Attachment 10

ZPER Sampling Plan

A.1 Definition

The population of interest for the Maternal Hospital-based Survey is all women who delivered a live-born infant in Puerto Rico during the surveillance period. To draw a sample from this population, a sampling frame must be identified. Since ZPER will employ hospital-based data collection, the operational target population is all women delivering a live birth in selected Puerto Rican hospitals during the surveillance period.

In practice, the hospital-based sampling frame is identified by the hospital's delivery log. In most hospitals the delivery log represents a complete and accurate account of all births occurring in the hospital. It will be important to insure that the delivery log in each participating hospital is reliable and kept up-to-date. If this is not the case, alternate procedures to identify women eligible for ZPER must be developed.

ZPER data collection will take place in all hospitals in Puerto Rico with 100 or more births per year. According to 2015 birth data, 36 hospitals meet this eligibility criteria and represent 98.5% of all Puerto Rico births.

In addition to a maternal hospital-based survey, a subset of fathers who are present at the time of the maternal survey will be invited to participate in a father hospital-based survey.

Additionally, a telephone follow-up survey will target a subset of the original female study participants at 2-10 months postpartum. For the telephone follow-up survey, the Puerto Rico Demographic Registry office will match information from ZPER respondents with the birth certificate record of their baby to obtain telephone numbers. The list of original study participants who can be matched to a birth certificate record will serve as the sampling frame.

A.2 Adjustments to the Sampling Frame

Inclusions and Exclusions. Exclusions to the ZPER sampling frame may be implicit or explicit. Because of the definition of the ZPER targeted population and the use of the hospital delivery logs as the sampling frames, certain mothers will *implicitly* be excluded from eligibility in the ZPER sample. An implicit exclusion is any restriction inferred by the definition of the targeted population (i.e., women who deliver at home or at hospitals with fewer than 100 births per year) or the choice of the hospital delivery log as the ZPER sampling frame (i.e., stillbirths and fetal deaths). All other exclusions arise from concerns or operational difficulties in sampling certain types of births, and are termed *explicit* exclusions.

i. Stillbirths and Fetal Deaths. By definition, the targeted population of ZPER is limited to pregnancies resulting in a live-born infant.

ii. Multiple Gestation Pregnancies. Mothers with a multiple gestation regardless of the order will be included in the sample if at least one infant is delivered in the sampling window. Since the survey questions are about the mother and her pregnancy and not specifically about the baby, the mother will complete one survey for this pregnancy, regardless of how many babies were delivered.

iii. Mothers Discharged Early From the Hospital or Otherwise Missed in the Hospital. For hospital-based supplementation, mothers discharged early from the hospital, transferred to another hospital, or otherwise missed in the hospital are excluded from follow-up.

iv. Deceased Infants. Mothers of deceased infants will be excluded from the sample. Based on 2015 neonatal mortality rates in Puerto Rico, approximately 26 infant deaths will occur during the study period. This number is too small to have any impact on the sample size calculations.

v. Mothers with Pregnancy Complications. Mothers who are ill or suffering from complications of pregnancy and delivery will not be excluded. Contact with these women may need to be delayed until their condition has improved. However, they should be approached before they are discharged to complete the survey.

vi. Mothers with Zika-affected or special needs infants. Mothers whose infants have microcephaly or other birth defects or special health conditions will not be excluded. If they are contacted, the interviewer should be aware of extra sensitivity of questions related to the baby's health.

A.3 Sampling Plan

A.3a 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.

A.3b 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:

$$\text{adjusted size} = n / (1 + (n/N)),$$

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

All fathers of an infant whose mother was sampled are eligible to participate in the hospital-based survey of fathers. However, given that fathers may not be present in the hospital at the exact time that the mother is interviewed, it is anticipated that roughly 50% of fathers are expected to be available to participate in the survey. No additional measures will be taken by the interviewers to pursue the father's in the event that they are not present at the time of the maternal interview.

The telephone follow-up survey sampling plan is based upon proportional sampling by region, among the original participants in the hospital based-survey. However, since proportional allocation by region is used, the formula for determining sample size reduces to that used for simple random sampling. A sample size of about 1068 (n = 1068) is necessary to estimate a prevalence for a dichotomous variable with a reasonable precision of 3% and a confidence level of 95%, assuming an infinitely large population size (N). The assumption of an infinitely large population will be violated since our population size is actually 7029. Since our desired sample size of 1068 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.

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

A.3c Steps for Establishing the Sample Rates

Hospital-based Survey

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.

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.

14. Fathers are eligible if the mother of their baby is sampled and they are in the hospital at the time of the maternal interview. There is not a separate sampling schedule for the father survey.

Telephone follow-up

1. Identify eligible mothers. Overall, between August 28 and November 26, 2016 there were 7029 resident Puerto Rico mothers who had a live birth. Of these, 3237 mothers gave birth on a ZPER sampling day. Of these, 2364 completed the ZPER survey.

