OXFORD UNIVERSITY PRESS | Journal of Medical Entomology

Understanding Mosquito Surveillance Data for Analytic Efforts: a case study

Journal:	*Journal of Medical Entomology*
Manuscript ID	JME-2020-0308.R2
Manuscript Type:	Review
Date Submitted by the Author:	n/a
Complete List of Authors:	Brown, Heidi; The University of Arizona Health Sciences, Epidemiology and Biostatistics Sedda, Luigi; Lancaster University, Lancaster Medical School Sumner, Chris; Yuma County Pest Abatement District, Pest Abatement Stefanakos, Elene; Yuma County Pest Abatement District, Pest Abatement Ruberto, Irene; Arizona Department of Health Services, Office of Infectious Disease Services Roach, Matthew; Arizona Department of Health Services, Office of Environmental Health
Please choose a section from the list:	Vector Control, Pest Management, Resistance, Repellents
Field Keywords:	Surveillance, Modeling, Mosquito Borne Diseases
Organism Keywords:	* Non-taxon specific

H. E. Brown
University of Arizona
Mel and Enid Zuckerman College of Public Health
Department of Epidemiology and Biostatistics
1295 N. Martin Ave.
Tucson, AZ 85724
Phone: 520/626-2262
Email: heidibrown@arizona.edu

Understanding Mosquito Surveillance Data for Analytic Efforts: a case study

H. E. Brown [1], L. Sedda [2], C. Sumner [3], E. Stefanakos [3], I. Ruberto [4], M. Roach [5]

1. University of Arizona, Mel and Enid Zuckerman College of Public Health, Department of Epidemiology and Biostatistics, 1295 N. Martin Ave., Tucson, AZ 85724
2. Lancaster University, Lancaster Medical School, Bailrigg campus, LA1 4YG, Lancaster, UK.
3. Yuma County Pest Abatement District, 2186 W. County 15th Street, Somerton, AZ 85350

4. Office of Infectious Disease Services, Arizona Department of Health Services, 150 N 18th avenue, Phoenix, AZ 85007
5. Arizona Department of Health Services, Office of Environmental Health, 150 N 18th Avenue, Phoenix, AZ 85007

Abstract:

Mosquito surveillance data can be used for predicting mosquito distribution and dynamics as they relate to human disease. Often these data are collected by independent agencies and aggregated to state and national level portals to characterize broad spatial and temporal dynamics. These larger repositories may also share the data for use in mosquito and/or disease prediction and forecasting models. Assumed, but not always confirmed, is consistency of data across agencies. Subtle differences in reporting may be important for development and the eventual interpretation of predictive models. Using mosquito vector surveillance data from Arizona as a case study, we found differences among agencies in how trapping practices were reported. Inconsistencies in reporting may interfere with quantitative comparisons if the user has only cursory familiarity with mosquito surveillance data. Some inconsistencies can be overcome if they are explicit in the metadata while others may yield biased estimates if they are not changed in how data are recorded. Sharing of metadata and collaboration between modelers and vector control agencies is necessary for improving the quality of the estimations. Efforts to improve sharing, displaying, and comparing vector data from multiple agencies are underway, but existing data must be used with caution.

Key Words: Mosquito-borne disease, vector surveillance, disease prediction, data sharing

Introduction

While the challenges to sharing vector data across agencies and with academic partners isn't new (Moise et al. 2019), it recently came to the fore during the Zika outbreak. In an age of improved data sharing, modelers from all disciplines wanted access to vector data to estimate risk for Zika given the presence of vectors. However, they were met with no centralized repository and inconsistencies in reporting across agencies.

In addition to emergent diseases, data challenges exist with endemic diseases like West Nile Virus (WNV) in the US. Focused studies have shown mosquito surveillance data is useful for predicting human WNV cases weeks in advance (Bolling et al. 2009, Kilpatrick and Pape 2013). Combining mosquito abundance with infection prevalence, the vector index, has proven useful for estimating human disease risk, but few areas have sufficiently complete data available to do so (Petersen 2019). Data completeness and consistency is further compounded when trying to merge data from multiple agencies.

Heterogeneity in data composition, quality, and spatial or temporal resolution challenge their use for display and modelling purposes. The pragmatic approach when estimating probabilities of species abundance or presence is to use the data as is, and, for example, acknowledging heterogeneity as random effects in mixed models or to create complex frameworks to account for biases (Fletcher et al. 2019, Manica et al. 2019). These modelling approaches can be applied only when complete information is provided with respect to the data collection and reporting process. Even then, the models are often too complex for decision making. Improving how predictive models are used by stakeholders for planning and response will come with sharing of standardized data and from collaboration with vector control agencies (Moy et al. 2018, Barker 2019). Such co-production of science has been shown to improve the data quality and utility of predictive disease risk models (Purse et al. 2020).

Along omni-comprehensive world species databases (the largest is GBIF https://www.gbif.org/), there have been initiatives to share mosquito and mosquito-borne disease data. These range from international data portals like VectorBase (vectorbase.org), national like VectorMap (http://vectormap.si.edu/) hosted by the Smithsonian Institute, and more local programs like VectorSurv (https://vectorsurv.org/) originating in California but expanding to other states. Related, ArboNET (https://wwwn.cdc.gov/arbonet/) is a passive disease surveillance system established in 2000 by the Centers for Disease Control and Prevention (CDC) in response to West Nile virus, which aims to facilitate arboviral data compilation and sharing (Petersen 2019).

Such collections of vector data may then be imported further into repositories aimed at reaching broader audiences. One example is the CDC's National Environmental Public Health Tracking (EPHT). EPHT was initiated as a comprehensive approach for collecting, compiling, and sharing data on environmental exposures, burdens, and diseases (Frumkin et al. 2008). Primarily, EPHT focuses on human disease outcomes, see for example (Eatman and Strosnider 2017), but some agencies are also including vector abundance and arbovirus surveillance data. Currently 26 state and local health departments share morbidity, mortality, and surveillance data via the EPHT portal (Centers for Disease Control and Prevention 2017). Many of these broader repositories make the data accessible internally to those providing data and externally through specific requests, reaching a broader group of potential users, but also moving the data further from its sources.

