

OPEN

# Error associated with estimates of Minimum Infection Rate for Endemic West Nile Virus in areas of low mosquito trap density

S. Chakraborty<sup>1</sup> & R. L. Smith<sup>2\*</sup>

West Nile Virus (WNV) is a mosquito-borne infection that can cause serious illness in humans. Surveillance for WNV primarily focuses on a measure of infection prevalence in the *Culex* spp. mosquitoes, its primary vectors, known as the Minimum Infection Rate (MIR). The calculation of MIR for a given area considers the number of mosquitoes tested, but not the relative effort to collect mosquitoes, leading to a potential underestimation of the uncertainty around the estimate. We performed Value of Information analysis on simulated data sets including a range of mosquito trap densities in two well-studied counties in Illinois between 2005 and 2016 to determine the relative error introduced into MIR associated with changing the density of mosquito traps. We found that low trap density increases the potential for error in MIR estimation, and that it does so synergistically with low true MIR values. We propose that these results could be used to better estimate uncertainty in WNV risk.

West Nile Virus (WNV) causes an infectious disease in birds, horses and humans that is transmitted by the bite of infected mosquitoes<sup>1</sup>. In 80% of human cases, the disease does not produce any symptoms, but in a subset of patients it can cause febrile illness, joint pain, fatigue and weakness and about 1 in 150 people develop encephalitis or meningitis<sup>1</sup>. The disease first emerged in the United States in 1999, following which it has spread across the contiguous states and is now considered established. The first report of WNV in the state of Illinois was among dead birds in 2001; it has since spread to almost all counties in the state. From 1999 until 2017, the CDC reports that Illinois has had 2,458 human cases of WNV, the 5<sup>th</sup> highest number in the nation, including 1,553 cases of neuro-invasive disease<sup>1</sup>. It is a nationally notifiable disease and the state of Illinois has a surveillance system to monitor this pathogen, resulting in reports in 2017 of 90 human cases, 8 human deaths, 25 birds with WNV positivity and 2,022 positive mosquito batches<sup>2</sup>.

Birds serve as the main reservoir hosts for WNV. The virus, which belongs to the *Flaviviridae* family, is transmitted by mosquitoes in the genus *Culex*. Several species of *Culex* mosquitoes, such as *C. tarsalis*, *C. quinquefasciatus*, *C. stigmatosoma*, *C. thriambus*, *C. pipiens*, and *C. nigripalpus*, have been found to be able to transmit WNV<sup>3</sup>. Mosquitoes become infected when they feed on birds that harbor the virus; they are then capable of transmitting WNV to humans and horses<sup>1</sup>.

Surveillance for WNV is conducted primarily by testing mosquitoes, monitoring birds particularly in the family *Corvidae*, and sero-surveillance of sentinel chicken flocks and equines. The Illinois Department of Public Health monitors WNV by testing groups of up to 50 mosquitoes, dead perching birds (such as crows, blue jays, and robins), and testing sick horses and humans with West Nile virus-like symptoms<sup>2</sup>. Since, individual collection and testing of mosquitoes can be an expensive and arduous process, pooled samples are used to detect the presence of pathogens within the species- a concept first introduced by Dorfman in 1943<sup>4</sup>. Mosquitoes are thus collected for testing by setting traps throughout the state; the placement of traps and number of mosquitoes submitted for testing is determined by local public health departments or mosquito abatement districts. The mosquito abundance in traps can be affected by factors such as temperature<sup>5</sup>, rainfall<sup>6</sup>, structure of urban landscapes<sup>7,8</sup>, vegetation<sup>9</sup> and climatic variability<sup>10</sup>. Mosquitoes from these traps are collected in pools of up to 50 for viral testing. The results of mosquito testing are used to calculate the Minimum Infection Rate (MIR) for West

<sup>1</sup>Program in Ecology, Evolution & Conservation Biology, University of Illinois at Urbana-Champaign, Champaign, IL, USA. <sup>2</sup>Department of Pathobiology, University of Illinois College of Veterinary Medicine, Urbana, IL, USA. \*email: [rlsdvm@illinois.edu](mailto:rlsdvm@illinois.edu)

