Surveillance for Emerging Threats to Mothers and Babies Network

Sampling and Weighting Methodology For End of Pregnancy and Infant Follow-Up Medical Record Abstraction

Last Updated: 2/27/2025

Division of Birth Defects and Infant Disorders

National Center on Birth Defects and Developmental Disabilities

Table of Contents

Background

SET-NET

The Surveillance for Emerging Threats to Mothers and Babies Network, or SET-NET is pregnant woman–infant linked longitudinal surveillance to understand the impact of emerging threats during pregnancy on pregnancy, birth, and early childhood outcomes. SET-NET is a preparedness model that can be further expanded if new threats emerge for pregnant women and their infants. In early 2020 it was rapidly adapted for COVID-19 surveillance.[1]

Medical records abstraction (MRA) to capture key clinical information on pregnant women and infants is fundamental to the SET-NET surveillance approach. However, given high COVID-19 case counts in some parts of the United States and limited resources, jurisdictions requested assistance from CDC to prioritize surveillance resources to capture data via MRA on a representative sample of included pregnancies as opposed to the entire cohort. Given the need to quickly report data for clinical decision-making and public health action and minimize the burden on already strained health departments, CDC developed a sampling approach to support health departments in the collection of population-based data. The sampling approach was expanded to SET-NET hepatitis C surveillance and may be expanded to other SET-NET exposures of interest. This document describes the methodology of the sampling approach for MRA, calculations of sampling weights, and population estimates for jurisdictions using end of pregnancy (EOP) sampling, infant follow-up (IFU) sampling, or both.

Surveillance Cohort

The population considered for surveillance through SET-NET varies by exposure. Exposure-specific inclusion criteria are provided in Table 1.

For SET-NET, data are collected on pregnant women with laboratory evidence of SET-NET exposures of interest. Infants born to women with exposures of interest during pregnancy may be monitored over time through SET-NET, even if the infant has no confirmed congenital infection, in order to support detection of long-term outcomes.

Modules and Data Sources

The SET-NET data system is organized into *general* variables and exposure-specific *modular* variables. The general variables pertain to all pregnant woman–infant pairs, regardless of the exposure of interest. Exposure-specific modular variables complement the general variables by providing information for pregnant woman–infant pairs about the exposure of interest. Modular variables were selected to align with existing data sources and published literature and were reviewed by a team of experts in obstetrics, pediatrics, epidemiology, and informatics with consideration for potential data capture. Together, general and modular variables align with key surveillance questions for each exposure, while striving to minimize burden and ensure quality data.[1]

The surveillance protocol focuses health department data collection efforts on medical records from hospitals and healthcare providers' offices (e.g., prenatal records, maternal hospitalization records, and infant follow-up medical records). Other data sources may include abstraction or linkage to records from routine case investigations and reports and vital statistics (birth and fetal death certificates). Linkage to data sources such as

[1] Woodworth KR, Reynolds MR, Burkel V, Gates C, Eckert V, McDermott C, Barton J, Wilburn A, Halai UA, Brown CM, Bocour A, Longcore N, Orkis L, Delgado Lopez C, Sizemore L, Ellis EM, Schillie S, Gupta N, Bowen VB, Torrone E, Ellington SR, Delaney A, Olson SM, Roth NM, Whitehill F, Zambrano LD, Meaney-Delman D, Fehrenbach SN, Honein MA, Tong VT, Gilboa SM. A Preparedness Model for Mother-Baby Linked Longitudinal Surveillance for Emerging Threats. *Matern Child Health J* 2021;25(2):198-206.

birth certificates is a common strategy to identify pregnancy status retrospectively for reported cases of infectious diseases.

Inclusion Criteria

The exposure-specific inclusion criteria are described in Table 1. This table may be expanded in the future to include additional SET-NET exposures.

Table 1. Inclusion criteria and priority populations for SET-NET cases.

Exposure	Inclusion criteria	Priority cases of interest
COVID-19	• Pregnant women who are SARS-CoV-2 RNA positive (laboratory-confirmed) in at least one clinical specimen at any point during pregnancy, up to and including the day of delivery, **AND** • Who reside in a participating jurisdiction **AND** • Who test positive during January 1, 2020, to December 31, 2021.	• Neonates who test positive for SARS-CoV-2 infection during the birth hospitalization or within 14 days of birth **AND** who are born to mothers who meet the inclusion criteria
Hepatitis C	• Pregnant women who are HCV RNA+ during pregnancy or prior to pregnancy, without evidence of treatment or clearance **AND** • Who reside in the jurisdiction **AND** • Whose date of pregnancy outcome is between January 1, 2018, and December 31, 2021 ***OR*** • Pregnancies that resulted in a child with any positive HCV RNA or IgM antibody test before 3 years of age **AND** • The child resides in the jurisdiction **AND** • The child was born between January 1, 2018 and December 31, 2021	• Children with any positive HCV RNA or IgM antibody test before 3 years of age AND their mother

Case Ascertainment

All SET-NET exposures are nationally notifiable diseases (Council of State and Territorial Epidemiologists case definitions: COVID-19, hepatitis C, syphilis), and case data are submitted through the National Notifiable Disease Surveillance System, or NNDSS, on an electronic report form specific to exposure. The NNDSS report forms include a pregnancy checkbox to identify pregnant cases; however, pregnancy status ascertainment typically requires case interview or medical chart review. The quality and accuracy of pregnancy status varies by exposure and jurisdiction. As such, most jurisdictions must rely on linkages between case surveillance and other available data sources to fully ascertain case counts. These additional data sources may include linkages of case surveillance systems to vital statistics data (such as birth certificates or fetal death certificates), linkages of case surveillance to prenatal screening records, or administrative data including hospital discharge data. For jurisdictions that are participating in sampling, the complete list of ascertained cases becomes the sampling frame from which to select cases for MRA. The unit of sample selection for the EOP sampling approach is the **pregnancy**. The unit of sample selection for the IFU sampling approach is the **pregnancy resulting in one or more live births**.

