CDC Foundation
Together our impact is greater

Assessing the Social Determinants of Health in Birthing Populations

Small Area Estimates of Select Social Determinants of Maternal and Child Health Indicators Using PRAMS Data

CDC Foundation Author

Michaela Curran, PhD

Project Partner

Centers for Disease Control and Prevention (CDC) Division of Reproductive Health

The social determinants of health are personal (e.g., income), familial (e.g., social support), work-related (e.g., scheduling autonomy), communal (e.g., community cohesion and support) and structural (e.g., federal or state policies) factors that constrain people's available choices for improving their health.[1] Public health professionals and local decision-makers need access to relevant local-level data about the social determinants of health to improve health outcomes for everyone. Detailed population health surveillance data about pregnancy and childbirth are typically obtained from large-scale survey efforts. However, sample sizes from these survey efforts are often too small to produce reliable direct estimates for lower levels of geographic aggregation (e.g., county, Census tract, etc.) or for subpopulations (e.g., certain racial or ethnic groups, gender identities, etc.) that are small within an area. Small area estimation (SAE) provides a potential solution for producing reliable estimates at lower levels of geographic aggregation or for small subpopulations using a model-based approach.

To help states and jurisdictions gain access to local-level data about maternal and child health, the Centers for Disease Control and Prevention's (CDC) Division of Reproductive Health team used Pregnancy Risk Assessment Monitoring System (PRAMS) Phase 8 (2017-2021) data on attitudes and experiences before, during and shortly after pregnancy to calculate state- and county-level estimates across 40 states and jurisdictions for nine indicators related to the social determinants of health: postpartum visit attendance, postpartum Medicaid insurance status (at the time of survey), postpartum depression as diagnosed by a healthcare provider, intimate partner violence in the 12 months before pregnancy, intimate partner violence during pregnancy, any intimate partner violence (before or during pregnancy), problems paying rent, mortgage or other bills in the 12 months before birth, whether a pregnant person moved in the 12 months before birth and whether a pregnant person was homeless in the 12 months before birth.

This document provides insight into the project's small area estimation process and offers lessons learned and promising actions for undertaking small area estimation projects. This project was part of the CDC Foundation **Improving Engagement in Community Level Data Collection** project, which was funded by Robert Wood Johnson. The views expressed here do not necessarily reflect the views of the Robert Wood Johnson Foundation.

The Small Area Estimation Project Process: An Overview

The small area estimation (SAE) process for this project proceeded in six steps. In the first step, a workgroup from the CDC's Division of Reproductive Health and the CDC Foundation senior statistician consulted on and selected indicators for estimation. The second step involved obtaining access to the needed data (county-level PRAMS data and National Center for Health Statistics (NCHS) National Vital Statistics System (NVSS) birth data). In the third step, the data was cleaned and prepared for estimation. The team estimated models and did internal validation for the estimates in the fourth step. Step five involved consulting with subject matter experts and the PRAMS team, on guidelines for interpreting small area estimates and suppression of estimates. In the sixth step, the PRAMS team released the estimates to the states and jurisdictions that participated by sharing their PRAMS data along with general guidance for the use of PRAMS small area estimates. This process is depicted in Figure 1.

Figure 1: The PRAMS Small Area Estimation Project Process

Step 1	Step 2	Step 3	Step 4	Step 5	Step 6
CDC and CDC Foundation Select Indicators for Small Area Estimation	Obtain Access to the Needed Data (PRAMS and NCHS Birth Data)	Clean and Prepare Data for Estimation	Estimate Models and Perform SAE Internal Validation	Consult on General Use Guidance, Guidelines	Release Estimates and General Use Guidance to Participating Sites

The full process required ongoing conversations between the CDC Foundation statistician and the PRAMS team, states and jurisdictions and subject matter experts. Engaging with these key partners throughout the process helped guide the project. In the sections that follow, the full process is described in detail, along with lessons learned.

1. Indicator Consultation and Selection

Indicator consultation began with compiling a list of social determinants of health variables available in PRAMS. These variables were compared across PRAMS Phases 8 and 9 via a crosswalk to ensure that the question wording was consistent across phases and that a given question would remain on the survey into the next phase for continuity. Once the crosswalk was created, a workgroup of staff from the CDC Division of Reproductive Health and the CDC Foundation senior statistician met and finalized the indicator selections. Ultimately, nine indicators were selected (see Table 1).

