



SAE: from experiment to production

Yongyi Min, Chief, Global SDG Monitoring Section,
UNSD

Haoyi Chen, Coordinator, Inter-Secretariat Working
Group on Household Surveys

Inter-Secretariat Working Group on Household Surveys

- Improve **coordination**: surveys within the country and efforts at the global level
- Advance (cross-cutting) survey **methodology**
- Enhance **communication and advocacy**

- Established at 46th Session of UNSC in 2015
- Current (rotating) co-chairs: UN Women and WB
- Secretariat: UNSD
- Members: 11 international agencies and 8 countries



New members

Welcome aboard!

As of August 2020, 8 Member States joined ISWGHS following the recommendations from IAEG-SDGs!

We look forward to your guidance and expertise.

Inter-agency and Expert Group on Sustainable Development Goal Indicators (IAEG-SDGs)

The 2030 Agenda for Sustainable Development

- A global blueprint for people, planet, prosperity, peace and partnerships, now and in the future
- 17 Goals, 169 targets and “Leaving no one behind” principle



The IAEG-SDGs :

- Composed of 28 Member States (and representatives of regional commissions, regional and international agencies and CSOs are observers)
- Developed the global indicator framework for SDGs (**231 indicators**)

IAEG-SDGs workstream on data disaggregation:

- Compilation of existing guidelines and methodologies on data disaggregation
- Preparation of Handbook on data disaggregation for SDGs
- Task Force on Small Area Estimation (joint with ISWGHS)

