



Food and Agriculture Organization
of the United Nations

**TWENTY-NINTH
SESSION OF**

**THE ASIA
AND PACIFIC COMMISSION
ON AGRICULTURAL
STATISTICS**

Ulaanbaatar, Mongolia
22-25 November 2021

**Guidelines on data
disaggregation for SDG
Indicators using
Survey Data**

APCAS29

Aida Khalil - Statistician
FAO - Office of the Chief Statistician

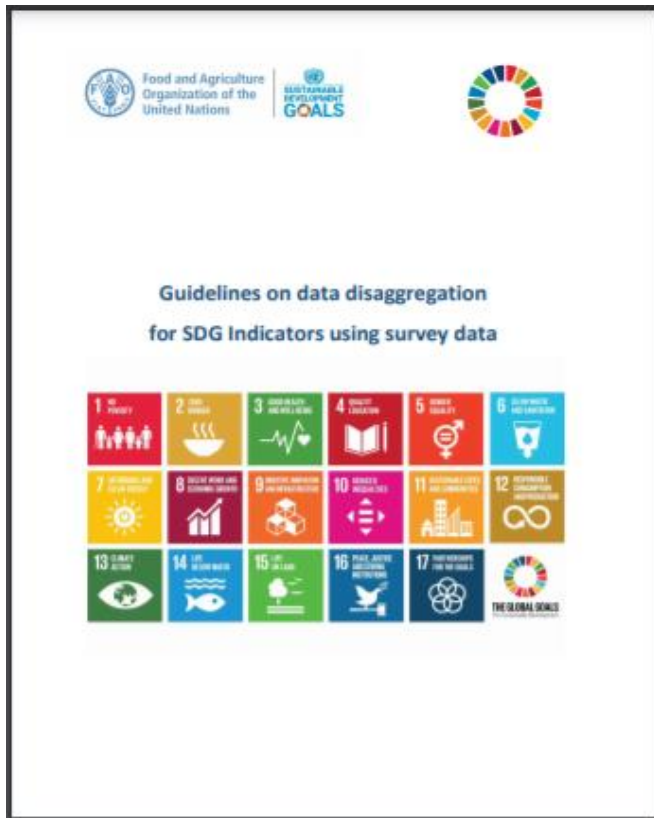
Data disaggregation and the SDGs

- 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: Need for more **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.”



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

Link to the Guidelines: <http://www.fao.org/documents/card/en/c/cb3253en/>

Relevance of the guidelines

- Approximately 1/3 of the Global SDG Indicators can be computed using survey data.
- 6 out of 21 SDG Indicators under FAO custodianship can rely on 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 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.
 - At the analysis stage: producing indirect estimates, coping with the little information available for “small areas” by borrowing strength from other data sources.

Addressing data disaggregation at sampling 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 sub-populations are not known in advance from the available sampling frame.

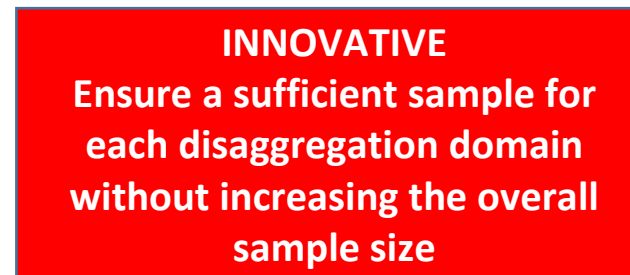
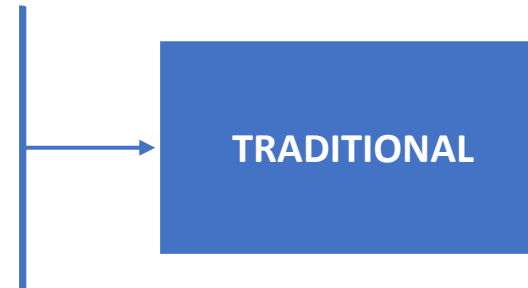
Addressing data disaggregation at sampling design stage (2)

Mainstreamed and innovative approaches to ensure sufficient sample size for each disaggregation domain are illustrated and discussed (with their pros and cons), including:

- Oversampling
- Deeper stratification
- Multiphase sampling with screening of respondents

- Marginal stratification designs
- Indirect sampling

Suitable software packages are suggested



Addressing data disaggregation at the analysis stage

- The guidelines illustrate alternative strategies for **direct domain sampling estimation**
- Most common **direct and indirect model-assisted domain estimators** are discussed, introducing their context of usability.
- **Small Area Estimation** (SAE) techniques are illustrated:
 - Presenting the process flow for their 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.