When using data from a compiled database, differences in reporting that may ultimately affect the model output may go undetected. In the US, for example, vector data is held in over 1000 different individual vector agencies across the country (Rund et al. 2019). Merging data across agency goes beyond which species and by what trap type. The complexities in how to merge begin to arise in just considering differences in trapping frequency or the locations of the site (e.g., urban vs agricultural). Most of these will be documented and can be addressed. What is not as often documented is the purpose for trapping and a detailed sampling design which would allow for a complete and transparent merging of data (Sedda et al. 2019). While clean, compiled vector data from multiple agencies is an objective, the lack of standardization between agencies would make such a database unhelpful, potentially even counter-productive (Yong 2017). This is especially true in the absence of robust metadata (Powers and Hampton 2019).

Herein, we use mosquito surveillance data from Arizona as a case study to highlight potential discrepancies in how data are reported between agencies. We acknowledge that many of JME's readership are aware of issues described herein, the call for standardization in reporting collection locations has been made elsewhere in JME (Foley et al. 2009). Yet merging mosquito surveillance data remains currently unfeasible because of the lack of standardization. Without discussion of these challenges even those of us who work closely with local vector control agencies may attempt to merge data erroneously, assuming that other agencies report in the very same way. Similarly, we hope that vector control agencies who don't work as closely with modelers may understand how their data are used and the significance of their data reporting. Whether using vector data for visualization or in prediction and forecasting models, these discrepancies must be addressed to insure the output's validity. Through an understanding of the origins and purposes of the data, better models and better predictions may be generated. We hope that this is a further step toward a unified mosquito surveillance protocol.

Modeling mosquito surveillance data

Vector data to support estimating vector-borne disease risk usually require spatial and temporal information to document changes in distribution across space and changes over time, both within a season or across years. This assumes some consistency or at least the knowledge of 1) the trap type, 2) whether the trap was set because of a suspected hotspot or as part of routine sampling, 3) the location and frequency for setting and servicing traps. Knowing the location and time of trapping events is a minimum requirement in spatio-temporal and population dynamics analyses, reflected in points 1 and 3. Point 2 is essential to correct for data inflation as expected in hotspots or for routine sampling when carried out in certain ecotypes. Variations on these aspects may produce dramatically different pictures of the distribution of mosquitoes and/or the risk for disease.

Data for this case study. We use the 2018 data from the local Arizona database for reporting to CDC's ArboNET. Prior to forwarding the data to the national ArboNET, the Arizona Department of Health Services (ADHS) receives and complies mosquito data from 19 vector control agencies across the 15 counties in Arizona. The state was moving toward VectorSurv as their centralized data repository at the time of this analysis. While this change will help to standardize many of the data collection inconsistencies described herein, data collected prior to the centralization likely continue to be requested. It will be even more relevant to be aware of some of the inconsistencies when the data are merged for longitudinal analyses and for trend analyses using historical data.

Using a data manipulation program written in SAS (SAS Institute Inc., Cary, NC), ADHS cleans and compiles data from agencies for inclusion into the local database which was maintained in Microsoft Access (Microsoft Corporation, Redmond, WA). WNV vector data were extracted and provided for this analysis as a Microsoft Excel spreadsheet (Microsoft Corporation, Redmond, WA). After removing 1,151 entries with no geographic positioning coordinates (1.6% of the total entries), there were 71,812 entries with 27 variables. When collapsing pools to trap nights, this further combines to 42,308 trap nights over 3,821 unique trap type/location entries recorded in 2018. Frequency comparisons were conducted in StataIC v15 (StataCorp, College Station TX).

Trapping Method. Traps that collect adult host seeking mosquitoes in the US commonly include CO_2 traps which may be enhanced with baits like octenol or light. For WNV surveillance, live bait in traps (e.g., bird-boxes) may also be used, but these are rare and were not in our dataset. Gravid females are often an important indicator of disease risk because they are more likely to have taken a blood meal (Williams and Gingrich 2007) and may be selected for using gravid traps. Immature mosquitoes may be collected via using a dipper in existing water habitat or through placement of ovi-traps, and reliably provide presence information but not abundance.

Trap types listed in this data set include primarily CO_2 traps (80% of all entries). The next most common were ovitraps (5.3%), Biogents Sentinel trap (1.8%), and Encephalitis Vector Survey trap (1.6%). Traps were defined as Other, in 8.1% of entries, while 3.1% of the entries did not list any trap type. Additional baits used, like octenol, were either not used or not reported in these data.

It is well established that the type of trap, when, and how it is placed influences the species and abundance of mosquitoes collected (Bidlingmayer 1967). Studies have shown that use of attractant (Meeraus et al. 2008), type of trap (DiMenna et al. 2006, Brown et al. 2008, Maciel-de-Freitas et al. 2008, Holderman et al. 2018), and location of trap (Anderson et al. 2004, DiMenna et al. 2006, Černý et al. 2011) may provide different depictions of mosquito abundance, species composition, and infection prevalence. Failure to account for attractant use and trap type information may result in biases or large errors in estimates of species presence and abundance depending upon the variety of traps deployed (Bidlingmayer 1967, Kline 2006). Because of the effects of trap type and bait used on the species observed, it would be necessary to confirm and account for such information where possible.

Reason for Trapping. In the data we reviewed, 74% of entries include a reason for trapping. Routine surveillance was the most commonly listed reason (85.2%), followed by complaint response (8.9%), surveillance (2.5%), enhanced surveillance (1.6%), response to human case (1.6%), or other (0.3%). Maricopa County, Arizona’s most populated county, accounted for 93% of the data in the 2018 database. The reason for trapping for 83% of that county’s data was "routine surveillance" with an additional 15% with no reason listed. In contrast, only 59.6% of the reported reason for trapping was routine surveillance of all other counties combined.