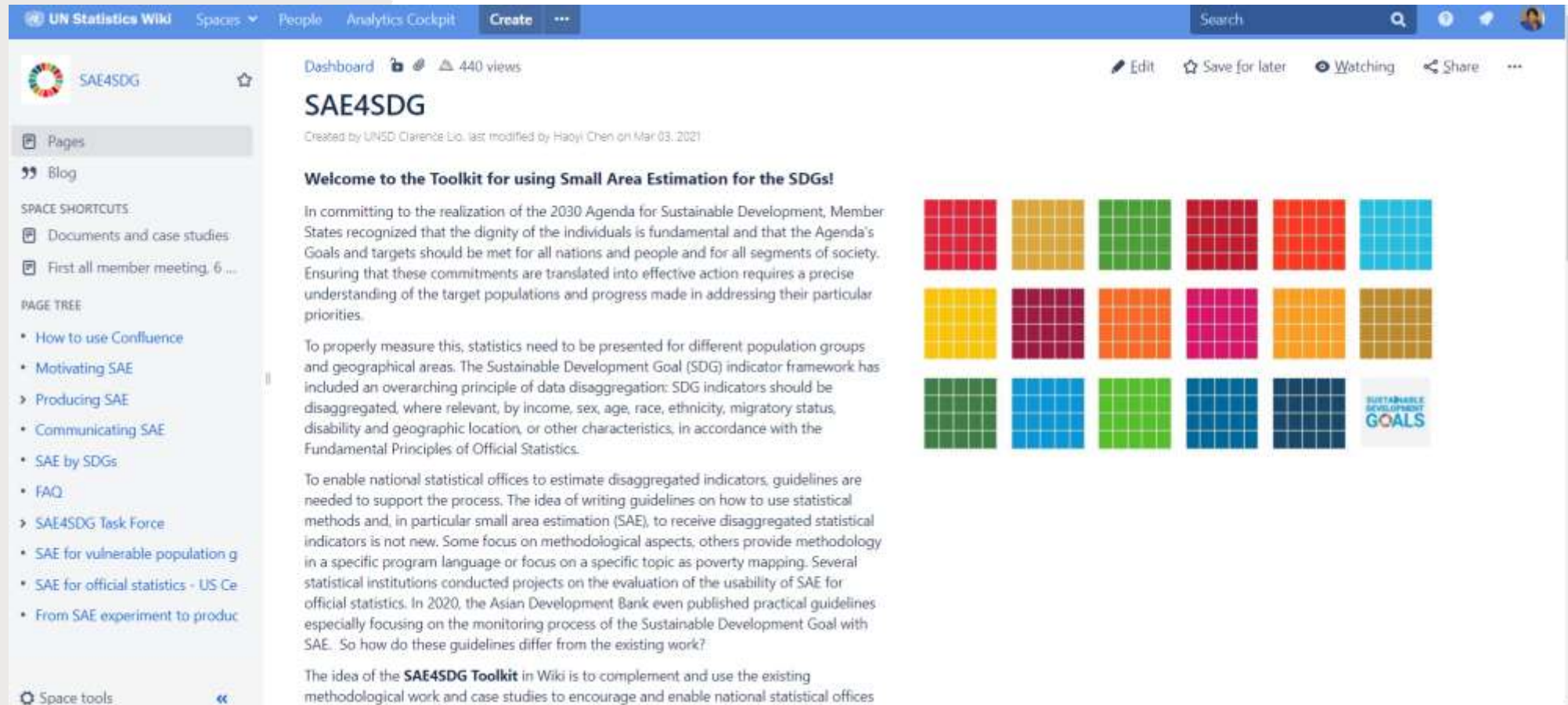


SAE Toolkit - overall objectives

- Using SAE methods to improve SDG data availability for vulnerable population groups – requested by IAEG-SDGs
- Offering practical guidance and country case studies
- Guiding on the enabling environment for using SAE for official data production
- Providing a space for partners to document and disseminate their SAE methodologies: transparency

Work modality

- A group of experts providing guidance
- Wiki-platform



The screenshot shows the UN Statistics Wiki page for SAE4SDG. The page title is "SAE4SDG" and it has 440 views. The page content includes a welcome message to the toolkit for using Small Area Estimation for the SDGs, a paragraph explaining the importance of disaggregated data, and a paragraph discussing the need for guidelines to support national statistical offices. The page also features a grid of colorful icons representing the Sustainable Development Goals.

UN Statistics Wiki Spaces People Analytics Cockpit Create ... Search

SAE4SDG ☆

Pages

Blog

SPACE SHORTCUTS

- Documents and case studies
- First all member meeting, 6 ...

PAGE TREE

- How to use Confluence
- Motivating SAE
- › Producing SAE
- Communicating SAE
- SAE by SDGs
- FAQ
- › SAE4SDG Task Force
- SAE for vulnerable population g
- SAE for official statistics - US Ce
- From SAE experiment to produc

Space tools

Dashboard 440 views

SAE4SDG

Created by UNSD Clarence Liu, last modified by Haoyi Chen on Mar 03, 2021

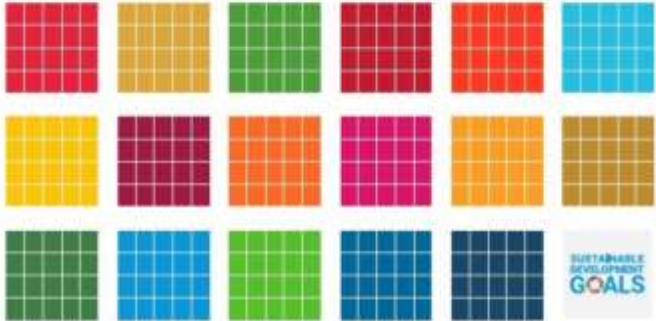
Welcome to the Toolkit for using Small Area Estimation for the SDGs!

In committing to the realization of the 2030 Agenda for Sustainable Development, Member States recognized that the dignity of the individuals is fundamental and that the Agenda's Goals and targets should be met for all nations and people and for all segments of society. Ensuring that these commitments are translated into effective action requires a precise understanding of the target populations and progress made in addressing their particular priorities.

To properly measure this, statistics need to be presented for different population groups and geographical areas. The Sustainable Development Goal (SDG) indicator framework has included an overarching principle of data disaggregation: 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.

