

Communities Organized to Prevent Arboviruses:
Assessment of Knowledge, Attitudes, and Vector Control Practices and
Sero-Prevalence and Incidence of Arboviral Infection in Ponce, Puerto
Rico (COPA Study)

Request for a Reinstatement (0920-1254)

Supporting Statement B

Program Contact

Thomas Daymude
National Center for Emerging and Zoonotic Infectious Diseases
Centers for Disease Control and Prevention
1600 Clifton Road, NE
Atlanta, Georgia 30333
Phone: (470) 553-3567
Email: qkh7@cdc.gov

Table of Contents

1. Respondent Universe and Sampling Methods.....	3
2. Procedures for the Collection of Information.....	7
3. Methods to maximize Response Rates and Deal with No Response.....	8
4. Tests of Procedures or Methods to be undertaken.....	9
5. Individuals Consulted on Statistical Aspects and Individuals Collecting and/or Analyzing Data.....	9

The collection involves statistical methods, and the results of this study will be of great relevance to provide evidence-based information regarding the burden of arboviral infections, perceptions related to arboviral disease and prevention, feasibility of a community-based vector control strategy and its effectiveness to decrease mosquito populations and vector borne disease incidence. The study will also be used as a platform to investigate emerging health threats in the region, including novel coronavirus 2019 (COVID-19). The information collected will be used to inform public health officials on the implementation and evaluation of mosquito control interventions and other efforts to reduce the burden of arboviruses and other emerging diseases.

One of the goals of the study is to impact the *Ae. aegypti* population sufficiently to prevent viral transmission to humans and reduce morbidity and mortality associated with vector-borne diseases. Data quality and monitoring reports will be generated monthly, and a project status report will be generated annually. After completion of the baseline in year 2, we will report on risk factors associated with arbovirus incidence and prevalence, attitudes towards traditional and novel vector control strategies, community attitudes and practices with regards to personal protection methods.

Prevalence rates (as indicated by a positive IgG result to dengue, Zika or chikungunya viruses) will be calculated for each cluster. With the resulting data, for a future vector control intervention evaluation, clusters will be paired based on prevalence rates and movement frequency; among each pair, with intervention and control status will be randomly assigned. Annually, the incidence rate in each cluster will be assessed through arboviral disease testing. At the end of the 3-year follow up after implementation of the intervention, comparisons will be made between the intervention and control clusters using paired t-tests to assess any difference between groups.

1. Respondent Universe and Sampling Methods

This project will include annual collection of behavioral data and biological specimens. The sero-surveys will be repeated every year to determine prevalence and incidence of arboviral infection. We estimate the true annual incidence of dengue is 3% and of Zika is about 1%.

Project success will be gauged by completion of baseline recruitment goal of 3,800 participants with 50% response rate during baseline recruitment and follow-up year replacement activities and 85% retention of participants between annual follow-ups.

Power Calculation

The project design aims to measure a difference of 50% between intervention and control communities over the course of the project period. Using simulations, we evaluated the power of several statistical methods (paired and unpaired t-tests and permutation testing) with varying incidence rates, cluster sizes, and intra-class correlation levels. The findings from the simulations indicated that even with low incidence, an intervention impact could be identified over a 2–3-year period with between 36–44 clusters in a 2-arm project. The selected communities will be randomly assigned in a 1:1 ratio as intervention and non-intervention.

Cluster Identification

We evaluated areas within the municipality of Ponce, Puerto Rico to identify potential project clusters. Areas were prioritized with high arboviral disease incidence rates, as identified through surveillance data from dengue cases during 2009–2013, as well as from Zika cases during 2016. A buffer area of

300mts was established between adjacent clusters with different intervention assignments to prevent the likelihood of intervention effects affecting another cluster. We attempted to create cluster areas based on existing community borders and natural divisions, such as rivers or major roads, to create eco-zones (areas where the likelihood of mosquito cross-contamination is low because of natural divisions). Areas perceived as unsafe for fieldwork by local collaborators were excluded from selection. The baseline study included 14 clusters. These clusters have been sub-divided into 38 to increase statistical power. Additional clusters may be added based on available resources. Covariate constrained randomization will be considered based on the findings from the baseline assessments to inform matching of the clusters. Matching variables will include those that may be potential confounders such as arbovirus seroprevalence, % population under 15 years of age, mosquito densities and human mobility patterns. Seroprevalence and incidence are not representative of the Ponce area. This information will not be generalized to Puerto Rico or Ponce. However, it will be used to identify risk factors for arboviral infection, and factors associated with higher levels of support for specific vector control strategies. Knowledge, attitudes and practices will also be used to inform future educational campaigns.