The reason for vector surveillance systems is to inform decisions about public health interventions. The reasons listed, however, indicate how implementation of traps set to protect human health may differ among agencies with further implications to mosquito abundance estimates. Traps set at the same, fixed location each year and repeatedly sampled over the season provide the most robust estimates of between and within year abundance. Traps set in response to citizen complaints produce high estimates of abundance and, if included in comparisons, may overestimate abundance, especially later in the season (Fig 1). Traps set in response to a human case may lead to biased estimates of vector abundance or arboviral infection prevalence because they likely represent hot spots of virus activity rather than normal variance. If the routine surveillance program is robust and the primary data source, the impact of using all available data may be minimal. In Figure 1 a, most (83.1%) of the data from this county are listed as routine sampling. The data restricted to only routine (solid boxes) and routine with response to case or complaint (dashed boxes) are similar. The lack of data from routine sampling later in the season is likely omission errors: coordinates for routine sites not marked as such in the database. In contrast (Fig 1 b), when estimating abundance using the data from all other counties where almost half of the trapping is for reasons other than "routine surveillance', not distinguishing by reason for trapping inflates the mosquito abundance (dashed boxes are higher than routine only). These data show an exception from early in the season where more than 100000 mosquitos were recorded in one trap on one night. For illustrative purpose, we also plot the comparison of only routine versus only response (complaint or human case related) trapping, Figure 1c. This further shows the inflated counts of trapping where complaints are registered or human cases occur.

The modeler interested in describing mosquito geographic dynamics will likely turn toward vector control data from multiple agencies for the longest period of time available. However, they run the risk of compounding uncertainty around trapping reasons both between agencies but also as agencies change their sampling strategy over time.

Vector control agencies focus sampling efforts to areas where the population they seek to protect are. This may leave sparsely populated areas undersampled. Surveillance designed to sample across high, medium and low population density areas are rare. As a result, description of mosquito geographic dynamics and movements to inform vector control in a highly populated area may be jeopardized by lack of information about potential immigration of mosquitoes from neighboring areas (Sedda et al. 2019).

Immigration and emigration of mosquitoes from an area to another can be estimated by using insect population models (Sedda et al. 2020). In case of immigration, new trapping locations can be adaptively added based on modeled mosquito migration patterns especially if sharing data from neighboring districts is possible and the data are comparable. Establishing new routine sites would then be determined by allocating the traps where immigration processes are taking place in combination with other priorities such as proximity to highly density areas. Agencies working to understand between year dynamics may wish to establish set surveillance locations for consistent intra-annual comparisons if they don’t already. At a minimum, including a reason for trapping, especially when in response to a case or nuisance is critical so that estimates could be adjusted accordingly.

Frequency of Trapping and Site Names. Trapping frequency varied by county and even within county. When evaluating trapping locations, 57.1% of the traps are located in Maricopa County and 32.3% are located in Pima County, Arizona's two most populated counties. Though they represent only 3.9% of all of the records entered, almost half (43.9%) of the sites were set for one night based on the coordinates reported in the database regardless of reason for trapping. As would be expected, locations listing "complaint response" as the reason for trapping were serviced less frequently, an average of 1.02 trap night (sd= 0.15, min= 1, max= 2) for the year. In contrast, traps labeled as "routine surveillance" were serviced an average of 14.38 trap nights (sd=16.56, min= 1, max= 43) over the year evaluated. Of CO_2 traps set for "routine surveillance," 24.6% of traps were set between 40 and 43 nights, 17.6% were set between 20-22 nights, and nearly half (43.9%) were set for only one trap night.

One explanation of the high occurrence of routine traps set for just one night, may be associated with how databases are designed. Error may occur if location coordinates are not linked to a specific site and an error of around 5m might be expected between readings taken with a smartphone (National Coordination Office for Space-Based Positioning, Navigation 2019). Thus, what is known to the vector control agency to be the same location, may be recorded as a cluster of close locations when identified by coordinates alone. In these data, 27.3% of trap with the same name had different GPS coordinates, even after removing sites with no name recorded.

Inclusion of a site name with permanent coordinates or standardization of terminology would resolve the normal error in location readings.

Alternatively, this might be indicative of differences in how *routine* is understood. Routine might be interpreted as sampling sites which are fixed locations that are sampled at regular intervals. Routine may also be interpreted as part of standard assessment of the mosquito fauna of the area, but not necessarily as specific fixed sites sampled at a fixed and regular frequency. Fixed surveillance sites, to revisit each year as part of routine vector surveillance can allow for comparison of trends over time.

Frequent trapping of at least every other week during the mosquito season is common in mosquito sampling designs and enables descriptions of seasonality and variation in mosquito abundance using a cost-effective approach (Vanlalruia et al. 2014). However, depending on the number of trapping locations, frequent trapping may not be sustainable for long surveillance campaigns over large areas. Allocation of carefully selected locations can provide better, cost-effective results especially if coupled with minimization of trapping uncertainties via spatially explicit data modelling, i.e. adaptive sampling (Fanshawe and Diggle 2012). Some analyses have shown that sampling over a period of two or more years may be useful to accurately infer seasonal and interannual cycles in mosquito abundance (Li et al. 2019). In a study of the association between *Aedes aegypti* vector abundance and dengue infection, mosquito data from longitudinal surveys were associated with infection risk, whereas cross-sectional vector indices were poor (Cromwell et al. 2017).

Routine surveillance trapping, that is a trap set in the same location annually and visited regularly within the mosquito season, provides estimates of seasonal dynamics in vector abundance. As described above, this has costs for the agency, especially when these are in addition to responding to nuisance complaints from community members about biting insects and enhanced trapping in response to human cases of arboviral disease. One solution might be to identify a subset of trapping locations allocated along an environmental gradient relevant to the region or as part of a systematic grid over the whole region based on historical data or important ecological delineation (Sedda et al. 2019) within a cost optimization framework (Longbottom et al. 2020). These locations would be part of the regular annual surveillance and serviced multiple times within the mosquito season. Many vector control agencies do this, but identifying these sentinel trapping locations explicitly in any shared dataset would be helpful to reduce the effect of systematic biases.