Sampling Methodology

Objective

The objective of the sampling approaches is for jurisdictions to collect a probability sample to obtain representative, precise estimates of mothers, pregnancy, and birth characteristics and outcomes, as well as longitudinal development characteristics of liveborn infants, among pregnancies with the exposure of interest.

Reporting Jurisdiction

Sampling occurs at the level of the reporting jurisdiction. Cities or counties that report exposure-specific pregnancy surveillance data separately from the state (e.g., California and Los Angeles County, Pennsylvania and city of Philadelphia, Illinois and city of Chicago) sample their target populations separately from the larger jurisdictional region. State jurisdictions remove cases reported through city or county jurisdictions from their sampling frames.

Sampling Stages

SET-NET is longitudinal, pregnant-woman-to-infant linked surveillance in which pregnant women with an infectious exposure are retrospectively identified. The pregnant woman is followed through their birth hospitalization, and the infant is followed for up to 2 years.

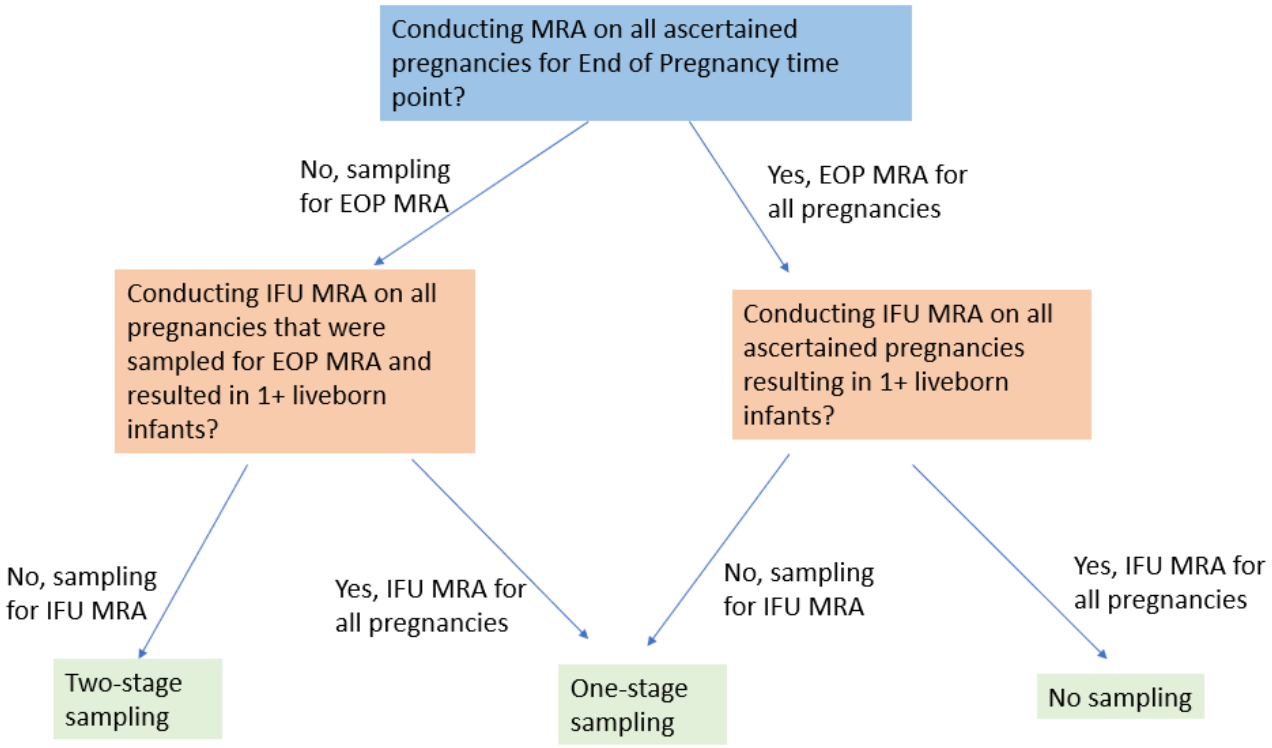
MRA occurs for the pregnant woman and the infant at EOP using the birth hospitalization medical records, which are obtained from birthing hospitals. MRA occurs for IFU from pediatric well visits, which are most often obtained from pediatrician practices. The SET-NET model presents two distinct time points for MRA: at the EOP/birth hospitalization time point, and for IFU. Sampling at these time points is called a sampling stage. CDC provided guidance to participating jurisdictions with different sampling approaches for COVID-19 and hepatitis C exposures.

COVID-19 sampling stages

Each jurisdiction selected one of four possible sampling approaches for COVID-19 surveillance MRA (see Figure 1):

1. No sampling, such that MRA occurs for every ascertained pregnancy and infant (i.e., a census approach)
2. One-stage sampling such that pregnancies are sampled for MRA for the EOP time point
3. One-stage sampling such that pregnancies resulting in one or more liveborn infants are sampled for MRA for the IFU time point
4. Two-stage sampling, such that the first stage is sampling pregnancies for MRA for the EOP timepoint, and the second stage samples of those pregnancies resulting in one or more liveborn infants for the IFU time point.

Figure 1. COVID-19 Sampling Approaches



Hepatitis C sampling stages

Each jurisdiction selected one of three possible sampling approaches for hepatitis C MRA:

1. No sampling, such that MRA occurs for every ascertained pregnancy and infant
2. One-stage sampling such that pregnancies are sampled for MRA for the EOP time point
3. One-stage sampling such that pregnancies resulting in one or more liveborn infants are sampled for MRA for the IFU time point

No jurisdiction used two-stage sampling for hepatitis C MRA.

