Panel Discussion Webinar FAQ: Small Area Estimation and the Future of Survey Research

1. **What are resources for learning skills for small area estimation that webinar attendees can use?**

RESPONSE: Recommended resources include the *Small Area Estimation* book by Rao and Molina (2015), along with the extensive references cited in the book, which provide a strong foundation in SAE theory and methods. Additional learning opportunities are available through short courses and conference presentations offered at the Joint Statistical Meetings (JSM), the Small Area Estimation (SAE) Conference, the World Statistics Congress, and other professional meetings. The International Association of Survey Statisticians (IASS) also hosts free monthly webinars on survey statistics topics, including small area estimation, with slides and recordings freely available on its website. For hands-on implementation, the R packages *sae* and *emdi* provide free tutorials and vignettes. Readers seeking a less technical introduction may find Liao, Zimmer, and Berzofsky (2021), *Small Area Estimation for the National Crime Victimization Survey: A Guide for Data Processing and Estimation Procedures*, to be a practical overview. Together, these resources cover both the theoretical foundations and practical implementation of SAE methods.

2. **What are the main challenges and risks of using nonprobability samples for SAE, and how can they be addressed?**

RESPONSE: The main challenges and risks of using nonprobability samples for SAE are the likelihood that the estimates or that the set of sample respondents are not representative of the population of interest and, therefore, there is a risk that estimates are biased. Unlike probability samples, where selection probabilities are known, nonprobability samples rely on assumptions and statistical adjustments to address selection bias. To mitigate potential bias, researchers typically use weighting adjustments based on known population benchmarks, similar to calibration or post-stratification methods used with probability. However, these adjustments are only effective if the variables used for weighting adequately account for the factors associated with both survey participation and the outcomes of interest. In many cases, bias arises from characteristics that are unobserved or unavailable so cannot be accounted for in the weighting model. To help mitigate this issue, researchers sometimes include survey items that are correlated with the outcome of interest and are also available in national probability surveys. Including these correlated auxiliary measures into weighting models can further help reduce the

potential for bias in the key outcomes, although it may not be possible to eliminate bias entirely. When using nonprobability samples for SAE, it is also important to account for the measurement error associated with estimates derived from these samples. The framework developed by Sen and Lahiri (2026) provides one approach for incorporating this additional source of uncertainty into SAE models.

3. **How can SAE provide impactful estimates that can help attend to the goals of the project?**

RESPONSE: Appropriate small area estimation (SAE) methodologies are available to address the project's goals by producing reliable estimates for domains or geographic areas that were not included in the original survey design or have insufficient sample sizes for direct estimation. Depending on the project's objectives, SAE methods can support not only the production of estimates but also the identification and ranking of areas with unusually high or low outcomes, providing actionable information for policy development, program planning, resource allocation, and targeted interventions.

4. **How can I get expertise in this field?**

RESPONSE: See Q1 for some reference material. Beyond the reading, there are often short courses on SAE at JSM or other statistical conferences. These courses usually offer an introduction into what SAE is, how to develop an SAE model, and what software is needed to conduct SAE. For a more in-depth understanding, consider taking a semester-long graduate course in SAE or related topics. such as hierarchical and mixed-effects modeling. Hands-on experience applying SAE methods to real survey data is also invaluable for developing practical expertise.

5. **From your experience, what are a few situations or conditions where SAE is not feasible?**

RESPONSE: SAE is generally less feasible or less reliable when there is little or no direct survey data available for the areas or domains of interest, especially if the target population is very small or highly specialized (e.g., a single tribe). In these situations, there may not be enough information to develop and validate a stable model, and the resulting estimates rely heavily on modeling assumptions.

Ultimately, the success of SAE depends on the availability of relevant survey and auxiliary data sources, modeling, statistical methodology and evaluation used. In a typical SAE project, outcome variable information is missing for many small areas. Appropriate SAE methods such as synthetic method exist to deal with this problem. In a typical SAE implementation, the importance of research cannot be overemphasized.

6. **Are public institutions using SAE? Will people see it as "making up numbers" when we are making modeled (not direct) estimates?**

RESPONSE: Yes. SAE is widely used by both Federal and State government agencies. Examples include SAMSHA's state level estimates of substance use and mental health measures, the National Cancer institute's small area estimates for cancer related measures, the U.S. Census Bureau's Small Area Income and Poverty Estimates (SAIPE) program, and the Ohio Department of Medicaid's Ohio Medicaid Assessment Survey county level estimates.

