Estimated risk of West Nile virus transmission through blood transfusion during an epidemic in Queens, New York City

Brad J. Biggerstaff and Lyle R. Petersen

BACKGROUND: Human West Nile virus (WNV) infection has been documented in the eastern United States since its discovery there in 1999. Epidemics of WNV encephalitis and meningitis raise concern that transmission of WNV may occur through voluntary blood donations.

STUDY DESIGN AND METHODS: Case onset dates from the 1999 Queens, NY, epidemic of WNV encephalitis and meningitis, and historic data on viremia in humans are used to estimate the number of cases that were viremic throughout the epidemic. Estimates of the inapparent-to-apparent WNV infection ratio, the proportion of asymptomatic infections reported in a seroepidemiologic survey coincident with the epidemic, and the population size are used to infer the WNV transfusion-transmission risk. Statistical resampling methods are used.

RESULTS: The maximum and mean risk of WNV transmission (/10,000) from donors in Queens were estimated as 2.7 (95% CI, 0.9-5.6) and 1.8 (95% CI, 1.4-2.2), respectively. The risk peaked in late August, with very low risk before August and after September.

CONCLUSION: Although most WNV-infected individuals have subclinical infections, these data suggest a low prevalence of viremia throughout the Queens epidemic and subsequent low risk of transmission of WNV by blood transfusion.

The West Nile virus (WNV) is a mosquito-borne flavivirus transmitted primarily among birds. Humans are incidental hosts. In the past decade, the WNV has caused large outbreaks of human encephalitis and meningitis in Europe,[1] the Middle East,[2] and Russia.[3] The virus was first recognized in the New World in 1999 when it caused an epizootic among birds and horses and an epidemic of meningitis and encephalitis in humans in the New York City metropolitan area.[4] Through 2001, avian mortality surveillance has documented geographic spread of WNV to approximately half the United States,[5] as well as to southeastern Canada. In 2001, human cases of WNV encephalitis or meningitis occurred in 10 states as well as the Grand Cayman Islands (CDC, unpublished data).

Epidemiologic investigations in Romania and the United States indicate that fewer than one percent of those infected with WNV develop encephalitis or meningitis, and approximately one-third develop a mild febrile illness; the remainder are asymptomatic.[1,6,7] The transient viremia after infection[8-10] and the high proportion of asymptomatic or mildly symptomatic infections raise concern about the potential for transfusion-related WNV transmission.[11] The risk of such transmission should be highest during epidemics, when the prevalence of viremia in the population would be highest. WNV antibody has been detected, however, in volunteer blood donations in an endemic region of France, even in the absence of a recognized human outbreak.[11]

ABBREVIATIONS: A = proportion of individuals who remain asymptomatic; EVC = estimated viremia curve; R = ratio of the number of unapparent or subclinical infections to the number of apparent infections; TVC = true viremia curve; WNV = West Nile virus.

From the Division of Vector-Borne Infectious Diseases, National Center for Infectious Diseases, CDC, Fort Collins, CO.

Address reprint requests to: Brad Biggerstaff, PhD, Division of Vector-Borne Infectious Diseases, CDC, PO Box 2087, Fort Collins, CO 80522–2087; bbiggerstaff@cdc.gov.

Received for publication February 22, 2002; revision received April 11, 2002, and accepted April 15, 2002.

TRANSFUSION 2002;42:1019-1026.

We used a statistical resampling approach to estimate temporal trends of the proportion of those infected who were viremic throughout an outbreak of WNV meningitis and encephalitis in Queens, NY, in 1999. We then used seroepidemiologic data from the 1999 Queens outbreak to estimate the risk of transfusion-related transmission in that outbreak setting.