Table 1: PRAMS SAE Indicator Selections

Variable Name	SAE Indicator Definition
Postpartum Visit Attendance	Prevalence of PRAMS respondents who had a postpartum visit 4-6 weeks after giving birth.

Postpartum Medicaid Insurance Status	Prevalence of PRAMS respondents who had Medicaid at the time of the survey.
Postpartum Depression	Prevalence of PRAMS respondents whose healthcare provider told them that they have postpartum depression.
Intimate Partner Violence before Pregnancy	Prevalence of PRAMS respondents who experienced intimate partner violence in the 12 months before pregnancy.
Intimate Partner Violence during Pregnancy	Prevalence of PRAMS respondents who experienced intimate partner violence during pregnancy.
Any Intimate Partner Violence	Prevalence of PRAMS respondents who experienced any (either before or during pregnancy) intimate partner violence.
Problems Paying Rent, Mortgage or Other Bills in the 12 months before birth	Prevalence of PRAMS respondents who had trouble paying their bills in the 12 months before birth.
Moved in the 12 months before Birth	Prevalence of PRAMS respondents who moved in the 12 months before birth.
Homeless in the 12 months before Birth	Prevalence of PRAMS respondents who were homeless in the 12 months before birth.

Three of the indicators were part of the PRAMS Standard Questionnaire, which meant that only a certain group of states and jurisdictions had data for those indicators ($N = 26$). The Standard Questionnaire indicators included problems paying rent, mortgage or other bills, moved in the year before birth and homeless in the year before birth. The remaining indicators were drawn from the PRAMS Core Questionnaire and included all PRAMS states and jurisdictions that chose to participate in the project ($N = 40$).

The bullet points below outline some lessons learned and promising actions for selecting SAE indicators:

- Before selecting an indicator, look at the amount of variation that it has because model performance can be impacted by low variation.[2]
    - It may be helpful to look at the amount of between-geography (e.g., state) variance for a proposed indicator because selecting indicators where very little between-geography variance exists can make small area estimation more difficult, as it relies on multilevel modeling (See Section 4 for additional details).

- Take a look at the sample sizes at each level (e.g., individual, county, state) for any potential indicator because the sample size can impact the confidence intervals for the estimates.
- A literature review can aid in indicator selection because the literature will provide guidance on the covariates to include in the model for a given indicator.
- Review guidance for small area estimate use in public health research to help with indicator selection. Kong and Zhang (2020) provide an excellent overview of SAE use in their article.[3] If other researchers have done SAE with the proposed data, their articles may also include guidance for use. The PRAMS SAE release general guidance included considerations for use.
- Consider overlap between all proposed data sources for each indicator. Some SAE methodologies draw on poststratification by key covariates using additional data sources, so it is important to ensure that the indicator is not missing data for these key covariates in any of the data sources.
- Checking the overall prevalence of a proposed indicator can be helpful, as well.[2] Some low prevalence indicators (≤ 10% in the general population) can experience issues with estimation, so if selecting a low prevalence indicator be aware that small area estimation might not perform well.

2. Obtaining Access to the Data

The small area estimation procedure used by the CDC Division of Reproductive Health (discussed in Section 4) required access to two data sources: PRAMS data[4] and NCHS restricted-use birth data.[5] Because county-level data are needed, gaining access to these two data sources involved two separate and distinct approval processes. Both processes required proposals that included a rationale for requesting the data, information about the team who will be working with the data and details about the methodology.

For this project, the CDC PRAMS team sent the SAE data proposal to the PRAMS sites (i.e., states and jurisdictions) for approval. Four of the sites requested some form of institutional review before granting approval. The process for submitting the NCHS proposal for the restricted-use birth data was similar to the process for submitting a proposal to PRAMS, and approval for the restricted-use birth data was determined by NCHS.

