



HHS Public Access

Author manuscript

Stat Commun Infect Dis. Author manuscript; available in PMC 2024 April 12.

Published in final edited form as:

Stat Commun Infect Dis. 2020 November 13; 12(1): . doi:10.1515/scid-2020-0001.

Estimation of time-dependent arbovirus infection risk in blood and tissue donations

Brad J. Biggerstaff*

Centers for Disease Control and Prevention, Division of Vector-Borne Diseases, Fort Collins, Colorado, USA

Abstract

West Nile virus (WNV) outbreaks raise the concern of WNV infection in donated blood and blood products destined for transfusion. We describe methods we developed to estimate time-dependent risk of WNV infection in donated blood, including improvements not previously detailed. The methods are then extended for use in estimation of the risk of WNV infection in donated cadaveric tissues by introducing stratification and stratum-specific weighting to address novel aspects of this application. Data from the WNV outbreak in Colorado in 2003 are used to estimate risk for donated cardiac tissue.

Keywords

blood safety; Poisson process; public health; tissue safety

Introduction

Relatively soon after the documentation in 1999 of West Nile virus (WNV) circulation in North America, it was recognized that during a WNV outbreak the probability or “risk” of WNV infection in blood donors may reach sufficiently high levels to pose a threat to the blood supply. In two papers, Biggerstaff and Petersen (2002, 2003) used surveillance data for WNV neuroinvasive disease (WNND; e.g., meningitis, encephalitis, acute flaccid paralysis) collected by local and state public health agencies and reported to the Centers for Disease Control and Prevention (CDC) along with exogenous information on WNV viral kinetics in blood, the inapparent-to-apparent infection ratio, and the overall rate of asymptomatic infection to estimate the time-dependent risk of donors being WNV-infected at donation in regions and during timeframes of WNV epidemic activity. In the flavivirus family, WNV is a mosquito-borne virus, infection with which may result in no illness (asymptomatic infection), a relatively mild febrile illness, and severe illness involving neurological processes that may lead to long-term incapacitation and death. More severe outcomes are less common, and severity tends to increase with age. By far the most common

This work is licensed under the Creative Commons Attribution 4.0 International License.

*Corresponding author: **Brad J. Biggerstaff**, Centers for Disease Control and Prevention, Division of Vector-Borne Diseases, Fort Collins, Colorado, USA, bkb5@cdc.gov.

Author contributions: All authors have accepted responsibility for the entire content of this manuscript and approved its submission.

Competing interests: Authors state no conflict of interest.

route of WNV transmission to humans is from the bite of an infected mosquito. Other modes of transmission have been documented, however, including laboratory exposure; from mother to baby during pregnancy, delivery, or breast feeding; and via blood transfusion or organ donation. Coincidentally, the first documented case of WNV transmission from donated, infected blood occurred in 2002, just weeks after the first estimates of such risk were published for the 1999 outbreak in Queens, New York City (Biggerstaff and Petersen 2002; Centers for Disease Control and Prevention 2002; Iwamoto et al. 2003).

Technical details of the initial statistical methods described here were used in the original paper for the Queens outbreak were presented in its appendix, with less formal description provided in the body of the text. Subsequent applications to other geographic areas, time scales, and viruses (Biggerstaff and Petersen 2003; Petersen, Tomashek, and Biggerstaff 2012; Shang et al. 2007, 2016) introduced improvements, extensions, and simplifications to the methods, though none of these papers contained detailed information on these changes from the original development. A more recent application (Biggerstaff and Petersen 2020) of the general methodological ideas to estimation of the risk of WNV infection in cadaveric tissue donations necessitated further methodological enhancements, and the present paper collects all of these updates into a single, detailed presentation.

We begin by providing background on the estimation goal and the information needed to reach it in Background & information required. Methodological and computational development then recapitulates and expands the appendix in the original paper for the Queens outbreak (Biggerstaff and Petersen 2002), with extensions to incorporate the asymptomatic proportion and uncertainty in this, as well as uncertainty in the inapparent-to-apparent infection ratio (Biggerstaff and Petersen 2003); and subsequently an extension to incorporate multi-year estimation (Petersen, Tomashek, and Biggerstaff 2012). Refinements for estimation of WNV in cadaveric tissue donations introduces new refinements to these methods to accommodate stratification and differential weighting by stratum to address specific characteristics of estimation for WNV infection in cadaveric tissue donations (Biggerstaff and Petersen 2020).

In Development using a Poisson process model for the underlying case onset times we detail how our resampling implementation can be viewed as assuming an inhomogeneous Poisson process model for the underlying WNND case onset times and estimating the associated cumulative rate function using nonparametric Gaussian kernel smoothing, which we indicated but did not detail in the Discussion of Biggerstaff and Petersen (2002). In Application: cardiac tissue, Colorado, 2003 we illustrate the updated methods on a real data set, the full analysis for which is reported in Biggerstaff and Petersen (2020). The paper closes with a summary of the computational resources used followed by a discussion section.

Background & information required

The recognition that the probability or risk of WNV transmission from donated blood may be unacceptably high during an active, regional outbreak of WNV was the motivation for the original applications discussed here. Because WNV activity is necessarily seasonal in

North America, being a mosquito-borne virus, estimation methods were developed to reflect this time-dependence. In the United States, WNND cases are reported through established public health surveillance to the CDC with accompanying information, including the date that the patient first exhibited symptoms. Because WNND is a severe illness, surveillance systems capturing its occurrence in the US are thought to be essentially complete, and so epidemic curves developed using these data can be used to characterize the population-level transmission dynamics in the general population. These WNND case onset data, therefore, form the basis for the estimation methods used here. For the example from the WNV outbreak in Colorado, 2003, the WNND case onset data can be summarized in an epidemic curve, shown in Figure 1 for cases aged 13–55 for the application we present.

Mostashari et al. (2001) reported an estimate of the ratio of the number of WNV infections to the number of WNND cases in a population subject to WNV transmission. Since we have available the WNND case onset dates from surveillance, we can use this *inapparent-to-apparent infection ratio* to estimate the number of infections in the population. This ratio is the key parameter in relating the observed surveillance data to population-level infections. Original applications used a single inapparent-to-apparent infection ratio for the whole population; more recent work (Carson et al. 2012; Williamson et al. 2016) has refined these estimates by age group and sex, which we are, therefore, now able to incorporate into our analyses, as we show below.