Cluster Size

We aim to recruit 3,800 participants. If participants are unavailable or decline participation in future years, replacement participants will be selected to maintain 3,800 participants. Using data from previous community surveys and census data in Puerto Rico, we estimate that we will recruit 1.5 participants per household, and that we will be able to recruit approximately 25% of homes visited due to refusals (15–20%), vacant homes (20%), and inability to contact residents (30–40%).

Household Selection & Replacement Strategy

Using housing structures as identified by ArcGIS, we will randomly select an initial list of double the number of homes needed to sample per cluster and evaluate the number of participants recruited through these homes. A replacement list will be generated by randomly selecting new structures not initially selected from all homes in the cluster. Replacement lists will be generated in this manner as needed until teams' complete interviews to maintain 3,800 participants.

Team Assignments and Data Collection Tools

Structures will be assigned to teams for recruitment. The selected structure ID numbers and locations will be loaded on to each tablet through an electronic household tracking tool, structure IDs will also be pre-loaded into REDCap (REDCap is a secure web application for building and managing online surveys and databases) for consent and interview information. Replacement structure IDs will be loaded on the tablets as needed, depending on the number of homes successfully recruited. Information about homes visited, structure status (vacant, inhabited, or not a home), refusals, and visit scheduling will be captured within an electronic household tracking tool app and uploaded daily to the encrypted file transfer protocol (eftp). REDcap will be used to capture eligibility and consent information, questionnaires, and specimen information. The information will be uploaded to the eftp daily. Paper laboratory forms will also be completed for each specimen to be delivered to the laboratory.

Analysis Plan

Progress reports are generated weekly, monthly, and annually to help monitor recruitment and data quality. Data quality and monitoring reports will be generated monthly, and a project status report will be generated annually. After the first year of data collection, prevalence rates (as indicated by a positive IgG result for dengue, Zika, or chikungunya viruses) were calculated for each cluster. With the resulting data, for a vector control intervention evaluation, clusters were paired based on prevalence rates and movement frequency; among each pair, intervention and control status were randomly assigned. Annually, the incidence rate in each cluster is assessed through arboviral disease testing. Using baseline and annual follow up data, we will report on risk factors associated with arbovirus incidence and prevalence, attitudes towards traditional and novel vector control strategies, community attitudes and practices with regards to personal protection methods and risk factors for and attitudes towards other emerging health threats (i.e., COVID-19). Following each year of the intervention implementation, comparisons will be made between the intervention and control clusters using paired t-tests to assess any difference between groups.

Sample size calculation

Overview

The objective of this community trial is to determine whether the community intervention, Wolbachia suppression, leads to a reduction in arboviral disease incidence. A community intervention, necessarily, is introduced to a contiguous geographic area or cluster of households. Therefore, a cluster randomized design will be employed, and treatment (control or intervention) will be randomly assigned to clusters within pairs in a matched design.

Power Calculation Considerations

Number of clusters and participants

Study resources and logistics will allow for the enrollment of about 3800 participants. The number enrolled will depend upon the number of clusters defined. As we increase the number of clusters, parts of the study region became ineligible. This is because some interventions, such as mosquito control, require a “buffer zone” so that distinct clusters receive only the treatment(s) they are designed to receive. We estimate our total sample size will be between 3500 and 4000 participants. Results for scenarios (cluster numbers and sizes) that have been explored and deemed feasible are included in this document.

Assumptions about incidence

Estimates

Arboviral disease incidence varies depending upon the disease. Baseline data have been collected in the study area and are shown in Table 1. Overall, 762 of 4550 (0.167) people tested positive for at least one of the three viruses at baseline. This data was collected after the introduction of Zika virus on the island and is more consistent with epidemic levels than endemic levels. Thus, we used estimates from 0.01 to 0.05 infections per person per year, which are more consistent with historical levels.

Intervention effect

The intervention will be considered successful if incidence is cut in half.

