# Detection of dengue virus in Aedes aegypti during an urban epidemic in Iquitos, Peru (December 2010 to March 2011): results and tables

Anna B. Kawiecki ORCID: 0000-0002-0499-2612

**Load libraries**

```r
# Handles relative file paths
library(here)

# Grammar of tables for publication-quality outputs
library(gt)

# Simplifies working with dates and times
library(lubridate)

# Spatial data handling and simple features for mapping
library(sf)

# Core tidyverse packages for data manipulation and visualization
library(tidyverse)
```

**Read in data**

```r
# Household-level data
m.ae.pcr <- readRDS(here("analysis", "data", "raw_data", "m.ae.pcr.rds") )
m.ae.mid.pcr <- readRDS(here("analysis", "data", "derived_data",
                             "household_level_data", "m.ae.mid.pcr.rds"))
m.surv <- readRDS(here("analysis", "data", "derived_data",
                       "household_level_data","m.surv.rds"))
m.h.surv <- readRDS(here("analysis", "data", "derived_data",
                         "household_level_data","m.h.surv.rds"))
h.surv <- readRDS(here("analysis", "data", "derived_data", "household_level_data",
                       "h.surv.rds"))
```

## RESULTS

**Average number of surveys per week.**

```r
m.surv.means <- m.h.surv %>%
  st_drop_geometry() %>%          # Remove geometry column
  filter(host == "mosquito") %>%  # Keep mosquito data only
  group_by(epiweek) %>%           # Group by epidemiological week
  summarise(n = n())              # Count surveys per week
```

```r
summary(m.surv.means$n)          # Summary of weekly survey counts
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   175.0   261.8   337.0   328.0   369.5   498.0
```

```r
l.surveys.week <- lm(n ~ 1, m.surv.means)   # Fit intercept-only model

confint(l.surveys.week, level = 0.95)       # 95% CI for mean surveys/week
```

```
##                 2.5 %   97.5 %
## (Intercept) 287.4761 368.5239
```

## Average number of Ae. aegypti females collected per week.

```r
# View summary statistics for Ae. aegypti females per survey-week
summary(m.surv$aa_female)
```

```
##     Min.  1st Qu.   Median     Mean  3rd Qu.     Max.
##   0.0000   0.0000   0.0000   0.5266   0.0000 127.0000
```

```r
# Fit intercept-only linear model for female counts
l.aa_f <- lm(aa_female ~ 1, m.surv)

# Compute 95% confidence interval for mean female count
confint(l.aa_f, level = 0.95)
```

```
##                  2.5 %    97.5 %
## (Intercept) 0.4710774 0.5821964
```

## Average number of Ae. aegypti females collected per week in houses with at least 1 infected female.

```r
m.surv.positive <- m.surv %>% filter(denv.house == "positive")
# View summary statistics for Ae. aegypti females per survey-week
summary(m.surv.positive$aa_female)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   1.000   1.000   2.000   5.187   5.000 127.000
```

## Average number of Ae. aegypti females collected per week in houses with no infected females.

```r
m.surv.negative <- m.surv %>% filter(denv.house == "negative")
# View summary statistics for Ae. aegypti females per survey-week
summary(m.surv.negative$aa_female)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   1.000   1.000   1.000   2.074   2.000  63.000
```

## How many surveys collected females that were infected?

```r
# Filter for surveys where DENV-positive Ae. aegypti were detected
m.house.infected <- m.ae.pcr %>%
  filter(denv.house == "positive") %>%
```

```
  # Flag duplicate locations
  mutate(dup = duplicated(location_code))

# Count number of surveys (rows)
nrow(m.house.infected)
```

## [1] 75

**How many households had DENV-infected females?**

```
# Number of unique household IDs where DENV-positive Ae. aegypti were detected
length(unique(m.house.infected$location_code)) #72
```

## [1] 72

# How many households had infected females found in multiple surveys?

```
# Subset for households with duplicated positive detections
m.house.infected.multi <- m.house.infected[which(m.house.infected$dup == TRUE),] %>%
  st_drop_geometry()  # Drop spatial geometry

# Count unique households with repeated positive detections
length(unique(m.house.infected.multi$location_code))
```