The bullet points below outline some lessons learned and promising actions for obtaining access to the data:

- Sites may have questions about the small area estimation procedure. Be prepared to describe SAE in accessible language and respond to questions about privacy and confidentiality as it relates to SAE.
- When providing details about the analysis and methodology, it is important to avoid jargon and explain it in a way that people who are not familiar with the methodology can understand.
- Institutional review boards (IRBs) may require a fee that must be paid prior to review. Read all IRB documentation carefully, and if there is no mention of a fee, contact the IRB directly to ask whether and when one is required.
- Some IRBs may exempt a small area estimate project from IRB review. Contact with IRB administrators can help give guidance on how a project should be submitted for review.
- For IRB submissions, be prepared to discuss technical aspects of the project (e.g., how and where data will be stored, whether data will be deidentified, etc.).
- It is important to conceptualize and provide a rationale for any data merges that may occur during an SAE project when preparing a proposal to obtain data. In the case of this project, a data merge between PRAMS data and NCHS birth data was required. This merge occurred at the county-level rather than the individual-level, but it was a requirement of the NCHS proposal and IRB submissions to discuss thoroughly any data linkages.
- If possible, work with state personnel on IRB submissions as they tend to understand their IRB processes well.

3. Cleaning and Preparing the Data for Estimation

Cleaning and preparing the data for estimation required recoding variables in the PRAMS and NCHS datasets and linking the two data sources. Cleaning the PRAMS data necessitated recodes to some of the county Federal Information Processing Standards (FIPS) codes for mother's resident county because they were not concordant with that of the NCHS data due to some sites using their own version of FIPS. It also required merging a complete list of counties into the PRAMS data because some counties had no PRAMS respondents, and the spatial smoothing procedure (discussed in detail in Section 4) entailed having a list of all counties.

For the NCHS birth data, a delimited data format suggested by NCHS was used when importing the data. The NCHS variable name for mom's county of residence was changed to match that of the PRAMS variable name for the data merge. Then, population counts by county and each covariate category were generated (e.g., the number of live births for a white, high school graduate mom with Medicaid in

Cook County, IL). These counts were merged with their matching PRAMS county and covariate category. The recoded covariate categories for the PRAMS variables and the NCHS birth variables had to match because they were the key on which the linkage occurred, and the variable categories had to be the same for the estimation process.

Because PRAMS sample sizes can be small if looking at one year of data only, years were grouped together to ensure adequate sample sizes. The 2016 data was excluded because of missing data on some of the indicators and key covariates. Data were modeled on two- (2020-2021) or three- (2017-2019) year time periods because single years of data did have an adequate sample size. A full sample version with all years (2017-2021) was also included as part of the estimation.

Cleaning and preparation also involved conceptualizing the models and ensuring that all potential covariates were recoded appropriately. Conceptualizing the models required knowledge of the literature about a given indicator. Table 2 lists the covariates utilized in the SAE models. Not all proposed covariates were included in the final SAE models. Their inclusion depended on whether they were present in both the PRAMS data and the NCHS birth data and whether they showed a significant association with each indicator during the estimation phase. The models started with maternal race/ethnicity, maternal education, paternal race/ethnicity, paternal education, infant's sex, maternal age (as a continuous variable) and insurance payer. The final models included maternal race/ethnicity, maternal education and insurance payer. The rest of the variables did not show enough of a significant association with any of the indicators to warrant inclusion in the final models.

Cleaning and preparing the data involved a multi-step process that included checkpoints and reviews of the literature to ensure that the data was cleaned and handled appropriately.

The bullet points below outline some lessons learned and promising actions for cleaning and preparing the data for estimation:

- Understanding how the data is collected, including aspects of survey design, is important for cleaning and preparing the data.
- It is important to review the analytic codebooks for all data sources involved in the project. For the PRAMS data, it may also help to look at the questionnaires to understand the wording of the questions and provide context for recodes.
- Because SAE often involves recoding and merging different data sources, it is important to include checkpoints in the cleaning process. Checkpoints can include cross-tabulation of the recoded variable with the original variable to

ensure recodes are correct, comparing records that fail to merge between sources and looking at logs produced by SAS, Stata or R during the recode and merge processes.
- Look to the literature to conceptualize the models and their covariates. With small area estimation, it is important not to take a "kitchen sink" approach with predictors and overfit the models.[6]
- When producing the population counts from other data sources, it is important to check them. The NCHS birth data for a given year is a large file, so potential problems can arise that are not as easy to identify as with smaller datasets where it is much less computationally intensive to identify problems. It is also important to check the merge between these data sources and identify records that failed to merge.