Computing multi-year incidence

If incidence in the control group is p per person per year, the probability that an individual will become infected over t years is $1 - (1 - p)^t$. For example, if incidence is 0.03 infections per person per year, the

probability that an individual will become infected during a 2-year period is $1 - 0.97^2 \approx 0.06$. Control and intervention clusters will be followed for at least two years following the full implementation of the intervention.

Table 1: Baseline IgM positivity data

	N	N IgM Positive	Proportion IgM Positive
Any arbovirus	4550	762	0.167
DENV	4550	19	0.004
ZIKV	4550	730	0.160
CHIKV	4550	32	0.007

Assumptions about other parameters

ICC

To comparing incidence rates or proportions, we use a method by Donner and Klar (2000) extended to the matched cluster case, described below. To compute the power using this method, knowledge of the intra-cluster correlation (ICC) is required. The ICC is defined as the proportion of total variation due to variation between clusters. Mathematically,

$$ICC = \frac{\sigma_B^2}{\sigma_B^2 + \sigma_W^2},$$

where σ_B^2 is the variance of incidence between clusters and σ_W^2 is the variance of incidence within clusters.

Suggested values for ICC can be found in literature. Campbell (2000) found that ICC was typically around 0.05 in primary care trials, while Donner and Klar (2000) note that in community randomized trials ICC was usually less than 0.01 and often near 0.001. Drawing from several field trials, Hayes and Bennett (1999) suggest ICC is often less than 0.25 and rarely exceeds 0.5 for health outcomes. Andersson et al. (2015) computed ICCs ranging from 0.03 to 0.08 for primary outcome variables in a community intervention study.

Blood samples from individuals in 12 clusters in the selected communities have been taken and are in the process of being tested for Zika virus IgM, dengue virus IgM, and chikungunya virus IgM. ICCs estimated from data thus far are of 0.02 (95% CI of 0.01 to 0.07), 0.003 (95% CI of 0.001 to 0.007), and 0.001 (95% CI of <0.001 to 0.01), for ZIKV, DENV, and CHIKV, respectively.

The largest upper bound on ICC estimated from baseline data is 0.07; an ICC of 0.08 was observed in Anderson et al. (2015). In our computations, we use ICCs that range from 0.01 to 0.1, though we expect that ICC in our community intervention study will be, at most, 0.07 and more likely in the 0.01 to 0.05 range.

Correlation induced by matching

We do not have any information to inform an estimate of the correlation induced by matching. Throughout, we assume an estimate of 0.2.

Method for Computing Power

The method we use to estimate power evaluates the difference of proportions in a matched cluster-pair design and is an extension of Donner and Klar (2000). Clusters are matched based on a characteristic (community prevalence of arbovirus infection) thought to be associated with expected infection incidence. Randomization occurs within the matched cluster pair and differences in incidence are computed within pairs and averaged. The variance of the mean difference is adjusted to account for ICC and for within-pair correlation. This variance is then used in a paired t-test. All hypothesis tests have an assumed Type I error of $\alpha=0.05$.

