October 14, 2019

2018 Ohio Pregnancy Assessment Survey

Methodology Report

Prepared for

Ohio Colleges of Medicine Government Resource Center
Attn: Timothy R. Sahr
1070 Carmack Road
Columbus, OH 43210
Telephone: (614) 366-3175
E-mail: timothy.sahr@osumc.edu

Prepared by

RTI International
3040 E. Cornwallis Road
Research Triangle Park, NC 27709

RTI Project Number 0216156.001.001

Table of Contents

Section **Page**

Appendices

Tables

Number **Page**

1

Project Overview

The 2018 Ohio Pregnancy Assessment Survey (OPAS) measured pre- and postnatal maternal health, health behaviors, and health system use associated with birth among mothers who recently gave birth to a live-born infant in Ohio. Sponsored by the Ohio Department of Health (ODH) and Ohio Department of Medicaid (ODM), OPAS data are used to help develop and assess programs designed to identify high-risk pregnancies and reduce adverse pregnancy outcomes. The OPAS also provides data on maternal and infant health in the Ohio Equity Institute (OEI) counties. The 2018 OPAS instrument included questions about experiences before, during, and after pregnancy, and utilized the Centers for Disease Control and Prevention's (CDC) Pregnancy Risk Assessment Monitoring System (PRAMS) core questionnaire. Topics included:

- attitudes and feelings about the most recent pregnancy;
- content and source of prenatal care;
- maternal alcohol and tobacco consumption;
- physical abuse before and during pregnancy;
- pregnancy-related morbidity;
- infant health care;
- infant feeding and sleeping practices;
- contraceptive use; and
- mother's knowledge of pregnancy-related health issues.

The population of interest for OPAS was all mothers who were residents of Ohio and delivered a live-born infant in Ohio during the surveillance period. The operational sampling unit for OPAS was infants who were born alive in Ohio to resident mothers during calendar year 2018. Ohio's vital records birth certificate file served as the best available source of the sampling frame representing live births and we used it for the OPAS.

Representatives from ODH, ODM, the Ohio Colleges of Medicine Government Resource Center (GRC), The Ohio State University College of Public Health (OSU CPH), and RTI International (RTI) formed a working group called the OPAS Executive Committee (OPAS EC). The OPAS EC met in fall 2017 to initiate the project. This collaboration continued through biweekly meetings, co-development of the survey instruments and methodological procedures for data capture, cleaning, and ongoing reporting of results.

As is standard, the study underwent an Institutional Review Board (IRB) determination. Study documents including the design, research protocol, and questionnaires were delivered to IRBs at ODH and at RTI. Both IRBs approved the study.

The goal of the 2018 OPAS was to use a mixed-mode data collection methodology to gather data from a statewide sample of women who had recently delivered babies; these data would provide high-quality information on maternal health, behaviors, and experiences during pregnancy and postpartum, similar to CDC's PRAMS. The 2018 OPAS was designed to ensure that county-specific estimates for three OEI counties individually and the remaining six OEI counties as a group could be made with sufficient precision. These counties and county groups are shown in ***Table 1-1***, and more details on the design and allocation are discussed in ***Section 2***.

Table 1-1. Definition of OEI Strata for the 2018 OPAS

OEI Stratum	County or County Group
1	Cuyahoga
2	Franklin
3	Hamilton
4	Other OEI cohort counties: Butler, Lucas, Mahoning, Montgomery, Stark, Summit
5	Remainder of Ohio

Section 2 describes the 2018 sample design, and ***Section 3*** describes the data collection process. ***Section 4*** describes sampling weights based on the design and weight adjustments, including adjustment for nonresponse and calibration by important domains to the population totals for 2018 births in Ohio. ***Section 5*** describes the data processing and data analysis process.

2

Sample Design

2.1 Stratification and Frame Preparation

As noted, the target population for the 2018 OPAS was limited to pregnancies resulting in a live-born infant. By using the birth certificate file as the sampling frame, the OPAS implicitly excluded stillbirths, fetal deaths, and induced abortions. The traditional reasons for using this file are twofold: (1) reporting systems for these outcomes are not routinely in place in many states, and (2) the standard definitions for these outcomes vary. Because of the importance of learning about infants who have died, they were included in the sampling frame.