4. Model Estimation and Validation

As a model-based approach, SAE can provide estimates for small areas of geographic aggregation or small subpopulations where direct estimation would produce unreliable estimates or fail because the sample sizes for a given area are too small to support it. It also can be used to produce estimates for demographic subpopulations with small sample sizes within an area.[7]

Unlike direct estimates, SAE can "borrow" statistical power from other domains (e.g., other small areas or subpopulations of interest), other data sources and variables or both.[8] Several different estimation methods fall under the umbrella of SAE. Most of them produce estimates by combining individual-level data, area-group composition population estimates (e.g., Census population estimates, estimates, National Center for Health Statistics birth certificate counts, etc.) and multilevel mixed effects regression techniques.[3]

This project drew on an SAE methodology called Multilevel Regression with Poststratification (or MRP) Logistic Regression.[9,10,11] MRP, as a first step, uses multilevel modeling to extract random effects for the small area aggregation of interest (e.g., state, county, etc.). The PRAMS SAE model nested individual PRAMS survey respondents in a hierarchical structure of counties and states and extracted the random effects for each county and state as empirical best linear unbiased predictors (EBLUPs). The model also extracted the fixed effects (i.e., the individual-level parameters) because they are later to be used to calculate the probability of an outcome for a given individual in a given county and state. Table 3 provides the command for multilevel models in commonly used statistical software. For logistic regression, these commands will allow you to specify the "binomial" family.

Table 2: Multilevel Model Commands in Statistical Software

Statistical Software	Command
R	*glm* or *glmer*
SAS	*GLIMMIX*
Stata	*meglm*

Counties that did not have any PRAMS respondents were subject to a spatial smoothing procedure where the random effects for bordering counties were averaged to create a random effect for the county with no respondents. This spatial smoothing procedure involved identifying the counties with missing random effects, identifying available (i.e., counties for the sites that opted to participate in the project) adjacent counties using the Census county adjacency file[12] and then generating an average random effect using all of the available adjacent counties.

The second step in MRP involves generating predicted probabilities or values for every potential individual in the population based on what is observed in the sample model. This is done by adding each state's and county's random effects to the fixed effects for each covariate category and the model intercept and then generating a predicted probability for each category within each state and county.

In the third step, the predicted probabilities created in the second step are applied to the population counts of the demographic groups of interest (e.g., Census, birth certificate data, etc.) and poststratification occurs. Poststratification by the demographic characteristics included in the multilevel model allows the analyst to calculate population-weighted probabilities, sum them across all possible demographic groups within the target geographic units and then divide the summed probabilities by the total number of births in those geographic units to obtain the final SAE.[9,10,11]

In the case of the PRAMS SAE, the population counts of interest included the covariates maternal race/ethnicity, maternal education and payer. The NCHS data was linked with the PRAMS data using these covariates (coded the same way). The PRAMS covariates provided the predicted probabilities for every potential individual in the population (e.g., the probability that a non-Hispanic white, high school graduate mom with Medicaid in a given county and state moved in the year prior to birth). The NCHS covariates provided the population counts (e.g., the number of non-Hispanic white, high school graduate moms with Medicaid in a given county and state) to create the population-weighted probabilities that are then summed over all the potential categories by state and county and divided by the total number of live births in each state and county to get the final estimate.

The PRAMS SAE model used the NCHS birth certificate data (which constitutes the entire population of live births) to generate live birth population counts for applying the estimates to the population and poststratification. It applied these counts to the state and county EBLUPs to generate predicted prevalence for a given outcome and then generated the final SAE via poststratification.[9] Figure 2 provides an overview of the MRP process used in the project.