Results

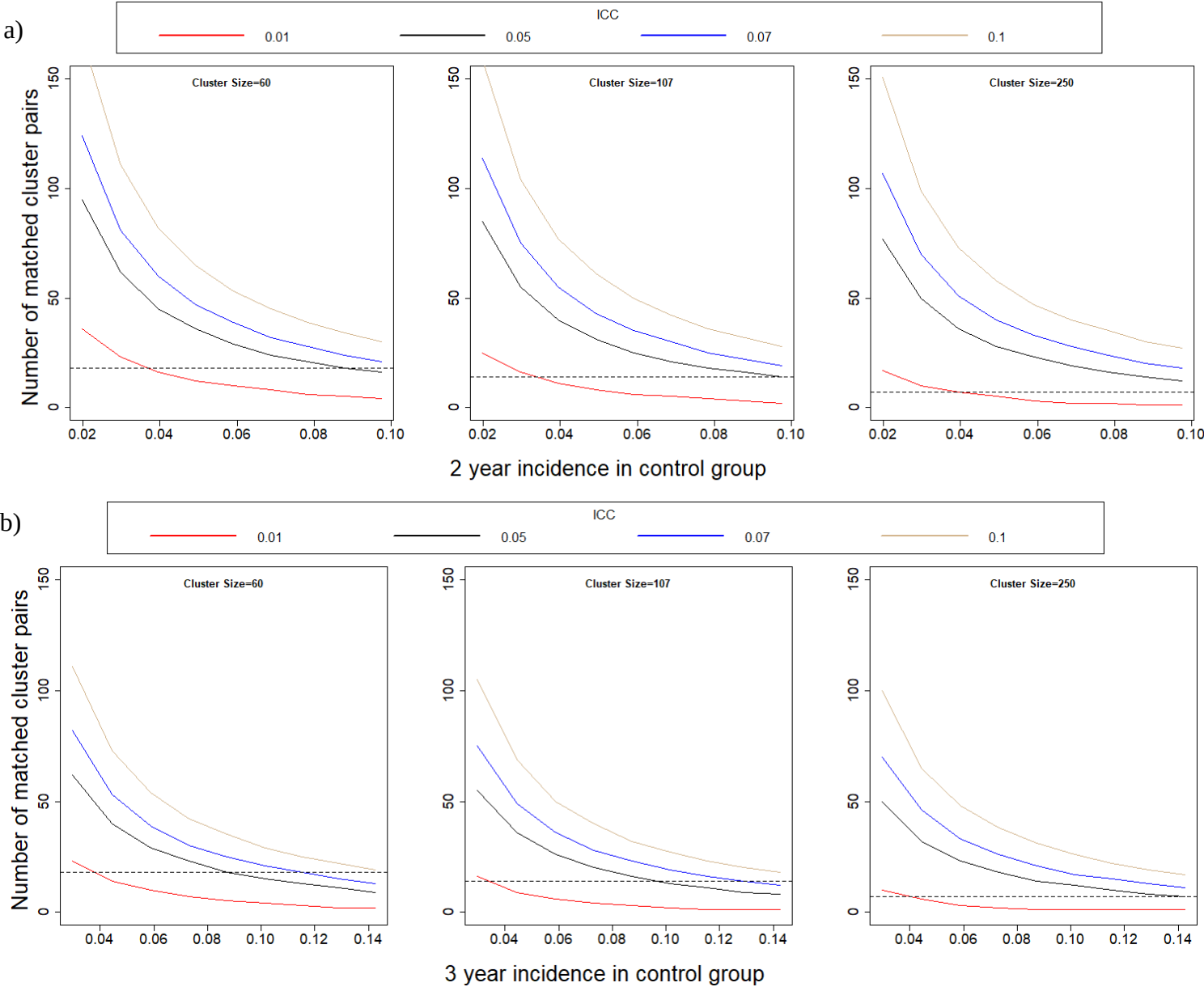
We have been able to define different cluster configurations and present the results for three of them:

- 7 pairs of matched clusters; 250 individuals per cluster (N=3500)
- 14 pairs of matched clusters; 107 individuals per cluster (N=2996)
- 18 pairs of matched clusters; 60 individuals per cluster (N=2160)

Figures 1 and 2 show the number of matched clusters required to achieve a power of 0.80 to detect a halving of the two-year and three-year incidence, respectively, for different values incidence and ICC.