When focusing specifically on estimation for WNV infection in blood and tissue donations, we restrict estimates of the number infections in the population in two ways. Because transmission of WNV occurs only when a donation contains or is *infected* with viable virus (termed *viremic* in blood), and because the duration of such infection is transient, estimates of *when* a potential donor was infected are necessary to characterize the risk. Using the WNND case onset times as anchor times, we use estimates of the duration from infection to symptom onset—the *incubation* period—and then estimates of the duration of infection to do this. Finally, because donors are necessarily asymptomatic—that is, they do not exhibit any symptoms of illness—and are screened for symptoms that would require deferral from either blood or tissue donation, we account for this using an estimate of the proportion of asymptomatic infections (Mostashari et al. 2001).

Methodological and computational development

We develop the methods considering specifically applications to estimation of WNV transmission risk via blood or tissue donation. The details, however, are written to be able to be applied more generally. The intent in the development, therefore, is that whenever densities or distributions or associated parameters are required to be specified, the user would specify such in the application under consideration. Whenever possible in application, we make distributional assumptions and specifications based on published data. For example, below we specify a Weibull distribution for the duration of WNV viremia based on data reported in Southam and Moore (1954), but in another application, a different distributional assumption for the duration of the pathogen's duration in blood or tissues may be more appropriate. In such case, in the algorithm presented in Estimation algorithm below, the reader should utilize the reader's own distributional model assumptions.

General approach, motivation, and estimation

For each $i = 1, 2, \dots, n$, let Y_i be observed WNNND case onset times (recorded to the day) in chronological order, and set $Y_1 = 0$ and $Y_n = T$. Further, for each case i , let V_{0i} be the duration from time of infection to symptom onset, and let V_{1i} be the duration of infection. Assume that the timing and duration of infection are independent of symptom onset time, and that the V_{0i} are independent and identically distributed, as are the V_{1i} ; finally assume that the V_{0i} are independent of the V_{1i} .

The case onset dates Y_i are recorded to the day, so to reflect the underlying continuous nature of the infection process in the environment and population we statistically smooth these observations by setting $X_i = Y_i + h\varepsilon_i$, where $h > 0$ is a smoothing parameter and the ε_i are independent standard normal deviates. We refer to “smoothing” here somewhat informally, though we note that in computation, generation of a value X_i by adding the scaled noise $h\varepsilon_i$ for a generated Gaussian deviate ε_i is equivalent to sampling from a Gaussian kernel density smooth of the original $\{Y_i\}$. Further, we note that specification of h is important and related to the problem of density estimation or Poisson intensity estimation, and we expand on this in Development using a Poisson process model for the underlying case onset times. We have used the value of Sheather and Jones (1991), though on occasions when this value for h was not computable due to insufficient data, we used the rule-of-thumb estimator given in Scott (1992). To preserve seasonality for multi-season data, we have estimated h using within-season information, for example by considering only day-of-year for multi-year observations of data with annual cycles, or by computing different h values for each year and averaging these or resampling from them.

Viewing the smoothed case onset times X_i as anchor times, the values $X_i - V_{0i}$ and $X_i - V_{0i} + V_{1i}$ represent the initial and terminal times of infection, respectively, for case i . For a given time, $t \in (0, T)$, then, case i is viremic at t whenever $X_i - V_{0i} < t < X_i - V_{0i} + V_{1i}$. We count the number of cases viremic at t using the function

$$V(t) = \sum_{i=1}^n I_{(X_i - V_{0i}, X_i - V_{0i} + V_{1i})}(t) \quad (1)$$

where $I_A(x) = 1$ if $x \in A$ and $I_A(x) = 0$ if $x \notin A$.

We interpret a particular realization of this random function $V(t)$ as representative of such realizations, where the random mechanisms are the environmental conditions and personal behaviors leading to exposure and subsequent advancement to infection in the population. The expected value $E[V(t)]$ is the mean of such a theoretical population and is the basis for the quantity we want to estimate.

As we saw, the inapparent-to-apparent infection ratio, ρ , with estimate $\hat{\rho}$, provides a means to inflate the observed counts of WNNND cases to estimate the number of infections in the population. In particular, writing u_+ for the next integer greater than u , the quantity $\hat{\rho}_+ V(t)$

estimates the number of individuals in the population who are viremic at time t . It will be convenient below to recast the quantity $\hat{\rho}_+ V(t)$ as follows. Since $V(t)$ counts the number of observed cases, $V(t)$ is a non-negative integer at each t . (In addition to the desire to produce an integer result, the use of $\hat{\rho}_+$ is expected to provide a conservative estimate, since $E[\hat{\rho}_+] \leq E[\hat{\rho}]_+$.) Then write $\hat{\rho}_+ V(t)$ as a simple expansion of the expression in Eq. (1),

$$\hat{\rho}_+ V(t) = \sum_{j=1}^{\hat{\rho}_+} \sum_{i=1}^n I_{(X_i - v_{0i}, X_i - v_{0i} + v_{1i})}(t) = \sum_{j=1}^{\hat{\rho}_+} \sum_{i=1}^n I_{(X_i^{(j)} - v_{0i}^{(j)}, X_i^{(j)} - v_{0i}^{(j)} + v_{1i}^{(j)})}(t)$$

which is just replication of $V(t)$ a total of $\hat{\rho}_+$ times, setting each $X_i^{(j)} = X_i$, etc.

Because we are interested in estimating the probability of a *donor* being viremic, and since donors are necessarily asymptomatic, we randomly truncate infection times at symptom onset for the proportion of infections expected to develop symptoms. This gives us, finally, an estimate of the number of *asymptomatic and WNV infected* individuals in the population at time t by adjusting $\hat{\rho}_+ V(t)$ to

$$\hat{\rho}_+ V^\star(t) = \sum_{j=1}^{\hat{\rho}_+} \sum_{i=1}^n I_{(X_i^{(j)} - v_{0i}^{(j)}, X_i^{(j)} - B_i^{(j)} [v_{0i}^{(j)} - v_{1i}^{(j)}])}(t) \quad (2)$$

where $B_i^{(j)}$ *iid* Bernoulli $(1 - p_a)$, and p_a is the probability that an infection is asymptomatic. This expression $\hat{\rho}_+ V^\star(t)$ is the statistic we base inference upon, and its expected value is the focus of our inference.