Figure 2: MRP Using PRAMS and NCHS Birth Data

Step 1
Generate Multilevel Models and Extract State and County Random Effects and Individual Fixed Effects Using PRAMS Data

Apply Spatial Smoothing Procedure for Counties with Missing Random Effects

Step 2
Generate Predicted Probabilities for Each Individual in the Sample by Adding the State and County Random Effects to the Fixed Effects for Each Covariate Category

Step 3
Predicted Probabilities Are Applied to Population Counts from NCHS Birth Data to Create Population-Weighted Probabilities

Population-Weighed Probabilities are Summed and Then Divided by Total Number of Births to Get the Final Small Area Estimate

Monte Carlo (MC) simulations were then used to generate means and confidence intervals (CIs) for the final estimates. Model-based estimates like SAE always are accompanied with some amount of uncertainty because models are an imperfect approximation of reality. MC simulation attempts to capture and estimate that uncertainty. It does so by drawing on the point estimates and their asymptotic covariance matrix to generate a distribution for the estimate.[9] The PRAMS SAE drew upon 1,000 simulations to generate the final means and CIs for the state and county estimates.

There is not a consensus as to whether including survey design or population weights is necessary in MRP models. Some researchers have argued that if you condition on the correct covariates in the model, accounting for survey design is not necessary. The empirical evidence appears to be mixed.[11,13]

The PRAMS SAE were not weighted because of the challenges associated with the complex survey design of PRAMS and multilevel models (i.e., levels in the model would need to map onto sampling units; weights would need to be declared at all levels, instead of final analytic weights; county-level weights would have to be constructed).

The “gold standard” for validity in small area estimation is external validation as gauged by comparing model-based estimates to direct estimates from a different, but similar data source. This project did not utilize external validation because of difficulties in identifying a similar enough data source. It would be possible to explore external validation as part of future SAE projects. For example, homelessness is becoming an important issue with the rise of housing costs, so that indicator could be informative in many ways. The U.S. Department of Housing and Urban Development (HUD) and community-based organizations have a data exchange that includes point in time counts by continuum of care (CoC) geographies. A less rigorous form of external validation involves a qualitative approach that unites community-based participatory research or community focus groups/interviews with the estimates. This could involve engaging with local community-based organizations or the community at large to ask them to what degree they feel that a given estimate is reflective of what they observe (or even personally experience) in their community. This is also a good way to leverage SAE for data-to-action initiatives.

State-level model-based estimates were compared to the direct estimates to gauge internal validity. Pearson's correlation coefficients were also utilized to compare model-based estimates to direct estimates at the state level.[9,11]

The bullet points below outline some lessons learned and promising actions for model estimation and validation:

- Remember that all covariates in the multilevel model must match the variables for the population counts. If maternal race/ethnicity, maternal education and payer are included in the model, then population birth counts for each of those variable categories must be included in the poststratification. If they do not match (e.g., a covariate is included in the multilevel model with no population count), a problem may occur when calculating the population-weighted probabilities.
- Some multilevel models (the first step in MRP) may not converge. This can be an issue especially with logistic multilevel models, and there are several reasons why it may happen. Some common reasons include not enough variance for the logistic regression to build a model, model complexity and excessive collinearity. Convergence failures require investigation to resolve. The first thing that the analyst should do is go back to the data itself and look for problems like coding issues, as well as investigate whether enough variance exists in the outcome to build a model.
- Sample size is important for MRP.[6] If the models do not perform well (e.g., convergence issues, large CIs, etc.), check whether it is possible to pool samples. For the PRAMS SAE, years of data were pooled to ensure a larger sample size.

- As MRP is a multi-step process, it is vital to perform checks at each step to ensure that the models run correctly, and the estimates appear reasonable.
- When possible, compare model-based estimates to survey weighted direct estimates. State level estimates should be relatively easy to produce. Large geographic areas with a large sample size could also be used to produce direct estimates for comparison.
- In some cases, spatial smoothing may not produce a random effect for a county with no respondents because none of the adjacent counties have a random effect. This does not occur often, but it may be possible to run the spatial smoothing procedure twice to pick up adjacent county estimates in the first run and then create a composite of those for a missing county. Keep in mind that this would create a composite estimate comprised of composite estimates. It is possible that doing a second smoothing would generate a weaker and potentially misleading random effect estimate.
- Explore the option of performing the spatial smoothing procedure on any county with a sample size below a given threshold (perhaps starting with 5-10 respondents). This may improve confidence intervals for the estimates.
- SAE may vary slightly depending on the statistical software used to generate the random effects. SAS uses residual subject-specific pseudo-likelihood (RSPL) for integrating the EBLUPs for the random effects, while Stata uses adaptive Gauss-Hermite quadrature and R uses Laplace or adaptive quadrature. SAEs generated in SAS tend to have narrower CIs versus those generated in Stata or R. The Stata and R processes appear to incorporate more uncertainty into the EBLUPs, so the standard errors tend to be larger, which translates into wider CIs for the final estimates.
- There are few diagnostic procedures to determine how well the multilevel models fit and no “gold standard” for determining how well the MRP method performs.
- Equations for MRP are detailed in Wang et al. 2022.[9] They may be helpful for understanding how the pieces of MRP fit together mathematically.