Missing Data and Empty Traps. Approximately half of the entries (47.6%) reported no mosquitoes collected. Nine of the 40 individuals who submitted reports, entered traps with zero counts. This indicates that empty traps are being recorded into this database, but inconsistently across agencies and individuals.

This inconsistency has important implications for the data analyst. The completeness of the data set is a measure of how long the data have been consistently reported and whether empty traps are reported. A zero count may occur when a trap is tampered with or malfunctions in some way. Alternatively, a zero count may occur when the trap functions properly but no mosquito is collected or, in some instances, when mosquitoes are collected, but not the target species.

Distinguishing between zero counts because a trap failed and traps with zero collected mosquitoes is essential to creating a better estimate of abundance and the beginning and end of the local mosquito season. As an example (Figure 2), we plotted abundance from week 23-27 for one submitting county which did consistently report empty traps. For one species, weekly trap counts both with the unaltered data and when artificially removing the empty traps are plotted. This boxplot over time shows abundance estimates that may be artificially inflated if empty traps are not included in the data. While this omission may be obvious to the analyst who explored their data before modeling, it cannot be remedied if empty traps are not recorded.

From a modeling point of view, zero counts can be addressed with overdispersed models when the zeros are real; Zero inflated models can help to address when the zeros are not associated with the mosquito distribution but rather trap malfunction (Warton 2005). However, this is only possible if zero-counts are recorded in the data. Training on data entry and setting mandatory data fields in the recording system may facilitate the completeness of reporting data. As databases become more automated, incorporated training and mandatory entry may reduce the problems associated with empty traps. When using data already collected, the procedure by which the submitting vector control personnel record empty and malfunctioning traps should be confirmed.

Pathogen Information. In Arizona, WNV (52%) and St. Louis encephalitis virus (47%), are the most common pathogens for arboviral testing. Dengue, chikungunya, and Zika viruses accounted for most of the rest. Reverse transcriptase-polymerase chain reaction (RT-PCR) is the most commonly used method of testing (35% of reported test types).

When a maximum number of mosquitoes are collected, mosquitoes are pooled into smaller groups for arbovirus detection. There is no established standard, but pools up to 50 or up to 100 are common. Pool size, mosquito species/genus composition and arboviral test results were documented in the Arizona database. For the data reviewed here, pools were of no more than 50 mosquitoes and usually consisted of a single species.

Arboviral information with mosquito abundance information can then be used to estimate transmission risk. Three common measures are minimum infection rate (MIR), maximum likelihood estimate (MLE), and the vector index. MIR is the ratio of positive pools to the number of mosquitoes tested and assumes only one positive mosquito in a positive pool (Reeves and Hammon 1962), whereas MLE estimates the most likely infection rate (Walter et al. 1980). The vector index is the product of a given species abundance multiplied by the proportion infected (Gujral et al. 2007). While these metrics are helpful in estimating human WNV risk (Kilpatrick and Pape 2013), they should be used with caution in areas of high transmission and low sample size (Gu et al. 2008, Bustamante and Lord 2010, Chakraborty and Smith 2019). It is worth noting that information like prevalence of the pathogen or of insecticide resistance can inform ad-hoc calculations for the size of the group during the pooling especially in presence of spatial heterogeneity which may require group size differing from trap to trap (Hepworth and Biggerstaff 2017).

Capturing Vector Control Information. Despite that data repositories like ArboNET were established to monitor disease not to assess mosquito control measures, data from these repositories are being used to describe mosquito distributions, arbovirus presence and disease risk. Refinements to these repositories include MosquitoNET, developed to also include insecticide resistance information. Including information on the magnitude and duration of intervention will likely further complicate standardization because of the wide range of approaches by organization, region and regulation. However, consideration of it would be useful to the researcher trying to interpret vector dynamics, to the citizens interested in vectors in their area, and also to vector control as to the lasting success of their efforts. Specifically, adulticide or larvicide, technician, equipment, pesticide mix, quantity, area treated, start/stop date and time, and description of the conditions when applied. These data were not captured in the database we reviewed. Newer repositories like VectorSurv designed with modeling in mind do include this information, with new sets of challenges to standardization.

In trying to understand the impact of vector control, the missing information about interventions can create challenges in estimating changing trends in mosquito abundance, especially given the current global rise in insecticide resistance in mosquitoes (Moyes et al. 2017). When and where adulticiding or larviciding is performed affects the mosquito abundance in an area. This is the goal of these activities. Including this information may be critical for research questions looking to understand drivers of the trends in mosquito abundance and evolution (Britch et al. 2010, Fouet et al. 2018) and to identify thresholds for action (Nasci and Mutebi 2019).

Standards in metadata. Using mosquito surveillance data for model inference and prediction is affected by the degree of standardization in how the data are collected, compiled and defined. This information is usually collected as metadata and includes qualitative information like trap type and the spatial allocation of trap across the landscape being monitored. Sometimes metadata can be employed in statistical frameworks as predictors, as random effects, or just to inform the model priors. The latter is the main element that allows for the estimation of data and model uncertainties of the predictions within the Bayesian paradigm (Lindley 1983). Machine learning methods, where algorithms learn from and act on data via transformations and segmentation, are gaining ground in inferential and predictive quantitative and translational methods (Toh et al. 2019). Meaningful associations and/or directionalities within black box machines require large, robust datasets, often necessitating combining across multiple vector control districts and/or multiple years of sampling. Thus increasing the need for robust metadata and standardization in reporting.

Biases and errors can only be controlled by the use of metadata directly into the model especially when dealing with heterogeneous data (Toh et al. 2019). One example of integrating spatially explicit data in order to provide a homogeneous health and environmental monitoring and subsequent analyses comes from Germany (Schröder 2006). Data standardization with robust metadata is fundamental for the efficiency of real-time or near-real time biosurveillance and early warning systems algorithms (Pollett et al. 2017).