The following exclusions were used when sampling 2018 births:

- out-of-state births to residents
- in-state births to nonresidents
- missing key information (such as mother's last name)
- delayed processing of birth certificates (>6 months after birth)
- all but one infant from twin and triplet births
- all infants from multiple-gestation births with plurality >3
- adopted infants
- surrogate births

Each month, the eligible birth records were stratified by birth weight (low, defined as under 2,500 grams, or normal/high/unknown, defined as larger than 2,500 grams or missing) and OEI county or county group (defined previously). Eligible birth records were those with three or fewer babies occurring in Ohio to Ohio residents. The primary month for eligible births was 2 months before the month in which the sample was selected. However, we compared the frame for the current month with the prior month's frame—if a birth record was new (on the frame) but its eligibility month had already occurred, it was included on the frame for the month on which it appeared as long as the birth was no more than 6 months prior. For example, the October sampling frame primarily contained August births as the eligible population, but also included any births from April, May, June, and July that had not previously appeared on the sample frame. Additionally, for births that included multiple babies, only one randomly selected baby from that birth was included on the frame. The frame did not contain information directly linking all babies from a single birth, but through name, birth date, address matching, and data cleaning, RTI was able to group together most babies by birth.

2.2 Allocation

As stated previously, one of the primary goals of the OPAS was to make county and county-group estimates for the OEI counties. To that end, more sample was allocated to the OEI counties than the remainder of the state. In addition, more of the sample was allocated to the low birth weight stratum than the normal/high/missing birth weight stratum to also ensure precision for low birth weight babies. ***Table 2-1*** shows the two-way stratification implemented for each monthly replicate with frame counts and sample counts by stratum. This table presents an overall sampling fraction here for illustrative purposes; the sample fractions for each monthly replicate varied somewhat as the sample was drawn over 12 replicates. Replicate sample sizes are shown in ***Table 2-2***.

Table 2-1. Design Strata

Birth Weight Stratum	OEI Stratum	Frame Count	Sample Count	Overall Sampling Fraction
Low	1 – Cuyahoga	1,444	1,108	0.767
	2 – Franklin	1,662	1,310	0.788
	3 – Hamilton	1,013	747	0.737
	4 – Rest of OEI	2,641	1,878	0.711
	5 – Rest of Ohio	4,459	2,427	0.544
Normal and high	1 – Cuyahoga	12,406	2,057	0.166
	2 – Franklin	16,567	2,304	0.139
	3 – Hamilton	9,643	2,667	0.277
	4 – Rest of OEI	25,983	2,657	0.102
	5 – Rest of Ohio	56,323	3,095	0.055

Table 2-2. Sample Sizes by Replicate

Replicate	Eligible Birth Month[a]	Frame	Sample
1	January	10,536	2,750
2	February	10,059	1,600
3	March	10,064	2,400
4	April	9,556	2,400
5	May	11,311	1,500
6	June	10,232	1,500
7	July	11,449	1,500
8	August	11,795	1,500

(continued)

Table 2-2. Sample Sizes by Replicate (continued)

Replicate	Eligible Birth Month[a]	Frame	Sample
9	September	11,064	1,500
10	October	10,322	1,500
11	November	10,573	1,500
12	December	10,671	600

[a] Eligible birth records not on the frame at the time of the birth month were included on the subsequent month's frame.

3

Data Collection

3.1 Measurement and Instrumentation

The survey instrument was developed jointly by the GRC, ODH, ODM, RTI, and the OPAS EC and included the core CDC Phase 8 questionnaire and selected standard CDC Phase 8 questions identified by ODH and ODM.

The 2018 OPAS used the Phase 8 core questionnaire with a few modifications requested by the project sponsors, ODH and ODM. ODM requested the insertion of one follow-up question regarding the length of Medicaid eligibility. ODH requested the deletion of three standard questions related to prenatal vitamins and employment and the addition of five optional standard questions related to breastfeeding, postpartum care, and prenatal care. A variety of standard questions were also included beyond these topics. The full instrument is found in Appendix A.

3.2 Multimode Data Collection Design

The 2018 OPAS allowed mothers of selected births to respond using three different possible modes (mail, telephone, and web) and in two different languages (English and Spanish). The 2018 OPAS used a push-to-web methodology.

