous government agencies have used these data as inputs for formulating and implementing poverty reduction programs. For example, the Philippine Department of Social Welfare and Development (DSWD) used the estimates to identify poor municipalities for its National Household Targeting System for Poverty Reduction (NHTS-PR) data collection.

Indicators	Disaggregation dimension	Data availability	Estimation approach	Model
1.1.1/1.2.1	Spatial	Unit-level survey and auxiliary data	Model-based estimation	ELL

Poverty estimation in Chilean comunas

To improve fund allocations among comunas, the Chilean Ministerio de Desarrollo Social (in the following the ministry) is required to provide poverty estimates for all 345 comunas in Chile which is the smallest territorial entity. After the evaluation of various options, the ministry decided to combine the National Socioeconomic Characterization Survey (CASEN), which is Chile's official data source for poverty statistics, with relevant administrative records. Since 2011, model-based poverty statistics are obtained for Chilean comunas.

Indicators	Disaggregation dimension	Data availability	Estimation approach	Model
1.2.1	Spatial	Unit-level survey and area-level auxiliary data	Model-based estimation	Arcsin-transformed area-level model

Small Area Income and Poverty Estimates (SAIPE) program by the U.S. Census Bureau

The SAIPE program produces small area estimates of income and poverty statistics for all school districts, counties, and states. The estimates are based on several data sources such as the American Community Survey and Federal Income Tax Returns. The produced indicators do not exactly follow the definition of the SDGs but the example is added since the SAIPE program is continuously improving their approach and the disaggregation is not only spatial but also by age groups.

Indicators	Disaggregation dimension	Data availability	Estimation approach	Model
	Spatial and age	Unit-level survey and area-level auxiliary data	Model-based estimation	Log-transformed area-level model

Mean income in Middle-layer Super Output Areas in England and Wales

Super Output Areas (SOAs) are a geographic hierarchy introduced for the reporting of small area estimates. The mean population of Middle-layer SOAs (MSOAs) ranges from a minimum of 5,000 to 7,200. To obtain different income estimates (e.g., equivalised and unequivalised) for the MSOAs, data from the Family Resources Data is combined with additional data including Census information, energy consumption and house price statistics using a linear mixed model.

The case study is explained on the homepage with more methodological details in the technical report.

Indicators	Disaggregation dimension	Data availability	Estimation approach	Model
	Spatial	Unit-level survey and area-level auxiliary data	Model-based estimation	Log-transformed unit-level model

Other examples: https://unstats.un.org/wiki/display/SAE4SDG/SAE+by+SDGs

ISTAT applications

- ✓ SAE to produce estimates of employment and unemployment rates for local labour market areas (LMA)

- ✓ Experimental SAE of a selection of labour market variables for cities and functional urban areas (FUA) based on a unit-level multivariate model. These are published as Italian sub-national statistics within the scope of an agreement between ISTAT and Eurostat.

The main indicators produced are:
Economically Active Population, total, and by sex
Economically Active Population aged 20-64, total, and by sex
Persons Unemployed, total, and by sex
Persons Employed aged 20-64, total, and by sex

Input data

The Italian Labour Force Survey (LFS) referring to year 2018 is the source of direct estimates whereas the Integrated System of Registers, in particular the Base Register of Individuals (RBI) and the Thematic Labour Registry (RTL), the Italian Ministry of Finance and the Revenue Agency provided auxiliary data (demographic, employment, social security and income information).

ISTAT case studies

Experimental studies to develop and propose SAE for SDG indicators

•Goal 1 - End poverty in all its forms everywhere

- Relative and Absolute Poverty indicators from EU-SILC and the Household Budget Survey (Laken indicators and others) for metropolitan cities and provinces.
- The challenge is how to choose the best method and the benchmarking approach.

•Goals 9 and 17 - Build resilient infrastructure, promote sustainable industrialization and foster innovation; Strengthen the means of implementation and revitalize the Global Partnership for Sustainable Development

- ICT indicators based on Multipurpose survey on everyday life aspect for metropolitan cities and provinces.
- Specific small area estimators can be applied like the design-based projection estimators, besides the classic SAE

- Goal 3 Ensure healthy lives and promote well-being for all at all ages
  - Health indicators based on European Health Interview Survey for Italian health districts/regions.
- Goal 5. Achieve gender equality and empower all women and girls
  - Direct and model-based estimates for indicators related to violence against women.

References and key readings

- Asian Development Bank (2020). Introduction to Small Area Estimation techniques: A practical guide for national statistical offices. https://www.adb.org/sites/default/files/publication/609476/small-area-estimation-guide-nsos.pdf

- Guidelines on small area estimation for city statistics and other functional geographies estimation. https://ec.europa.eu/eurostat/web/products-manuals-and-guidelines/-/ks-gq-19-011

- Guidelines on SAE ESSnet on SAE (2012) https://ec.europa.eu/eurostat/cros/system/files/WP6-Report.pdf

- Rao and Molina (2015). Small Area Estimation, New York: John Wiley & Sons.

- Tzavidis et al. (2018), From start to finish: a framework for the production of small area official statistics J. R. Statist. Soc. A (2018) 181,Part4,pp.927–979

- UNSTATS wiki page/ The SAE Toolkit) https://urlsand.esvalabs.com/?u=https%3A%2F%2Funstats.un.org%2Fwiki%2Fdisplay%2FSAE4SDG%2FSAE4SDG&e=17c5563b&h=502428d8&f=n&p=y