To enable national statistical offices to estimate disaggregated indicators, guidelines are needed to support the process. The idea of writing guidelines on how to use statistical methods and, in particular small area estimation (SAE), to receive disaggregated statistical indicators is not new. Some focus on methodological aspects, others provide methodology in a specific program language or focus on a specific topic as poverty mapping. Several statistical institutions conducted projects on the evaluation of the usability of SAE for official statistics. In 2020, the Asian Development Bank even published practical guidelines especially focusing on the monitoring process of the Sustainable Development Goal with SAE. So how do these guidelines differ from the existing work?

The idea of the **SAE4SDG Toolkit** in Wiki is to complement and use the existing methodological work and case studies to encourage and enable national statistical offices



SAE by SDGs

Created by Ann-Kristin Kreutzmann, last modified by Haoyi Chen just a moment ago

This page gives a small guide on how to start a SAE case study and collects case studies for the Sustainable Development Goals. Case studies are not available for all SDGs yet, but more cases can be added continuously.

- [How to start a SAE case study](#)
- [Case studies for the estimation of disaggregated SDG indicators](#)
- [Case study submission](#)
- [References](#)

How to start a SAE case study

The guidelines give an overview of literature, available software and basic topics in small area estimation. While the concrete specification of a case study will vary among different applications, the same questions need to be answered.

	Questions
User needs <ul style="list-style-type: none">• Goal• Indicator of interest• Disaggregation level	<ul style="list-style-type: none">• What are the key policies or funding decisions?• What questions need to be answered?• What are you trying to measure?• What type of indicator is the indicator of interest?• What is the relevant dimension of disaggregation?
Data availability	<ul style="list-style-type: none">• Which survey data is available for the estimation of the indicator?• What are the data challenges?• Which additional data sources can be used?
SAE methods/Specification	<ul style="list-style-type: none">• Which SAE approach can be used based on the inputs above?• Which approaches are available in statistical software?• What are the available expertise to do the computation, analysis and interpretation?
Model validation	<ul style="list-style-type: none">• What is the plan for data validation?• What data sources for benchmarking?• Any plans for external review process?
Model refining	<ul style="list-style-type: none">• Plan to refine the model
Extend case study to official production	<ul style="list-style-type: none">• Plan or a roadmap to extend the case study for official data production?

Case studies for the estimation of disaggregated SDG indicators

In the following, case studies are summarized for the different SDGs. The descriptions are short and usually refer to a publicly available longer description of the study. The tables sum up user needs (indicators and disaggregation level), data availability and the specified estimation approach. The idea is to learn from other cases since some problems occur in different applications and thus for some problems, solutions may be found in another application.

Goal 1. End poverty in all its forms everywhere

[> Case studies](#)

SAE for official statistics

Dashboard / SAE4SDG   29 views

From SAE experiment to production: the enabling environment

Created by Haoyi Chen, last modified on Apr 21, 2021

Small area estimation has been in the field for many years but using it for official data production is still uncommon. It is important to understand the underlying reasons for the slow onset of SAE in the official data arena and identify "non-tech areas that should be emphasized as creating an "enabling environment" for small area estimation.

- Challenges in using SAE for official data production
- Enabling environment to enable the use of SAE for official data production
 - Establishing a clear and focused objective that links SAE to data use for policymaking
 - Fostering an environment for research and development
 - Government commitment and sustainable financial support to SAE experimentation and production
 - Design-based versus model-based estimates: a changing culture in the national statistical offices
 - Usable input data for SAE
 - [Maintaining a high and fit-for-purpose quality standard](#)
 - Collaboration
 - Capacity building
 - Disclosure control
 - Transparency in releasing methodology and communicating quality
- Practical way forward: from experimental statistics to official statistics

Challenges

- Lack of support from upper management (resources)
- Lack of technical capacity
- Lack of proper input data
- Unsure about the use of model-based estimates
- Difficult to communicate the method and results

Government commitment and legal mandate

- Requirement of disaggregated data by law, to distribute funding
- Building a team

Input data

- Data access
- Data quality

Collaboration

- Researchers
- Other government agencies and private sector
- Other data community: IT/cloud infrastructure, processing and technical capacity
- Within NSO:
 - Subject-matter experts
 - Geospatial experts

Capacity building

- What is the most effective way?