Our computation of an estimate of $E[\hat{\rho}_+ V^\star(t)]$ proceeds using statistical simulation, using both the bootstrap and Monte Carlo integration. (We previously described the Monte Carlo integration as akin to imputation, as this approach samples from or imputes values from specified densities in computing integrals against these distributions.) The motivation for this approach is seen in the following. For each $i = 1, 2, \dots, n$, assume that $X_i \sim F_i^X$, for cumulative distribution functions (CDF) F_i^X . Write F_i for the CDF of Y_i . Then, assuming $\hat{\rho}$ is a consistent estimator of ρ and is independent of all other random variables (which will be the case when estimated from an exogenous source), with $E[\hat{\rho}_+] = \rho_+$,

$$\begin{aligned}
 E[\hat{\rho}_+ V^\star(t)] &= \sum_{i=1}^n E[\hat{\rho}_+] P[X_i - V_{0i} < t < X_i - B_i(V_{0i} - V_{1i})] \\
 &= \sum_{i=1}^n \rho_+ P[t + B_i(V_{0i} - V_{1i}) < X_i < t + V_{0i}] \\
 &= \sum_{i=1}^n \rho_+ (P[X_i < t + V_{0i}] - P[X_i < t + B_i(V_{0i} - V_{1i})]) \\
 &= \sum_{i=1}^n \rho_+ \left\{ \int_0^\infty F_i^X(t + v_0) f_{V_0}(v_0) dv_0 \right. \\
 &\quad \left. - \int_0^\infty \int_0^\infty \sum_{b=0}^1 F_i^X(t + b[v_0 - v_1]) p_a^b (1 - p_a)^{1-b} f_{V_0}(v_0) f_{V_1}(v_1) dv_0 \, dv_1 \right\} \\
 &= \rho_+ \left\{ \int_0^\infty G_X(t + v_0) f_{V_0}(v_0) dv_0 \right. \\
 &\quad \left. - \int_0^\infty \int_0^\infty \sum_{b=0}^1 G_X(t + b[v_0 - v_1]) p_a^b (1 - p_a)^{1-b} f_{V_0}(v_0) f_{V_1}(v_1) dv_0 \, dv_1 \right\} \\
 &= \rho_+ \left\{ \int_0^\infty G_X(t + v_0) f_{V_0}(v_0) dv_0 \right. \\
 &\quad \left. - p_a \int_0^\infty \int_0^\infty G_X(t + v_0 - v_1) f_{V_0}(v_0) f_{V_1}(v_1) dv_0 \, dv_1 - (1 - p_a) G_X(t) \right\}
 \end{aligned}
 \tag{3}$$

where $G_X(u) = \sum_{i=1}^n F_i^X(u)$ and noting that ρ_+ is computable from the distribution of $\hat{\rho}$.

Computation of the integrals in last expression may proceed directly, either analytically or numerically, if we assume particular distributions for the F_i^X , f_{V_0} , and f_{V_1} , and values for p_a and ρ . A sensible, computationally accessible alternative is to couple the bootstrap with Monte Carlo integration. We enumerate the computational steps involved after listing our distributional assumptions. The estimates for the F_i , and so for G , follow from the use of smoothed bootstrap resampling of the original onset time observations Y_i and subsequent smoothing to X_i ; a specific motivation for this is detailed in Development using a Poisson process model for the underlying case onset times below. In cases when estimates for the mean incubation period are unavailable, we assume a quartic density for V_0 , so that

$$f_{V_0}(v_0) = \frac{15}{8(d-c)} \left(1 - \left[\frac{2v_0 - (c+d)}{d-c} \right]^2 \right)^2 I_{[c,d]}(v_0)$$

where $d > c > 0$ are chosen to correspond to the distribution for duration from infection to symptom onset. We adopted the quartic density when the only information we had on this variable was that the incubation period is roughly c to d days. The density f_{V_0} is symmetric with mean $(c+d)/2$ and variance $(d-c)/20$. Sampling from the quartic may be done using, for example, acceptance-rejection sampling. When more information regarding the distribution of V_0 is available, for example if there are published estimates of the mean and standard deviation of onset duration, we use a gamma density for $f_{V_0}(v_0)$.

In the original application for WNV transfusion risk via blood transfusion (Biggerstaff and Petersen 2002), when considering how to utilize available information for the distribution of

V_1 to inform specification of $f_{V_1}(v_1)$, the density representing the duration of WNV infection, we resampled from the observed values directly reported in Southam and Moore (1954) as representative of the distribution (so we employed bootstrapping). To be able to expand our methodology, we subsequently adopted the Weibull density for this duration, and we estimated the shape and scale parameters by matching moments. We use the Weibull below, noting little difference in our experience when the parameters for the Weibull are specified to match the sample mean and sample variance of the data in Southam and Moore (1954).

Because we use reported estimates $\hat{\rho}_a$ and $\hat{\rho}$, we incorporate uncertainty in these estimates directly into the Monte Carlo integration as well. In particular, we assume a beta density for $\hat{\rho}_a$ and a gamma density for $\hat{\rho}$.

Estimation algorithm

With the distributional assumptions and approximations given, we compute the estimate of $E[\hat{\rho}_+ V^*(t)]$ as follows. For each of $j = 1, 2, \dots, J$ simulations:

- (1) Resample with replacement n values from the originally observed case onset dates Y_1, Y_2, \dots, Y_n ; write these $Y_k^{(j)}$ for $k \in \{1, 2, \dots, n\}$; throughout this algorithm, k indexes the n resampled cases.
- (2) Compute the smoothing parameter h_j using the n values $Y_k^{(j)}$.
- (3) Generate n independent standard normal deviates, $\varepsilon_k^{(j)}$, and then compute $X_k^{(j)} = Y_k^{(j)} + h_j \varepsilon_k^{(j)}$.
- (4) Generate n independent values $V_{0k}^{(j)}$ from an assumed distribution $f_{V_0}(v_0)$.

In Biggerstaff and Petersen (2002) we used a quartic density with $a = 1$ and $b = 5$ and in subsequent applications used a gamma density with mean incubation period, $\hat{\mu}_0$, and variance, σ_0^2 , with these moments the same as for the stated quartic based on references given in (Biggerstaff and Petersen 2002). To be conservative, we began to use the a gamma density to permit longer incubation periods than permitted by the quartic, which is bounded.

- (5) Generate n independent values $V_{1k}^{(j)}$ from a specified distribution $f_{V_1}(v_1)$.

We use a Weibull density with mean $\hat{\mu}_1$ and variance $\hat{\sigma}_1^2$, where these are specified from available estimates.

- (6) Generate a value $\hat{\rho}^{(j)}$ from an assumed distribution $f_{\hat{\rho}}(r)$; round this value up to the next integer and write as $\hat{\rho}_+^{(j)}$.

We use a Weibull density with mean $\hat{\mu}_\rho$ and variance $\hat{\sigma}_\rho^2$, where these are specified from available estimates.