In this paper, we use Arizona mosquito data to highlight discrepancies that may arise due to variability in the collection, reporting and use of mosquito surveillance data. We expect that readers might feel, “but that isn’t how we do it” as they read this case-study. Thus further supporting the need for clarity in what data are reported. Understanding the variability may help to prioritize standardization efforts and for the inclusion and use of metadata. Time and resource constraints, however will continue to limit what data each agency can report (Lindsey et al. 2012). Addressing these gaps and insuring useful predictive models will come with the increasing trends for data sharing data and centralized repositories.

Conclusion

This case-study was motivated to support efforts by health departments to more openly share vector control data with academic and community partners on repositories like EPHT. Data at the level provided within ArboNET would allow for spatial and temporal analysis of trends, comparisons across trap type and reason, and potential development of early warning with the inclusion of pathogen data. We describe expected variability in data collected across agencies as it relates to potential impacts for modeling work. As portals like CDC's ArboNET or EPHT increasingly become resources for national level data, the challenges to maintaining the integrity of the data becomes more salient.

The aim of data repositories and data sharing is to build better understanding of mosquito-borne disease risk. Vector control uses their data to promote human health by identifying areas with high mosquito activity and responding. Modelers are more likely to use the data to understand drivers of high mosquito activity, with the same goal of protecting human health. When using mosquito surveillance data, ideally the researchers would be working with the agency collecting the data. In the absence of such collaboration, we hope the above information will help clarify the need for standardization and metadata. Coordination across agencies and data aggregation would allow for increased capacity to develop usable models of vector control, vector seasonality, and predictions of entomologic risk now and under future climates.

Conflicts of Interest. None to disclose.

Authors (HEB, MR, and IR) were partially supported by the Centers for Disease Control and Prevention cooperative agreements CDC-RFA-EH16-1602: Enhancing Community Resilience by Implementing Health Adaptations (Award 5 NUE1EH001318) and CDC-RFA-EH17-1702: Enhancing Innovation and Capabilities of the Environmental Public Health Tracking Network (Award 1 NUE1EH001339). LS work was supported by the Wellcome Trust Seed Award (212501/Z/18/Z).

References

Anderson, J. F., T. G. Andreadis, A. J. Main, and D. L. Kline. **2004**. Prevalence of West Nile Virus in tree canopy-inhabiting *Culex pipiens* and associated mosquitoes. Am J Trop Med Hyg. 71: 2–9.

Barker, C. M. 2019. Models and surveillance systems to detect and predict West Nile Virus outbreaks. J. Med. Entomol. 56: 1508–1515.

Bidlingmayer, W. L. 1967. A comparison of trapping methods for adult mosquitoes: species response and environmental influence. J. Med. Entomol. 4: 200–220.

Bolling, B. G., C. M. Barker, C. G. Moore, W. J. Pape, and L. Eisen. **2009**. Seasonal patterns for entomological measures of risk for exposure to *Culex* vectors and West Nile virus in relation to human disease cases in Northeastern Colorado. J. Med. Entomol. 46: 1519–1531.

Britch, S. C., K. J. Linthicum, W. W. Wynn, T. W. Walker, M. Farooq, V. L. Smith, C. A. Robinson, B. B. Lothrop, M. Snelling, A. Gutierrez, and H. D. Lothrop. **2010**. Residual mosquito barrier treatments on U.S. military camouflage netting in a southern California desert environment. Mil. Med. 175: 599–606.

Brown, H. E., M. Paladini, R. A. Cook, D. Kline, D. Barnard, and D. Fish. **2008**. Effectiveness of mosquito traps in measuring species abundance and composition. J. Med. Entomol. 45: 517–521.

Bustamante, D. M., and C. C. Lord. **2010**. Sources of error in the estimation of mosquito infection rates used to assess risk of arbovirus transmission. Am. J. Trop. Med. Hyg. 82: 1172–1184.

Centers for Disease Control and Prevention. **2017**. Better Information For Better Health. (https://www.cdc.gov/nceh/tracking/BetterInfoBetterHealth.htm). Accessed 2019-11-18

Černý, O., J. Votýpka, and M. Svobodová. **2011**. Spatial feeding preferences of ornithophilic mosquitoes, blackflies and biting midges. Med. Vet. Entomol. 25: 104–108.

Chakraborty, S., and R. L. Smith. **2019**. Error associated with estimates of Minimum Infection Rate for endemic West Nile virus in areas of low mosquito trap density. Sci. Rep. 9: 19093.

Cromwell, E. A., S. T. Stoddard, C. M. Barker, A. Van Rie, W. B. Messer, S. R. Meshnick, A. C. Morrison, and T. W. Scott. **2017**. The relationship between entomological indicators of *Aedes aegypti* abundance and dengue virus infection. PLoS Negl. Trop. Dis. 11: e0005429.

DiMenna, M. A., R. Bueno, Jr., R. R. Parmenter, D. E. Norris, J. M. Sheyka, J. L. Molina, E. M. LaBeau, E. S. Hatton, and G. E. Glass. **2006**. Comparison of mosquito trapping method efficacy for West Nile virus surveillance in New Mexico. J. Am. Mosq. Control Assoc. 22: 246–253.

Eatman, S., and H. M. Strosnider. **2017**. CDC's national environmental public health tracking program in action: Case studies from state and local health departments. J. Public Heal. Manag. Pract. 23: S9–S17.

Fanshawe, T. R., and P. J. Diggle. **2012**. Adaptive Sampling Design for Spatio-Temporal Prediction, pp. 249–268. *In* Spat. Des. John Wiley & Sons, Ltd, Chichester, UK.

Fletcher, R. J., T. J. Hefley, E. P. Robertson, B. Zuckerberg, R. A. Mccleery, and R. M. Dorazio. **2019**. A practical guide for combining data to model species distributions. Ecology. 100: e02710.

Foley, D. H., R. C. Wilkerson, and L. M. Rueda. **2009**. Importance of the "What," "When," and "Where" of Mosquito Collection Events. J. Med. Entomol. 46: 717–722.

Fouet, C., P. Atkinson, and C. Kamdem. **2018**. Human interventions: Driving forces of mosquito evolution. Trends Parasitol. 34: 127–139.