MATERIALS AND METHODS

General approach

Our study population was the population of Queens County (one of five counties that comprise New York City) during the 1999 outbreak. Of the 62 individuals with severe WNV neurologic disease, 32 lived in Queens. We first estimated how many of the 31 infected individuals with known symptom onset dates were viremic at each time point throughout the outbreak period. This was done using a statistical resampling approach that incorporated these 31 symptom onset dates, an assumed distribution of the length of time between the onset of viremia and the onset of symptoms, and an estimated distribution of the length of viremia. The assumed distribution of the duration between onset of viremia and symptom onset was derived from historic data on the incubation period of WNV-related disease and the timing of viremia onset relative to symptom onset during this incubation period.[12] The estimated distribution of the duration of viremia was derived using data from a large, human volunteer study of experimental WNV infection.[10] Assuming that the dates of infection of these 31 individuals with known symptom onset dates were similar to those of all Queens residents who became infected that year, we then used this estimate of the number of the cases with viremia over time to infer the risk of tranfusion transmission of WNV over time by using the population size of Queens and seroepidemiologic survey results[6] that provide estimates of the proportion of WNV-infected individuals who develop severe neurologic disease and of the proportion of asymptomatic WNV infections.

Study population and data collection

Onset dates for WNV encephalitis and meningitis cases for the 1999 outbreak were obtained from surveillance data reported to CDC by the New York State and New York City Departments of Health (Grant L. Campbell, written communication, October 2001). We restricted our study population to Queens, the epicenter of the outbreak and the location of a seroepidemiologic survey conducted toward the end of the outbreak.[6] Symptom onset times (in days) of the 31 cases are shown in the pin plot in Fig. 1, with time $t = 0$ corresponding to August 2, 1999, the date of the first reported onset, and time $t = 41$ corresponding to the last reported onset on September 12, 1999. The duration of this epidemic was thus considered to be 41 days.

Historic data on WNV viremia distribution in humans

Several studies have reported on the course of WNV infection and viremia in humans. The time from inoculation to symptom onset, the incubation period, is not precisely known but seems to be relatively short (approx., 2-6 days).[8,13,14] There is approximately a 1- to 2-day lag between inoculation with the virus and the detection of virus in the blood so that the duration of viremia relative to symptom onset is roughly 1 to 2 days shorter than the incubation period.[8,10]

In an experimental study of WNV inoculation of humans with cancer, Southam and Moore[10] provide data concerning the duration of viremia, stratified by the severity of WNV-associated illness. We used those data from individuals with demonstrable viremia who were in the lowest of the five disease categories, showing no symptoms. Those exhibiting symptoms or with more serious WNV-related disease would likely be excluded from blood donation. Further, individuals with more severe underlying conditions were found to have more severe WNV-related disease and would be less likely to be rep-

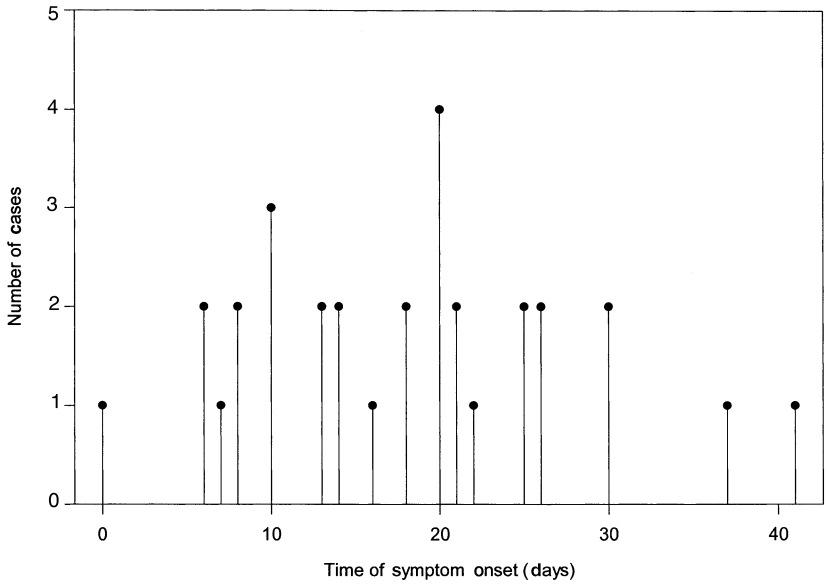

Fig. 1. Symptom onset times of 31 individuals with WNV encephalitis or meningitis, Queens, New York, 1999. Time 0 is August 2, 1999.

resentative of a healthy blood donor population. The data we used are given in the pin plot in Fig. 2. There are 19 observations, with a mean of 6.2 days and a SD of 2.9 days; the median is 6 days, with a range of 1 to 11 days.