- (7) Replicate each the resampled case infection time, $(X_k^{(j)} - V_{0k}^{(j)}, X_k^{(j)} - V_{0k}^{(j)} + V_{1k}^{(j)})$, $\hat{\rho}_+^{(j)}$ times; identify these replicates using index r with superscripts (j, r) rather than (j) .
- (8) Generate a value $\hat{\rho}_a^{(j)}$ from an assumed distribution $f_{\hat{\rho}_a}(p)$.

We use a beta distribution with mean $\hat{\mu}_{\rho_a}$ and variance $\hat{\sigma}_{\rho_a}^2$, where these are specified from available estimates.

- (9) Generate $n \times \hat{\rho}_+$ independent values $B_k^{(j,r)}$ from the Bernoulli distribution with probability $1 - \hat{\rho}_a^{(j)}$. This will allow independent truncation of $100(1 - \hat{\rho}_a^{(j)})\%$ of the total number of generated “population” infections’ WNV infection times in simulation j .

With these values in-hand for each simulation $j = 1, 2, \dots, J$, compute on a grid of values $0 = t_1 < t_2 < \dots < t_g < \dots < t_G = T$, for $g = 1, 2, \dots, G$,

$$E[\widehat{\rho_+ V^*}(t_g)] = \frac{1}{J} \sum_{j=1}^J \sum_{k=1}^n \sum_{r=1}^{\hat{\rho}_+} I(x_k^{(j,r)} - v_{0k}^{(j,r)}, x_k^{(j,r)} - b_k^{(j,r)} | v_{0k}^{(j,r)} - v_{1k}^{(j,r)})(t_g) = \frac{1}{J} \sum_{j=1}^J W_j(t_g)$$

This curve is an average of J simulations of step functions, $W_j(t_g)$, where the heights of the steps are the estimated number of asymptomatic, viremic individuals in the population at each time point, where we “connect” or linearly interpolate values between grid points. For simplicity, write $E(t) = E[\widehat{\rho_+ V^*}(t)]$.

Confidence bands

We compute confidence bands based on the simultaneous bootstrap- t approach used in intensity estimation for Poisson processes, as derived in Cowling, Hall, and Phillips (1996). We use the Poisson approximating standard deviation and compute

$$T_j(t) = \frac{W_j(t) - E(t)}{\sqrt{W_j(t)}}$$

for $j = 1, 2, \dots, J$. To compute $100(1 - \alpha)\%$ simultaneous confidence bands $C_{1-\alpha}$ over $t \in (0, T)$, find q_1 and q_2 such that

$$P[q_1 \leq W_j(t) \leq q_2, \forall t \in (0, T) | \mathbf{X}] = \alpha$$

and

$$P[W_j(t) \leq q_1, \forall t \in (0, T) | \mathbf{X}] = P[W_j(t) \geq q_2, \forall t \in (0, T) | \mathbf{X}],$$

for equal-tailed bands, where \mathbf{X} represents all the data and exogeneous information available. Given these constraints, set

$$C_{1-\alpha} = \{(t, y): t \in (0, T), \max[0, E(t) - q_2 \hat{\sigma}] < y < E(t) - q_1 \hat{\sigma}\},$$

where $\hat{\sigma} = \sqrt{E(t)}$.

Finally, the sought-after, time-dependent risk of a WNV-infected donation is computed directly as the estimated number of such infections, $E(t)$, divided by the population size, N , and we write $P(t) = E(t)/N$. Similarly, the confidence bands for the risk are computed as $C_{1-\alpha}/N$.

Summary measures

As part of the simulation estimation, we also estimate the mean risk over the course of the outbreak by computing the average height of the risk curve. We do this by computing, for each simulation j , the area under the curve $W_j(t)$ using Simpson's Rule, then dividing this by the duration of the outbreak, T . The mean of the simulated areas gives the final estimate of the average risk, and we use the 2.5th and 97.5th quantiles of these simulated areas as 95% confidence intervals for the average risk. Lastly, we estimate the maximum risk over the course of the outbreak by reading off the maximum height of the estimated curve, along with a 95% confidence interval for this maximum from the confidence bands for the curve; the location of the maximum gives an estimate of the date of maximum risk.

Refinements for estimation of WNV in cadaveric tissue donations

In certain applications, as with the motivating example here of estimation of WNV infection in cadaveric tissues, stratification by variables for which key parameters vary should improve estimation. In the present example, the inapparent-to-apparent infection ratio is known to vary by age, with symptomatic infection (and severity of illness) more likely with increasing age. A straightforward refinement of the methods in Methodological and computational development to incorporate this is to use stratification as follows. Assume that the population is stratified into S strata of sizes N_s , for $s = 1, 2, \dots, S$, and let $N = \sum_s N_s$ be the population size. Indicate case onset times, etc., in stratum S by adding this index to the subscript throughout, so write, for example, Y_{si} , X_{si} , $V_s(t)$, $V_s^*(t)$, ρ_s , $W_{sj}(t)$, $E_s(t)$, $P_s(t)$. Stratum-specific estimates $P_s(t) = E_s(t)/N_s$ are readily computed by restricting the methods of Methodological and computational development to strata of interest.

Combining the strata for a population-level estimate of the risk may proceed naturally by weighting the stratum-specific estimates by relative population size, N_s/N . This weighting provides an overall estimate of the number of asymptomatic and infected donors. In some applications, however, further weighting to refine these estimates to individuals who either are more likely to donate, if this information is available, or who are actually eligible to donate may be warranted. This is the case in the cadaveric tissue donation application, where the proportions of asymptomatic and infected individuals are not expected to be the same as or even equally proportional among age groups, since donors must be recently deceased. Because death rates differ by age and there are available age group-specific inapparent-to-apparent infection ratio estimates, we are able to include weighting by stratum-specific death rates to account for the different probabilities of asymptomatic and infected individuals becoming deceased (and so eligible tissue donors). For $s = 1, 2, \dots, S$, let δ_s be the stratum-specific (age group-specific) annual death rates, and let $\Delta_s = \frac{T}{365}\delta_s$ be the death rates adjusted to the observed outbreak time duration of interest (where we have assumed that

T is in days, and note that $\frac{T}{365}$ will be greater than 1 for multi-year estimates, and we can naturally adjust for leap years). For each stratum, then, $\Delta_s E_s(t)$ is an estimate of the number of asymptomatic, WNV-infected potential cadaveric donors at time t , where we do note that the use of death rates this way is intended to reflect the cumulative death rates over the whole of the observed outbreak $(0, T)$, apportioned uniformly over the time of interest; time-dependent death rates $[\delta_s(t)]$ could be incorporated directly if available. The corresponding stratum estimates for the risk of an asymptomatic, viremic donor are therefore $\Delta_s E_s(t)/N_s$, so that these are combined to the population level, incorporating relative population size, as

$$P(t) = \sum_{s=1}^S \frac{N_s \Delta_s E_s(t)}{N} = \frac{1}{N} \sum_{s=1}^S \Delta_s E_s(t) = \frac{1}{NJ} \sum_{s=1}^S \sum_{j=1}^J \Delta_s W_{sj}(t) \quad (4)$$

where we use for convenience the same number of simulations J for each stratum. From the expression (4), we see that the order of computation—by stratum or by simulation—does not matter, and our implementation is to compute by simulation. Doing this facilitates computation of bootstrap confidence bands using $P(t)$, which are computed as in Methodological and computational development based on the resampled estimates $W_j(t) = \sum_s \Delta_s W_{sj}(t)$.