5. Suppression Guidelines

Suppression guidelines were developed based on guidance from the literature, subject matter experts and PRAMS staff. As of this writing, no universal standard exists for data suppression for small area estimates. This is not surprising because all survey data are different, and as a result, small area estimation processes can vary. Finding appropriate thresholds for data that is “borrowing power” from other data sources or nearby geographies within a data set is not as straightforward as with direct estimates. Further, suppression based on number of respondents can be difficult to determine because of this borrowing power aspect of SAE. Suppression guidelines can be difficult to develop.

Ultimately, the suppression guidelines utilized in this project drew on direct-estimate guidance from NCHS/National Health and Nutrition Examination Survey (NHANES) for presentations and publications.[14] This guidance relies on absolute and relative confidence intervals. The guidance is to suppress an estimate when the absolute CI width is greater than or equal to 30% of the estimate (i.e., the point estimate in this case) and to suppress an estimate when the absolute CI width is less than 30% but the relative CI width (defined as the (absolute width/proportion (point estimate)) *100) is greater than 130% of the estimate.

The bullet points below outline some lessons learned and promising actions for developing suppression guidelines:

- It is important to keep in mind that the NCHS/NHANES suppression guidelines are for presentations and publications. Guidance based off of them do not preclude sharing estimates with local public health officials or decision-makers with important caveats that will be outlined in Section 6.
- Guidelines based on absolute and relative CI are sensitive to prevalence, which presents a problem for low-to-middle prevalence indicators (e.g., intimate partner violence, postpartum depression, etc.). This is important to keep in mind when considering potential guidelines for suppression.
- An alternative approach considered in this project was to utilize relative standard error (RSE).[15] The recommendation is to suppress if the N is smaller than 50 or the relative standard error is greater than 30. This is a general suppression recommendation (i.e., not necessarily optimized for SAE) for the Behavioral Risk Factor Surveillance System (BRFSS). Relative standard error is defined as the (standard error/percentage (point estimate)) *100. This approach can be difficult when using Monte Carlo simulations to produce point estimates because the simulations add an additional layer of complexity with defining which standard error and which point estimate to utilize.
- Suppression of small area estimates can involve two factors: suppression for privacy and confidentiality reasons and suppression due to estimate performance. Determining whether one or both is important for a project can help guide the discussion.

6. Releasing Estimates and General SAE Use Guidance

The final step involved releasing the estimates to the PRAMS sites that participated. A co-occurring project also estimated the following variables (postpartum visit and postpartum depression), so only one set of estimates were released to participating jurisdictions. Another variable (postpartum Medicaid insurance status) did not perform well in models. Ultimately for this project, the indicators that were released included problems paying rent, mortgage or other bills, moved before birth, homeless before birth, intimate partner violence before pregnancy, intimate partner violence during pregnancy and any intimate partner violence.

With the estimates, several considerations for use were provided to the sites and limitations were outlined based on best practices from the PLACES: Local Data for Better Health platform. [3,16] The considerations for use included:

- There are few diagnostic procedures to determine how well the multi-level models fit and no "gold standard" for determining how well the MRP method performs.
- There are few data sources to help validate county level estimates.
- Estimates, 95% confidence limits and absolute and relative confidence limit widths can be reviewed to inform discussions around suppression and potential caution with use of estimates.
- **Avoid** use of small area estimates for overall public health rankings or policy and program analysis and evaluation because they are not able to capture local policy or intervention effects. Sites could draw on small area estimates for establishing a baseline for policies or programs (preferably in conjunction with other data sources, if available) and conduct their own local surveys for evaluation purposes.
- Estimates can be used by states and localities to
    - Better understand geographic disparities of specific behavioral or health indicators,
    - Help identify prevalent health issues,
    - Be used to inform and implement targeted tailored prevention activities and
    - Help establish health objectives and support public health decision-making.
- Because year was not included as a covariate in the MRP models, using the SAE to assess trends is not recommended.