Statistical approach

We now describe the statistical method we used; a formal development of the method is given in Appendix 1.

Estimating the proportion of cases with viremia throughout the outbreak. Our strategy was to view the symptom onset times of the cases (n = 31) as anchor times, then to use the information about how virema relates to symptom onset to estimate the number of cases with viremia at any time *t* during the outbreak. Then, using this information and information on the unapparent-to-apparent WNV infection ratio (R), the population size, and the proportion of individuals infected who remain asymptomatic (A), we estimated the risk of WNV transmission by transfusion from a unit of blood donated at time *t* during the epidemic.

We used Monte Carlo simulation to estimate the number of cases with viremia at a fixed time *t* by simulating for each case onset time an associated viremia time span, then counting the number of cases with viremia at time *t*. Because the individual case onset times are recorded to the day as discrete times, but the underlying infection process is instead continuous, we first smoothed the observed case onset times by adding a smoothing component. Next, the simulated viremia time spans were computed for each case. To do this, the onset of viremia, relative to the case onset time, was chosen by taking a random sample from an assumed distribution (Appendix 1) for the duration from inoculation to symptom onset, based on historic information as noted above. The duration of viremia was then chosen by taking a random sample (with replacement) of the duration times given in Fig. 2. A graphic example of this procedure is given in Fig. 3. We have assumed in this procedure that the relative timing and duration of viremia for a case is independent of the symptom onset time.

This Monte Carlo sampling process was repeated 1000 times, and the resulting counts of the number of cases viremic at any time *t* were averaged. Considering these counts for all times *t* throughout the outbreak yielded a curve representing the expected number of cases with viremia over the course of the epidemic. We call this curve the estimated viremia curve (EVC).

We computed two summary measures of the EVC to aid interpretation, the maximum value and the mean value over the duration of the epidemic. The maximum is simply the point at which the curve is highest, and the timing of this maximum tells us when the largest proportion of viremic cases occurred. The mean, computed by dividing the area under the EVC by the duration of the epidemic, provides a measure of the expected proportion of viremic cases over the whole course of the epidemic.

Confidence bands for the true viremia curve (TVC) around the EVC can be computed in various ways. We used a simultaneous percentile-*t* approach, the details of which are in Appendix 1. A 95-percent CI for the true maximum is read from the confidence bands for the TVC, while the 95-percent CI for the true mean is computed using a percentile-*t* approach similar to that for the TVC.

Inference to the general population. To estimate the number of viremic individuals in the population at any time, we multiplied the EVC by R at time *t*. We assumed that R was constant across time. We then estimated the proportion of the population who were viremic at time *t* as R $\times$ EVC/P, where P is the population size. The numerator of R was not known during the outbreak; however, we used R = 140:1 based on results from a seroepidemiologic survey conducted in a 4.8-km^2 area of Queens at the ebbing of the 1999 oubreak.[6]

Computation of WNV transmission risk from blood transfusion. Assume that the rate of blood donation is constant over the course of the epidemic, and assume that potential blood donors have the same risk of infection with WNV as the general population.