2. Link mothers with birth certificates. ZPER participants must be matched to their infant's birth certificate in order to collect current contact information. The Puerto Rico vital registry office is performing the linkage based on 5 fields, mother's name, mother's date of birth, baby's date of birth, hospital of delivery, and method of delivery. Of all ZPER participants, only 8 were unable to be matched, a 99.7% matching rate. These ZPER participants who are successfully matched to their infant's birth certificate record form the sampling frame for the ZPER follow-up survey.

3. Determine the desired number of respondents in the sample. This number is based upon the desired precision level for study estimates. For the follow-up, a precision level of +/- 3% is desired. An estimate of a proportion based upon 1068 respondents will have a 95% confidence interval of +/- 3%.

4. Compute the Finite Population Correction (FPC), if applicable.

5. Estimate the participation rate of ZPER telephone follow-up data collection. Based on previous hospital-based surveillance in Puerto Rico that included a telephone follow-up component, a 60% response rate is expected.

6. Compute final adjusted sample size. Divide the FPC corrected sample size by the estimated response rate to determine the final sample size.

7. Based on the distribution of all births during the ZPER study period by region, proportionally allocate the total sample among the 8 regions.

8. Complete Table A.1 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.1 to Table A.2.

10. Divide the final adjusted sample size by the number of ZPER participants to compute the sampling fraction for each region.

11. From the sampling fraction, determine the operational sample size. The operational sample size, 1437, is the expected sample size based on the number of ZPER participants in each region. Note that in the Metro region, all ZPER participants linked with a birth certificate will be included.

Table A.2 Calculation of Sample Size – Hospital-based Survey

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 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	

All fathers of the infants of sampled women are eligible to participate if they are present at the time of the maternal interview. Because it is estimated that only 50% of eligible fathers will be present at the time of the maternal interview, the operational sample size for the father survey is 2760 x 0.50, or 1,380.

Table B.1 Calculation of Telephone Follow-up Sample Size

*Numbers adjusted for multiple births and non-residents of Puerto Rico.

Region	All Live Births* (8/28 – 11/26, 2016)	Proportion of Births	ZPER Participants (8/28 – 11/26, 2016)	Adjustments for Unmatched Records	Estimated Unadjusted Sample Size for ZPER Follow-up Study	FPC Adjusted Sample Size#	Estimated Sample Size Adjusted for Nonresponse
Aguadilla	188	2.7%	124	124	29	25	42
Arecibo	979	13.9%	390	390	149	129	215
Bayamon	682	9.7%	343	340	104	90	150
Caguas	1,126	16.0%	378	377	171	149	248
Fajardo	133	1.9%	121	121	20	18	29
Mayaguez	682	9.7%	348	347	104	90	149
Metro	2,199	31.3%	373	372	290	290	483
Ponce	1,040	14.8%	274	274	158	137	229
Total	7,029	100%	2,351	2,345	1,068	927	1,545

Table B.2 ZPER Sampling Fractions and Estimated Sample Sizes

Region	**ZPER Participants Adjusted for Matching (from Table 4.1) N**	**Estimated Adjusted Sample Size (from Table 4.1) n**	**Sampling Fraction f = n/N**	**Operational Sample Size**
Aguadilla	124	42	0.32	42
Arecibo	390	215	0.56	215
Bayamon	340	150	0.44	150
Caguas	377	248	0.67	248
Fajardo	121	29	0.24	29
Mayaguez	347	149	0.44	149
Metro	372	483	1.30	372
Ponce	274	229	0.84	229
Total	2,345	1,545		1,434

A.4 Selection of Sample

Proportional sampling within each stratum is used for drawing the sampling schedule based on the time of birth. The time and date of birth is written on the hospital delivery log. All births for a particular region that fall within the pre-established sampling time intervals for that region are selected for the study provided they do not satisfy an exclusion criterion. Where possible, sampling intervals will consist of complete days (midnight to midnight) for ease of selection. However, there may be some regions where the sampling interval will include partial days. The sampling schedule is designed to be balanced by weeks in the surveillance period and days within each week.

Based on the desired sample size and the number of live births occurring at each hospital for each region, sampling fractions can be computed. The length of the sampling interval will vary by region and is determined from the sampling fractions. For a multiple birth, the mother is selected only once. The selection procedures must satisfy the probability requirements of the sample. The sample is chosen so that, within each region, each record has an equal probability of being selected. Based on these probabilities, weights can be determined for island-wide estimates.

Because of shorter hospital delivery stays and earlier discharges of mothers, the hospital delivery log must be frequently monitored during defined sampling intervals. No more than 24 hours should lapse between the beginning of the sampling interval and when the delivery log is checked. Similarly no more than 24 hours should lapse between checks of the delivery log within a sampling interval. For regions with 100% sampling, birth logs should be checked a minimum of every other day.

For the telephone survey, proportional sampling is used for drawing the sample. All matched birth certificates will be first separated by region. Within each region the birth certificates will be sorted by hospital and date of birth within hospital. The sampling rates for each region will be applied to randomly select the birth certificates for the follow-up. For the Metro region, all matched births will be included in the sample. For the other regions, the sample is chosen so that, within each region, each record has an equal probability of being selected. Based on these probabilities, weights can be determined for island-wide estimates.