Development using a Poisson process model for the underlying case onset times

In the Monte Carlo evaluation of the integrals in the expression for $E[\hat{\rho}_+ V^*(t)]$ in Methodological and computational development, we used Gaussian smoothing of the observed Y_i and bootstrap resampling (so, effectively, the smoothed bootstrap (Davison and Hinkley 1997, § 3.4)) as essentially the estimation of the cumulative distribution functions F_i^X , and so of $G_X(u) = \sum_i F_i^X(u)$. On its face, this does not assume a particular distributional form for the stochastic process $\{Y_i\}$.

We noted without detail in the Discussion section of Biggerstaff and Petersen (2002), however, that this approach can be seen as assuming that the process $\{Y_i\}$ follows an inhomogeneous Poisson process in which a nonparametric, kernel smoothing estimate for the cumulative rate function of the process is used. For completeness, we now detail this observation here.

Begin by assuming that the WNN case onset dates $\{Y_i\}$ (ignoring the rounding of the observations to the day) are observations from an inhomogeneous Poisson process with rate function $\lambda(t)$ and cumulative rate function $\Lambda(t) = \int_0^t \lambda(u) du$. For this model, the CDF $F_i(t)$ of Y_i is (Liao 2014)

$$F_i(t) = 1 - e^{-\Lambda(t)} \sum_{j=0}^{i-1} \frac{[\Lambda(t)]^j}{j!} = 1 - \Pi(i-1; \Lambda(t))$$

introducing the notation $\Pi(x; \Lambda(t))$ for the CDF of a Poisson random variable with rate $\Lambda(t)$. From this it follows that

$$G(t) = \sum_{i=1}^n F_i(t) = n \left[1 - \Pi(n-1; \Lambda(t)) + \frac{\Lambda(t)}{n} \Pi(n-2; \Lambda(t)) \right] \quad (5)$$

Direct estimation of $G(t)$, therefore, reduces to estimation of $\Lambda(t)$. Various options are available for this, of course, including parametric regression, nonparametric spline regression, and nonparametric kernel smoothing. Employing the last of these and using a Gaussian kernel gives

$$\hat{\Lambda}(t) = \frac{1}{n} \sum_{i=1}^n \frac{1}{u} \phi\left(\frac{t - Y_i}{u}\right)$$

for smoothing parameter $u > 0$ and standard normal density function $\phi(\cdot)$. Use of $\hat{\Lambda}(t)$ in Eq. (5), then, provides an estimate of $G(t)$, which is $G_x(t)$ above after the estimation of $\Lambda(t)$, as included in Eq. (3). In doing this, we then may employ bootstrap resampling of the Y_i for estimation using $\hat{\Lambda}(t)$ in the Monte Carlo computations described above, so that $\hat{\Lambda}(t)$ is computed for each such bootstrap sample. Such sampling is, however, the same as sampling from the kernel smoothing estimate $\hat{\Lambda}(t)$ and then computing the integral in Eq. (3). Further, sampling from a Gaussian-smoothed kernel estimate of the Y_i is equivalent to sampling from the original data (Y_i) and adding scaled Gaussian noise, with scale equal to the smoothing parameter in the kernel smoothing estimate; this is exactly the smoothed bootstrap (e.g., Davison and Hinkley 1997, §3.4).

Operationally, then, using bootstrap resampling and then smoothing the resampled values using Gaussian noise and then using this in the Monte Carlo integrations in Eq. (3) is equivalent to assuming an inhomogeneous Poisson Process for the $\{Y_i\}$ and then estimating the cumulative rate function using a Gaussian kernel smoother, then evaluating the integral in Eq. (3).

Computer implementation

The methods presented were implemented in mostly purpose-written software using the statistical package R (R Core Team 2018) <https://www.R-project.org/>. Also used in these analyses were the boot package (Canty 2017), and the chron package (James 2018) Software is available from the author upon request. To implement the methods in other settings, data with disease case onset times and summary estimates of the modeling parameters—incubation period, inapparent-to-apparent infection ratio, viremia duration, asymptomatic

rate—are needed as inputs. Stratum-specific information for the case data and modeling parameters is required to implement the weighting methods described.

Application: cardiac tissue, Colorado, 2003

We implement the updated methods with the application introduced above, estimation of the risk of WNV infected cardiac tissue during the Colorado WNV outbreak in 2003, which remains at the high end of state-level WNV outbreak experience (<http://www.cdc.gov/westnile/statsmaps/>). Figure 1 shows the number of WNND cases aged 13–55 years by date of symptom onset in Colorado in 2003. In this case, a total of 358 WNND cases were reported to CDC in this age range, for an outbreak lasting 99 days. Cardiac tissue donations are restricted to donors under 56 years of age, and because there were few WNND cases aged less than 13 years, these were analyzed separately in Biggerstaff and Petersen (2020). The data in Figure 1 therefore provide time-dependent information on the WNV transmission activity in Colorado throughout 2003 for potential cardiac tissue donors.

As detailed in Biggerstaff and Petersen (2020), estimates of the inapparent-to-apparent infection ratios, $\hat{\rho}_s$, and their standard errors are available by age group stratum (Carson et al. 2012), as reported in Table 1. Note that these age grouping were selected to reflect available data and risk categories used for cardiac tissue by tissue banking industry and regulatory agencies. Also shown are the total numbers of WNND cases reported in these age groups, the annual death rates for these age groups for Colorado (<http://wonder.cdc.gov/>), and the total population for these age groups.