For the push-to-web methodology,

1. All sample members in the monthly birth cohort were mailed an initial invitation to complete the survey on the web. The initial invitation included a token incentive of $1 and instructions for completing the computer-assisted web interview (CAWI) survey.

2. After approximately 10 to 14 days, each nonresponding sample member received a second invitation via the mail to complete the CAWI survey.

3. After approximately another 10 to 14 days, each nonresponding sample member received a full-survey mailing packet that included the mail survey, business reply envelope, and instructions for completing the CAWI survey. The letter noted that those who completed the CAWI survey would receive a $5 completion incentive.

4. A final full-survey mailing was sent to all nonrespondents 10 to 14 days after the first full-survey mailing, including instructions for completing the CAWI survey.

5. Computer-assisted telephone interviews (CATI) were attempted with all those who did not complete a web or mail survey beginning approximately 60 days after the initial mailing. Calls were attempted for at least 30 days.

The initial contact included a token incentive and cover letter asking for the mother's cooperation, described the procedures for completing the online questionnaire, and provided a telephone number for mothers who wished to obtain additional information.

The second contact included a revised letter further eliciting the mother's cooperation and described the procedures for completing the online questionnaire.

The third and fourth mailings were the full questionnaire packet. The packets contained several items. First, the letter was modified slightly for the final mailings, primarily by adding an additional appeal for response to the web survey. Second, the packet contained an informed consent document. This document described the OPAS in more detail, explained how the mother was chosen and why, elicited the mother's cooperation, described the procedures for filling out the questionnaire (including the time required and the potential sensitivity of some questions), explained that there was no penalty for not participating or not answering all the questions, described the confidentiality of the answers, and provided a telephone number to obtain additional information. For minor participants, the informed consent document also mentioned that any instances of child abuse that are discovered must be reported to the proper authorities according to Ohio law. Third, the questionnaire booklet was included. Only English-language questionnaires were mailed to respondents; a phone number was provided for women to request a paper version of the Spanish questionnaire. A self-addressed return envelope with postage was provided for the easy return of the questionnaire. Fourth, a question-and-answer brochure was added to provide additional information about the OPAS. This brochure contained answers to the most frequently asked questions about the survey and may have been an important tool to convince the mother to participate. Lastly, if a sampled mother had multiple births, a card was included that stated "Some of the questions are about mothers and some of the questions are about babies. For the questions about babies, please answer for Baby _____." This card was intended to reinforce that the mother should be answering questions for only the selected baby.

RTI collected and processed mail questionnaires with TeleForm, an optical scanning technology system. TeleForm reduces costs by minimizing manual efforts in collecting data from forms, handling peak volumes of incoming forms, archiving and retrieving electronic images of scanned forms, automatically receiving data via scanning, and automatically sending confirmations and data forms. Scannable forms increase cost-effectiveness and improve data quality by eliminating errors associated with manual data entry. The layout and formatting of the paper surveys were in accordance with the recommended formatting for the PRAMS.

Completed mail surveys went through the standard data receipt process, which included logging and batching procedures that document when materials were received. The central receipting control system included a receipt log and generated status reports that allowed staff to quickly and easily assess progress within and across waves. This receipt log documented inbound data, formed the basis for RTI's chain-of-custody tracking, and allowed staff to quickly determine when and what materials had been received relative to a case or the wave sample as a whole.

All CATI and CAWI surveys were programmed into Voxco call center software. Questionnaire coding for the mixed-mode collection was identical in terms of questionnaire items and response options.

The Voxco system allows (1) the seamless merge of mixed-mode collected data; (2) real-time data checks for outliers, missing data, input errors, and intentionally skewed data; (3) monitoring of CATI and CAWI interviews; (4) automatic skip protocols; (5) interactive annotated notes per interview if problems develop; (6) instant data backup to secured backup server; and (7) real-time data quality review.

The 2018 OPAS consisted of 60 questions with an average completion time of approximately 25 minutes for the mail survey, 30 minutes for the CATI survey, and 18 minutes for the CAWI survey. The questionnaire for all modes was the same but responding to the survey by phone interview ran slightly longer than either the paper or web interview. All introductory letters were printed with English on one side and Spanish on the other. Spanish-translated survey materials and the CATI were available to participants who requested a Spanish survey.