County	County		Proportion of Traps Sampled					
	Cook	DuPage	50	60	70	80	90	100
Number of Weeks	240	182	422	422	422	422	422	422
Number of Positive Pools per Week	21 [3, 103]	4 [1, 17]	5 [1, 23]	6 [1, 28]	7 [1, 33]	7 [1, 37]	8 [1, 42]	9 [1, 47]
Total Number of <i>Culex</i> Mosquitos per Week	12381 [7130, 18232]	2796 [2075, 3681]	2677 [1360, 6972]	3220.5 [1635, 8390]	3770 [1907, 9771]	4308 [2182, 11236]	4850 [2478, 12610]	5423 [2780, 14041]
MIR	0.0022 [0.0003, 0.0045]	0.0014 [0.0003, 0.0080]	0.0019 [0.0002, 0.0078]	0.0019 [0.0002, 0.0078]	0.0019 [0.0003, 0.0079]	0.0019 [0.0003, 0.0079]	0.0019 [0.0003, 0.0079]	0.002 [0.0003, 0.0078]
Number of Traps Used per Week	373 [229, 506]	92 [67, 116]	84 [46, 202]	101 [55, 242]	118 [64, 282]	134 [73, 322]	151 [82, 363]	168 [91, 402]
Density of Traps (traps/mi <sup>2</sup> )	0.39 [0.24, 0.54]	0.28 [0.20, 0.36]	0.16 [0.11, 0.23]	0.19 [0.13, 0.28]	0.22 [0.15, 0.33]	0.26 [0.17, 0.37]	0.29 [0.2, 0.42]	0.32 [0.22, 0.47]
Probability of False Negative	NA	NA	0.09	0.06	0.04	0.03	0.01	NA

**Table 1.** Description of observed and simulated data (median [interquartile range (IQR)]) used for estimating baseline MIR and to determine effect of lower density sampling. The probability of a false negative was calculated as the number of iterations in which MIR was 0 when  $MIR_{100}$  was greater than 0.

Nile Virus, which is defined as the number of positive pools of a particular mosquito species over a defined time period and area divided by the total number of mosquitoes in those pools. The underlying assumption of MIR is that there is just one infected individual within a pool of mosquitoes. Another method used by researchers to detect the infection rate is using the Maximum Likelihood Estimate (MLE) which is defined as the infection rate most likely observed given the testing results and an assumed probabilistic model (i.e., binomial distribution of infected individuals in a positive pool)<sup>11</sup>. An increase in MIR estimates often is assumed to increase risk of disease transmission to humans. The Centers for Disease Control & Prevention have shown that MIR is an important indicator in WNV surveillance systems that can be helpful in predicting patterns in virus activity, and thereby human cases in a given area<sup>12</sup>. Previous studies that have used MIR to detect WNV are Bernard *et al.*<sup>13</sup>, Kulasekara *et al.*<sup>14</sup> and Hadler *et al.*<sup>15</sup>. Therefore, prediction of MIR is a public health priority in areas endemic for WNV.

Under-sampling is believed to be a significant obstacle in developing robust prediction models for MIR. Under-sampling could be caused by not having enough traps set out within the geographic area of study, by not testing all mosquitoes collected in the traps, by low mosquito abundance in the traps, or by a combination of any of those factors. Additionally, MIR may not be the best measure of calculation if the virus to be detected is common or if pool sizes are too large<sup>11</sup>. Without sufficient data, MIR estimates are likely to be inaccurate and appropriate public health efforts will be difficult to determine. However, there is no current method for determining the error introduced into MIR by under-sampling.

Value of information analysis is a quantitative method to estimate the return on investment (value) produced by research<sup>16</sup>. This concept can be applied to compare the results of using an artificially reduced data set with that of the full data set to determine how much data are required to meet a specific criterion. The Value of Information (VOI) approach is increasingly becoming a useful tool with applications in prioritizing research decisions<sup>17</sup>, economic design of clinical trials<sup>18</sup>, as well as in treatment interventions<sup>19</sup> and social sciences<sup>20</sup>. As the roots of VOI lie in statistics and has wide ranging applications, it is therefore used as a guiding concept in this paper to deal with error associated with MIR. In this case, the value we are interested in estimating is the accuracy and precision of the MIR estimate, while the information we are using is varying mosquito trapping densities.

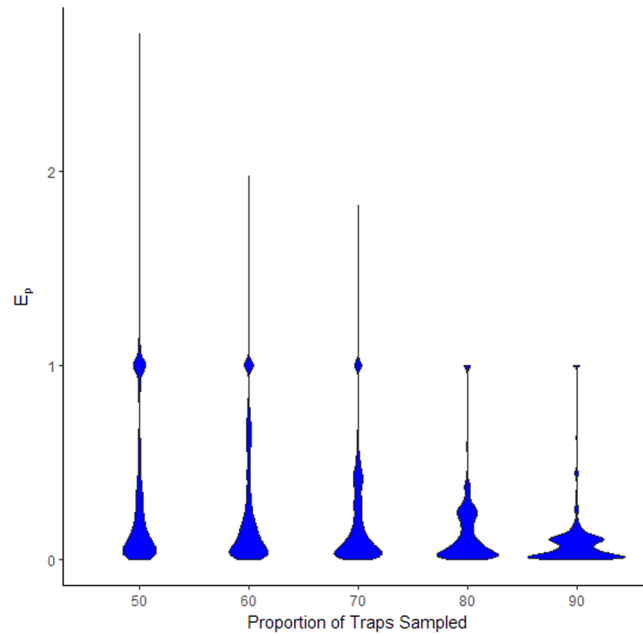
The objective of this study was to determine the error associated with calculation of WNV MIR at the county level in cases of low mosquito trap density. This analysis considered data obtained from mosquito traps that were set in different locations in Illinois between the periods of 2005–2016. Utilizing the value of information concept as outlined above, we determined the impact of low trap density (under-sampling) in the accuracy of MIR results.

