Why is SAE important for SDG data disaggregation

This section briefly explains in which cases small area estimation methods may be useful. It includes a short explanation why disaggregation is needed with regards to the SDGs and what the shortcomings of direct approaches may be.

- Demand for more disaggregated data for SDG monitoring
- Limitation of direct estimation from household surveys
- Small area estimation for SDG data disaggregation

Demand for more disaggregated data for SDG monitoring

The need for disaggregated statistical indicators can have various roots. In these guidelines, the reason for disaggregation is the monitoring process of the Sustainable Development Goals.

In 2015, the Member States of the United Nations committed themselves to the 2030 Agenda for Sustainable Development that contains 17 Sustainable Development Goals (SDGs) and 169 targets. Tracking progress on the SDGs at a global level requires statistical indicators and its sound interpretation. Additional to the global monitoring purpose, high quality, timely and reliable data have the potential to support evidence-based policy decisions and to design allocation programs in social, economic and environmental fields at the national level.

During the process national statistical offices (NSOs) and other government agencies of the Member states were asked to provide a national indicator framework, optimally including a set of Global Monitoring Indicators harmonized across countries for comparability. The majority of indicators should be based on official data sources such as administrative data, censuses, registers, or household surveys, while new data sources were also discussed and evaluated.

One of the core principles of the 2030 Agenda for Sustainable Development is to leave no one behind which includes to identify the ones furthest behind and ensure that the targets are met for all individuals. Therefore, one part of the review processes is the provision of data that is high-quality, accessible, timely, reliable and disaggregated by income, sex, age, race, ethnicity, migration status, disability, and geographic location, and other characteristics relevant in national contexts and for the specific target. The disaggregation of data for SDG indicators ensures the monitoring of inequalities.

To identify if targets of the SDGs are achieved for all relevant groups, indicators that can be disaggregated are preferred for the monitoring process. The key dimensions for disaggregation include:

1. Characteristics of the individual or household such as sex, ages, income, disability, religion, ethnicity and indigenous status,
2. Economic activity,
3. Spatial dimensions such as metropolitan areas, urban and rural, or districts.

For more information, which disaggregation is suggested for which indicator, the reader is referred to the E-Handbook of the SDGs and to the Compilation on Data Disaggregation Dimensions and Categories for Global SDG Indicators which provides a complete overview of all disaggregation dimensions and categories relevant to all SDG Indicators.

A few examples of SDG indicators are presented below, underlining the data disaggregation dimensions required by the indicators. Note that the highlighted disaggregation dimensions are those are specified by the indicators for the global SDG indicator framework. Countries would need to adjust the data requirements to meet the national and local policy needs. For example the required data disaggregation dimensions for SDG indicator 11.2.1 are sex, age and person with disabilities. Location, even though is not specified in the indicator, can be quite relevant for policy implementation.

Goal 1: End poverty in all its forms everywhere

**Target 1.1:** By 2030, eradicate extreme poverty for all people everywhere, currently measured as people living on less than $1.25 a day.

**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 8: Promote sustained, inclusive and sustainable economic growth, full and productive employment and decent work for all

**Target 8.5:** By 2030, achieve full and productive employment and decent work for all women and men, including for young people and persons with disabilities, and equal pay for work of equal value

**Indicator 8.5.2:** Unemployment rate, by sex, age and persons with disabilities

Goal 11. Make cities and human settlements inclusive, safe, resilient and sustainable

**Target 11.2:** By 2030, provide access to safe, affordable, accessible and sustainable transport systems for all, improving road safety, notably by expanding public transport, with special attention to the needs of those in vulnerable situations, women, children, persons with disabilities and older persons

**Indicator 11.2.1:** Proportion of population that has convenient access to public transport, by sex, age and persons with disabilities

---

**Mapping of the SDG indicators: the role of household surveys**

The Inter-Secretariat Working Group on Household Surveys carried out a mapping exercise and identified around one-third of SDG indicators can be derived from household surveys. Such assessment was submitted to the UN Statistical Commission at its 50th session in 2019. More information is available [here](#).

---

**Direct and indirect estimator**

A **direct estimator** uses values of the variable of interest only from the time period of interest and from units in the domain of interest.

An **indirect estimator** uses values of the variable of interest from a time period other than that of interest and/or from a domain other than that of interest.

Limitation of direct estimation from household surveys

According to an assessment made by the Inter-Secretariat Working Group on Household Surveys (ISWGHS), about one-third of the global SDG indicators can be derived from household surveys. (United Nations, 2019) For example, information on labour, health, education, poverty and other dimensions of wellbeing can all be derived from household surveys.

