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)
• 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
SAE by SDGs

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.

<table>
<thead>
<tr>
<th>User needs</th>
<th>Questions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Goal</td>
<td>- What are the key policies or funding decisions?</td>
</tr>
<tr>
<td>Indicator of interest</td>
<td>- What questions need to be answered?</td>
</tr>
<tr>
<td>Disaggregation level</td>
<td>- What are you trying to measure?</td>
</tr>
<tr>
<td></td>
<td>- What type of indicator is the indicator of interest?</td>
</tr>
<tr>
<td></td>
<td>- What is the relevant dimension of disaggregation?</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Data availability</th>
<th>Questions</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>- Which survey data is available for the estimation of the indicator?</td>
</tr>
<tr>
<td></td>
<td>- What are the data challenges?</td>
</tr>
<tr>
<td></td>
<td>- Which additional data sources can be used?</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>SAE methods/Specification</th>
<th>Questions</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>- Which SAE approach can be used based on the inputs above?</td>
</tr>
<tr>
<td></td>
<td>- Which approaches are available in statistical software?</td>
</tr>
<tr>
<td></td>
<td>- What are the available expertise to do the computation, analysis and interpretation?</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Model validation</th>
<th>Questions</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>- What is the plan for data validation?</td>
</tr>
<tr>
<td></td>
<td>- What data sources for benchmarking?</td>
</tr>
<tr>
<td></td>
<td>- Any plans for external review process?</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Model refining</th>
<th>Questions</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>- Plan to refine the model</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>External case study to official production</th>
<th>Questions</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>- Plan or a roadmap to extend the case study for official data production</td>
</tr>
</tbody>
</table>

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

From SAE experiment to production: the enabling environment

Created by Hacyi 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

Examples: 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 Statline (2012), a set of guidelines were provided that can be used to avoid 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 that official statistics give a
d. Objective data used to compute the model should be related to the subject of the statistic of interest. The model should only be used model, 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.
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 est. 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. c. Standard. Model based methods that are used at Statistics Netherlands should follow any general consensus in the literature on simil
SAE methodologies used by countries and international agencies

Introduction

In recent years, various progressions in small area estimation for vehicle accidents in the United States have been made by the National Highway Traffic Safety Administration (NHTSA) through the Small Area Estimation (SAE) Program. This program, led by the US Census Bureau, provides annual estimates of vehicle crashes and fatalities that are used for decision-making at all levels of government. The purpose of this report is to provide an overview of the methodologies used by countries and international agencies for small area estimation.

Methodologies

1. **US Census Bureau**
   - **Overview**: The Census Bureau provides small area estimates of vehicle crashes and fatalities using a variety of statistical techniques. These estimates are used for decision-making at all levels of government.
   - **Techniques**: The Census Bureau uses a combination of statistical techniques, including small area estimation, to provide accurate estimates of vehicle crashes and fatalities.
   - **Application**: These estimates are used by policymakers, researchers, and the public to develop policies and make informed decisions.

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

**Answer**: Prior to SAE1, all local live income and poverty information can only be shown from the deidentified census long form. This means that small area estimates on poverty are only available in 10-yearly broad categories for the number of children aged 5 to 17. Using statistical techniques like small area estimation (SAE), it is possible to produce more accurate estimates of poverty at a smaller scale. This is achieved by estimating the number of people in poverty on the basis of the most reliable data available, which is provided by the US Census Bureau.

Input data

SAE requires the following data to construct small area estimates:
- **SAE data**
- **Supplementary data**
- **Local area data**

These data are used to construct small area estimates of vehicle crashes and fatalities.

Put data quality reflection

Small area estimates are subject to various sources of error, including limited data, sampling error, and model error. The Census Bureau uses statistical techniques to account for these errors and produce estimates that are as accurate as possible.

Adjournment on the made and estimates

The results of small area estimates are subject to various sources of error, including limited data, sampling error, and model error. The Census Bureau uses statistical techniques to account for these errors and produce estimates that are as accurate as possible.

Future work on SAE

The Census Bureau will continue to refine and improve their small area estimation techniques, using the latest statistical techniques to produce more accurate estimates of vehicle crashes and fatalities.

References

Frequently asked questions

- How to deal with districts/cities that have zero sample points?
- How to work with different area delineation between two data sources?
- When integrating survey and population censuses, how do we deal with the difference in...
- How NGOs 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?