## Results

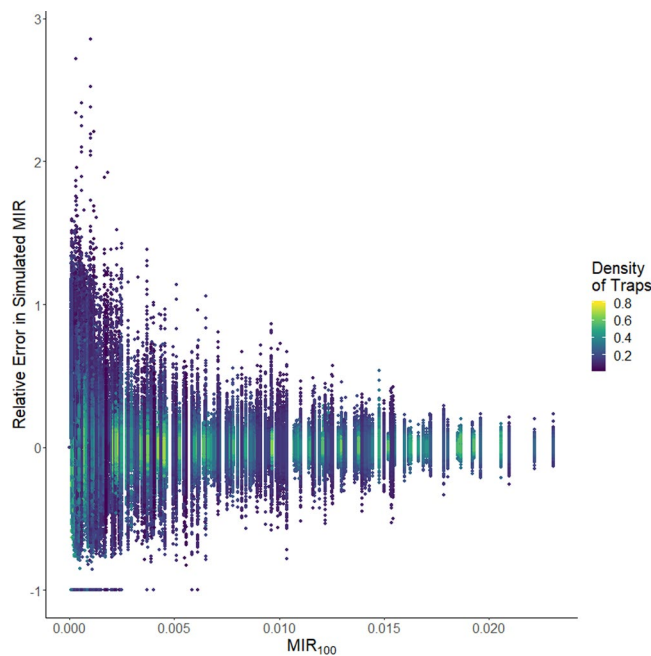
After selecting only weeks in which at least 50 pools of mosquitos were tested and at least one pool was positive, a total of 240 weeks of data were available for Cook County and 182 weeks of data were available for DuPage County (Table 1).

After randomly sampling subsets of mosquito trap data, the absolute relative error in estimated MIR for a county,  $E_p = \left| \frac{MIR - MIR_{100}}{MIR_{100}} \right|$ , was clearly skewed, with a high concentration near 0 and a long tail (Fig. 1). There was a secondary peak in frequency near  $E_p = 1$  due to the bounded distribution of MIR, which cannot fall below 0; any iteration in which no positive traps were sampled (probabilities shown in Table 1) would result in an  $E_p$  value of 1. This also resulted in a skew in the relative error,  $\frac{MIR - MIR_{100}}{MIR_{100}}$  (Fig. 2) towards more positive outliers.

The distribution of relative error in MIR was clearly wider when the density of traps decreased, and also when the observed  $MIR_{100}$  was low (Fig. 2). Results of a lognormal regression to determine the effect of trap density and  $MIR_{100}$  on relative absolute error showed that there was a significant synergy between the variables (Table 2). Likewise, the range of the 95% confidence interval around the MIR estimate tended to increase more when the density of the traps decreased (Fig. 3), but a higher estimate of  $MIR_{100}$  was associated with a higher range around the MIR estimate, although this effect was somewhat decreased with high trap density (Table 3). Information about model fit can be found in the Supplementary Information.



**Figure 1.** Distribution in simulated absolute relative error ( $E_p = \left| \frac{MIR_p - MIR_{100}}{MIR_{100}} \right|$ ) in MIR associated with sampling different proportions of mosquito trap data in Cook and DuPage counties on a weekly basis between 2005 and 2016. Each proportion of traps was randomly selected from all trap data available in a given week for 50 iterations.



**Figure 2.** Simulated relative error in MIR for Cook and DuPage counties created by randomly sampling only a subset of mosquito trap data. Color indicates the number of traps per square mile in the simulated data.  $MIR_{100}$  is the observed MIR using all data.