Target Population and Sampling Frame

Target Population

The target population for the SET-NET exposure of interest should include all pregnancies meeting the inclusion criteria described in Table 1 for the EOP timepoint. For the IFU timepoint, the target population is limited to pregnancies meeting the inclusion criteria described in Table 1that resulted in one or more liveborn infants.

Sampling Frame of Pregnancies with Exposures of Interest

Jurisdictions construct sampling frames according to the inclusion criteria shown in Table 1. Sampling frames comprise the full list of cases ascertained through data linkages and other ascertainment methods. There are various ways for jurisdictions to identify pregnancies, and jurisdictions are not limited to one method for identification of pregnant women. Some examples are

1. **Jurisdictional case surveillance data for pregnant cases** (pregnancy status directly indicated on NNDSS case report form or reportable disease registries)
2. **Data linkages to confirm pregnancy status** (linking billing data, prenatal screening, or other data sources to case surveillance data)
3. **Data linkages to birth outcome data** (linking birth certificates, fetal death certificates, administrative databases, or other data sources to case surveillance data)

Gaps in the Sampling Frame

Jurisdictions considered the implications for their generalizability from each ascertainment method including timeliness and completeness of case ascertainment. Approaches that link to datasets for births may introduce a time lag. Jurisdictions confirming pregnancy status using data linkages need to determine whether cases that do not link are part of their sampling frames (e.g., women indicated as pregnant on the case report form who do not link to vital records). For those cases that are not linked, jurisdictions could determine how to account for these cases in their sampling approach.

Sample Requirements

The SET-NET sampling designs have four requirements:

1. Random selection must be used in each sampling step so that every eligible pregnancy in the sampling frame has a non-zero chance of selection into the sample.
2. The probability of selection for every pregnancy must be known and retained in the final analytic files.
3. Jurisdictions must have unique identifiers for every sampled pregnancy, and these identifiers must be retained in the final analytic files.
4. The sample must be selected using simple random sampling at separate intervals, without replacement, at one or more separate time intervals during the surveillance period (discussed in the next section).

Selection of the Sample

Sample Size

Each jurisdiction determined the sample size based on their capacity for conducting MRA.

Sampling Intervals

CDC's recommended sampling approach differed by exposure of interest. For COVID-19, sampling of pregnancies for MRA occurred at regular intervals throughout the reporting period, rather than waiting until the surveillance period ended. Based on their capacity, jurisdictions determined regular, appropriate intervals for sample selection. For HCV, the intervals were set to full calendar-year birth cohorts. For example, all pregnancies with the pregnancy outcome occurring in 2018 were in a single interval.

After selecting the sample for an interval, all cases from that interval become ineligible for sampling in any subsequent intervals. Each sample should pull from cases in the sampling frame that were not included in a previous sampling frame. Therefore, any given pregnancy only has one opportunity to be sampled.

Priority Cases

CDC requested ***complete ascertainment of select cases for MRA***. These cases were considered to have priority outcomes of interest. Priority outcomes must have low frequency and importance to answer key surveillance questions. The priority outcomes for each SET-NET exposure are noted in Table 1.

Some jurisdictions also defined their own priority outcomes of interest and, as such, conducted MRA for all pregnancies with priority outcomes. Those priority outcomes included specific adverse outcomes, such as

stillbirths, maternal deaths, infant deaths, and infants with postnatal infection. **For IFU,** stillbirths and infant deaths before maximum infant follow-up age preclude infant follow-up and therefore are not included as priority outcomes for jurisdictions sampling at the IFU timepoint.

Identifying priority outcomes implies complete census (i.e., a sampling fraction of 1.0) of these priority outcome cases and only a random sample of other pregnancies. The sample is drawn after the priority outcome cases are removed.

Random Sample and Stratified Random Sample

After priority cases are identified and removed from the sampling frame, jurisdictions used simple random sampling to select pregnancies for MRA. Random selection provides the best method to obtain a representative sample. A smaller number of jurisdictions conducted stratified random sampling to ensure adequate representation of a given subgroup of cases. For example, a jurisdiction may have wanted to stratify its random sample by maternal race and ethnic subgroups and thus may have chosen to apply a higher sampling fraction to less prevalent strata.

Sample Weighting

CDC calculates sampling weights following each quarterly SET-NET data submission window. The sampling weights adjust each record such that, as a whole, a jurisdiction's submitted cases represent the jurisdiction's total sampling frame. All submitted pregnancies receive a sampling weight, even pregnancies from those jurisdictions that conduct MRA on all cases (i.e., census approach). Jurisdictions that conducted sampling for MRA submitted documentation of their sampling interval, which included total cases, number of priority cases, number of records eligible for selection, the number of cases selected, and whether medical records were available and abstracted for each case.

Weight calculations

The general formulas for sampling weights are shown in Formulas 1, 2, and 3. The total sampling weight, *w*, is calculated per sampling stage, i.e., there is a weight for EOP and a weight for IFU. The formulas below use the general nomenclature *w*, which is later clarified as w_{eop} or $w_{ifu,t}$ to accurately indicate sampling stage and timepoint.

The total weight, *w*, comprises the product of a selection weight (*w1*, shown in Formula 1) and a LTFU weight (*w2*, shown in Formula 2). The selection weight reflects the probability of selection for each pregnancy, and the LTFU weight reflects a selected pregnancy's probability of having MRA completed. Pregnancies with completed MRA receive a total sampling weight (Formula 3), which is simply the product of the selection weight and the LTFU weight.

The same weight calculations apply to sampling approaches using a single stratum and interval and also to sampling approaches using multiple strata and/or multiple intervals.