3.3 Telephone Interviewing and Training

Telephone follow-up began after the last questionnaire was mailed. RTI used a variety of sources of telephone numbers to obtain a valid number for a mother. We used a telephone directory service and a batch identification search process. Calls to a number were staggered over different times of day and different days of the week. The calling period for a batch ran until the baby (who was the subject of the interview) was no more than 9 months old. If necessary, RTI made up to 10 call attempts to a number to reach a mother; note that a call attempt means that the telephone number was called and may not have resulted in an actual potential respondent contact/conversation.

The 2018 OPAS initially used interviewers who had been trained previously to implement the 2017 OPAS. To supplement interviewer attrition, additional telephone interviewer training occurred on April 19, 2018. Interviewers were introduced to the instrument, trained in the fundamentals of phone survey research, and engaged in several mock-interview situations. After the training, all interviewers were certified by taking a quiz on content and processes associated with the survey. In total, eight additional interviewers and quality specialists were trained for the 2018 CATI data collection. The interviewer training manual can be found in Appendix B.

In terms of CATI contact, the OPAS telephone interviewing procedures required contact with participants or access to identifiable information about participants. As such, the OPAS Protocol, *Humans Subjects Training Manual*, and *Telephone Interviewer Manual* included procedures and guidance for protecting human participants in the OPAS. For example, when making telephone call attempts, leaving answering machine messages, practicing refusal conversion, encountering sensitive situations, and implementing procedures to handle emergency situations, OPAS interviewers had to adhere to the protocol and interviewing guidelines to:

- obtain informed consent and fully inform the participants about the study,
- protect the participants' confidentiality by not revealing the nature of the survey to other household members,
- administer the survey correctly and accurately, and

- respect the participants' right to withdraw from the research without penalty of any kind.

The OPAS interviewers were also trained and familiar with the *Telephone Interviewer Manual* for the following procedures:

1. Leaving answering machine messages
2. Conducting refusal conversion
3. Handling sensitive situations
4. Implementing procedures to handle emergency situations
5. Implementing procedures and referrals for complaints and concerns

Telephone interviewers received ongoing training and monitoring by GRC staff and RTI. This training included the interview process, adverse event protocols, question-and-answer protocols, and OPAS response assistance protocols. Additionally, adverse event and question-and-answer protocols were in place for the mail returned surveys (relating to data entry findings) and CAWI surveys.

3.3 Incentive Experiment

To maximize the effectiveness of incentives for future iterations of OPAS, the OPAS team embedded an incentives experiment examining the impact of offering different pre- and post-incentive amounts. The goal was to determine which combination of pre- and post-incentives would achieve the highest response rate without drastically increasing the cost-per-complete. During 3 months of data collection (February, March, April births) a total of 6,400 sampled mothers were randomized to one of eight different experimental groups. Each group was assigned a different combination of pre-incentive ($1, $5, $10) and post-incentive ($0, $5, $10). For all other birth months, the incentives provided were the standard $1 pre-incentive, $5 post-incentive.

The experiment determined that the ideal combination for OPAS is a $10 pre-incentive and no post-incentive. This combination resulted in a 45-50% increase in the response rate with a minimal increase in cost. Appendix C provides the full report and recommendations from the experiment.

3.4 Data Collection Timeline

Table 3-1 provides the data collection timeline for the 2018 OPAS. The process for the first replicate of data collection began in May 2018, and call attempts on the final replicate were completed by June 2019.