To produce statistical indicators, one approach is direct estimation, i.e. the estimation of indicators solely based on survey data. Direct estimators are conventionally used by national statistical offices, based on the application of weights to the survey sample units belonging to the small area or domain of interest. Depending on the survey characteristics, various features have to be taken into account for the estimation such as sampling design, survey weights, and imputation. In these guidelines, the focus is not on direct estimation (see further literature in the information box). Instead, the limitations will be shown to motivate the usage of combining survey information with other data sources.

A typical household survey is planned for providing estimates with controlled precision at a certain level of geographical disaggregation (areas) or for specific target populations (domains). However, on any other disaggregated geographical level, or for other sub-populations the survey may lack information due to small or nonexistent sample points.

The figures below show maps of direct estimates for a proportion calculated for regions in Colombia corresponding to the geographical disaggregation level at which the survey was planned (administrative level 1) and for a lower regional level (administrative level 2). Grey colored administrative level 2 regions are those where no survey data are available. Please note that the presented values are obtained based on synthetic data and only serve illustrative purposes. The values are no real estimates for Colombian regions.

Direct estimates for planned geographical disaggregation, administrative level 1

Direct estimates for administrative level 2

**Direct estimation/Survey methodology**

Some literature about survey sampling and direct estimation and its distinction to indirect estimation:


Small area estimation for SDG data disaggregation

The maps in the figures above reveal a situation in which:

- 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**.

To address these issues, the following solutions may be possible:

1. Design the data collection in a way that there is enough information to produce reliable indicators at any desired disaggregation level which usually means an increase of sample size and costs.
2. Use statistical methods that add information to the existing survey information.

Option one is ideal but costly. For the second option, **small area estimation** is one approach among other statistical methods. SAE methods combine survey data with an additional data source, generally through a modelling procedure. The additional data source can be population census, administrative data or alternative data sources such as mobile phone or satellite data.

As pointed by Rao and Molina (2015), small area estimation may be useful when the **direct estimator does not reach a user-specified level of precision**. Small area estimator is an estimator developed to reach this level of precision on the disaggregation level of interest. Furthermore, small area estimators can provide predictions for domains where no sample information is available.

By constructing the models strength is borrowed over space and/or time. Therefore, the small area estimation procedures constitute a model-based approach and the corresponding estimators are known as model-based estimators. In contrast, estimation of large sampling errors. For more information on survey design and total survey errors including coverage error, nonresponse error, measurement error, processing error and sampling error, see e.g., Lohr (2010), Sampling: Design and Analysis.

To summarize, small area estimation (SAE) methods can be useful if

- No other data source is available to produce estimate for the indicator.
- The statistical indicator is based on survey data, and
- There is not enough or no information for the disaggregation level of interest if relying only on survey data, and
- Suitable data source is available to be combined with survey data to construct SAE models.

In addition to the above considerations, an important element to consider when deciding whether SAE should be used for more disaggregated data is policy-relevance and availability of resources. SAE method requires tailored approach for each indicator, while one large household survey can typically produce estimates for a large number of indicators. From this angle, SAE is not really a "less-expensive alternative" to direct estimates obtained from household surveys. An effective way of using SAE is to produce data that cannot be obtained from other sources for a focused list of indicators such as the SDG indicators.

Of course, availability of source data, validity of the right variables for SAE modelling and finding the right model are all crucial elements to make SAE work. Once a valid model is established through rounds of testing, countries would also need to ensure there is sufficient resources for sustainability. More discussion on how to move from SAE experiment to official data production is available later.

What does unreliable mean in SAE?

A common way to measure **reliability** of a SAE estimate is using the coefficient of variation (CV).

It is a measure of the variability of the estimate (the standard error) as a percentage of the estimate and can be used to derive a confidence interval.

Unreliable in the SAE context means that the coefficient of variation of the estimates is above a pre-defined/user-specified threshold.

“Roughly, an estimate with a coefficient of variation less than 10% is considered reliable, whereas estimates with high coefficient of variation (i.e., greater than 20%) are unreliable.” (Asian Development Bank, p.17, 2020)

However, it is important to note that an estimator could also be unreliable because of large sampling errors. For more information on survey design and total survey errors including coverage error, nonresponse error, measurement error, processing error and sampling error, see e.g., Lohr (2010), Sampling: Design and Analysis.

Source