Formula 1. Selection weight (w1)

w1 is the inverse of the probability of selection:

$$p1_{ij} = \text{Prob(selection | interval } i \text{ and stratum } j) = \frac{number\ of\ cases\ selected\ for\ MRA\ for\ interval\ i\ and\ stratum j}{number\ of\ total\ cases\ in\ sampling\ frame\ for\ interval\ i\ and\ stratum j}$$

Therefore:

$$w1_{ij} = \frac{1}{Prob(selection\ |interval\ i\ and\ stratum\ j)}$$

Formula 2. LTFU weight (w2)

*w*2 is the LTFU weight for observations in sampling interval *i* and stratum *j*. The denominator is the number of cases with completed MRA.

$$w2_{ij} = \frac{number\ of\ cases\ selected\ for\ MRA\ for\ interval\ i\ and\ stratum\ j}{number\ of\ cases\ with\ MRA\ completed\ for\ interval\ i\ and\ stratum\ j}$$

Formula 3. Total sampling weight (w)

w is the total sampling weight for observations in sampling interval *i* and stratum *j* with completed MRA.

$$w_{ij} = w1_{ij} \times w2_{ij}$$

Census Approach

This is the approach for jurisdictions not sampling for MRA. Jurisdictions implementing the census approach do not send additional documentation to CDC; all pregnancies have a total sampling weight (*w*) of 1.0. CDC assumes any missingness is completely at random, no LTFU weight is calculated, and each record represents one case. The cases are assigned a stratum of 1 for all records. For jurisdictions conducting a full census approach for MRA, each case is weighted to represent only itself, and the sum of the cases is the size of the population of interest in the jurisdiction. It is assumed that every submitted record has completed MRA.

- Selection weight: *w1* = 1.0
- LTFU weight: *w2* = 1.0
- Total sampling weight: *w* = *(w1* × *w2)* = 1.0

Simple Random Sample and Stratified Random Sample Approaches

Most jurisdictions are conducting simple random sampling and not stratified random sampling. For these jurisdictions, there is only one jurisdictional stratum *j* for all pregnancies within the jurisdiction. For jurisdictions that conduct stratified random sampling, the selection and the nonresponse weights will be unique to each stratum *j* per interval *i*. All jurisdictions that are sampling for MRA will receive the total sampling weight calculation. Formulas 1 and 2 are the same for all pregnancies without priority outcomes (regardless of interval or stratum), but pregnancies with priority outcomes use a slightly different calculation for LTFU weights (*w2*) shown in Formula 4. For pregnancies with priority outcomes, the numerators and denominators are pooled across all intervals. This is because the pregnancies with priority outcomes are, by definition, uncommon and low in number; many intervals will not have any.

Formula 4. LTFU weight for pregnancies with priority outcomes (w2)

*w*2 is the LTFU weight for observations across all pregnancies with priority outcomes. The denominator is the number of cases with completed MRA.

$$w2 = \frac{number\ of\ pregnancies\ with\ priority\ outcomes\ in\ jurisdiction\ (all\ intervals)}{number\ of\ pregnancies\ with\ priority\ outcomes\ with\ MRA\ completed\ (all\ intervals)}$$

Pregnancies without priority outcomes

- Selection weight: *w1* = inverse probability of selection (Formula 1)
- LTFU weight: *w2* (Formula 2)
- Total sampling weight: *w* = *w1* × *w2* (Formula 3)

Pregnancies with priority outcomes

Jurisdictions should select all priority pregnancies with certainty such that the probability of selection is 1.0. The LTFU weight for priority outcomes is similar to the LTFU weight shown in Formula 2; however, priority pregnancies are pooled over intervals because they are so rare. The selection weight reflects that these cases were selected with 100% certainty.

- Selection weight: *w1* = 1.0
- LTFU weight: *w2* (Formula 4)
- Total sampling weight: *w* = *w1* × *w2*

Final Analytic Weights

Formulas 1 through 4 calculate weights for a single sampling stage, i.e., EOP or IFU. The final analytic weights depend on the sampling approach, specifically the number of sampling stages. For example, analyses of EOP data use the total EOP weight (w_{eop}) only, but analyses of IFU data are more complicated because the final analytic weights for IFU are often partially dependent on EOP weights. The IFU analytic weights are the product of the IFU total sampling weight (w_{ifu}) and the selection weight of the EOP weight ($w1_{eop}$):

Formula 5. Analytic IFU weight ($w_{ifu\_analytic}$)

$w_{ifu\_analytic}$ is the product of the EOP selection weight ($w1_{eop}$) and the total IFU weight (w_{ifu})

$$w_{ifu\_analytic} = w1_{eop} \times w_{ifu}$$

The final analytic weights are described by sampling approach here and summarized in Figure 2:

1. Census approach (no sampling)

For jurisdictions that conduct MRA on all pregnancies at both the EOP and IFU stages, the final analytic weight for each pregnancy is w_{eop} x w_{ifu} = 1.

2. One stage sampling such that pregnancies are sampled for MRA at the EOP timepoint

For jurisdictions that conduct MRA on a random sample of pregnancies only for EOP, the analytic weights reflect w_{eop} and assume a certainty IFU weight (w_{ifu} = 1).

- EOP analytic weight: w_{eop}
- IFU analytic weight: $w_{ifu\_analytic} = w_{eop} \times 1 = w_{eop}$

3. One stage sampling such that pregnancies resulting in one or more live births are sampled for MRA at the IFU timepoint

For jurisdictions that conduct MRA on all pregnancies at the EOP time point and a random sample of pregnancies resulting in one or more live births for the IFU timepoint, the analytic weights reflect a certainty EOP weight (w_{eop} = 1) and use the IFU total sampling weight.