The equation from Table 2 could be used to calculate the expected potential absolute relative error (ARE) around an observed MIR by using the equation

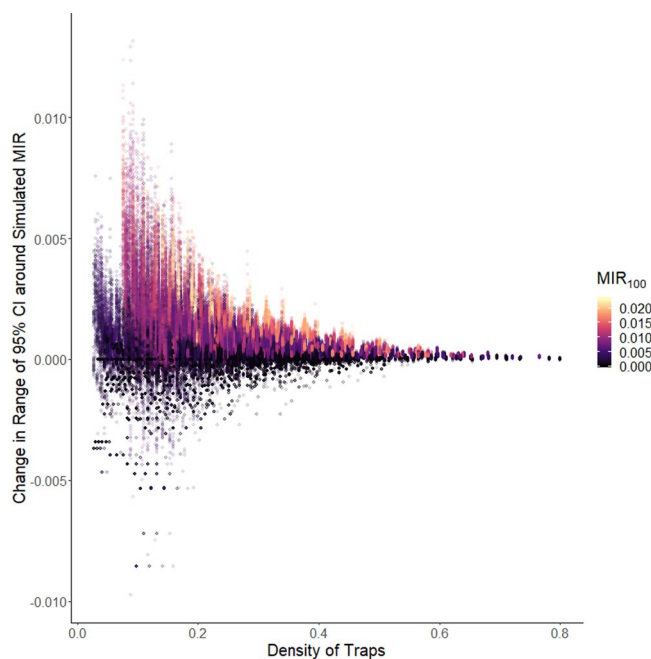
$$E_p = \exp[\beta_0 + \beta_1MIR + \beta_2Density + \beta_3MIR * Density] - 0.00001$$

Variable	Coefficient ( $\beta$ )	95% CI
Intercept	-0.24	-0.34, -0.15
$MIR_{100}$	-130	-150, -120
Trap Density	-5.9	-6, -5.7
$MIR_{100}$ *Trap Density	-1.0	-16, 14

**Table 2.** Results of lognormal regression for absolute relative error in estimated MIR,  $\log(MIR + 0.00001) \sim \beta_0 + \beta_1 MIR_{100} + \beta_2 Density + \beta_3 MIR_{100} * Density + \varepsilon_i$ .

Variable	Coefficient ( $\gamma$ )	95% CI
Intercept	0.00043	0.00035, 0.0005
$MIR_{100}$	0.29	0.28, 0.3
Trap Density	-0.0013	-0.0013, -0.0012
$MIR_{100}$ *Trap Density	-0.69	-0.7, -0.68

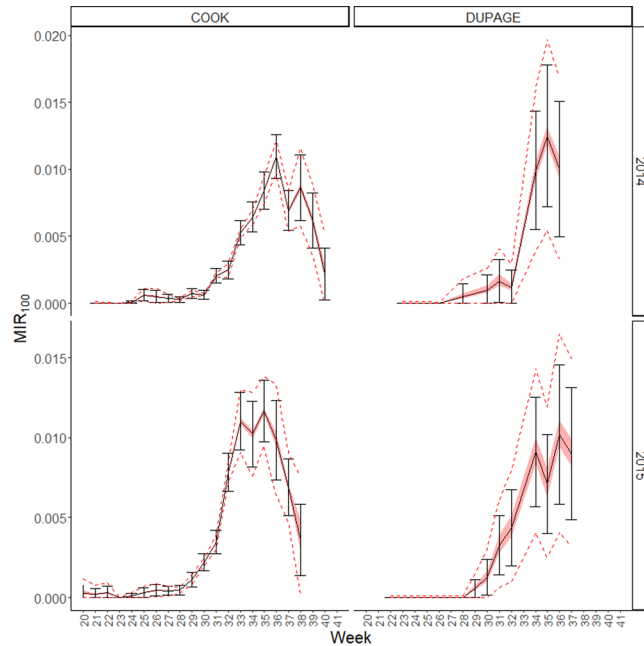
**Table 3.** Results of linear regression for range of the 95% confidence interval around estimated MIR,  $(Range - Range_{100}) \sim \gamma_0 + \gamma_1 MIR_{100} + \gamma_2 Density + \gamma_3 MIR_{100} * Density + \varepsilon_i$ .



**Figure 3.** Change in the range of the 95% confidence interval around the simulated MIR for Cook and DuPage counties created by randomly sampling only a subset of mosquito trap data, as a function of the number of traps per square mile in the simulated data. Color represents  $MIR_{100}$ , the observed MIR using all data.

where  $E_p = \left| \frac{MIR - MIR_{100}}{MIR_{100}} \right|$ . The expected potential range of  $MIR_{100}$  can be calculated as  $\frac{MIR}{1 \pm E_p}$ . Likewise, the expected change in the 95% confidence interval range can be calculated from the equation in Table 3. When the regression equation from Table 2 is applied to the observed data for Cook and DuPage counties, the range of predicted values is shown to be small (Fig. 4): at current trap density, the median range around the MIR is 0.18, which is only smaller than the observed 95% confidence interval, which has a median of 2.1. This is likely due to the high trap density, large number of pools, and high  $MIR_{100}$  in Cook County for this period. When the model is used to predict potential for error in example years from Will, McHenry, and Lake Counties (Fig. 5), where trap density is lower, it is seen that the error associated with MIR calculated from all traps is quite high; the median range around the MIR is 2.0. This is only slightly less than the observed 95% confidence interval, which has a median of 3.8, likely due to the small number of pools tested.