## [1] 2

**Number of days between household surveys in households where infected females were found repeatedly:**

```
# Get survey dates for the 1st household with repeated infection
m.house.infected.multi.1 <- m.house.infected %>%
  filter(location_code == m.house.infected.multi$location_code[[1]] ) %>%
  dplyr::select("date.survey") %>%
  st_drop_geometry()

# View the selected survey dates
m.house.infected.multi.1
```

## # A tibble: 2 x 1
##   date.survey
## * <date>
## 1 2011-03-11
## 2 2011-03-15

```
# Calculate the number of days between two surveys (hardcoded)
days_diff.1 <- as.numeric(difftime(ymd("2011-03-11"), ymd("2011-03-15"), units = "days"))

# View result
days_diff.1 # 4
```

## [1] -4

```
# Get survey dates for the 2nd household with repeated infection
m.house.infected.multi.2 <- m.house.infected %>%
  filter(location_code == m.house.infected.multi$location_code[[2]] ) %>%
  dplyr::select("date.survey") %>%
  st_drop_geometry()

# View the selected survey dates
m.house.infected.multi.2
```

```
## # A tibble: 3 x 1
##   date.survey
## * <date>
## 1 2010-12-14
## 2 2011-01-07
## 3 2011-01-14
```

```
# Calculate days between two surveys
days_diff.2 <- as.numeric(difftime(ymd("2010-12-14"), ymd("2011-01-14"), units = "days"))

# View result
days_diff.2 # 31
```

```
## [1] -31
```

## Average percent of infected females in households with infected females

```
# Average percent of infected females in households with one or more infected females
mean(m.house.infected$perc.pos) #68.07393
```

```
## [1] 68.07393
```

## Number of the DENV positive females that were not dissected

```
m.ae.mid.pcr.pos <- m.ae.mid.pcr %>%
  st_drop_geometry() %>%                # Remove spatial geometry
  filter(!is.na(mosquito_id)) %>%       # Exclude records with missing mosquito IDs
  filter(pcr == "positive") %>%         # Keep only PCR-positive results
  group_by(n.body.parts.tested,         # Group by number of body parts tested,
           pcr.positive.body,           # PCR-positive body part,
           date) %>%                    # and date of survey
  summarise(n = n()) %>%                # Count number of records per group
  filter(n.body.parts.tested == 1 |     # Keep if only 1 body part tested,
         is.na(n.body.parts.tested))    # or if the entire body was tested
```

```
## `summarise()` has grouped output by 'n.body.parts.tested', 'pcr.positive.body'.
## You can override using the `.groups` argument.
```

```
# Calculate how many positive samples had abdomen tested separately (out of 128 total)
128 - sum(m.ae.mid.pcr.pos$n)
```

```
## [1] 123
```

## % females with detectable DENV in the head-thorax-legs-wings

```r
# Summarize PCR-positive mosquitoes tested on 2 body parts
m.ae.pcr.mid.sum <- m.ae.mid.pcr %>%
  st_drop_geometry() %>%                    # Drop spatial geometry
  filter(pcr == "positive" &                # Keep PCR-positive cases
         n.body.parts.tested == 2) %>%      # Only where 2 body parts were tested
  filter(pcr.positive.body !=               # Exclude where whole mosquito tested
           "abdomen or thorax") %>%
  mutate(dissemination.evidence = case_when(  # Classify dissemination
    pcr.positive.body == "abdomen" ~
      "No evidence of dissemination",
    T ~ "Evidence of dissemination"
  )) %>%
  group_by(dissemination.evidence) %>%      # Group by dissemination class
  summarise(n = n())                        # Count observations per group

# Proportion with evidence of dissemination
m.ae.pcr.mid.sum[
  which(m.ae.pcr.mid.sum$dissemination.evidence ==
          "Evidence of dissemination"),
][["n"]] / sum(m.ae.pcr.mid.sum$n)
```

```
## [1] 0.804878
```