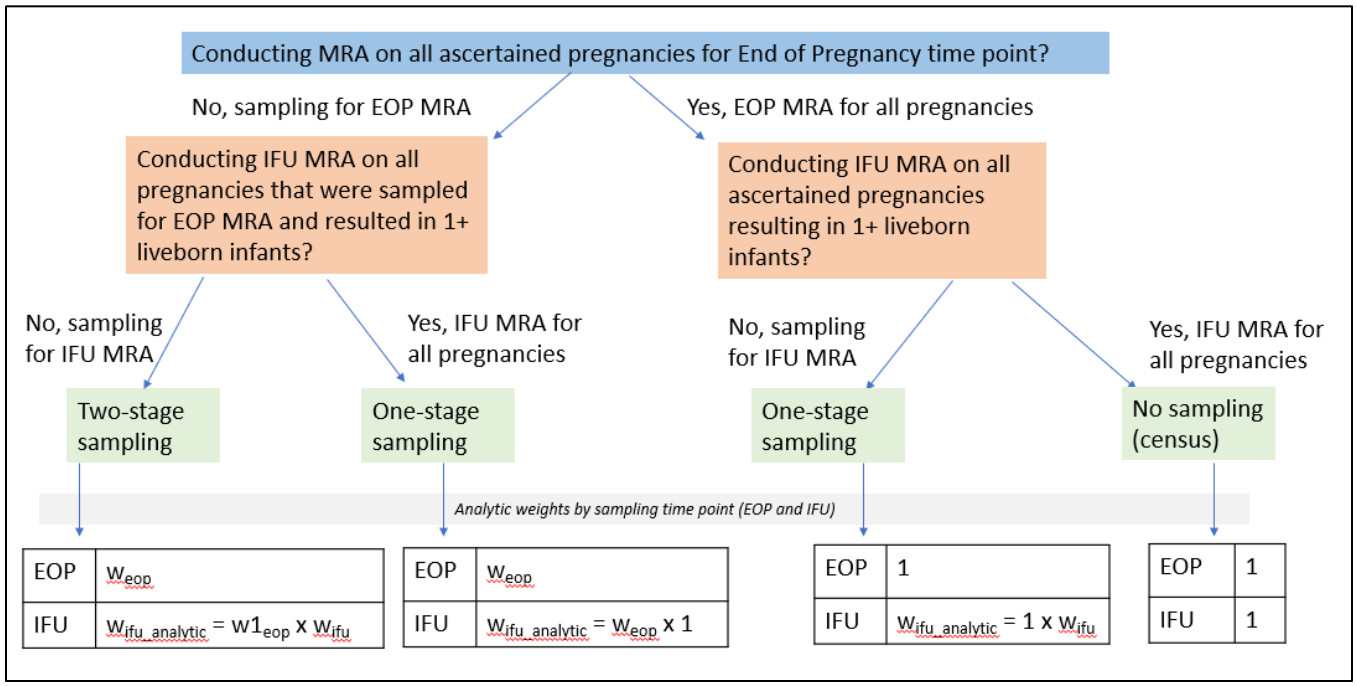
- EOP analytic weight: 1
- IFU analytic weight: $w_{ifu\_analytic} = 1 \times w_{ifu}$

4. Two-stage sampling of pregnancies at the EOP time point and again at the IFU time point

For COVID-19 jurisdictions that conduct two-stage sampling, such that the first stage is sampling pregnancies for MRA for the EOP timepoint, and the second stage is sampling pregnancies resulting in one or more liveborn infants for the IFU time point, the analytic weights are as follows:

- EOP analytic weight: w_{eop}
- IFU analytic weight: $w1_{eop}$ (EOP selection weight) × w_{ifu} (Formula 5).

Figure 2. Analytic Weights by Sampling Stage and Time Point



IFU Laboratory Data Analytic Weight (COVID-19 only)

The mechanism for submitting infant COVID-19 laboratory data differed across participating jurisdictions and affected the IFU analytic weights for analyses using those data as the primary exposure or outcome. It was necessary to create a separate analytic weight specific to IFU laboratory analyses, w_{ifu_lab}.

There were four possible ways for jurisdictions to obtain infant COVID-19 laboratory records, resulting in four different equations for calculating w_{ifu_lab}. Each are described below and summarized in Figure 3.

1. The jurisdiction obtained infant laboratory data only during IFU MRA. In this approach, the laboratory data are available for all infants with completed MRA. Therefore, the IFU lab weight is equal to the IFU MRA weight:

$$w_{ifu\_lab} = w_{ifu\_analytic}$$

2. The jurisdiction linked to and submitted infant laboratory data for all liveborn infants to pregnant women meeting the case inclusion criteria. This is a census of infant COVID-19 laboratory data, so the IFU lab weight is a certainty weight of 1.