Model-based estimates are valuable when direct survey estimates are unreliable or unavailable because of small or zero sample sizes. They "borrow strength" by combining survey data with auxiliary information related to the outcome of interest. For areas with larger sample sizes, modeled estimates are expected to be close to the direct estimates, while for areas with small or no samples, they rely more heavily on the auxiliary variables and model assumptions. Therefore, selecting appropriate auxiliary variables and using sound statistical methods are essential.

Some users may perceive modeled estimates as "making up numbers." However, SAE uses well-established statistical methods to combine available survey data with auxiliary information to produce estimates that are generally more reliable than unstable direct estimates. Transparency about methodology, assumptions, limitations, and uncertainty helps build confidence in the results.

7. **How important is SAE currently and how should statisticians and young researchers skill up and push the boundaries of knowledge.**

RESPONSE: SAE is currently important and its methods are only becoming increasingly important as surveys need to scale back the amount of direct data collection they can conduct due to higher costs and budget cuts. SAE methods will only gain importance as a tool to mitigate the challenges facing large survey collections as time goes on. See Q1 and Q4 for how statisticians and young researchers can skill up and push the boundaries of their knowledge.

8. **Is this helpful to analyze data for small areas - such as census tracts - that are subject to small-number suppression?**

RESPONSE: Yes. SAE can be used to produce estimates for small geographic areas, such as Census tracts, and can help address issues related to small sample sizes and data suppression. However, its effectiveness depends on the availability of sufficient auxiliary data at the Census tract level and on having survey data with Census tract identifiers. If key

auxiliary variables are unavailable because of suppression, or if respondents cannot be linked to Census tracts, the resulting SAE models may not be reliable. In practice, SAE generally performs better at geographic levels, such as counties, where more auxiliary information is available.

9. **How do you handle gaps for populations that are missing or undercounted in the data? e.g. Tribal, FQHCs are often absent**

RESPONSE: SAE can help quantify and adjust for undercount when there are data available to model the undercount (e.g., through administrative records, post-enumeration surveys, or other auxiliary data). However, it cannot fully address populations that are systematically missing from all available data sources.

10. **How can smaller communities use regional data to figure out local patterns?**

RESPONSE: SAE can help smaller communities identify local patterns by using statistical models that borrow strength from data across areas and incorporate community-level auxiliary variables that are correlated with the outcome of interest to produce more reliable local estimates. The effectiveness of this approach depends on the availability of high-quality auxiliary data and geographic identifiers at the desired level of estimation.

11. **Data validation is challenging with SAE and there are challenges with data suppression. How do you ensure reliability?**

RESPONSE: SAE relies on statistical model assumptions, so careful model validation is essential. Established diagnostic and validation methods, such as those described by Rao and Molina (2015), can be used to assess model performance. For a few successful implementations of SAE methods, including public policy settings such as the SAIPE program of the US Census Bureau and Chilean Poverty Mapping, see Pratesi (2016). One practical approach is to compare the direct and model-based estimates by plotting the ratio of the direct estimate to the modeled estimate against the sample size (or the precision of the direct estimate). As sample size increases and the direct estimates become more reliable, the ratio is expected to approach 1, indicating good agreement between the direct and model-based estimates. Internal and external validation using independent data sources, when available, can provide additional evidence of model performance. When data are suppressed or direct estimates are unavailable, validation becomes more challenging, making model diagnostics, sensitivity analyses, and transparent reporting of assumptions and uncertainty especially important.

12. Has this method been used for homelessness?

RESPONSE: Yes. Although SAE has not been widely used for estimating homelessness, there are published examples of its application to produce state- and county-level estimates of the homeless population (Hudson and Vissing 2010). Whether SAE is appropriate depends on the available data, the geographic areas or population domains of interest, and the availability of high-quality auxiliary variables that are strongly associated with homelessness. As with other SAE applications, the reliability of the estimates depends on the quality of both the survey data and the auxiliary information used in the model.

13. What type of location-based data, sample size, and type of survey do I use to produce subnational and local estimates?

RESPONSE: It depends on the geographic level of the survey and the desired level of estimation. For example, a national survey (e.g., NHIS) can be used to produce state-level estimates using SAE. County-level estimates may also be feasible, although they are generally more challenging because many counties have few or no sampled respondents. Likewise, a state-based survey can be used to produce county- or subcounty-level estimates.