Table 3-1. Data Collection Timeline

2018 OPAS	Jan	Feb	Mar	Apr	May	June	Jul	Aug	Sep	Oct	Nov	Dec
Wave	B25	B26	B27	B28	B29	B30	B31	B32	B33	B34	B35	B36
	2750	1600	2400	2400	1500	1500	1500	1500	1500	1500	1500	600
1st Web Invite	3/16/2018	4/13/2018	5/11/2018	6/15/2018	8/3/2018	8/17/2018	9/19/2018	10/12/2018	11/26/2018	1/7/2019	1/25/2019	2/20/2019
2nd Web Invite	3/30/2018	5/4/2018	5/25/2018	6/29/2018	8/20/2018	8/31/2018	10/3/2018	10/26/2018	12/11/2018	1/21/2019	2/8/2019	3/6/2019
1st Mail Packet	4/24/2018	5/17/2018	6/11/2018	7/12/2018	8/31/2018	9/20/2018	10/17/2018	11/9/2018	12/26/2018	2/4/2019	2/22/2019	3/20/2019
2nd Mail Packet	5/11/2018	6/1/2018	6/25/2018	7/31/2018	9/20/2018	10/4/2018	10/31/2018	11/27/2018	1/10/2019	2/18/2019	3/8/2019	4/3/2019
Non-Response Follow-Up	5/25/2018	6/15/2018	7/9/2018	8/14/2018	10/4/2018	10/18/2018	11/14/2018	12/11/2018	1/17/2019	3/4/2019	3/22/2019	4/17/2019
Finish Calling	6/15/2018	7/6/2018	7/30/2018	9/4/2018	10/25/2018	11/8/2018	12/5/2018	1/1/2019	2/7/2019	3/25/2019	6/17/2019	6/17/2019

Table 3-2 shows the number of completed interviews by design stratum and mode and by replicate and mode.

Table 3-2. Completed Interviews for the 2018 OPAS by Design Stratum and Sample Replicate

	Mail	Telephone	Web	Total
Overall	1,801	1,057	3,982	6,840
Design Stratum				
Low birth weight—Cuyahoga County	66	81	177	324
Low birth weight—Franklin County	89	91	188	368
Low birth weight—Hamilton County	36	42	110	188
Low birth weight—Other OEI cohort counties	122	125	293	540
Low birth weight—Remainder of Ohio	251	118	444	813
Normal and high birth weight—Cuyahoga County	157	104	426	687
Normal and high birth weight—Franklin County	214	137	502	853
Normal and high birth weight—Hamilton County	253	141	556	950
Normal and high birth weight—Other OEI cohort counties	237	124	580	941
Normal and high birth weight—Remainder of Ohio	376	94	706	1,176
Replicate				
1—January births	272	163	505	940
2—February births	163	82	398	643
3—March births	231	140	566	937
4—April births	225	111	640	976

(continued)

Table 3-2. Completed Interviews for the 2018 OPAS by Design Stratum and Sample Replicate (continued)

	Mail	Telephone	Web	Total
5—May births	140	105	247	492
6—June births	143	98	221	462
7—July births	118	95	251	464
8—August births	108	91	275	474
9—September births	123	55	272	450
10—October births	109	36	263	408
11—November births	120	53	254	427
12—December births	48	28	91	167

3.5 Response Rates

Table 3-3 provides the unweighted and weighted response rates for the 2018 OPAS using the American Association of Public Opinion Research response rate 1 (RR1). As shown in Equation 3.1, the calculation for RR1, the minimum response rate, is the number of completed interviews (I) divided by the number of interviews (complete plus partial [P]) plus the number of no-interviews (refusal [R] and break-off plus noncontacts [NC] plus others [O]) plus all cases of unknown eligibility (unknown if housing unit [UH] plus unknown other [UO]).

$$RR1 = \frac{(I)}{(I+P) + (R + NC + O) + (UH+UO)} \quad (3.1)$$

Table 3-3. Unweighted and Weighted Response Rates for the 2018 OPAS by Design Stratum and Sample Replicate

	Unweighted RR	Weighted RR
Overall	33.8	34.9
Design Stratum		
Low birth weight—Cuyahoga County	29.2	28.8
Low birth weight—Franklin County	28.1	28.3
Low birth weight—Hamilton County	25.2	25.3
Low birth weight—Other OEI cohort counties	28.7	27.9
Low birth weight—Remainder of Ohio	33.5	33.0
Normal and high birth weight—Cuyahoga County	33.4	32.4
Normal and high birth weight—Franklin County	37.0	35.7

(continued)

Table 3-3. Unweighted and Weighted Response Rates for the 2018 OPAS by Design Stratum and Sample Replicate (continued)

	Unweighted RR	Weighted RR
Normal and high birth weight—Hamilton County	35.6	34.0
Normal and high birth weight—Other OEI Cohort counties	35.4	34.7
Normal and high birth weight—Remainder of Ohio	38.0	36.5
Replicate[a]		
1—January births	34.2	35.1
2—February births[b]	40.2	42.7
3—March births[b]	39.0	41.5
4—April births[b]	40.7	42.4
5—May births	32.8	34.6
6—June births	30.8	33.8
7—July births	30.9	34.3
8—August births	31.6	33.8
9—September births	30.0	31.5
10—October births	27.2	31.3
11—November births	28.5	30.3
12—December births	27.8	28.8

[a] Eligible birth records not on the frame at the time the birth month became available were included on the subsequent month's frame.