The gamma density was assumed for the incubation period, with mean $\hat{\mu}_0 = 3$ ($SD\hat{\sigma}_0 = 4/7$) days (Biggerstaff and Petersen 2002). The mean duration of WNV infection in cardiac tissue was estimated as $\hat{\mu}_1 = 9.39$ ($SD\hat{\sigma}_1 = 16.68$) days based on a Weibull survival model fit to data recorded in Southam and Moore (1954) (see Biggerstaff and Petersen 2020). Additionally, assuming the asymptomatic rate of $\hat{\mu}_{pa} = 0.79$ ($SD\hat{\sigma}_{pa} = 0.13$) as estimated in Mostashari et al. (2001), we apply the simulation algorithm described above and employ the refinements detailed in Refinements for estimation of WNV in cadaveric tissue donations using $J = 1000$ simulations to produce the estimated risk of WNV-infected cardiac tissue donations shown in Figure 2. The solid black line is the risk curve, while the black, dashed lines are the 95% confidence bands. To illustrate the variability in the simulations, 100 of the simulated curves $W_j(t)$ are shown in gray.

Finally, the average risk of WNV-infected cardiac tissue donations over the duration of this outbreak was 6.3 (95% CI 3.1, 9.6) per 10,000 deaths, with a maximum risk of 12.3 (95% CI 6.8, 20.5) per 10,000 deaths on August 10.

Discussion

While the original methods described here were developed for WNV, the underlying ideas were general. Subsequent applications to severe acute respiratory syndrome (SARS) virus (Shang et al. 2007), chikungunya virus (Brouard et al. 2008; Liumbruno et al. 2008), dengue virus (Seed et al. 2009; Petersen, Tomashek, and Biggerstaff 2012), hepatitis A

virus (Perevosnikovs et al. 2010), and Ross River virus (Shang et al. 2016) show that the methods may be applied in situations for which sufficient information is available on viral/pathogen kinetics, symptomatic rates, inapparent-to-apparent infection ratios, and, in some cases, surveillance coverage. In at least one application (Seed et al. 2009), the methods were applied in real time to guide decisions on suspension and continuance of blood donation collection during an outbreak. Most of these applications have been retrospective, so that case surveillance has been completed before analysis. In real-time application, time for disease progression and reporting delays of identified cases to public health surveillance systems can be expected to cause under-representation of real-time disease case counts. Using observed counts without correction for this would result in under-estimation, since too few cases would be identified in real-time because of these delays. Statistical methods of prediction of occurred but not yet reported events may be considered in such cases to attempt to adjust for this expected downward bias (see, e.g., Lawless 1994).

Application of these methods yields estimates of risks of infected donations before any post-donation testing or processing. In the US, for example, blood donations are now routinely tested seasonally for WNV, using policies implemented shortly after risk estimates were published and WNV transfusion transmission was documented (U. S. Food and Drug Administration 2009). Further, donated tissues undergo post-donation processing to reduce or eliminate transmission of multiple pathogens. Risk estimates produced using the methods developed here are intended to inform applicable industries, public health officials, and regulatory agencies concerning these risks to guide requirements for intervention strategies and policies. If estimates of the impact of post-processing are available, they may be naturally applied to results of the present analyses to characterize post-processing risks.

One assumption we made was that individuals share common distributions for both the incubation period and the viral infection duration. This is likely not strictly the case in either the donor or the general population, as there may be biological factors that mediate these for different individuals or groups of individuals. For WNV there is not currently evidence that systematic differences by known factors for these quantities would be on a scale to impact the general conclusions drawn from the methods used here, which take account of known uncertainties directly in the estimation. Should such become available, the estimation framework presented allows direct incorporation of such information in an easily implemented way. We also assumed that durations of onset and of infection were the same for individuals who remain asymptomatic or who develop symptoms, as there are not available separate estimates for these durations between these two groups; should such estimates become available, they can be incorporated into the methods presented here easily using the stratification mechanism introduced. Further, if any of the parameters differ by identifiable strata (say, by sex or by age), then one may naturally incorporate any available estimates or information into the framework we introduce here for weighted, stratified analyses, as well. In the absence of such detailed information, however, we may view these assumptions as reflecting a population average distribution across any such relevant factors.

In our approach we relied on Gaussian kernel smoothing to underlie the estimation of time-dependent risk curves. Naturally, other methods of smoothing, such as spline smoothing, semi- or fully parametric regression, or wavelet basis smoothing methods may have been

used. We adopted the approach we present due to the the direct and ready interpretability of the conversion of the initial, discrete observations to continuous “observations” (Y_i to X_i) and to utilize the bootstrap methods in (Cowling, Hall, and Phillips 1996). Evaluation of other methods of smoothing in the risk estimation context considered here may prove an interesting future research topic.

The analysis estimating risk of WNV infection in donated cardiac tissue during the Colorado outbreak of 2003 was part of a more complete analysis reported in Biggerstaff and Petersen (2020). Interested readers may refer there for estimates for blood, and cardiac, musculoskeletal, and soft tissues for outbreaks covering the contiguous US, 1999–2014; Colorado 2003; and Dallas-Fort Worth 2012.

We have detailed and updated our approach for time-dependent risk estimation for arbovirus infection of blood donations, and to address unique aspects of risk estimation for cadaveric tissue donations, we expanded the methods to include stratification and weighting. We expect that these refinements will provide greater flexibility and increased precision for researchers employing this general approach for a general variety of pathogens.

Acknowledgment:

The author thanks the American Association of Tissue Banks for the suggestion to pursue this work, particularly Mr. Frank Wilton, former AATB President and CEO, and also the Scientific and Technical Affairs Committee of the American Association of Tissue Banks for technical guidance, particularly the former committee chair, Dr. Jeffrey Cartmell. Further, the author thanks Dr. Lyle R. Petersen, Division Director, Division of Vector-Borne Diseases, CDC, for his collaboration on these projects and for encouragement in writing this methods paper. Thanks also to the ArboNET team at CDC for curating the data used in this study and to the SciComp team at CDC for access to their computational resources. Finally, thank you to Dr. Mark Delorey of DVBD/CDC for his careful review of the manuscript. The findings and conclusions herein are those of the authors and do not necessarily represent the official position of the US Centers for Disease Control and Prevention.

Research funding:

None declared.