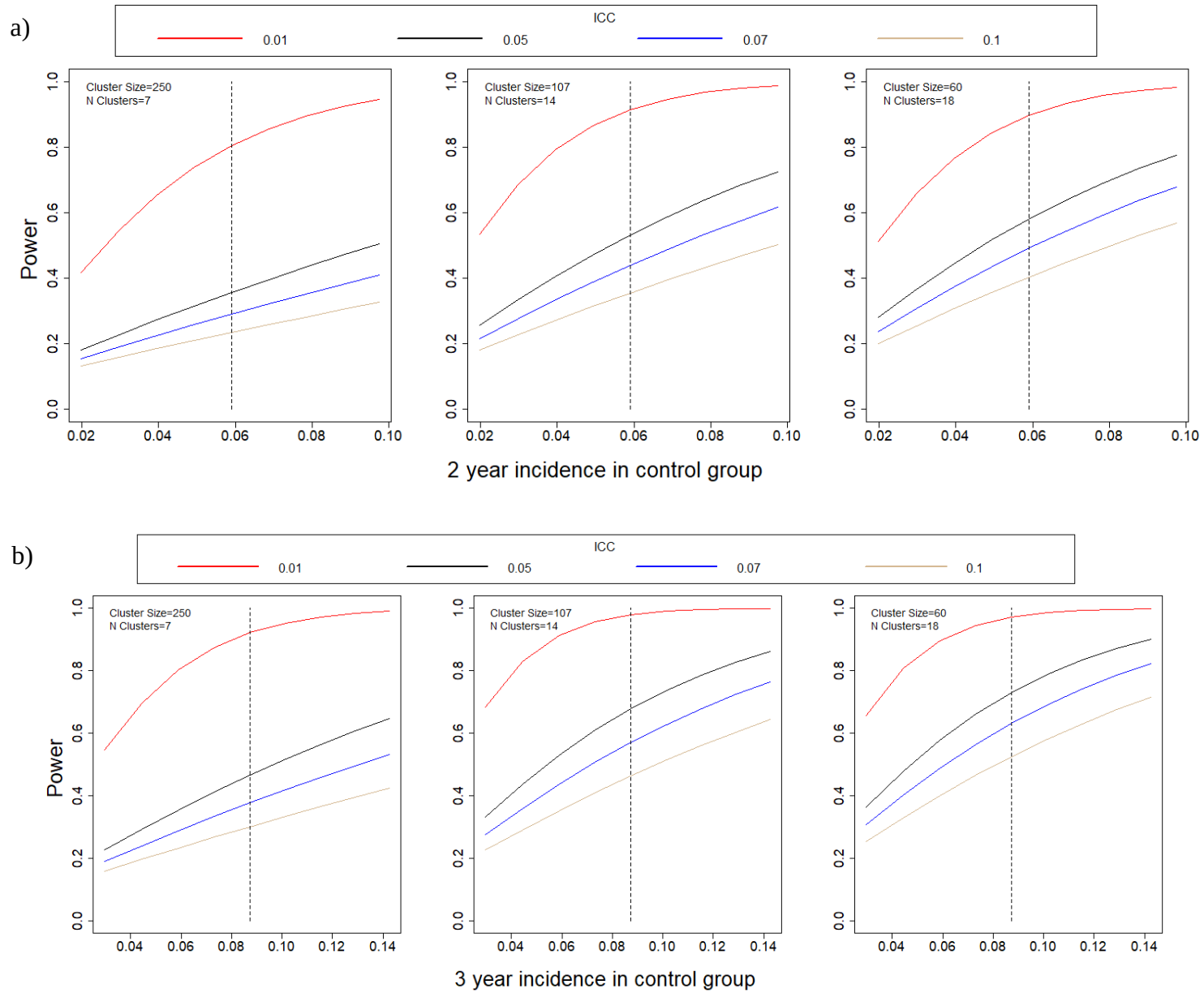
Figures 3 and 4 show the power to detect a halving of the two-year and three-year incidence, respectively, for different values of incidence and ICC.

Figure 1: Number of clusters needed to detect, with 80% power, a halving of the 2-year (a) and 3-year (b) incidence when compared to the control group. The dashed horizontal line corresponds to the number of matched clusters we can create, given the cluster size.



:

Figure 2: Power to detect a halving of the 2-year (a) and 3-year (b) incidence when compared to the control group. The dashed line is the expected 2-year incidence, assuming no outbreak.



It is clear that ICC drives the power estimates. If ICC is close to the initially assumed 0.01, power for the comparison of incidences is well over 0.80. If it is closer to 0.07, as in the Andersson study, power decreases dramatically, and the study, as designed will be underpowered.

Estimates of ICC using existing baseline data varied but were all 0.02 or less. If ICC is 0.01, results indicate the power to detect the desired reduction in proportion infected over a 2-year period is 0.78 with our initial design of 7 clusters per arm. This increases to 0.90 for 14 clusters per arm (107 per cluster) and to 0.88 for 18 clusters per arm (60 per cluster). The baseline study included 14 clusters. Based on power calculations, these clusters have been sub-divided and additional clusters added to reach a total of 38 clusters.

2. Procedures for the Collection of Information

Housing structures will be assigned to distinct teams for recruitment. The selected structure ID numbers and locations will be loaded on to each tablet through an electronic household tracking tool app; structure IDs will also be pre-loaded into REDCap for consent and interview information. Replacement structure IDs will be loaded on the tablets as needed, depending on the number of homes successfully recruited. Information about homes visited, structure status (vacant, inhabited, or not a home), refusals, and visit scheduling will be captured within an electronic household tracking tool app and uploaded weekly to a secure network drive. REDCap will be used to capture eligibility and consent information, questionnaires, and specimen information. The information will be uploaded to the secure eftp. Paper lab forms will also be completed for each specimen to be delivered to the lab.

Being a resident is defined by having slept in the house for at least four of the past seven nights. The questionnaire section will vary depending on age of each participant. A questionnaire with general household questions will be administered to one household representative in each home with one or more participants. This representative should be 21 years or older or an emancipated minor. If all eligible household members are non-emancipated minors, a household member over the age of 50 may act as household representative and complete this section of the survey only. Non-emancipated minors that live without their parent/guardian or any adult may answer the questionnaire with household questions if their parent/guardian has consented to their participation. The individual and mobility questionnaire will be administered to all participants. The assessment of knowledge, attitudes, and practices questionnaire will be administered to all participants 14 years of age and older. A vector control intervention questionnaire will be administered to the household representative. The knowledge, attitudes, and practices questionnaire will be focused on vector control, healthcare-seeking behavior, and disease occurrence. We will collect demographic information (e.g., age, sex, duration of time residing in cluster) and information on recent illnesses from all participants via the individual questionnaire. Parents or guardians will serve as proxy respondents for children age <7 years. The questionnaires will be administered after written consent and verbal assent (when appropriate) from those present in the household at the time of the visit.

We will ask participants if they have been ill with arbovirus and COVID-19 like illness (i.e., fever, rash, joint pain, and conjunctivitis) in the past year. If so, we will collect details on the symptoms experienced during their illness. An acute illness surveillance (AIS) project component is being implemented to better identify and assess the incidence of incident arboviral disease and COVID-19 among COPA participants. This additional weekly activity will use an automated text-messaging system to ask COPA household representatives and other household adults who consent to receive text messages if any COPA participants in the household have experienced fever or other COVID-like symptoms in the past 7 days. Project staff will contact households in which one or more participants reported symptoms to schedule an appointment to collect samples for arbovirus and SARS-CoV-2 molecular testing and to administer a AIS questionnaire about symptoms, exposure and health seeking behaviors. From previous febrile surveillance studies, we expect approximately 40% of household adults will respond to text messages each week and 10% of COPA participants will report acute symptoms and agree to a sample collection visit each year.

Participants with a positive SARS-CoV-2 molecular test will be contacted by phone 2-4 weeks later for a COVID-19 case follow-up questionnaire on symptoms, health care seeking, potential exposures, and outcomes of SARS-CoV-2 infection. We are expecting that 20% of participant that report symptoms will have a positive COVID-19 result and respond to this follow-up questionnaire. The questionnaires will be administered to eligible residents of selected households. At the time of the questionnaire administration, ~15 mL of blood will be collected to conduct serological testing of arboviruses for a sero-survey and if they report COVID-19 like symptoms a nasal sample will be collected.

From year two follow up visits and on, a survey will include the topics in year one, co-morbidities, mental health, and alcohol use and a brief mobility assessment. The mobility assessment will be conducted to determine the main places visited by each participant in the past week. Participants will be prompted to recall the locations visited between 6am and 8pm each day, and the time spent at each location, using a tablet-based data collection tool that includes maps. These data will be used to determine the proportion of time spent in intervention and non-intervention communities and will help explain the level of protection by the intervention.

GPS coordinates will also be collected for each household visited to later assess for potential clustering of arboviral infections within communities.

The sero-survey, individual, and vector control intervention questionnaires will be repeated every 12 months after the initial assessment, up to a period of 5 years. OMB clearance will be requested for three years, and amendments submitted as appropriate. Questionnaire data will be directly entered into REDcap. In cases where data collection using electronic devices is not possible, the data will be collected on paper. When paper forms are used, they will be entered into the database either daily or as a group at the close of data collection; 10% of entered forms will be re-checked to identify any problems with data entry accuracy that must be addressed.

Data will be stored at Ponce Health Sciences University and CDC's Dengue Branch; only members of the study will have access to recordings, written notes and transcripts. Participants' names, addresses and telephone numbers will be collected in case we need to contact them later in the study. This

information will be kept secure in password protected computers and locked cabinets. Based on human subject requirements, paper records will be kept for the duration of the study and at least three years after that. After this period, they will be archived or destroyed according to federal records management guidelines.

3. Methods to maximize Response Rates and Deal with No Response

Follow up participants will be called on 3 occasions during different times and days to maximize the chance of contacting them. For participant replacement purposes houses will be visited three times (one of these times being a Saturday), unless the resident refuses participation.