Addressing data disaggregation at the analysis stage (2)

- **Projection estimator** (Kim and Rao, 2012) is introduced, discussed and experimented on actual survey data.

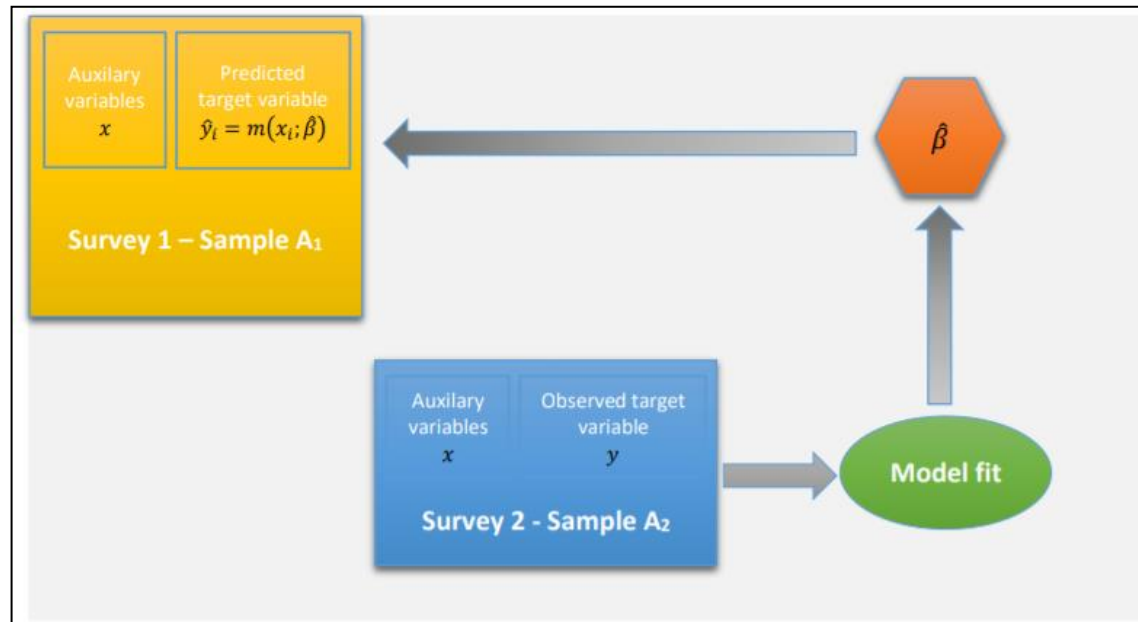
It allows producing disaggregated indicators by the **joint use of two sample surveys**:

- **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);
- **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 data disaggregation at the analysis stage (3)

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

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



A practical application based on SDG Indicator 2.1.2

The indirect estimation approach based on the projection estimator 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).

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

Two data sources:

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

Results are presented along with accuracy measures of indirect estimates.

Assessing estimates accuracy

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.

Methods and 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 **only on models using auxiliary variables**.
- Global Variance: when **model-based approaches are used jointly with** inference based on the **sampling design**.

Key takeaways and recommendations

- SDG Indicators should be disaggregated by relevant disaggregation dimensions. Each indicator has its specific requirements in terms of data disaggregation for Global reporting. In addition, NSOs should engage with national policy makers in order to identify disaggregation dimensions to be prioritized and planned at the survey design stage.
- Unplanned disaggregation dimensions: indirect estimation techniques can be adopted to integrate survey data with additional data sources such as census data, administrative registers and geo-spatial databases. The adoption of indirect estimation approaches allows increasing estimates accuracy keeping survey costs under control.
- FAO should continue developing tools and indirect estimation methods to integrate alternative data sources to disaggregate SDG indicators under its custodianship
- FAO can provide technical assistance to member countries for the production of SDG indicators disaggregated by relevant dimensions using direct and indirect estimation approaches.

Thank you!

ClaraAida.Khalil@fao.org