## TABLES

### Table 1: Summary of adult mosquito surveys

```r
table1 <- m.ae.pcr %>%
  st_drop_geometry() %>%                    # Remove geometry column
  group_by(result, ae.f.survey) %>%         # Group by survey outcome
  summarise(n.houses = n(),                 # Count of houses per group
            n.aa_f = sum(aa_female),        # Total female mosquitoes
            n.aa_f.test = sum(n),           # Number tested by PCR
            n.aa_f.denv = sum(n.pos.mosq)) %>%# Number DENV-positive
  ungroup() %>%
  mutate(n.total.surveys = sum(n.houses),        # Total surveys = 9405
         n.total.worked.surveys =
           sum(n.houses[result != "No survey"]),    # Succesful surveys = 6720
         n.total.worked.aa_f.surveys =
           sum(n.houses[ae.f.survey !=
             "No Ae.ae females collected"]),        # Houses where females were collected
         n.total.collected.aa_f =
           sum(n.aa_f, na.rm = TRUE),             # Collected females
         n.total.tested.aa_f =
           sum(n.aa_f.test, na.rm = TRUE),        # Tested females
         n.total.positive.aa_f =
           sum(n.aa_f.denv, na.rm = TRUE)) %>%    # Positive females
  dplyr::select(-c("result", "ae.f.survey",
                   "n.houses", "n.aa_f",
                   "n.aa_f.test", "n.aa_f.denv")) %>%   # Drop raw cols
  gather() %>% distinct() %>%                          # Reshape, remove duplicates
```

```r
  mutate(statistic = case_when(
          key == "n.total.surveys" ~ "Adult surveys",
          key == "n.total.worked.surveys" ~
            "Successful adult surveys",
          key == "n.total.worked.aa_f.surveys" ~
            "Adult surveys with female collections",
          key == "n.total.collected.aa_f" ~
            "Female adults collected",
          key == "n.total.tested.aa_f" ~
            "Female adults tested by PCR",
          key == "n.total.positive.aa_f" ~
            "Female adults DENV positive by PCR")) %>%
  mutate(group = case_when(
    str_detect(statistic, "survey") == TRUE ~ "Adult surveys",
    T ~ "Female adults"                     # All others go to this group
  )) %>%
  group_by(group) %>%  #
  mutate(
    perc = case_when(
      group == "Adult surveys" ~ value / 9405 * 100,
      group != "Adult surveys" &
        statistic != "Female adults DENV positive by PCR" ~
        value / 3539 * 100,
      T ~ value / 2795 * 100)) %>%
  mutate(perc = round(perc, digits = 1)) %>%  # Round percentages
  mutate(perc = gsub(100, "", perc)) %>%      # Remove any 100 string
  mutate(value = case_when(                   # Combine value and percent
    perc != "" ~ paste(value, "(", perc, "%", ")"),
    T ~ as.character(value)
  )) %>%
  rename('Number (%)' = value) %>%            # Rename final column
  dplyr::select(-c("key", "perc")) %>%        # Drop unneeded columns
  group_by(group) %>%
  gt(rowname_col = "statistic")               # Format for display table

## `summarise()` has grouped output by 'result'. You can override using the
## `.groups` argument.
table1  # Print table
```

|  | Number (%) |
|---|---|
| **Adult surveys** |  |
| Adult surveys | 9405 |
| Successful adult surveys | 6720 ( 71.5 % ) |
| Adult surveys with female collections | 1553 ( 16.5 % ) |
| **Female adults** |  |
| Female adults collected | 3539 |
| Female adults tested by PCR | 2795 ( 79 % ) |
| Female adults DENV positive by PCR | 128 ( 4.6 % ) |

```r
# Save as CSV for external use
write.csv(table1, here("analysis", "outputs", "tables", "table1.csv"))
```

**Table 2: Vertebrate host species identified from Ae. aegypti blood meals collected in Iquitos, December 2010- January 2011**

```r
# Create a data frame with species and counts
table2 <- data.frame(Species= c("Human", "Avian","Dog","Cat"),
                      Number= c(245,2,2,1)) %>%
  mutate(
    perc = Number / sum(Number) * 100 ) %>%        # Calculate percentages
  mutate(
    value = paste(Number, "(", perc, "%", ")")) %>% # Combine count & percent
  rename('Number (%)' = value) %>%                  # Rename final column
  dplyr::select(-c("Number", "perc")) %>%           # Drop intermediate columns
  gt(rowname_col = "Species") |>                    # Create formatted gt table
  tab_header(
    title = md("Table 2: Host species identification") # Add table title
  )

table2  # Display the table
```