References

- Biggerstaff BJ, and Petersen LR. 2002. “Estimated Risk of West Nile Virus Transmission through Blood Transfusion during an Epidemic in Queens, New York City.” *Transfusion* 42: 1019–26. [PubMed: 12385413]
- Biggerstaff BJ, and Petersen LR. 2003. “Estimated Risk of Transmission of the West Nile Virus through Blood Transfusion in the US, 2002.” *Transfusion* 43: 1007–17. [PubMed: 12869104]
- Biggerstaff BJ, and Petersen LR. 2020. “Estimated Risk of West Nile Virus Infection in Cadaveric Tissue Donations.” Unpublished
- Brouard C, Bernillon P, Quatresous I, Pillonel J, Assal A, De Valk H, and Desenclos JC, A Workgroup, Quantitative Estimation of the Risk of Blood Donation Contamination by Infectious. 2008. “Estimated Risk of Chikungunya Viremic Blood Donation during an Epidemic on Reunion Island in the Indian Ocean, 2005 to 2007.” *Transfusion* 48 (7): 1333–41. [PubMed: 18298600]
- Canty A and Ripley B. 2017. *boot: Bootstrap R (S-Plus) Functions R package, version 1.3–20.*
- Carson PJ, Borchardt SM, Custer B, Prince HE, Dunn-Williams J, Winkelman V, Tobler L, Biggerstaff BJ, Lanciotti R, Petersen LR, and Busch MP. 2012. “Neuroinvasive Disease and West Nile Virus Infection, North Dakota, usa, 1999–2008.” *Emerging Infectious Diseases* 18: 684–6. [PubMed: 22469465]

- Cowling A, Hall P, and Phillips MJ. 1996. "Bootstrap Confidence Regions for the Intensity of a Poisson Point Process." *Journal of the American Statistical Association* 91: 1516–24.
- Davison A, and Hinkley D. 1997. *Bootstrap Methods and Their Application*, Cambridge Series in Statistical and Probabilistic Mathematics Cambridge: Cambridge University Press.
- U. S. Food and Drug Administration. 2009. *Guidance for Industry: Use of Nucleic Acid Tests to Reduce the Risk of Transmission of West Nile Virus from Donors of Whole Blood and Blood Components Intended for Transfusion* Also available at <https://www.fda.gov/media/124300/download>.
- Centers for Disease Control and Prevention. 2002. "West Nile Virus Infection in Organ Donor and Transplant Recipients—Georgia and Florida, 2002." *MMWR Morb Mortal Wkly Rep* 51 (35): 790. <https://www.cdc.gov/mmwr/preview/mmwrhtml/mm5135a5.htm>. [PubMed: 12227442]
- Iwamoto M, Jernigan DB, Guasch A, Trepka MJ, Blackmore CG, Hellinger WC, Pham SM, Zaki S, Lanciotti RS, Lance-Parker SE, DiazGranados CA, Winquist AG, Perlino CA, Wiersma S, Hillyer KL, Goodman JL, Marfin AA, Chamberland ME, and Petersen LR and The West Nile Virus in Transplant Recipients Investigation Team. 2003. "Transmission of West Nile Virus from an Organ Donor to Four Transplant Recipients." *New England Journal of Medicine* 348: 2196–203. [PubMed: 12773646]
- James D, and Hornik K. 2018. *chron: Chronological Objects which can Handle Dates and Times*. R package, version 2.3–53
- Lawless J 1994. "Adjustments for Reporting Delays and the Prediction of Occurred but Not Reported Events." *Canadian Journal of Statistics* 22: 15–31.
- Liao M 2014. *Applied Stochastic Processes* Boca Raton, Florida: CRC Press.
- Liumbruno GM, Calteri D, Petropulacos K, Mattivi A, Po C, Macini P, Tomasini I, Zucchelli P, Silvestri AR, Sambri V, Pupella S, Catalano L, Piccinini V, Calizzani G, and Grazzini G. 2008. "The Chikungunya Epidemic in Italy and its Repercussion on the Blood System." *Blood Transfus* 6: 199–210. [PubMed: 19112735]
- Mostashari F, Bunning ML, Kitsutani PT, Singer DA, Nash D, Cooper MJ, Katz N, Liljebjelke KA, Biggerstaff BJ, Fine AD, Layton MC, Mullin SM, Johnson AJ, Martin DA, Hayes EB, and Campbell GL. 2001. "Epidemic West Nile Encephalitis, New York, 1999: Results of a Household-Based Seroepidemiological Survey." *Lancet* 358: 261–4. [PubMed: 11498211]
- Perevoscikovs J, Lenglet A, Lucenko I, Steinerte A, Payne Hallstrom L, and Coulombier D. 2010. "Assessing the Risk of a Community Outbreak of Hepatitis a on Blood Safety in Latvia, 2008." *Euro Surveill* 15: 19640. [PubMed: 20739001]
- Petersen LR, Tomashek KM, and Biggerstaff BJ. 2012. "Estimated Prevalence of Dengue Viremia in Puerto Rican Blood Donations, 1995 through 2010." *Transfusion* 52: 1647–51. [PubMed: 22304614]
- R Core Team. 2018. *R: A language and environment for statistical computing*. R Foundation for Statistical Computing Vienna, Austria. URL <https://www.R-project.org/>.
- Scott DW 1992. *Multivariate Density Estimation. Theory, Practice and Visualization* New York: Wiley.
- Seed CR, Kiely P, Hyland CA, and Keller AJ. 2009. "The Risk of Dengue Transmission by Blood during a 2004 Outbreak in Cairns, Australia." *Transfusion* 49: 1482–7. [PubMed: 19389025]
- Shang G, Biggerstaff BJ, Richardson AM, Gahan ME, and Lidbury BA. 2016. "A Simulation Model to Estimate the Risk of Transfusion-Transmitted Arboviral Infection." *Transfusion and Apheresis Science* 55: 233–9. [PubMed: 27474684]
- Shang G, Biggerstaff BJ, Yang B, Shao C, and Farrugia A. 2007. "Theoretically Estimated Risk of Severe Acute Respiratory Syndrome Transmission through Blood Transfusion during an Epidemic in Shenzhen, Guangdong, china in 2003." *Transfusion and Apheresis Science* 37: 233–40. [PubMed: 18036985]
- Sheather S, and Jones C. 1991. "A Reliable Data-Based Bandwidth Selection Method for Kernel Density Estimation." *Journal of the Royal Statistical Society: Series B* 53: 683–90.
- Southam CM, and Moore AE. 1954. "Induced Virus Infections in Man by the Egypt Isolates of West Nile Virus." *The American Journal of Tropical Medicine and Hygiene* 3: 19–50. [PubMed: 13114588]

Williamson PC, Custer B, Biggerstaff BJ, Lanciotti RS, Sayers MH, Eason SJ, Dixon MR, Winkelman V, Lanteri MC, Petersen LR, and Busch MP. 2016. "Incidence of West Nile Virus Infection in the Dallas-Fort Worth Metropolitan Area during the 2012 Epidemic." *Epidemiology and Infection* 145: 2536–44. [PubMed: 26829991]