Frumkin, H., J. Hess, G. Luber, J. Malilay, and M. McGeehin. **2008**. Climate change: The public health response. Am. J. Public Health. 98: 435–445.

Gu, W., T. R. Unnasch, C. R. Katholi, R. Lampman, and R. J. Novak. **2008**. Fundamental issues in mosquito surveillance for arboviral transmission. Trans. R. Soc. Trop. Med. Hyg. 102: 817–822.

Gujral, I. B., E. C. Zielinski-Gutierrez, A. LeBailly, and R. Nasci. **2007**. Behavioral risks for West Nile virus disease, northern Colorado, 2003. Emerg. Infect. Dis. 13: 419–425.

Hepworth, G., and B. J. Biggerstaff. **2017**. Bias correction in estimating proportions by pooled testing. J. Agric. Biol. Environ. Stat. 22: 602–614.

Holderman, C. J., S. A. Gezan, A. E. S. Stone, C. R. Connelly, and P. E. Kaufman. **2018**. Mosquitoes (Diptera: Culicidae) collected from residential yards and dog kennels in Florida using two aspirators, a sweep net, or a CDC trap. J. Med. Entomol. 55: 230–236.

Kilpatrick, A. M., and W. J. Pape. **2013**. Practice of epidemiology predicting human West Nile Virus infections with mosquito surveillance data. Am. J. Epidemiol. 178: 829–835.

Kline, D. L. 2006. Traps and trapping techniques for adult mosquito control. J. Am. Mosq. Control Assoc. 22: 490–496.

Li, R., L. Xu, O. N. Bjørnstad, K. Liu, T. Song, A. Chen, B. Xu, Q. Liu, and N. C. Stenseth. **2019**. Climate-driven variation in mosquito density predicts the spatiotemporal dynamics of dengue. Proc. Natl. Acad. Sci. U. S. A. 116: 3624–3629.

Lindley, D. V. 1983. Theory and practice of bayesian statistics. Stat. 32: 1–11.

Lindsey, N. P., J. A. Brown, L. Kightlinger, L. Rosenberg, and M. Fischer. **2012**. State health department perceived utility of and satisfaction with arboNET, the U.S. national arboviral surveillance system. Public Health Rep. 127: 383–390.

Longbottom, J., A. Krause, S. J. Torr, and M. C. Stanton. **2020**. Quantifying geographic accessibility to improve efficiency of entomological monitoring. PLoS Negl. Trop. Dis. 14: e0008096.

Maciel-de-Freitas, R., R. C. Peres, F. Alves, and M. B. Brandolini. **2008**. Mosquito traps designed to capture *Aedes aegypti (*Diptera: Culicidae) females: Preliminary comparison of Adultrap, MosquiTRAP and backpack aspirator efficiency in a dengue-endemic area of Brazil. Mem. Inst. Oswaldo Cruz. 103: 602–605.

Manica, M., B. Caputo, A. Screti, F. Filipponi, R. Rosà, A. Solimini, A. della Torre, and M. Blangiardo. **2019**. Applying the N-mixture model approach to estimate mosquito population absolute abundance from monitoring data. J. Appl. Ecol. 56: 2225–2235.

Meeraus, W. H., J. S. Armistead, and J. R. Arias. **2008**. Field comparison of novel and gold standard traps for collecting *Aedes albopictus* in northern Virginia. J. Am. Mosq. Control Assoc. 24: 244–248.

Moise, I. K., L. C. Zulu, D. O. Fuller, and J. C. Beier. **2019**. Persistent Barriers to Implementing Efficacious Mosquito Control Activities in the Continental United States: Insights from Vector Control Experts, pp. 3–20. *In* Curr. Top. Neglected Trop. Dis. IntechOpen.

Moy, B., R. Harrigan, and H. Godwin. **2018**. Building capacity to support the use of geospatial modeling for vectorborne disease control: West Nile virus as a case study. J. Environ. Health. 80: 24–31.

Moyes, C. L., J. Vontas, A. J. Martins, L. C. Ng, S. Y. Koou, I. Dusfour, K. Raghavendra, J. Pinto, V. Corbel, J. P. David, and D. Weetman. **2017**. Contemporary status of insecticide resistance in the major *Aedes* vectors of arboviruses infecting humans. PLoS Negl. Trop. Dis. 11: e0005625.

Nasci, R. S., and J. P. Mutebi. **2019**. Reducing West Nile virus risk through vector management. J. Med. Entomol. 56: 1516–1521.

National Coordination Office for Space-Based Positioning, Navigation, and T. 2019. GPS.gov: GPS Accuracy. (https://www.gps.gov/systems/gps/performance/accuracy/).

Petersen, L. R. 2019. Epidemiology of West Nile virus in the United States: Implications for arbovirology and public health. J. Med. Entomol. 56: 1456–1462.

Pollett, S., B. M. Althouse, B. Forshey, G. W. Rutherford, and R. G. Jarman. **2017**. Internet-based biosurveillance methods for vector-borne diseases: Are they novel public health tools or just novelties? PLoS Negl. Trop. Dis. 11: e0005871.

Powers, S. M., and S. E. Hampton. **2019**. Open science, reproducibility, and transparency in ecology. Ecol. Appl. 29: e01822.

Purse, B. V., N. Darshan, G. S. Kasabi, F. Gerard, A. Samrat, C. George, A. T. Vanak, M. Oommen, M. Rahman, S. J. Burthe, J. C. Young, P. N. Srinivas, S. M. Schäfer, P. A. Henrys, V. K. Sandhya, M. M. Chanda, M. V. Murhekar, S. L. Hoti, and S. K. Kiran. **2020**. Predicting disease risk areas through co-production of spatial models: The example of Kyasanur Forest Disease in India’s forest landscapes. PLoS Negl. Trop. Dis. 14: e0008179.

Reeves, W. C. and W. M. Hammon. **1962**. Epidemiology of the arthropod-borne viral encephalitides in Kern County, California, 1943-1952, by W. C. Reeves and W. McD. Hammon. In collaboration with W. A. Longshore, Jr., H. E. McClure and A. F. Geib. Berkeley.