Table 2: Host species identification

|  | Number (%) |
|---|---|
| Human | 245 ( 98 % ) |
| Avian | 2 ( 0.8 % ) |
| Dog | 2 ( 0.8 % ) |
| Cat | 1 ( 0.4 % ) |

```r
# Save table as CSV
write.csv(table2, here("analysis", "outputs", "tables", "table2.csv"))
```

**Table 3: Adult mosquito surveys outcomes by surveillance strategy**

```r
table3 <- m.surv %>%
  # Remove geometry column
  st_drop_geometry() %>%
  # there are 2 females that weren't tested, so we have to calculate incidence only from those that wer
  group_by(pos.case.contact) %>%
  summarise(
    # Number of surveys (houses worked)
    n.surveys= n(),
    # Total number Ae. aegypti collected (males + females)
    n.aa.total= sum(aa_total, na.rm= TRUE),
    # Number of Ae. aegypti females collected
    n.aa.f= sum(aa_female, na.rm= TRUE),
    # Number of Ae. aegypti females tested by PCR
    n.aa.f.test= sum(n, na.rm= TRUE),
    # Number of Ae. aegypti females tested by PCR that were DENV positive
    n.aa.f.denv = sum(n.pos.mosq, na.rm = TRUE),
    # Average number of total Ae. aegypti per house
    avg.aa.total= mean(aa_total, na.rm= TRUE),
    # Average number of Ae. aegypti females per house
    avg.aa.f= mean(aa_female, na.rm= TRUE),
```

```r
  ) %>%
  # Select relevant columns for table
  dplyr::select(c("pos.case.contact", "n.surveys", "n.aa.f.test","n.aa.f.denv", "avg.aa.f")) %>%
  # Calculate DENV % prevalence
  mutate(perc.aa.f.denv= (n.aa.f.denv/n.aa.f.test)*100) %>%
  # Rename suveillance strategy labels
  mutate(pos.case.contact= case_when(
    pos.case.contact=="other surveillance" ~ "Other surveillance",
    T~ "Positive case-contact"
  )) %>%
  # Rename columns for clarity
  rename('Average number of females/survey'=avg.aa.f,
         'Total number of surveys'=n.surveys,
         'Total number of females tested'=n.aa.f.test,
         'Total number of DENV positive females'=n.aa.f.denv,
         'Vector DENV prevalence (%)'=perc.aa.f.denv,
         'Surveillance strategy'=pos.case.contact ) %>%
  # Convert to long format for table display
  pivot_longer(
    cols = -'Surveillance strategy',
    names_to = "variable",
    values_to = "Value"
  ) %>%
  # Format numeric values to 1 decimal
 mutate(across(where(is.numeric), ~ round(.x, 2))) %>%
   # Keep only necessary columns
  dplyr::select(variable, 'Surveillance strategy', Value) %>%
  # Order variables logically
  mutate(variable= factor(variable,
                          levels = c( 'Total number of surveys',
                                      'Total number of females',
                                      'Total number of females tested',
                                      'Total number of DENV positive females',
                                      'Average number of females/survey',
                                      'Vector DENV prevalence (%)'))) %>%
  # Arrange rows by variable order
  arrange(variable) %>%
  # Group by variable for table formatting
  group_by(variable) %>%
  # Create gt summary table
  gt(rowname_col = "Statistic") %>%
  # Add table title
  gt::tab_header(
    title = md("Table 3: Adult mosquito surveys outcomes by surveillance strategy")
  )

table3
```

Table 3: Adult mosquito surveys outcomes by surveillance strategy

| Surveillance strategy | Value |
| --- | --- |
| Total number of surveys | |
| Other surveillance | 6541.00 |

| | |
|---|---|
| Positive case-contact | 179.00 |
| **Total number of females tested** | |
| Other surveillance | 2650.00 |
| Positive case-contact | 145.00 |
| **Total number of DENV positive females** | |
| Other surveillance | 123.00 |
| Positive case-contact | 5.00 |
| **Average number of females/survey** | |
| Other surveillance | 0.51 |
| Positive case-contact | 0.97 |
| **Vector DENV prevalence (%)** | |
| Other surveillance | 4.64 |
| Positive case-contact | 3.45 |

```r
# Save table to CSV
write.csv(table3,here("analysis", "outputs", "tables", "table3.csv"))
```