Regardless of the survey, respondents must have geographic identifiers at the level of interest, and high-quality auxiliary variables must be available for the same geographic units. Although probability-based surveys are generally preferred, there is no minimum sample size required for SAE. However, as the sample size decreases, the estimates rely more heavily on the statistical model and auxiliary information, making careful model specification and validation increasingly important.

14. How do we utilize positively skewed data for weight adjustments in small area estimation, specifically the indirect method?

RESPONSE: The approach depends on whether the skewness is in the survey weights or in the outcome being estimated. If the outcome is positively skewed, it may be appropriate to apply a transformation (e.g., a log transformation) or use an SAE model that accommodates skewed distributions, rather than relying on a model that assumes normality. See Chen et al (2026) for an application. If the skewness is caused by extreme survey weights, weight trimming or calibration may be considered to reduce the influence of extreme weights while maintaining representativeness. Regardless of the approach, model diagnostics and sensitivity analyses should be conducted to evaluate the impact of the transformation or weighting adjustment on the resulting small area estimates.

15. **How large population-wise should a "small area" be in general?**

RESPONSE: There is no minimum or maximum population size that defines a "small area." In small area estimation, an area is considered "small" when reliable direct survey estimates cannot be produced because of limited sample sizes. In practice, the feasibility of SAE depends less on the population size itself than on the availability of sufficient survey data, geographic identifiers, and high-quality auxiliary variables to support the statistical model.

16. **Will this presentation address disclosure issues and how to solve them in the realm of SAE?**

RESPONSE: The panel did not have the opportunity to discuss disclosure issues in detail. In practice, researchers developing SAE models typically need access to geographic identifiers (e.g., state or county FIPS codes) in the survey data, which often requires access to restricted-use data.

One advantage of SAE is that it produces model-based estimates rather than releasing individual-level survey records. As a result, agencies can disseminate estimates for small areas or populations without releasing confidential microdata, thereby substantially reducing disclosure risk. However, model-based estimates should still undergo appropriate disclosure review before public release, particularly for very small geographic areas or populations.

17. **How can we begin to consider integrating AI into SAE in response to budget constraints and hard to reach foreign populations?**

RESPONSE: AI has the potential to support, rather than replace, traditional SAE and other sophisticated statistical models. In the near term, AI can be used to help identify and gather auxiliary information from multiple sources, improve data integration and quality, and automate parts of the model development and validation process. These capabilities may reduce the cost and effort required to develop SAE models, making them more accessible to smaller agencies with limited resources.

For hard-to-reach or underrepresented populations, AI may also help identify additional data sources or improve data linkage. However, the reliability of SAE estimates will continue to depend on the quality and representativeness of the underlying data and the validity of the statistical models. AI cannot compensate for systematic under-coverage or poor-quality data.

18. How would you advocate for the use of SAE to clients who are on the fence about using this method?

RESPONSE: When discussing with clients on the fence about incorporating SAE into their design and analysis plans, we may want to emphasize that SAE is a well-established statistical approach used by many federal and state agencies to produce reliable estimates for small geographic areas and population subgroups when direct survey estimates are unavailable or too imprecise. SAE improves direct estimates by borrowing strength from related data and auxiliary information to improve precision. We also want to stress the importance of using appropriate auxiliary variables, validating the models, and being transparent about the methodology and uncertainty. Ultimately, the decision should be guided by whether the resulting model-based estimates are sufficiently reliable and fit for the intended purpose.

19. How software dependent is this methodology?

RESPONSE: SAE is more dependent on the statistical model than on the software being used. Some standard SAE methods can be implemented in existing R packages, making it straightforward to fit commonly used area-level and unit-level models. For more complex applications or customized methodologies, analysts may need to develop their own code in R, SAS, or other statistical software. Regardless of the software, it is important to understand the underlying model and estimation method to ensure the results are appropriate and comparable across implementations.

20. How can you apply small area estimation to sample survey estimates?

RESPONSE: Small area estimation (SAE) uses sample survey data together with auxiliary information to produce more reliable estimates for small geographic areas or population domains. When individual-level survey data is available, unit-level SAE models can be applied directly. When only direct survey estimates are available, area-level SAE models use those estimates as inputs and borrow strength from auxiliary variables through statistical modeling to improve precision. In many applications, the final estimates combine information from both the direct survey estimates and the model. Rao and Molina (2015) provide a comprehensive review of SAE methods and applications.