



Food and Agriculture
Organization of the
United Nations

SUSTAINABLE
DEVELOPMENT
GOALS

Guidelines on data disaggregation for SDG indicators using survey data

Office of the Chief Statistician -FAO

Clara Aida Khalil – Statistician
ClaraAida.Khalil@fao.org



July 20-23rd, 2021

FAO-OEA/CIE-IICA Working group on agricultural and livestock statistics for Latin America and the Caribbean
30th Session
Virtual Host: INEC Costa Rica

Data disaggregation and the SDGs

- With the adoption of the 2030 Agenda for Sustainable Development, Member States have pledged to leave no one behind (LNOB) and reach the furthest behind first:
 - This calls for more granular and **disaggregated data** than currently available in most countries.

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 (2)

Recognizing the many statistical challenges posed by data disaggregation, the IAEG-SDG has formed a **Working Group on Data Disaggregation**, with the objective of strengthening national capacities and develop the necessary statistical standards and tools to produce disaggregated data.

This led to:

- The development of a **minimum disaggregation set** and the compilation of **categories and dimensions of data disaggregation** currently in place and planned by custodian agencies.
- A comprehensive summary of **disaggregation standards and classifications** for all SDG Indicators.
- A compilation of **policy priorities** by disaggregation dimension.
- A compilation of **methods and tools** for data disaggregation.

Guidelines on data disaggregation for SDG Indicators



As a member of the **WG on data disaggregation** and the **TA on SAE**, the FAO has developed «**Guidelines on data disaggregation for SDG Indicators using survey data**».

Main objectives of the Publication:

- Offer methodological and practical guidance for the production of **direct** and **indirect** disaggregated estimates of SDG Indicators.
- Provide tools to assess estimates accuracy and present strategies for data integration, including **small area estimation** (SAE) methods

Relevance of the guidelines

- Approximately 30% of the Global SDG Indicators are based on survey data
- 7 out of 21 SDG Indicators under FAO custodianship can be computed using data from household and/or agricultural surveys.

ISSUE ADDRESSED: The use of traditional sampling techniques imposes limitations on the production of disaggregated data and reliable estimates for small sub-populations.

Innovative techniques that could address some of these issues are far from being mainstreamed in National Statistical Offices.

Data disaggregation with sample surveys

The guidelines in a nutshell:

- Direct estimates of an indicator for a given sub-population: based only on sample information from the sub-population itself. Two main issues:
 - Sampling size often not large enough to guarantee reliable estimates for small domains;
 - Possibility of having non sampled sub-domains.
- These issues can be addressed:
 - At design stage: adopting sampling designs that guarantee an observed set of sampling units for every sub-population for which disaggregated data must be produced.
 - At the analysis stage: producing indirect estimates, coping with the little information available for “small areas” by borrowing strength from other sources of data.

Addressing data disaggregation at the design stage

In order to produce **direct disaggregated estimates**, sampling strategies should ensure the presence of a sufficient number of sampling units in each disaggregation domain:

- **Straightforward** when the number of units belonging to a given sub-population can be determined from the sampling frame. In these cases, the main issue is the selection of the degree of oversampling to apply.
- **More complex** when members of rare sub-populations are not known in advance from the available sampling frame. A variety of methods can be used in these situations: large-scale screening, disproportionate stratified sampling, two-stage sampling, multiple frame sampling.

Addressing data disaggregation at the design stage (2)

The guidelines illustrate alternative sampling strategies for direct domain sampling estimation:

- Most common **domain estimators** are discussed, introducing their context of usability.
- **Model-assisted** and **model-based** estimation approaches are illustrated.

Mainstreamed and innovative approaches to address data disaggregation at sampling design stage are discussed (with their pros and cons):

- Oversampling
- Deeper stratification
- Multiphase sampling with screening of respondents
- Marginal stratification designs
- Indirect sampling

Suitable software packages are suggested

Assessing estimates accuracy

Tools to assess the accuracy of direct estimates are provided:

- **Sampling variance estimation:** to be used when computation of indicators is based on the inferential properties of repeated sampling schemes.
- **Model Variance:** suitable when estimation relies on models using auxiliary variables.
- **Global Variance:** when model-based approaches are used jointly with inference based on the sampling design.

The publication emphasizes the importance of **estimating** and **disseminating** accuracy measures:

- To enable users assessing the fitness for use of an estimate.
- To build public trust in data and their use.

