Zika Postpartum Emergency Response Survey – Sampling and Methods

A. Sampling Plan

Sampling Scheme for ZPER. For ZPER surveillance, there is a particular interest from a public health perspective in making inferences by geographic region. Some regions do not represent a large portion of a Puerto Rico's overall population. To make inferences about specific subpopulations and make comparisons among several subpopulations, women in those subpopulations (commonly called *strata*) will need to be oversampled.

The main advantage of *stratified sampling* is that for a given overall sample size, stratifying will permit separate estimates of subgroups of interest and permit comparisons across these subgroups. The sampling plan is designed so that inferences about prevalence rates for maternal behaviors and knowledge of Zika can be estimated with sufficient precision both overall in Puerto Rico and within selected strata. For ZPER, 8 geographic regions will serve as the strata. The 8 regions are Arecibo, Aguadilla, Bayamon, Caguas, Fajardo, Mayaguez, Metro, and Ponce. The sampling will be further stratified by hospital, although proportional allocation will be used (each hospital within a region will have the same sampling fraction). Unlike region, hospital is not a subgroup of analysis interest.

Determining Overall Sample Size.   Required sample sizes for the questionnaire are determined in relation to the given proportion that is being estimated, at a given level of precision, and with a given level of statistical confidence. For specified levels of precision and confidence, the sample size required is at its maximum when the advance estimate (the number used in sample size calculations) of the proportion being estimated equals 0.50. ZPER data are used in estimates of proportions of risk factors that range from common (such as delivery paid for by Medicaid) to rare (such as a confirmed Zika diagnosis). Using 0.50 in sample size calculations leads to sufficiently large sample sizes, whatever the true population proportions are for the various risk factors.

The Zika sampling plan is based upon stratified sampling by hospital within region. However, since proportional allocation by hospital is used, the formula for determining sample size for stratified sampling reduces to that used for simple random sampling. Based on the stratification measures found above, a sample size of about 400 (n = 400) is necessary in each stratum to estimate a prevalence for a dichotomous variable with a reasonable precision of 5% and a confidence level of 95%, assuming an infinitely large population size (N). The assumption of an infinitely large population will be violated in the oversampled strata. In any stratum where our desired sample size of 400 comprises more than 5% to 10% of the population, it is appropriate to apply the finite population correction (FPC). The FPC will reduce the desired sample sizes in such cases without compromising the precision of the estimates.

The formula for FPC is:

adjusted size= n / (1 + (n/N)),

where n=desired sample size,

N=population size**.**

Mothers in some hospitals may be more difficult to contact than mothers in others. Thus, actual stratum sample sizes must be larger than theoretically needed to achieve a given level of statistical power.

Based on the estimated stratum-specific response rates, the stratum-specific sample sizes will be inflated to ensure an adequate number of responses for analysis. Based on previous hospital-based surveillance in Puerto Rico and the US-Mexico border, a 90% response rate is assumed across all strata.

Births in Puerto Rico have been steadily declining in recent years. The most recent birth data by hospital available is for 2015. Since births have continued to decline, a sampling rate based on 2015 birth distributions would not achieve the desired sample size. Therefore, it is necessary to account for the declining birth rate. Based on estimated birth data for 2016[1], **Table A.1** describes the drop in birth rates from 2012 to 2016. An adjustment factor of 1.16 will be used to estimate the number of 2016 births in each region.

Table A.1 Overall Births in Puerto Rico, 2012 -2016

Year	Total Births	Percent Decline	Adjustment Factor (2015 births/2016 births)
2012	38,900	---	
2013	36,578	6.0%	
2014	34,485	5.7%	
2015	31,227	9.4%	
2016 (estimated)	28,000	10.3% (est.)	1.16

Steps for Establishing the Sample Rates

1. Establish the distribution of births in Puerto Rico by hospital. Obtain a list of births by hospital and identify within which health region each hospital is located. This list was provided by the Puerto Rico Health Department for 2015 births. Determine which hospitals have a sufficient number of births to support ZPER surveillance.