$$w_{ifu\_lab} = 1$$

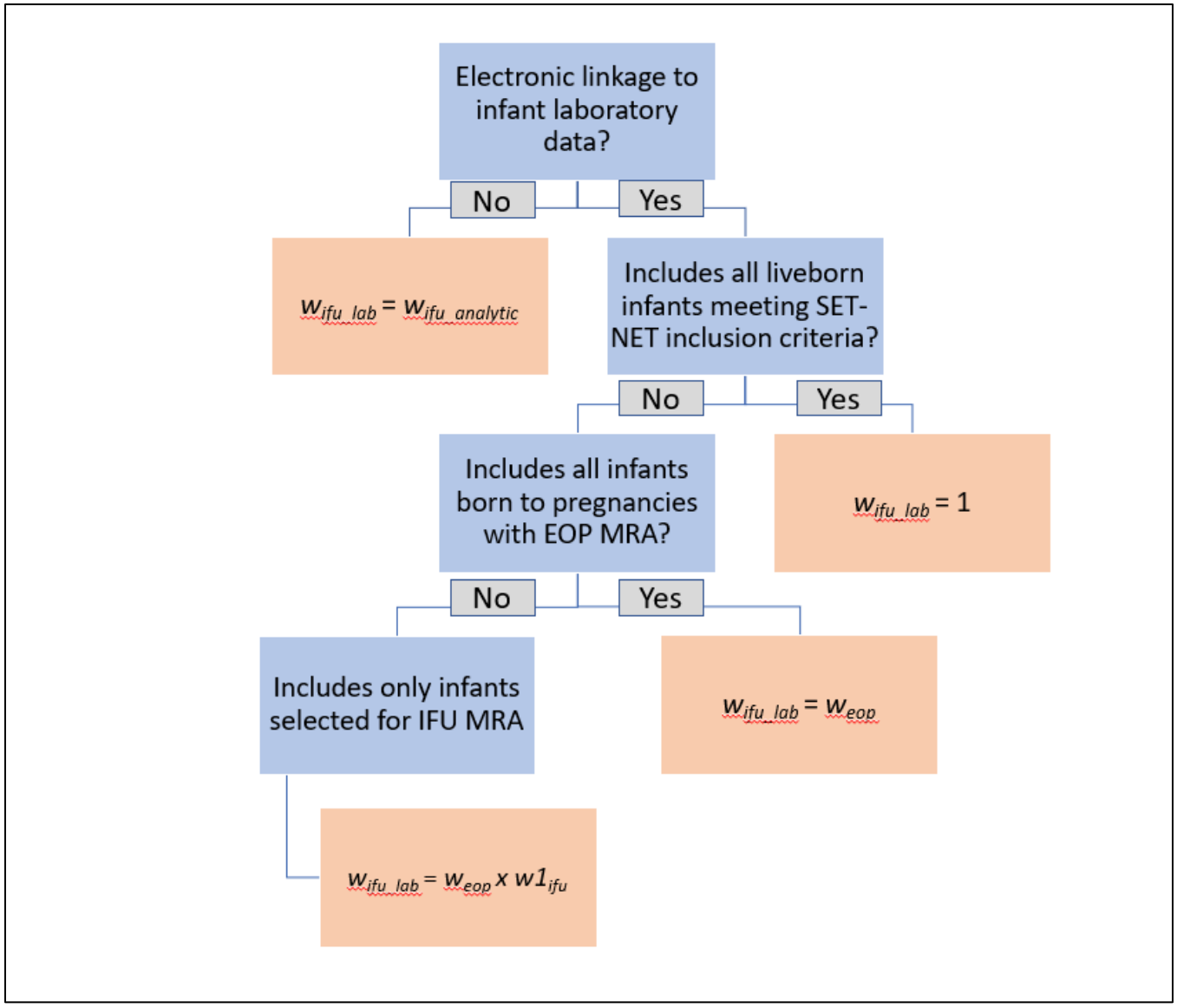
3. The jurisdiction linked to and submitted infant laboratory data for all liveborn infants to a pregnant woman selected for EOP MRA. In this approach, the IFU laboratory weight is equivalent to the EOP weight:

$$w_{ifu\_lab} = w_{eop}$$

4. The jurisdiction linked to and submitted the infant laboratory data for all infants selected for IFU MRA. In this approach, the laboratory data are available for all sampled infants even if MRA was not completed, so the IFU LTFU weight (w2ifu) is not incorporated.

$$w_{ifu\_lab} = w_{eop} \times w1_{ifu}$$

Figure 3. Analytic weights for analysis of COVID-19 IFU laboratory data



Finite Population Corrections

CDC recommends considering a finite population correction (FPC) when analyzing weighted SET-NET data aggregated across multiple jurisdictions. The FPC can be used to adjust standard error estimates for participating jurisdictions reporting data on more than 5% of the population (e.g., some jurisdictions may be reporting more than 99% of the population). For those jurisdictions, it is more appropriate to analyze the data as a population without replacement and with population totals included for each jurisdiction and strata. Qualifying jurisdictions provided the information for the FPC, either total case counts or an estimate of their total case counts, in their linkage documents provided with each data submission.

Limitations

This sampling approach has some limitations. First, CDC and jurisdictions collaborated to ensure the sampling frames were as complete as possible so that these data represented the population of interest under surveillance. However, given issues with completeness and accuracy of pregnancy status across data sources and jurisdictions, CDC allowed jurisdictions to make local decisions regarding their development of their sample frame based on their knowledge and experience with the accuracy and completeness of data sources. Therefore, gaps in the sampling frame that vary across jurisdictions still may exist. For COVID-19 SET-NET data, CDC conducted sensitivity analyses of two jurisdictions sending linked vital statistics data for all cases in their jurisdiction and compared to weighted estimates calculated from their subset of sampled cases. For the combined jurisdictions, the 95% confidence intervals from the sampled data included the population estimate from the full cohort (e.g., all cases identified through linked birth certificates) for 92% of the maternal variables.

Second, because CDC guided jurisdictions to partition their sampling into intervals to allow for staff to begin MRA, intervals with partial MRA are adjusted for LTFU at the time of weighting, and these same intervals may be updated later when medical records are abstracted. In addition, intervals without any MRA cannot be weighted and are omitted from weighted datasets until abstraction begins. Thus, reports using the interim weighted dataset are considered preliminary, and findings may be updated as MRA are completed for the entire interval and subsequently for the entire cohort for an exposure. However, these interim analyses are critical for informing clinical decision-making and public health action, and CDC will continue to monitor this approach and ensure conclusions are based on the best available data.

Summary

The COVID-19 pandemic stretched health department capacity to conduct medical records abstraction on all cases in SET-NET. Although this approach was originally developed for COVID-19, its application was adapted for hepatitis C surveillance, using birth-cohort intervals, to allow for sampling of liveborn infants for IFU MRA. The sampling approach allows for the collection of population-based data while balancing the capacity and resources of health departments to conduct quality data collection from medical records. The approach also allows flexibility, so the jurisdictions, in consultation with CDC, were able to decide on their sampling approach, including identification of priority cases of interest or stratifications that might be useful to inform their local programmatic needs. As SET-NET was developed to be a preparedness network, this sampling approach to collect population-based pregnant woman–infant linked longitudinal surveillance may have applications to other emerging threats and future responses.