[b] The incentive experiment was conducted during this month.

4

Weighting

4.1 Sampling Weight

The sampling weight for each selected birth record was the inverse of its probability of selection (the probability of selection is also called a sampling fraction). Because of the two-way stratification of OEI county group by birth weight in the OPAS, the 10 different probabilities of selection varied somewhat by the replicate in which the record was selected (i.e., across months).

4.2 Adjustment for Nonresponse

If a sampled mother did not respond to the survey, the sampled record was considered a unit nonrespondent. Biased estimates due to nonresponse occur if respondents and nonrespondents differ on key outcomes or on characteristics associated with key outcomes. For example, if mothers of one education level are less likely to respond to the OPAS than mothers of a higher education level, then, without any adjustment, survey estimates correlated to mother's educational attainment could be biased. Thus, RTI adjusted the sampling weights of respondents to reduce the potential for nonresponse as described below.

RTI investigated variables available on the OPAS frame, such as maternal age, education, marital status, trimester of first visit for prenatal care, parity, race, Hispanic ethnicity, unknown birthweight, and Women, Infants, and Children (WIC) enrollment, to determine if they should be included in a nonresponse adjustment model. We used classification and regression trees (Breiman, Friedman et al., 1984), also known as CART, to identify variables available on the frame for both respondents and nonrespondents for which there was differential nonresponse, and we used these variables to adjust sampling weights for nonresponse. Variables used to form the nonresponse classes for the 2018 data were mother's marital status; WIC enrollment; father's race; mother's age; previous live births now living; father's age; previous cesarean; and total number of prenatal care visits; .

Consistent with the 2016 and 2017 OPAS, the 2018 OPAS employed a model-based adjustment using SUDAAN's PROC WTADJUST (RTI International, 2012). Weighting classes that rely on too many variables can result in too many or too small weighting classes, which would reduce the efficiency of this adjustment (Kim, Li et al., 2007). Additionally, once classes are identified, if they are determined to be too small, the process by which collapsing occurs to ensure a sufficient number of respondents and nonrespondents can be time consuming. Using a model-based adjustment instead allows for the adjustments to be constrained within a defined set of values to maintain control over the variation of the weight adjustment. WTADJUST is an application of the generalized exponential model (GEM) of Folsom and Singh (2000). The design strata and sample replicate were included in the model regardless of whether CART identified them as key variables in differential nonresponse.

Equation 4.1 shows the weighting process for the nonresponse adjustment. Let w_{hi} be defined as the sampling weight for the *i*th birth record in stratum *h*. Then, the nonresponse-adjusted weight is defined as

$$NRW_{hi} = \begin{cases} W_{hi} \times A_{hi} & \text{Respondent} \\ 0, & \text{Nonrespondent} \end{cases} \tag{4.1}$$

where A_{hi} is the model-based nonresponse adjustment for the *i*th birth record in stratum *h*.

4.3 Final Calibration to Frame Totals

The final adjustment to the weights accounted for any errors in the frame and ensured that the weighted respondent characteristics matched the frame characteristics as closely as possible. This final calibration encompassed the PRAMS-recommended "adjustment for omissions in the sampling frame" while also allowing for additional variables to be specified in the model such that the final weight sums match the control totals on all domains.

Omissions in the sampling frame (noncoverage) can occur for a small number of birth records that take longer to appear on the frame than most others and may not have been on the frame at the time of sample selection. Although "late" birth records are part of the sample frame for a birth month if they are no more than 6 months away from the birth month and have not previously appeared on the frame, others may not be on the frame in time to be eligible for sample selection.