The limitations included:

- Because of sample size limitations, we did not calculate single year estimates.
- Although we pooled years of survey data, the sample size in some counties remained small (< 10), which posed some challenges in modelling.
- The small area estimation process does not reduce any non-sampling errors in the survey data, such as non-response or recall biases. To the extent they exist, these errors remain within the modeling process.
- Small area estimates are model-based and are subject to measurement error (as well as self-report bias). Margins of error differ by indicator, geographic area and scale. As a best practice, PRAMS jurisdictions should examine and report these margins of error. While some researchers have drawn on small area estimates in statistical models to estimate the association between them and a given dependent variable, it is important to keep in mind that any measurement error in the small area estimate will be an additional source of bias in such a model.
- Estimates did not account for geographic covariates (e.g., health care access or hospital density).
- Final analytic weights were not included in MRP models, and there is no clear consensus on the need to use population weights.

References

1. Bird CE, Rieker PP. *Gender and Health: The Effects of Constrained Choices and Social Policies*, New York: Cambridge University Press, 2008.
2. Pickering K, Scholes S, Bajekal M. "Synthetic estimation of healthy lifestyles indicators: Stage 2 report." London: National Centre for Social Research, 2004.
3. Kong AY, Zhang X. "The Use of Small Area Estimates in Place-Based Health Research." *Am J Public Health*. 2020 Jun;110(6):829-832. doi: 10.2105/AJPH.2020.305611.
4. https://www.cdc.gov/prams/php/data-research/index.html
5. https://www.cdc.gov/nchs/nvss/nvss-restricted-data.htm#anchor_1553801903
6. Ornstein JT. "Getting the Most Out of Surveys: Multilevel Regression and Poststratification" in *Causality in Policy Studies: A Pluralist Approach*, A Damonte, F Negri, eds. Cham: Springer, 2023.
7. Ghosh M. "Small area estimation: its evolution in five decades." *Statistics in Transition New Series*. 2020; 21(4):1-22.
8. Jiang J, Rao JS. "Robust Small Area Estimation: An Overview." *Annu. Rev. Stat. Appl.* 2020. 7:337–60.
9. Wang, Y, et al. "US county-level estimation for maternal and infant health-related behavior indicators using pregnancy risk assessment monitoring system data, 2016–2018." *Popul Health Metrics*. 2022; 20, 14.
10. Zhang X, et al. "Multilevel regression and poststratification for small-area estimation of population health outcomes: a case study of chronic obstructive pulmonary disease prevalence using the behavioral risk factor surveillance system." *Am J Epidemiol*. 2014;179(8):1025-1033.
11. Zhang X, et al. "Validation of Multilevel Regression and Poststratification Methodology for Small Area Estimation of Health Indicators from the Behavioral Risk Factor Surveillance System." *Am J Epidemiol*. 2015;182(2):127-137.
12. https://www.nber.org/research/data/county-adjacency
13. Zhang X, et al. "Multilevel Regression and Poststratification (MRP) for Small Area Estimation: An Application to Estimate Health Insurance Coverage Using Geocoded American Community Survey." In *JSM 2017 – Survey Research Methods Section*. Joint Statistical Meetings; 2017:2734-2746.
14. https://wwwn.cdc.gov/nchs/nhanes/tutorials/reliabilityofestimates.aspx
15. Pizzonia C. "Comparing BRFSS Suppression Guidelines for Statistical Reliability: An Example Using Maine BRFSS Data." https://www.northeastbranchasm.org/neb/wp-

content/uploads/2017/12/Pizzonia_BRFFS_Suppression_NEEpi.pdf. (Accessed March 22, 2024)
16. https://www.cdc.gov/places/faqs/using-data/index.html