Appendix A: Glossary of Terms

The following terms are terms used throughout the SET-NET sampling process.

Case: A pregnant woman with the exposure of interest during pregnancy based on inclusion criteria.

Case Report Form (CRF): The method for case surveillance reports from jurisdictions to CDC via the National Notifiable Disease Surveillance System (NNDSS) for the exposure of interest. The CRF for selected exposures may capture pregnancy status on the CRF, although quality and completeness may vary by exposure and jurisdictional capacity to conduct interviews or medical chart review.

Census: The total number of cases ascertained.

Infant: The live birth resulting from the pregnancy meeting inclusion criteria for surveillance. For the purpose of this document, stillborn infants are not included when the term "infant" is used.

Infant Follow-Up (IFU): Data collected at specified time intervals from the medical records of an infant's well child visit.

Interval: Specific time point that a jurisdiction sets as their time frame for selection of sampled cases.

Jurisdiction: State, local, and territorial health departments.

Lost-to-follow-up (LTFU): When a case has been selected for MRA, if that record cannot be found, or any other reason why that case did not receive MRA.

Medical record abstraction (MRA): Collecting data from a medical record.

Pregnant woman: The pregnant woman included in the surveillance.

Pool: The total eligible population that a jurisdiction will draw their sample from.

Priority case: A case with a certain selected outcome of interest such that all pregnancies with this outcome are selected for MRA.

Sampling frame: The list of eligible cases from which the sample is selected for a specified interval.

Selection weight: The inverse of the probability of selection.

Target population: The entire population that the sampled data are meant to generalize.

Total sampling weight: The selection weight multiplied by the LTFU weight.

Appendix B: Weighting Examples

Example A1: One-stage sampling for EOP

For this example, a fictional jurisdiction identified 1,515 pregnancies meeting the inclusion criteria. The jurisdiction determined they could complete EOP MRA for 100 cases per interval and took unstratified random samples from two intervals. In interval 1, they found 1010 cases meeting the inclusion criteria. Of these, they identified 10 pregnancies with priority outcomes and completed MRA for seven pregnancies with priority outcomes; then they randomly sampled 100 pregnancies without priority outcomes and completed MRA for 80 of those cases. In interval 2, they found 505 cases meeting the inclusion criteria, of which 5 had priority outcomes. They randomly sampled 100 of the pregnancies without priority outcomes and completed MRA for 50. They located records for all 5 pregnancies with priority outcomes and completed MRA for all of them.

Table A1. Example Data for Weighting One-stage EOP sampling approach

	Interval 1	Interval 2
	End of Pregnancy	End of Pregnancy
Total cases meeting inclusion	1010	505
Sampling Frame (nonpriority cases)	1000	500
Sampled Cases	100	100
Selection Probability	0.10	0.20
Sampled Cases with Completed MRA	80	50
Priority Cases	10	5
Priority Cases with Completed MRA	7	5

After receiving their data, the SET-NET team would calculate the weights using the following procedures:

Example A1 Interval 1 non-priority EOP weights:

$$w1_{eop\_1} = \frac{1}{Prob(selection\ |interval\ i)} = \frac{1}{.10} = 10.00$$

$$w2_{eop\_1} = \frac{number\ of\ cases\ selected\ for\ MRA\ for\ interval\ i}{number\ of\ cases\ with\ MRA\ completed\ for\ interval\ i} = \frac{100}{80} = 1.25$$

$$w_{eop\_1} = w1_{eop\_1} * w2_{eop\_1} \quad = 10.00 \times 1.25 = \ 12.50$$

Example A1 Interval 2 non-priority EOP weights:

$$w1_{eop\_2} = \frac{1}{Prob(selection\ |interval\ i)} = \frac{1}{.20} = 5.00$$

$$w2_{eop\_2} = \frac{number\ of\ cases\ selected\ for\ MRA\ for\ interval\ i}{number\ of\ cases\ with\ MRA\ completed\ for\ interval\ i} = \frac{100}{50} = \ 2.00$$

$$w_{eop\_2} = w1_{eop\_2} \times w2_{eop\_2} \quad = 5.00 \times 2.00 = \ 10.00$$

Example A1 priority EOP weights (pooled over all intervals):

$$w1_{eop\_pri} \ = \ 1.00$$

$$w2_{eop\_pri} \ = \frac{number\ of\ pregnancies\ with\ priority\ outcomes\ (all\ intervals)}{number\ of\ pregnancies\ with\ priority\ outcomes\ with\ EOP\ MRA\ completed\ (all\ intervals)} = \frac{15}{12} = 1.25$$

$$w_{eop\_pri} = \ w1_{eop\_pri} \ \times w2_{eop\_pri} \ = \ 1.00 \ \times 1.25 = 1.25$$

Applying weights to the dataset makes the cases with completed MRA equal to total cases meeting inclusion criteria:

Interval 1 non-priority end of pregnancy cases: $80 \times 12.50 = 1000$
Interval 2 non-priority end of pregnancy cases: $50 \times 10.00 = 500$
Total End of Pregnancy priority cases : $12 \times 1.25 = 15$
Total weighted case count : $1000 + 500 + 15 = 1515$

Two-stage sampling: EOP and IFU

In this example, the fictional jurisdiction from example A1 completed EOP MRA. With their remaining resources, they determined they could complete medical record abstraction for infant follow-up for about half of the pregnancies with EOP MRA. They drew a random sample from the pregnancies selected for EOP MRA that ended in live birth. In interval 1, out of the 1010 pregnancies meeting the inclusion criteria, there were 970 pregnancies resulting in at least one live birth including the 10 priority cases. The sampling frame becomes the non-priority pregnancies sampled for EOP MRA that ended in a live birth, so the jurisdiction sampled 48 of these pregnancies for IFU MRA and searched for the infant records for the 10 priority pregnancies. They located and abstracted 39 out of 58 total infant records. In interval 2, there were 455 live births, including 5 priority cases. Of the interval 2 pregnancies selected for EOP MRA, 90 resulted in a live birth. From this sampling frame of 90, they sampled 45 and completed MRA on 20 of the sampled non-priority and all 5 of the priority.