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

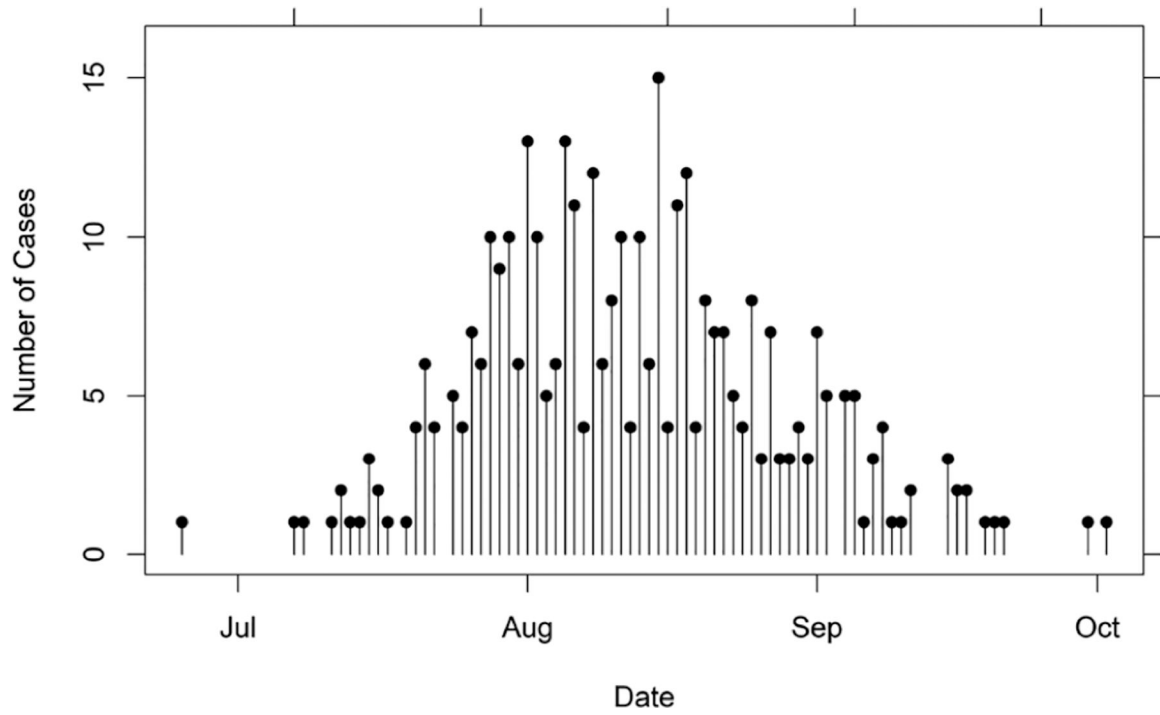


Figure 1: Number of WNND cases aged 13–55 years by date of symptom onset, Colorado, 2003.

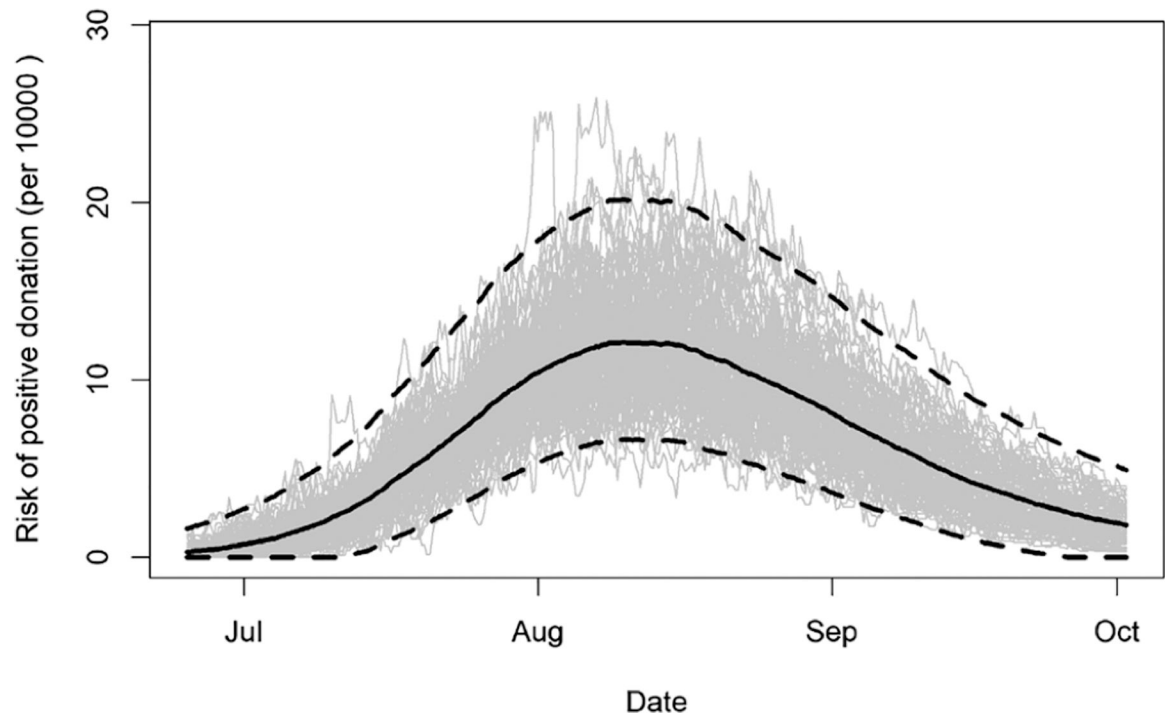


Figure 2: Estimated risk of WNV-infected cardiac tissue among potential donations, Colorado, 2003, shown as the black, solid line, with 95% confidence bands shown as black, dashed lines. The gray lines are a sample of the simulated risk curves, $W_j(t)$.

Table 1:

Age group-specific WNND counts for Colorado 2003; inapparent-to-apparent infection ratios, $\hat{\rho}$, and their standard errors (SE) (Biggerstaff and Petersen 2020); annual death rates, δ , per 1,000 population; and total population (Pop).

Age group	WNND	$\hat{\rho}$ (SE)	δ
13–20	22	742.57 (96.92)	0.58
21–40	134	526.90 (58.39)	1.12
41–55	202	286.23 (19.52)	3.86
Total	358	Pop = 2,911,979	

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript