**Communities Organized To Prevent Arboviruses:**
**An evaluation of an intervention in Ponce, Puerto Rico (COPA)**

# Request for OMB approval of an Ongoing Information Collection in Use without an OMB Number

**March 25, 2019**

**Supporting Statement B**

**Contact:**
Lee Samuel
National Center for Emerging and Zoonotic Infectious Diseases
Centers for Disease Control and Prevention
1600 Clifton Road, NE
Atlanta, Georgia 30333
Phone: (404) 718-1616
Email: llj3@cdc.gov

# Table of Contents

The collection involves statistical methods and the results of this study will be of great relevance to provide evidence-based information regarding the feasibility of a community based vector control strategy and its effectiveness to decrease mosquito populations and vector borne disease incidence. Results and analysis will be used to update and refine relevant recommendations regarding vector control strategies for Puerto Rico.

The goal of the program is to impact the Ae. aegypti population sufficiently to prevent virus 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 4,000 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 each cluster to prevent the likelihood of intervention or non-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 40 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 4,000 participants. If participants are unavailable or decline participation in future years, replacement participants will be selected to maintain 4,000 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 4,000 participants.

*Team Assignments and Data Collection Tools*

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, structure IDs will also be pre-loaded into EpiInfo (EpiInfo is a CDC application that allows data entry using electronic devices) 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). Epi Info will be used to capture eligibility and consent information, household questionnaires, individual 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 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. 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, clusters will be paired based on prevalence rates and human 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.

*Sample size calculation*

**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, 466 of 2788 (0.17) people tested positive for at least one of the three viruses. This incidence, however, is more consistent with epidemic levels than endemic levels. We use 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 a 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 |
|---|---|---|---|
| DENV | 2528 | 12 | <0.01 |
| ZIKV | 2638 | 452 | 0.17 |
| CHIKV | 2755 | 14 | 0.01 |

**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 $\sigma_B^2$ is the variance of incidence between clusters and $\sigma_W^2$ 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 CV 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 positivity. 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.

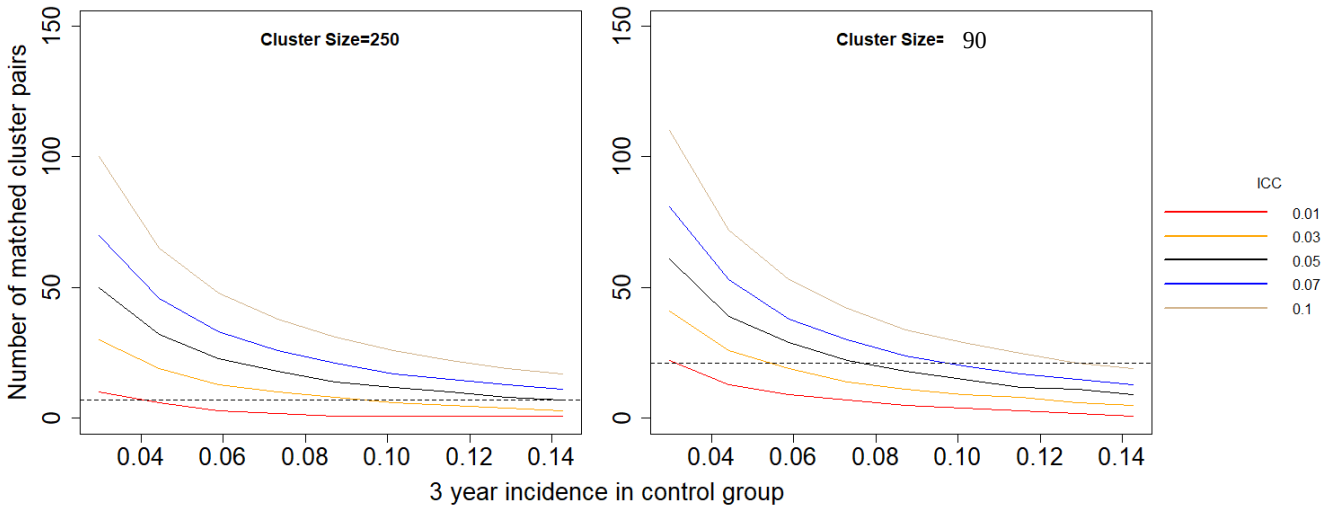
**Method for Computing Power**

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 pair 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.

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

- 7 pairs of matched clusters; 250 individuals per cluster (N=3500)
- 21 pairs of matched clusters; 90 individuals per cluster (N~4,000)

Figures 1. 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.




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 40 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 EpiInfo 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. Epi Info will be used to capture eligibility and consent information, household questionnaires, individual 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 and day of birth 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. A questionnaire on socio-demographic information will be administered to all participants. The assessment of knowledge, attitudes, and practices questionnaire will be administered to all participants 7 years and older with question adapted for ages: 7-11 (younger child), 12-13 (older child), 14-50 (adult). A vector control intervention questionnaire will be administered to all participants 21 years or older born on an odd numbered day of the month. 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 household (and individual) questionnaires.

Parents or guardians will serve as proxy respondents for children aged <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-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. The questionnaires will be administered to all 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.

At the year two follow up visit, a survey will be administered including the themes in year one, co-morbidities, mental health, drug 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 EpiInfo. 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 these period, they will be archived or destroyed according to federal records management guidelines.

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

Houses will be visited three times (one of these times being a Saturday), unless the resident refuses participation. A flyer (Attachment 8) with an overview of the project and a phone number that can be used to schedule an appointment at a more convenient time will be left in the mailbox, when no one is home.

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 questionnaire was conducted with communications staff and local residents of the region. While PHSU was leading this study in 2018, the procedures were piloted with 69 households and 128 participants in the neighborhoods of El Tuque and Pastillo. As a result, changes were introduced to the study forms such as shortening the questionnaire, adding reasons for refusal to the tracking form, retraining on labeling of lab specimens.

## 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 |
| Luisa I. Alvarado | PHSU/SLEH | lalvarado@psm.edu | 787-840-2575 |
| Jetsimary García | PHSU/PRI | jegarcia16@stu.psm.edu | 787-840-2575 |
| Eli Rosenberg | U. of Albany | erosenberg2@albany.edu | 518-486-9667 |
| Mariely Linares | PHSU | mlinares@psm.edu | 787-221-7460 |
| José Martinez | PHSU | jlmartinez@psm.edu | 787-614-3515 |
| José Molina | PHSU | jmolina@psm.edu | 939-350-6393 |
| Maritza Rentas | PHSU | mrentas@psm.edu | 787-216-9715 |
| Gladys González | PHSU | glgonzalez@psm.edu | 787-221-7460 |
| Jennie Nazario | PHSU | jnazario@psm.edu | 787-840-2575 |
| Elaine Martinez | PHSU | eimartinez@psm.edu | 787-840-2575 |
| Ana Perez | PHSU | aperez@psm.edu | 787-840-2575 |
| Emma Ramirez | PHSU | eramirez@psm.edu | 787-840-2575 |
| Genessis Negron | PHSU | gnegron@psm.edu | 787-840-2575 |
| Joaris Soto | PHSU | Jsoto15@stu.psm.edu | 787-840-2575 |
| Julia Herrera | PHSU | jherrera@psm.edu | 787-840-2575 |

| | | | |
|---|---|---|---|
| Lismarie Santiago | PHSU | lisantiago@psm.edu | 787-840-2575 |
| Robert Rodriguez | PHSU | rrodriguez@psm.edu | 787-840-2575 |
| Suheil Albizu | PHSU | salbizu@psm.edu | 787-840-2575 |
| Tatiana Morales | PHSU | tmorales@psm.edu | 787-840-2575 |
| Vanessa Rodriguez | PHSU | | 787-840-2575 |
| Veronica Rodriguez | PHSU | vrodriguez@psm.edu | 787-840-2575 |
| Wanda Vargas | PHSU | wvargas@psm.edu | 787-840-2575 |
| 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 |
| Emma Little | CDC | nna1@cdc.gov | 787-706-2399 |
| Adriana Romero | CDC | odo3@cdc.gov | 787-706-2399 |
| Carmen Perez | CDC | cnp8@cdc.gov | 787-706-2399 |
| Coral Rosado | CDC | nsj3@cdc.gov | 787-706-2399 |
| Karla Marrero | CDC | ipg5@cdc.gov | 787-706-2399 |
| Claudia Colon | Contractor/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 |
| Tyler Sharp | CDC | iyp4@cdc.gov | 787-706-2399 |
| Brenda Torres | Contractor/CDC | xfu1@cdc.gov | 787-706-2399 |

| | | | |
|---|---|---|---|
| Laura Zambrano | CDC | xbs6@cdc.gov | 787-706-2399 |
| Mark Delorey | CDC | esy7@cdc.gov | 787-706-2399 |
| Ronald Cancel | Contractor/CDC | ngt5@cdc.gov | 787-706-2399 |
| Oscar Padro | CDC | fmx3@cdc.gov | 787-706-2399 |