One factor of interest to mosquito control officials is the ability to detect a rise in MIR. We examined the simulated data for false negatives, circumstances in which the observed  $MIR_{100}$  was non-zero but the simulated  $MIR_p$  was zero (Fig. 6). As with the absolute relative error, the probability of a false negative was significantly increased with low trap density and with low  $MIR_{100}$ , and the effect of trap density and  $MIR_{100}$  was synergistic (see Supplementary information). However, the effect of  $MIR_{100}$  was much greater than that of trap density.



**Figure 4.** Predicted error associated with decreased West Nile Virus sampling in Cook and DuPage County for 2 example years. Black solid lines show the observed MIR with error bars showing the 95% confidence intervals, while the red shaded bar is the predicted error around the mean and the red dashed lines show the predicted 95% confidence intervals around the upper and lower bounds of the potential MIR estimate.

## Discussion

There are various environmental factors that contribute to arbovirus transmission. Low levels of the virus might thrive within the host and reservoir populations and the exact conditions that lead to widespread outbreaks are difficult to pinpoint. Therefore, constant surveillance of mosquitoes and the sentinel organisms are a critical aspect of public health activities especially for West Nile Virus (WNV).

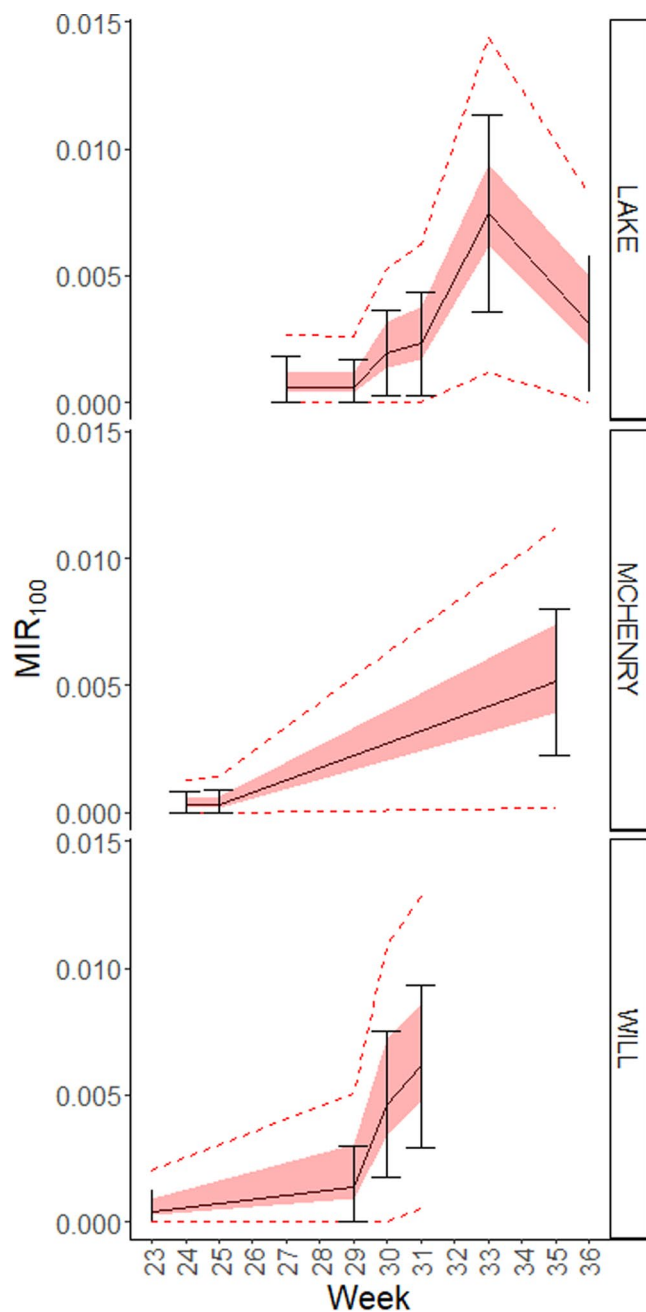
We found that if the true MIR is low, higher trap density is needed to accurately estimate MIR. This is intuitive; the sample size necessary to find a disease in a population is inversely proportional to the prevalence of the disease. Current estimates of MIR error are based on basic statistical theory, with sample size (total number of pools) responsible for much of the estimation<sup>21</sup>. This may result in underestimation of error around low MIR values and areas with low trap density but high trap numbers (such as a large geographical area).

Our study also provides an algorithm by which MIR error can be estimated. We believe that this error calculation can be incorporated into public health planning, giving decision makers a better sense of the potential range in MIR. Importantly, we have found that the probability of failing to detect a non-zero MIR was significantly impacted by the density of traps, especially when the actual MIR is low.

Our study was limited by the use of observed MIR, rather than true MIR, to calculate error. However, the mosquito surveillance systems of Cook and DuPage counties are extremely comprehensive, and the resulting calculated MIRs are likely to approach the true value in most cases. In fact, our model predicts low error in Cook County with only 10% of existing trap data, indicating that the full data set may be sufficient for the purposes of this study.

Researchers working on surveillance of arboviruses are constantly trying to optimize the tools and methods to better estimate disease risk and transmission in order to inform public health measures. Under-sampling has previously been identified as one of the most common sources of error in determining the mosquito infection rates<sup>22</sup>. For instance, DeFelice *et al.* (2017) used data assimilated from MIR and human case reports to model WNV transmission in New York and to generate retrospective forecasts of past WNV outbreaks in Long Island<sup>23</sup>. The authors realized for the model to work effectively in predicting outbreaks, it relied on timely availability of mosquito infection rate data, which can vary depending on the methods used for mosquito sampling. Bustamante and Lord (2010) also discussed other sources of error that are introduced while conducting mosquito surveillance such as temperature, trapping methods used for mosquito sampling, assays used for virus detection and the MIR vs MLE approach<sup>24</sup>. Thus, they suggest using other surveillance indicators such as historical baseline data and mosquito population size along with MIR for determining arboviral disease transmission.

The results of this study demonstrate that WNV surveillance in mosquitoes can be affected by under-sampling. However, the effect of this under-sampling on error in MIR estimates may now be calculated using our approach. It is important to note that we assumed that the spatial distribution of the traps was not a factor; this is a simplification that should be addressed in future studies. Gu *et al.* (2008) provide recommendations to better estimate MIR when there are not enough samples or in areas with low level transmission, recommending either “targeted surveillance (increased sampling at locations of higher transmission likelihood) or estimating MIR during periods of high transmission, thereby shifting from detection of mosquito infection to estimation of the transmission intensity, while expanding the number of sampling sites to evaluate the range of arboviral transmission”<sup>25</sup>. The placement of traps should also take into account landscape features and other factors likely to affect surveillance efforts.



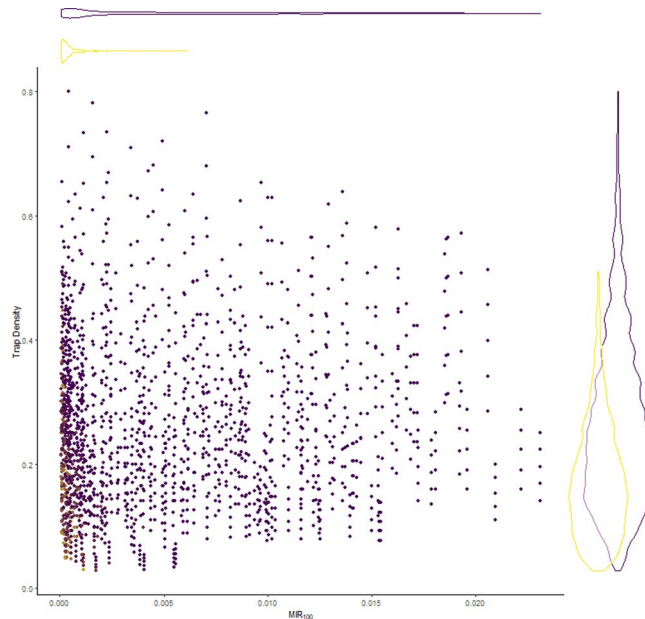
**Figure 5.** Predicted MIR error associated with trap density in Will, McHenry, and Lake Counties for one example year each. Black solid lines show the observed MIR with error bars showing the 95% confidence intervals, while the red shaded bar is the predicted error around the mean and the red dashed lines show the predicted 95% confidence intervals around the upper and lower bounds of the potential MIR estimate.

Public health surveillance of diseases like West Nile virus and other mosquito-borne infections is essential and can be conducted smoothly via concerted efforts of public health agencies, health departments and research institutions. The setup of adequate number of traps in different counties to regularly monitor the MIR in the mosquito populations is critical so that necessary public health efforts can be initiated before outbreaks occur. This paper shows that in areas where trap density is low, the method utilized here can be used to detect accurately the error in MIR, which can inform appropriate disease control and prevention measures.

### Methods

Data were obtained from the Illinois Department of Public Health mosquito surveillance database, which collects mosquito trap testing information from major stakeholders such as public health departments and mosquito abatement districts in the state of Illinois. Mosquito trap data for 4 counties in Illinois (DuPage, Cook, Will, and Lake) were obtained for the years 2005 to 2016. Trap density (traps per square mile) was calculated for each week





**Figure 6.** Simulated trap density in which the simulated MIR was 0 (in yellow) or non-zero (in blue) when the observed  $MIR_{100}$  was non-zero.

by dividing the number of traps tested by the total area of the county. All analyses were performed in R<sup>26</sup>. Analysis was performed using data from Cook and DuPage counties; all other county data were used for illustration of potential impact. Only weeks in which at least 50 pools were tested were included in the analysis.

For each county in each week, all data from  $p$  percent of traps ( $p \in \{50\%, 100\%\}$ ) were removed at random and the remaining trap data were used to calculate the simulated  $MIR_p$  using the binGroup package<sup>27</sup>. Simulated  $MIR_p$  was compared to the observed MIR with 100% of data,  $MIR_{100}$ , for that county-week combination, and absolute relative error was calculated as  $E_p = \left| \frac{MIR_p - MIR_{100}}{MIR_{100}} \right|$ . This was repeated 50 times for each county-week combination. Weeks in which  $MIR_{100}$  was 0 were removed from the analysis.

The error  $E_p$  was examined visually for each level of  $p$  and determined to be log-normally distributed with zero-inflation using the fitdistrplus package<sup>28</sup>. Due to the zero-inflation, we performed shifted logistic transformation by adding a conservative number (0.00001) to all  $E_p$  prior to log transformation. The impact of trap density and  $MIR_{100}$  on  $\log(E_p)$  was analyzed using mixed linear regression modeling with the lme4 package<sup>29</sup>, using the county-week combination as a random effect to account for repeated sampling. Two-way interactions were included, and all effects were considered significant at the  $\alpha = 0.05$  level. The 95% confidence interval around  $MIR_{100}$  and  $MIR_p$  were calculated using the binGroup package<sup>27</sup>, and the difference in the confidence interval range was calculated as  $range_{100} - range_p$ . The impact of trap density and  $MIR_{100}$  on the difference in the confidence interval range was analyzed using mixed linear regression modeling, as described above. All figures were created using the ggplot2 package<sup>30</sup>.

## Data availability

All data and analysis scripts are available at [https://github.com/rlsdvm/MIR\\_VOI](https://github.com/rlsdvm/MIR_VOI).

Received: 7 June 2019; Accepted: 27 November 2019;

Published online: 13 December 2019

## References

- Centers for Disease Control and Prevention. West Nile Virus-Symptoms, Diagnosis and treatment. Available at, [www.cdc.gov/westnile/symptoms/index.html](http://www.cdc.gov/westnile/symptoms/index.html) (2018).
- Illinois Department of Public Health. West Nile Virus (WNV). Available at, <http://www.dph.illinois.gov/topics-services/diseases-and-conditions/west-nile-virus> (2018).
- Colpitts, T. M., Conway, M. J., Montgomery, R. R. & Fikrig, E. West Nile Virus: biology, transmission, and human infection. *Clin. Microbiol. Rev.* **25**, 635–648 (2012).
- Dorfman, R. The Detection of Defective Members of Large Populations. *Ann. Math. Stat.* **14**, 436–440 (1943).
- Beck-Johnson, L. M. *et al.* The importance of temperature fluctuations in understanding mosquito population dynamics and malaria risk. *R. Soc. open Sci.* **4**, 160969 (2017).
- Epstein, P. R. Climate change and emerging infectious diseases. *Microbes Infect.* **3**, 747–754 (2001).
- Deichmeister, J. M. & Telang, A. Abundance of West Nile virus mosquito vectors in relation to climate and landscape variables. *J. Vector Ecol.* **36**, 75–85 (2010).
- Keating, J. *et al.* Spatial and temporal heterogeneity of Anopheles mosquitoes and Plasmodium falciparum transmission along the Kenyan coast. *Am. J. Trop. Med. Hyg.* **68**, 357–365 (2003).
- Gardner, A. M. *et al.* Terrestrial vegetation and aquatic chemistry influence larval mosquito abundance in catch basins, Chicago, USA. *Parasit. Vectors.* **6**, 9 (2013).

10. Chaves, L. F. *et al.* Climatic variability and landscape heterogeneity impact urban mosquito diversity and vector abundance and infection. *Ecosphere* **2**, 1–21 (2011).
11. Gu, W., Lampman, R. & Novak, R. J. Problems in estimating mosquito infection rates using minimum infection rate. *J. Med. Entomol.* **40**, 595–596 (2003).
12. Centers for Disease Control and Prevention. West Nile Virus in the United States: Guidelines for Surveillance, Prevention, and Control. Available at, <https://www.cdc.gov/westnile/resources/pdfs/wnvguidelines.pdf> (2013).
13. Bernard, K. A. *et al.* NY State West Nile Virus Surveillance Team. West Nile virus infection in birds and mosquitoes, New York State, 2000. *Emerg. Infect. Dis.* **7**, 679–685 (2001).
14. Kulasekera, V. L. *et al.* West Nile virus infection in mosquitoes, birds, horses, and humans, Staten Island, New York, 2000. *Emerg. Infect. Dis.* **7**(4), 722 (2001).
15. Hadler, J. *et al.* West Nile virus surveillance in Connecticut in 2000: an intense epizootic without high risk for severe human disease. *Emerg. Infect. Dis.* **7**(4), 636 (2001).
16. Wilson, E. C. F. A practical guide to value of information analysis. *PharmacoEconomics* **33**, 105–121 (2015).
17. Minelli, C. & Baio, G. Value of Information: A Tool to Improve Research Prioritization and Reduce Waste. *PLoS Med* **12**(9), e1001882 (2015).
18. Claxton, K. & Posnett, J. An economic approach to clinical trial design and research priority-setting. *Health economics.* **5**(6), 513–24 (1996).
19. Dong, H., Coyle, D. & Buxton, M. Value of information analysis for a new technology: computer-assisted total knee replacement. *Int. J. Technol. Assess. Health Care.* **23**(3), 337–42 (2007).
20. Mohseninejad, L., van Baal, P. H., van den Berg, M., Buskens, E. & Feenstra, T. Value of information analysis from a societal perspective: a case study in prevention of major depression. *Value Health.* **16**(4), 490–7 (2013).
21. Bross, I., Anderson, R. L. & Bancroft, T. A. Statistical Theory in Research. I.: Basic Statistical Theory, Statistical Theory in Research. II, Analysis of Experimental Models by Least Squares. *Q. Rev. Biol.* **29**(1), 100–101 (1954).
22. Katholi, C. R. & Unnasch, T. R. Important experimental parameters for determining infection rates in arthropod vectors using pool screening approaches. *Am. J. Trop. Med. Hyg.* **74**(5), 779–785 (2006).
23. DeFelice, N. B., Little, E., Campbell, S. R. & Shaman, J. Ensemble forecast of human West Nile virus cases and mosquito infection rates. *Nat. Commun.* **8**, 14592 (2017).
24. Bustamante, D. M. & Lord, C. C. Sources of error in the estimation of mosquito infection rates used to assess risk of arbovirus transmission. *Am. J. Trop. Med. Hyg.* **82**(6), 1172–1184 (2010).
25. Gu, W., Unnasch, T. R., Katholi, C. R., Lampman, R. & Novak, R. J. Fundamental issues in mosquito surveillance for arboviral transmission. *Trans. Royal Soc. Trop. Med. Hyg.* **102**(8), 817–822 (2008).
26. The R Development Core Team. R: A Language and Environment for Statistical Computing. **0** (2009).
27. Zhang, B., Bilder, C., Biggerstaff, B.J., Schaarschmidt, F. & Hitt, B. binGroup: Evaluation and Experimental Design for Binomial Group Testing (2018).
28. Delignette-Muller, M. & Dutang, C. fitdistrplus: An R package for fitting distributions. *J. Stat. Softw.* **64**, 1–34 (2015).
29. Bates, D. *et al.* Linear mixed-effects models using Eigen and S4 (2014).
30. Wickham, H. ggplot2: Elegant Graphics for Data Analysis. (Springer-Verlag, 2016).

## Acknowledgements

The authors would like to acknowledge the contribution of Dr. Marilyn O'Hara Ruiz, who made essential contributions to the conception and implementation of this study prior to her death. This publication was supported by Cooperative Agreement #U01 CK000505, funded by the Centers for Disease Control and Prevention. Its contents are solely the responsibility of the authors and do not necessarily represent the official views of the Centers of Disease Control and Prevention or the Department of Health and Human Services.

## Author contributions

R.L.S. conceived the study and conducted the statistical analysis. S.C. conducted the VOI analysis. Both S.C. and R.L.S. interpreted the results and wrote the manuscript.

## Competing interests

The authors declare no competing interests.

## Additional information

**Supplementary information** is available for this paper at <https://doi.org/10.1038/s41598-019-55632-7>.

**Correspondence** and requests for materials should be addressed to R.L.S.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2019