Quality standard

- Quality assurance
- External evaluation

SAE for official statistics – national examples



Example: United States SAIPE Program

In September 1994, the Congress passed the Improving America's Schools Act and signed it into law (PL 103-382). Title I of the law specifies the distribution of Federal funds to school districts based largely on "the number of children aged 5 to 17, inclusive, from families below the poverty level on the basis of the most recent satisfactory data, ..., available from the Department of Commerce."

This law further requires that in Fiscal Year 1997, the Secretary of Education use updated data on poor children for counties and, beginning in Fiscal Year 1999, updated data for school districts, published by the Department of Commerce, unless the Secretaries of Education and Commerce determine that the use of updated population data would be "inappropriate or unreliable."

It also directs the Secretary of Education to fund a National Academy of Sciences panel to provide advice on the suitability of the Census Bureau estimates for use in allocating funds.

Source: Small Area Income and Poverty Estimates (SAIPE) Program, Origins of the Project

Challenges in using SAE for official data production From National Statistical Offices

- "We did an experiment using small area estimation method for poverty but the results were not consistent with our own estimates so we did not pursue it again."
- "We do not have good input data source for SAE - census data are outdated and administrative data sources do not have good coverage and are lack of proper auxiliary variables."
- "SAE method is complicated and we are not comfortable with independently developing the method"
- "It is very difficult to convince the managers to use model-based estimates."

Model-based estimates at Statistics Netherlands

In a more recent paper from Statistics Netherlands (Buelens, Wolf and Zeelenberg, 2016), a set of guidelines were provided that can be used to evaluate model-based estimates. Those interested in more details should refer to the original paper.

1. General principle. The general principle when using model based estimation in official statistics, is the principle that official statistics give a de
a. Objectivity: data used to estimate the model should be related to the subject of the statistic of interest. The model should only be used to estimate the present, but estimation should not exceed the present.
b. Reliability: failure of the model should not lead to changes in the (conclusions based on the) estimate of the statistical phenomenon. It should be possible to replace the model by another model without changing the estimate.
2. The use of models.
a. Goal. The goal of using model based estimation should be to estimate data that is not available, and as such to improve the overall estimation.
b. Data. Models are used to estimate missing data. Both for fitting the model as well as for the final estimation procedure, only data that are available should be used.
c. Standard. Model based methods that are used at Statistics Netherlands should follow any general consensus in the literature on similar methods.
d. Model selection. Alternative models should be considered, in order to find the most appropriate model. With model selection, the simulation should be used to evaluate the models.

SAE methodologies used by countries and international agencies

Dashboard / SAE4SDG   38 views

SAE practices

Created by Haoyi Chen, last modified on May 04, 2021

Asian Development Bank

FAO

UNICEF

US Census Bureau

Asian Development Bank

Created by Haoyi Chen, last modified on May 04, 2021

Brief introduction of the organisation

ADB is committed to achieving a prosperous, inclusive, resilient, and sustainable Asia and the Pacific, while sustaining its efforts to eradicate extreme poverty. Established in 1966, it is owned by 48 members—43 from 11

A description of the SAE work within the organisation

In 2017, the Asian Development Bank (ADB) launched the Data for Development project which aims to support the statistical capacity of national statistics offices (NSOs) in Asia and the Pacific, helping them comply with the Sustainable Development Goals (SDGs). This component focuses on strengthening the capacity of NSOs to generate fine-grained data for policies and programs targeted to vulnerable sectors of society. One of the outputs of this component is a guide on disaggregation of official statistics, which includes an inventory of various small area estimation (SAE) methodologies to yield granular data for official statistics compilation. The guide explains SAE techniques with examples of how the easily accessible R analytical platform can be used to implement them, particularly to estimate indicators on poverty, employment, and health outcomes.