Rund, S. S. C., I. K. Moise, J. C. Beier, and M. E. Martinez. **2019**. Rescuing troves of hidden ecological data to tackle emerging mosquito-borne diseases. J. Am. Mosq. Control Assoc. 35: 75–83.

Schröder, W. 2006. GIS, geostatistics, metadata banking, and tree-based models for data analysis and mapping in environmental monitoring and epidemiology. Int. J. Med. Microbiol. 296: 23–36.

Sedda, L., E. R. Lucas, L. S. Djogbnou, A. V. C. Edi, A. Egyir-Yawson, B. I. Kabula, J. Midega, E. Ochomo, D. Weetman, and M. J. Donnelly. **2019**. Improved spatial ecological sampling using open data and standardization: An example from malaria mosquito surveillance. J. R. Soc. Interface. 16: 20180941.

Sedda, L., B. M. Taylor, A. E. Eiras, J. T. Marques, and R. J. Dillon. **2020**. Using the intrinsic growth rate of the mosquito population improves spatio-temporal dengue risk estimation. Acta Trop. 208: 105519.

Toh, T. S., F. Dondelinger, and D. Wang. **2019**. Looking beyond the hype: Applied AI and machine learning in translational medicine. EBioMedicine. 47: 607–615.

Vanlalruia, K., N. Senthilkumar, and G. Gurusubramanian. **2014**. Diversity and abundance of mosquito species in relation to their larval habitats in Mizoram, North Eastern Himalayan region. Acta Trop. 137: 1–18.

Walter, S. D., S. W. Hildreth, and B. J. Beaty. **1980**. Estimation of infection rates in populations of organisms using pools of variable size. Am. J. Epidemiol. 112: 124–128.

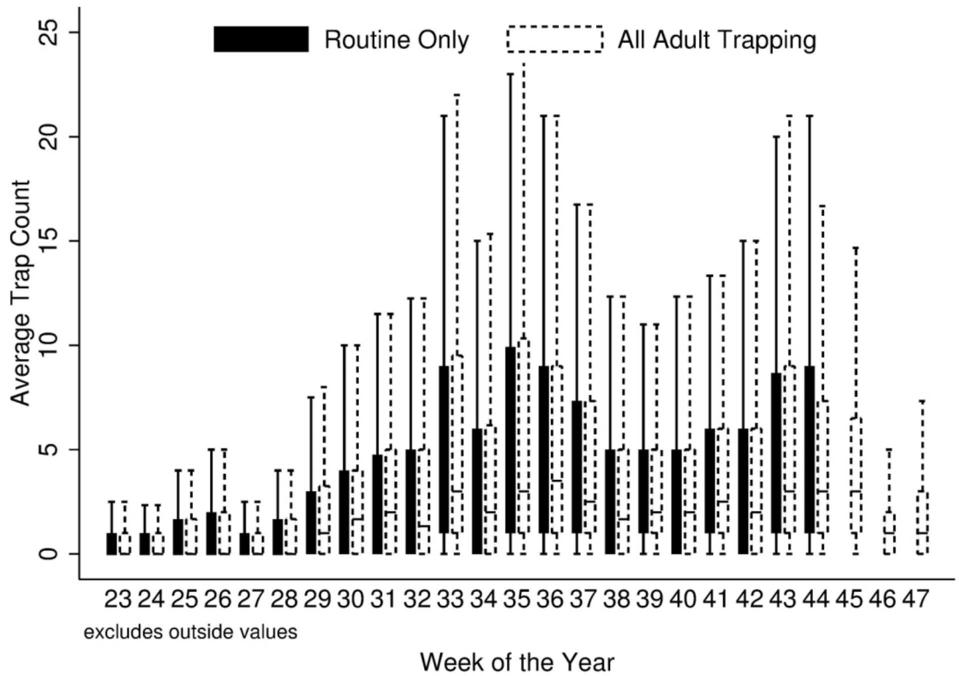
Warton, D. I. 2005. Many zeros does not mean zero inflation: Comparing the goodness-of-fit of parametric models to multivariate abundance data. Environmetrics. 16: 275–289.

Williams, G. M., and J. B. Gingrich. **2007**. Comparison of light traps, gravid traps, and resting boxes for West Nile Virus surveillance. J. Vector Ecol. 32: 285.

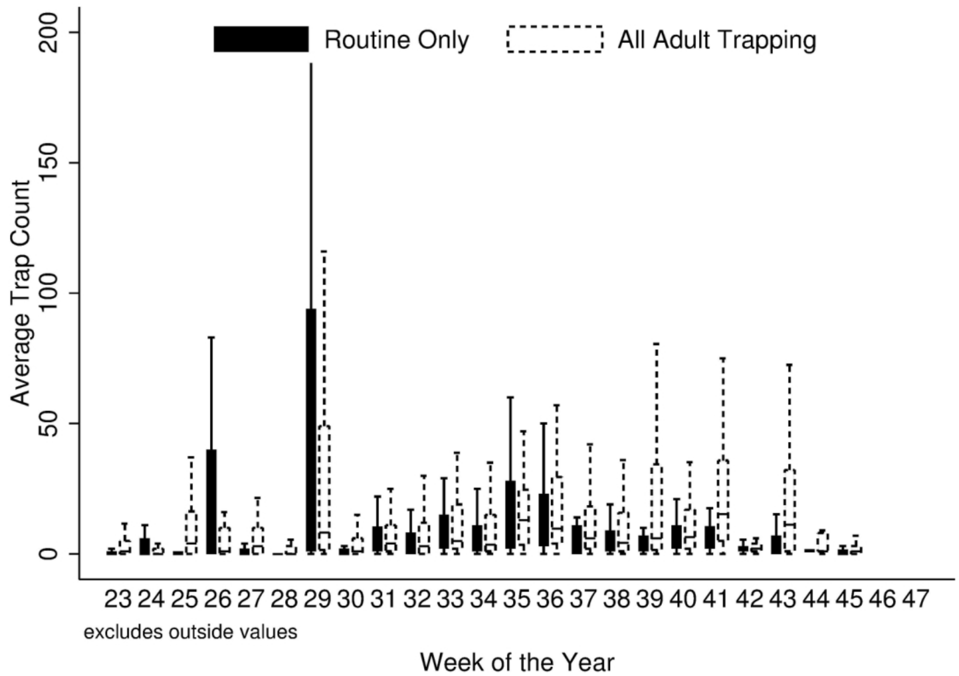
Yong, E. 2017. The case for sharing all of America's mosquito data. Atl.