The PRAMS methodology uses a weighting class adjustment to adjust weights for noncoverage, similar to the nonresponse weight adjustment. For the reasons described previously for the nonresponse weight adjustments, the 2018 OPAS instead used a model-based approach, implementing a GEM model with SUDAAN's PROC WTADJUST. In addition to including variables related to noncoverage, the model also included additional variables from the OPAS frame that may not necessarily be related to noncoverage but may be related to key outcomes of interest. The final calibration model included mother's race/ethnicity, mother's age (categorized), preterm birth indicator, WIC status, and sampling stratum.

The final calibrated and nonresponse-adjusted weight, FW_{hi} , is defined in Equation 4.2,

$$FW_{hi} = NRW_{hi} \times C_{hi} \tag{4.2}$$

where NRW_{hi} is the nonresponse-adjusted weight from (4.1) and C_{hi} is the model-based calibration adjustment for the *i*th birth record in stratum *h*.

4.4 Quality Control Checks and Missing Data

The OPAS benefits from a rich list frame with a large amount of complete information for the eligible population; however, even in the final population file for 2018, a small number of records had a value of "unknown" for some calibration variables, including mother's ethnicity (HISP_MOM) with 41

(0.20%) unknown and mother's race (MRACE*) with 10 missing (015%). The indicator variable "preterm" (based on GEST_COMB) was missing for 2 cases (0.03%). The calibration variable with the highest number of unknown records in the population was WIC with 85 missing (0.42%). Instead of imputing this missing frame data, the small number of cases with unknown values for these variables in the sample were calibrated to the small number of cases with unknown values on the frame. Thus, no imputation was done for any variables.

Throughout the weighting process, the weight sums were checked after each adjustment, key outcomes of interest such as mother's smoking status and prenatal intendedness were monitored for any large changes, and the unequal weighting effects were checked.

Because the OPAS is a single-stage design with no clustering, the main source for design effect (i.e., inflation of variance due to the complex sample design) comes from the variation in the weights. This can be checked by calculating the unequal weighting effect (UWE), which measures the weight variation across strata. The UWE in domain *g*, when all respondents in a stratum have the same sampling weight, is defined as:

$$UWE_g = \frac{n_g \sum w_i^2}{(\sum w_i)^2} = \frac{n_g \sum_{h=1}^{H} (n_h w_h^2)}{[\sum_{h=1}^{H} (n_h w_h)]^2} \quad (4.3)$$

where n_g is the sample size in a geographic area or demographic domain g, n_h is the number of respondents in stratum h, and w_h is the sampling weight for stratum h in the sample. The UWE was monitored throughout the weighting process for key subgroups of interest (e.g., by racial/ethnic group) and it was determined no intervention (i.e., weight trimming or limits to the adjustment factors) was needed.

5

Data Processing and Analysis

5.1 Data Processing

RTI did a small amount of data processing to ensure internal consistency within a completed survey. Most of the data editing occurred for surveys completed by mail, because skip patterns and bounds on response options were programmed into the CATI and web surveys.

For some completed mail surveys, the mother responded with information in more than one type of unit, for example, providing height information in feet/inches as well as centimeters or weight information in pounds as well as kilograms. For height and weight, the English units were kept if the two responses were equal; otherwise, we compared them to see which one was more biologically plausible (e.g., a mother who reported being 6′5″ and 5′7″). For length-of-time questions for which more than one answer was given, the response in years was kept for length of time on Medicaid (Q11), the response in weeks was kept for timing of prenatal care (Q14), and the response in weeks was kept for length of breastfeeding (Q41). All invalid responses were deleted (set to missing).

Some responses to open-ended numeric questions on the mail survey also included character data such as "+" or "-." These characters were removed to make the data numeric-only, consistent with CAWI and CATI.

Some responses were inconsistent with values in the Vital Statistics data (i.e., the sampling frame), including infant date of birth (IDOB_DAT) and mother's date of birth (MDOB_DAT). In places where an obvious inconsistency occurred (e.g., the month and day were written in reverse order), the survey values were cleaned to match the values on the frame (BC_IDOB and BC_MDOB) but some cases remained where the two were not equal.

There were also some implausible values for "today's date" (i.e., survey completion date). Where possible, these were edited if month and day were flipped, there was an obvious error (e.g., June 31), or the year could be corrected based on when the replicate was in the field. A handful of cases had unedited values for "today's date" which led to a large value for infant age at survey completion.