Reference:

- Asian Development Bank. Introduction to Small Area Estimation Techniques: A Practical Guide for National Statistics Offices

Future work on SAE

The guide compiles various SAE techniques and worked examples on how to implement the methodology, which were covered in a series of country training workshops provided to the staff of several national statistics offices to meet the disaggregated data requirements of the SDGs. Furthermore, since its publication in May 2020, several researchers and academics have reported the usefulness of the guide in their work.

Moving forward, the team will continue exploring potential areas of collaboration with national statistical systems who may need technical assistance in building capacity on the application of SAE methods.

 Like  Share this to the blog



Write a comment...

US Census Bureau

Created by Haoyi Chen, last modified on May 04, 2021

Introduction

One of the most famous programmes on small area estimation for official statistics is the Small Area Income and Poverty Estimates (SAIPE) Program led by the US Census Bureau. SAIPE provides annual estimates of income and poverty for small areas. This page provides a brief discussion with the SAIPE team at the US Census Bureau as well as other reference materials.

How to motivate SAE - how did you convince the government to use small area estimates?

Answer: Prior to SAIPE, all local level income and poverty information can only be produced from the decennial census long-form. This means that small area estimates on poverty is only available every 10 years, based largely on "the number of children aged 5 to 17, inclusive, from families below the poverty level on the basis of the most recent satisfactory data ... available from the Department of Commerce." This law of the Department of Commerce, unless the Secretaries of Education and Commerce determine that the use of updated population data would be "inappropriate or unreliable." It also directs the Secretary of Education to

From the description above, three distinct features stand out:

1. A legal act is in place that requires that the Secretary of Education distribute Federal funds based on data produced at county and school district level, unless data are "inappropriate or unreliable".
2. The legal act also specifies that such data should be produced by the Department of Commerce that houses the US Census Bureau
3. Funding of an external expert panel to provide quality check

Therefore this is really a "top-down" approach where the law requires that quality data are to be used for policymaking, distributing Federal fund in this case. The program is well-funded because of the legislative

Input data

Surveys that provide poverty data: Current Population Survey (CPS) through 2004 and American Community Survey starting in 2005.

Administrative data:

- US Federal income tax data
- Supplemental Nutrition Assistance Program (SNAP) participants data
- Supplemental Security Income (SSI) reciprocity rate

Data from the Census Bureau Population Estimates Program are used to construct denominators of several of the regression covariates.

Reference: An Overview of the US Census Bureau's Small Area Income and Poverty Estimates (SAIPE Program). Bell, Basal and Maples, 2015

Input data quality reflection

The quality of the input data is important. One administrative data that was considered but not used is the Free and Reduced-Price Lunch Data. Studies showed such data are not sufficiently precise for formal use in official statistics. The reflection is on how household surveys could be better designed to allow good small area estimation. For example, CPS sample that collected poverty data are relatively small and for some small geographic

Adjustment made on the model and estimates

Adjustments of small area estimates are made near final to reflect model and administrative data errors. One's definition of SAIPE estimation has made many changes in its model and estimation

Frequently asked questions

SAE4SDG / Dashboard / SAE4SDG / FAQ  

FAQ

 **Panel** | How to deal with districts/cities that have zero sample points?

|

 **Panel** | How to work with different area delineation between two data sources

 **Panel** | When integrating survey and population censuses, how do we deal with the difference in ...

 **Panel** | How NSOs that develop SAE are organized internally?

Consultations

- Key SAE experts: consultation meeting organized by JPSM Technical Group on SAE
- Emails and focus-group discussions
 - Australia
 - Canada
 - Chile
 - Indonesia
 - Philippines
 - UK
 - US
 - Viet Nam
- Next steps:
 - Approaching more countries and document the challenges/lessons learned
 - Present a paper during the next SAE conference
 - Finalise the first stage of the Toolkit; advocating the usefulness of SAE but underline the important aspects to be considered
 - Organise small technical group discussion
 - Remote sensing?

Questions for you

-
- Has your office worked with small area estimation, through an experiment or for official data production?
 - What do you consider as the most important elements for successful use of SAE for official data?
 - What are the most challenging aspects?
 - Anything that the United Nations or other development agency could do to help?