Figure 1. Weekly mosquito abundance counts plotted by week to compare restricting to only those traps characterized in the database as "routine surveillance" (dark bars) with all other adult trapping (dashed bars). (a) Is from one county with a robust sampling program. This county accounts for 93% of the reported data in the database, of which 83.1% of the data was labeled as routine surveillance. (b) Is for all other counties combined, where 59.6% of the data were labeled as "routine surveillance". (c) Is a comparison of only routine surveillance (solid) and only response trapping (dashed). Note that this is for illustrative purposes and includes all mosquito species together and all trap type except ovi-traps. Outliers are not plotted.

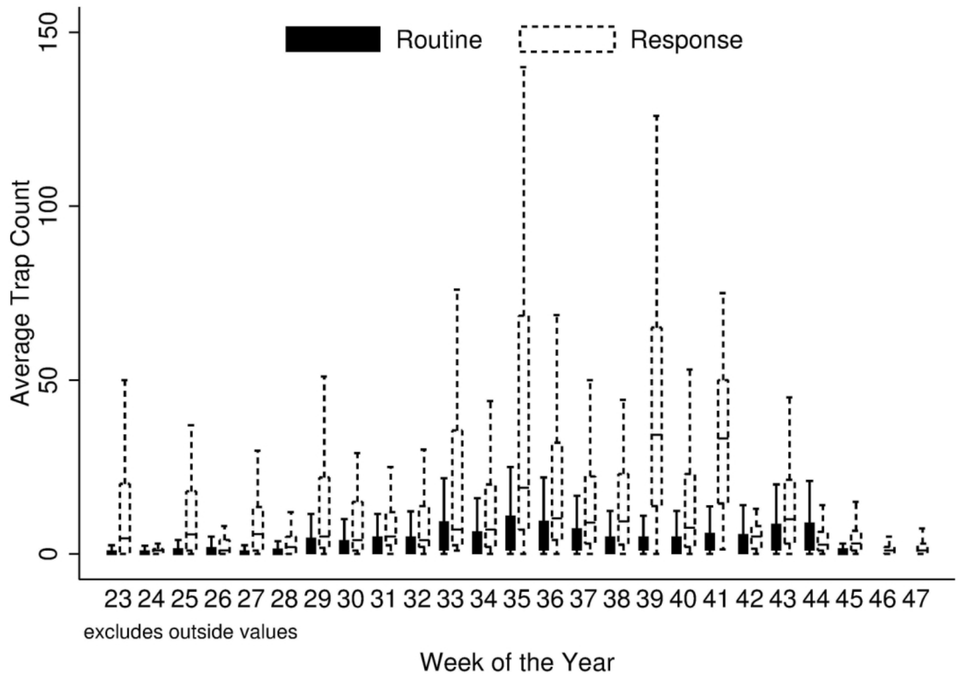
Figure 2. Weekly mosquito abundance counts plotted by week for one submitting county to illustrate the impact of recording empty traps. Dark bars indicate weekly trap counts with all data included. Dashed bars indicate weekly trap counts when zero values (e.g., empty traps) were artificially removed from the dataset. Outliers are not plotted.



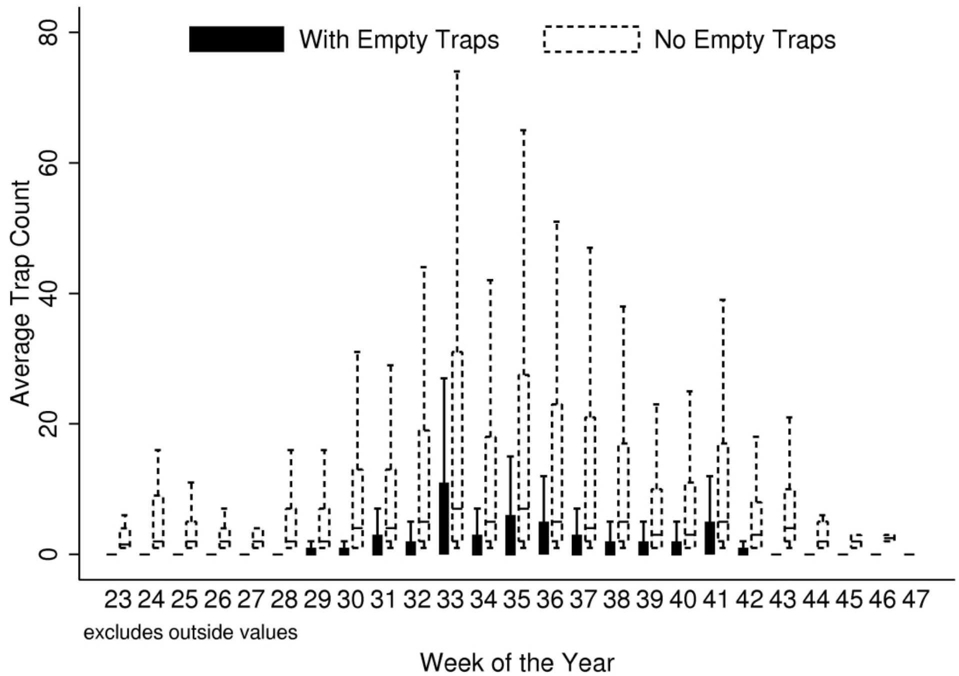
139x101mm (300 x 300 DPI)



139x101mm (300 x 300 DPI)



139x101mm (300 x 300 DPI)



139x101mm (600 x 600 DPI)