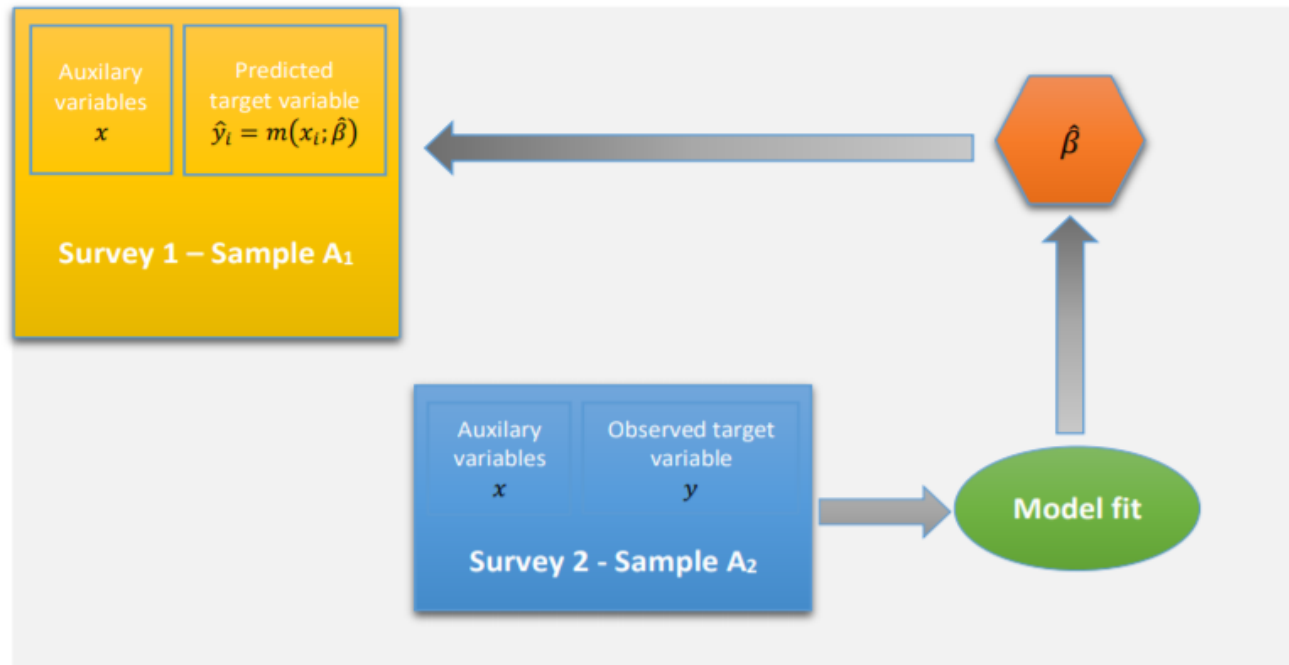
Addressing disaggregation at the analysis stage

- At the analysis stage, data disaggregation can be addressed adopting **indirect estimation approaches** - including **SAE** techniques - coping with the little information available for so-called small areas by borrowing strength from additional sources.
- **Projection estimator** (Kim and Rao, 2012): model-assisted approach that allows integrating data from two sample surveys, where the first survey, is characterized by a large sample A_1 , but only collects auxiliary information or variables of general use (e.g. socio-economic variables); while the second survey has a smaller sample A_2 but collects information on the target variable y , along with the same set of auxiliary variables available from A_1 .

Addressing disaggregation at the analysis stage (2)

The total of variable y in the disaggregation domain d can be obtained as

$$\hat{Y}_{PR,d} = \sum_{i \in A_1} w_{i1} m(x_i; \hat{\beta}) \gamma_{id}$$



Source: FAO, 2020.

A practical application based on [SDG Indicator 2.1.2](#)

The approach has been adopted to produce disaggregated estimates of [SDG Indicator 2.1.2](#) on the Prevalence of Moderate and Severe Food Insecurity based on the Food Insecurity Experience Scale (FIES).

Two data sources:

- Malawi's Fourth Integrated Household Survey (IHS4) 2016-17
- Malawi FIES survey module collected through the Gallup World Poll – 2016

Objective: Estimate Indicator 2.1.2 by sex, age class, and income quintile.

The guidelines present results along with their accuracy measures

An introduction to Small Area Estimation

The guidelines also provide an [introduction to small area estimation methods](#) (SAE) by:

- Presenting the process flow for [SAE implementation](#);
- Providing an overview of main [unit-level](#) and [area-level](#) approaches;
- Indicating main references on the topic;
- Giving tools to assess the quality of small area estimates.

Way forward

Starting from this work, the FAO is:

- Implementing additional case studies based on the [projection estimator](#) and [SAE techniques](#) for other countries and other FAO-relevant SDG Indicators;
- Developing methodologies and guidelines to integrate survey data with additional data sources: [census data](#), [administrative data](#), [geo-spatial information](#).
- Trying to engage countries for the provision of technical assistance on data disaggregation for SDG Indicators.

Thank you