Skip pattern rules implemented for CATI/CAWI were also applied to the mail surveys in post-processing. That is, where a mother should have skipped a question or questions based on a prior answer but did not, the variables that should have been skipped were reset to missing. There were a few exceptions to this rule:

- A response to Q7 (mark all that apply for a list of types of health care visits) overwrote Q6 (health care visits in 12 months before pregnancy).
- For Q10, a response of "didn't go for prenatal care" was overwritten if a mother's response to the earlier question about health insurance during prenatal care indicated that she had

insurance for her prenatal care (Q9), reported at least one type of insurance, and did not endorse the option "I did not go for prenatal care."

- If the response to "Is your baby alive now?" (Q35) was "No" but "Is your baby living with you now?" (Q36) was "Yes," the remaining responses were checked to determine if baby is alive or not and these variables (Q35, Q36) were adjusted accordingly.

- If a mother reported having a postpartum visit (Q52 = "Yes") but there was a response to barriers to postpartum visit (Q53) and there was no response items discussed with health care provider at postpartum visit (Q54) then Q52 was reset to "No" (i.e., assume no postpartum visit).

Final cleaned and formatted datasets were created in both SAS and Stata.

5.2 Data Analysis

Because the OPAS is a stratified design with variable weights, properly specifying the relevant sample design estimation in the software program used for analysis is imperative. Below are two sections of examples for SAS, SUDAAN, Stata, and R. Both sections of examples assume access to the analysis weight variable (WTANAL) and the variable that consolidates the two stratification variables into one ten-level variable (SUDSTRAT). The second section on doing variance estimation with the finite population correction (fpc) also assumes access to the stratum totals (TOTCNT).

5.2.1 Variance Estimation without FPC

SAS

```
proc surveyfreq data = opas_wtdat_2018;
      strata SUDSTRAT;
      weight WTANAL;
```

SUDAAN

```
proc crosstab design = wr data = opas_wtdat_2018 filetype = sas notsorted;
   nest SUDSTRAT; *combined stratum variable for SUDAAN;
   weight WTANAL;
```

Stata

```
Svyset [pweight=WTANAL], strata(SUDSTRAT) vce(linearized)
```

Then use the svy prefix for analysis commands.

R

```
library(survey)
#Note that SUDSTRAT needs to be a factor in R
dstrat <- svydesign(id=~1,strata=~SUDSTRAT, weights=~WTANAL, data=
opas_wtdat_2018)
```

5.2.2 Variance Estimation with FPC

SAS

```
proc surveyfreq data = opas_wtdat_2018 TOTAL=totals;
      strata SUDSTRAT;
      weight WTANAL;
```

This code assumes there is a second dataset called 'totals' which contains the strata variable SUDSTRAT and the SAS variable _TOTAL_ which takes the value of TOTCNT for each stratum from the OPAS dataset.

SUDAAN

```
proc crosstab design = wor data = opas_wtdat_2018 filetype = sas notsorted;
   nest SUDSTRAT; *combined stratum variable for SUDAAN;
   weight WTANAL;
   totcnt TOTCNT;
```

Stata

```
Svyset [pweight=WTANAL], strata(SUDSTRAT) fpc(TOTCNT)
```

Then use the svy prefix for analysis commands.

R

```
library(survey)
#Note that SUDSTRAT needs to be a factor in R
dstrat <- svydesign(id=~1, strata=~SUDSTRAT, weights=~WTANAL, fpc=~TOTCNT,
data=opas_wtdat_2018)
```

References

Breiman, L., J. Friedman, C. J. Stone, and R. A. Olshen (1984). *Classification and regression trees*. Boca Raton, FL: CRC Press.

Dillman, D., J. Smyth, and L. Christian (2014). *Internet, phone, mail, and mixed-mode surveys: The tailored design method*. 4th edition. Hoboken, NJ: John Wiley & Sons.

Folsom, R. E. and A. C. Singh (2000). *The generalized exponential model for sampling weight calibration for extreme values, nonresponse, and poststratification*. Proceedings of the American Statistical Association, Survey Research Methods Section.

Kim, J. J., J. Li, and R. Valliant (2007). "Cell collapsing in poststratification." *Survey Methodology* **33**(2): 139-150.

RTI International (2012). "SUDAAN: Statistical software for weighting, imputing, and analyzing data, Release 11." Research Triangle Park, NC: RTI International.