Participants will be provided with a token of appreciation of 20 dollars. The blood draw is required for participation. Participants are expected to complete the survey but can refuse to answer any question they do not wish to answer. For participants under 7 years of age, the parent will receive the full token of appreciation.

4. Tests of Procedures or Methods to be undertaken

Cognitive testing of the questionnaires is conducted annually with communications staff and local residents of the region.

5. Individuals Consulted on Statistical Aspects and Individuals Collecting and/or Analyzing Data

Name	Affiliation	Email	Phone
Vanessa Rivera-Amill	PHSU/PRI	vrivera@psm.edu	787-840-2575
Mariely Linares	PHSU	mlinares@psm.edu	787-221-7460
José Molina	PHSU	jmolina@psm.edu	939-350-6393
Gladys González	PHSU	glgonzalez@psm.edu	787-221-7460
Lissette Rodriguez	PHSU	lirodriguez@psm.edu	787-840-2575
Jurinet Negron	PHSU	jnegron@pesm.edu	787-840-2575
Ana Perez	PHSU	aperez@psm.edu	787-840-2575
Emma Ramirez	PHSU	eramirez@psm.edu	787-840-2575

Julia Herrera	PHSU	jherrera@psm.edu	787-840-2575
Lismarie Santiago	PHSU	lisantiago@psm.edu	787-840-2575
Suheil Albizu	PHSU	salbizu@psm.edu	787-840-2575
Tatiana Morales	PHSU	tmorales@psm.edu	787-840-2575
William Ramirez	PHSU	wiramirez@psm.edu	787-840-2575
Veronica Rodriguez	PHSU	vrodiguez@psm.edu	787-840-2575
Marjorie Martinez	PHSU	marjmartinez@psm.edu	787-840-2575
William Gonzalez	PHSU	wigonzalez@psm.edu	787-840-2575
Aimara Medina	PHSU	aimedina@psm.edu	787-840-2575
Nicole Leon	PHSU	nleon@psm.edu	787-840-2575
Carolina Torres	Contractor/CDC	qsc4@cdc.gov	787-706-2399
Wilmarie Rivera	Contractor/CDC	qjf7@cdc.gov	787-706-2399
Brenda Torres	Contractor/CDC	xfu1@cdc.gov	787-706-2399
Ronald Cancel	Contractor/CDC	ngt5@cdc.gov	787-706-2399
Cindia Diaz	Contractor/CDC	prp8@cdc.gov	787-706-2399
Jose Ruiz	Contractor/CDC	pmn3@cdc.gov	787-706-2399
Eduardo Rivera	Contractor/CDC	itl0@cdc.gov	787-706-2399
Coral Rosado	Contractor/CDC	nsj3@cdc.gov	787-706-2399
Stella Lebron	Contractor/CDC	smx5@cdc.gov	787-706-2399
Nicole Medina	Contractor/CDC	moi6@cdc.gov	787-749-6124
Eli Rosario	Contractor/CDC	nvg9@cdc.gov	787-532-4603
Olga Lorenzi	CDC	oal9@cdc.gov	787-706-2399
Gabriela Paz Bailey	CDC	gmb5@cdc.gov	787-706-2399
Liliana Sanchez	CDC	naq5@cdc.gov	787-706-2399
Kyle Ryff	CDC	xak2@cdc.gov	787-706-2399

Laura Adams	CDC	ipb2@cdc.gov	787-706-2399
Dania Rodríguez	CDC	ivn0@cdc.gov	787-706-2248
Chelsea Major	CDC	lhi5@cdc.gov	787-706-2399
Claudia Colon	CDC	ouh3@cdc.gov	787-706-2399
Janice Perez	CDC	jpq9@cdc.gov	787-706-2399
Aidsa Rivera	CDC	erj2@cdc.gov	787-706-2399
Mark Delorey	CDC	esy7@cdc.gov	787-706-2399
Oscar Padro	CDC	fmx3@cdc.gov	787-706-2399
Eli Rosenberg	U. of Albany	erosenberg2@albany.edu	518-486-9667
Marijulie Martinez	ORISE Fellow/CDC	ssv3@cdc.gov	787-706-2399
Nicole Perez	ORISE Fellow/CDC	xhp4@cdc.gov	787-706-2399