2. Select the hospitals where data collection will occur. Criteria for hospital selection should be defined. For ZPER, all hospitals with at least 100 births per year will be included. These 36 hospitals account for 98.5% of all Puerto Rican births.

3. Calculate the number of eligible mothers. Mothers giving birth to twins result in two births, but only one eligible mother. The multiple birth rate is approximately 2% of total births. Thus the number of eligible mothers can be estimated as 99% of the total births.

4. Adjust for estimated declines in the birth rates from 2015 to 2016. From table A.1 the adjustment factor is 1.16. Divide by 1.16 to get adjusted eligible mothers.

---

[1] New York Times http://www.nytimes.com/2016/04/15/health/zika-virus-pregnancy-delay-birth-defects-cdc.html

5. Determine the desired number of respondents in the sample. This number will be based in part upon costs and resources, but is often chosen to be 400, as an estimate of a proportion based upon 400 respondents will have a 95% confidence interval of +/- 5%.

6. Compute the Finite Population Correction, if applicable.

7. Estimate the completion rate of hospital-based data collection. Based on previous hospital-based surveillance in Puerto Rico and the US-Mexico border, a 90% response rate is assumed across all strata. Divide the FPC Corrected Sample Size by the estimated response rate to determine the final sample size.

8. Complete Table A.2 using the result of steps 3 through 7 to fill in the appropriate columns.

9. Carry the adjusted population size and estimated adjusted sample size from Table A.2 to Table A.3.

10. Compute the population size for the 3-month surveillance period by dividing the annual population size by 4.

11. Divide the final adjusted sample size by the 3-month expected population to compute the sampling fraction.

12. From the sampling fraction, determine the number of days over the 3 month (91 day) surveillance period during which sampling will be conducted. Note that in two regions, Fajardo and Aguadilla, all women giving live birth during the surveillance period will be included. The operational sample size is the expected sample size based on the number of births actually occurring in each region.

13. Using the number of days during which sampling will be conducted, determine the intervals for sampling. Using the example of 2 out of every 15 days, the two days should be randomly chosen at one hospital. For nearby hospitals, the sampling days can be shifted by one or two days to distribute the workload. Care should be taken to vary the days of the week that sampling occurs throughout the surveillance period. Specific sampling schedules by region will be provided by CDC.

Table A.2 Calculation of ZPER Sample Size

Puerto Rico 2015 Births by Region

Region	Total Number of Live Births	Live Births in Eligible Hospitals	Adjustments for Multiple Births & Declining Birth Rates	Estimated Unadjusted Sample Size	FPC Corrected Sample Size[#] (Respondents)	Estimated Sample Size Adjusted for Nonresponse
Aguadilla	971	971	829	400	270	300
Arecibo	3883	3883	3314	400	357	397
Bayamon	2987	2987	2549	400	346	384
Caguas	4842	4842	4132	400	365	405
Fajardo	715	715	610	400	242	268
Mayaguez	3036	3036	2591	400	347	385
Metro	9534	9534	8137	400	381	424
Ponce	4783	4783	4082	400	364	405
Other	476					
Total	31,227	30,751	26,244	3200	2671	2968

Table A.3 ZPER Sampling Fractions and Estimated Sample Sizes

Region	ZPER Adjusted Population Size (from Table 4.2)	ZPER Estimated Population Size During 3 Month Study Period	Estimated Adjusted Sample Size (from Table 4.2)	$f = n/N$	Operational Sample Size	f in days (number of days to sample out of 91 days)
Aguadilla	829	207	300	1.00	207	91
Arecibo	3314	828	397	0.48	397	43
Bayamon	2549	637	384	0.60	384	55
Caguas	4132	1033	405	0.39	405	35
Fajardo	610	153	268	1.00	153	91
Mayaguez	2591	648	385	0.59	385	54
Metro	8137	2034	424	0.21	424	19
Ponce	4082	1021	405	0.40	405	36
Total	26,244	6561	3003		2760	