Table A2. Example Data for Weighting Two-stage sampling approach

	Interval 1		Interval 2	
	End of Pregnancy	Infant Follow-Up	End of Pregnancy	Infant Follow-Up
Total cases meeting inclusion	1010	970 (resulting in 1+ live births)	505	455 (resulting in 1+ live births)
Sampling Frame (nonpriority cases)	1000	96 (based on cases sampled in EOP stage)	500	90 (based on cases sampled in EOP stage
Sampled Cases	100	48	100	45
Selection Probability	0.10	0.5	0.20	0.5
Sampled Cases with Completed MRA	80	32	50	20
Priority Cases	10	10 (resulting in 1+ live births)	5	5 (resulting in 1+ live births)
Priority Cases with Completed MRA	7	7	5	5

After receiving their data, the SET-NET team would calculate the second stage weights for IFU MRA using the following procedures:

Example A2 Interval 1 non-priority IFU weights:

$$w1_{ifu\_1} = \frac{1}{Prob(selection\ |interval\ i)} = \frac{1}{.5} = 2.00$$

$$w2_{ifu\_1} = \frac{number\ of\ cases\ selected\ for\ MRA\ for\ interval\ i}{number\ of\ cases\ with\ MRA\ completed\ for\ interval\ i} = \frac{48}{32} = 1.50$$

$$w_{ifu\_1} = w1_{ifu\_1} \times w2_{ifu\_1} = 2.00 \times 1.50 = 3.00$$

Example A2 Interval 2 non-priority IFU weights:

$$w1_{ifu\_2} = \frac{1}{Prob(selection\ |interval\ i)} = \frac{1}{.5} = 2.00$$

$$w2_{ifu\_2} = \frac{number\ of\ cases\ selected\ for\ MRA\ for\ interval\ i}{number\ of\ cases\ with\ MRA\ completed\ for\ interval\ i} = \frac{45}{20} = 2.25$$

$$w2_{ifu\_2} = w1_{ifu\_2} \times w2_{ifu\_2} = 2.00 \times 2.25 = 4.50$$

*Example A2 IFU **Priority** weights (pooled over all intervals):*

$$w1_{ifu\_pri} = 1.00$$

$$w2_{ifu\_pri} = \frac{number\ of\ pregnancies\ resulting\ in\ live\ birth\ with\ priority\ outcomes\ (all\ intervals)}{number\ of\ pregnancies\ resulting\ in\ live\ birth\ with\ priority\ outcomes\ with\ IFU\ MRA\ completed\ (all\ intervals)} = \frac{15}{12} = 1.25$$

$$w_{ifu\_pri} = w1_{ifu\_pri} \times w2_{ifu\_pri} = 1.00 \times 1.25 = 1.25$$

After calculating w_{eop} and w_{ifu} , weighting would produce the following nonpriority weights for $w_{ifu\_analytic}$:

Interval 1 nonpriority analytic IFU: $w_{ifuanalytic\_1} = w1_{eop\_1} \times w_{ifu\_1} = 10.00 \times 3.00 = 30.00$
Interval 2 nonpriority analytic IFU: $w_{ifu\_analytic\_2} = w1_{eop\_2} \times w_{ifu\_2} = 5.00 \times 4.50 = 22.50$
Total priority analytic IFU : $w_{ifu\_analytic\_p} = w1_{eop\_pri} \times w_{ifu\_pri} = 1.00 \times 1.25 = 1.25$

Applying weights to the dataset makes the cases with completed MRA equal to total cases meeting inclusion criteria and resulting in a live birth:

Interval 1 non-priority infant follow-up cases: $32 \times 30.00 = 960$
Interval 2 non-priority infant follow-up cases: $20 \times 22.50 = 450$
All priority infant follow-up cases: $12 \times 1.25 = 15$
Total weighted case count: $960 + 450 + 15 = 1425$

Additionally, weighting would produce an w_{ifu_lab} weight for nonpriority pregnancies depending on the following conditions:

1. Laboratory data obtained with completed infant MRA, $w_{ifu\_lab} = w_{ifu\_analytic}$
   Interval 1: $w_{ifu\_lab\_1} = w_{ifu\_analytic\_1} = 30.00$
   Interval 2: $w_{ifu\_lab\_2} = w_{ifu\_analytic\_2} = 22.50$
2. Laboratory data linked for all infants, $w_{ifu\_lab} = 1$
   Interval 1: $w_{ifu\_lab\_1} = 1$
   Interval 2: $w_{ifu\_lab\_2} = 1$
3. Laboratory data linked for cases with maternal data, $w_{ifu\_lab} = w_{eop}$
   Interval 1: $w_{ifu\_lab\_1} = w_{eop\_1} = 12.50$
   Interval 2: $w_{ifu\_lab\_2} = w_{eop\_2} = 10.00$
4. Laboratory data linked for entire IFU sample, $w_{ifu\_lab} = w_{eop} \times w1_{ifu}$
   Interval 1: $w_{ifu\_lab\_1} = w_{eop\_1} \times w1_{ifu\_1} = 12.50 \times 2.00 = 25.00$
   Interval 2: $w_{ifu\_lab\_2} = w_{eop\_2} \times w1_{ifu\_2} = 10.00 \times 2.00 = 20.00$