This latter assumption was supported by the findings of the Queens serosurvey, which showed constant WNV antibody prevalence among

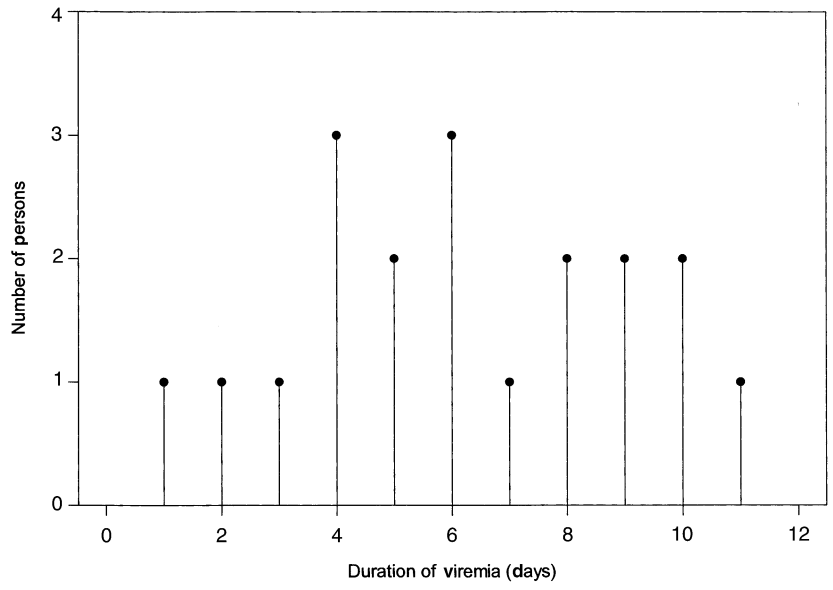

Fig. 2. Duration of viremia for 19 patients after intentional infection with WNV.[10]

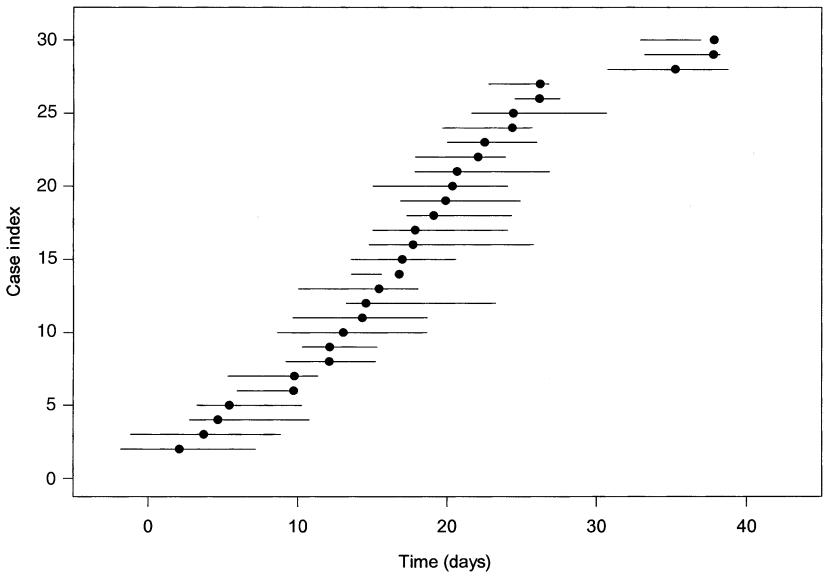

Fig. 3. Results of one simulation used to compute the EVC for 31 individuals with WNV encephalitis or meningitis during the 1999 Queens epidemic. Each dot represents the symptom onset time for one individual and each corresponding horizontal line represents the duration of viremia. Time 0 is August 2, 1999. The number of horizontal lines intersecting a vertical line drawn at a given time point is the estimated number of individuals viremic at that time point for this realization.

adults of all ages and both sexes.[6] The value of A multiplied by the number of viremic individuals at time t estimates the number of individuals who pose a risk of transmitting WNV through blood donation at time t. The estimate of the proportion of infected individuals in the Queens seroepidemiologic survey with febrile illness during the outbreak was 0.32.[6] We therefore used a value of 0.68 (1 – 0.32) for A, assuming that 0.32 of viremic individuals would not donate blood or would be excluded from donation because of clinical symptoms. Finally, we assumed that transfused blood components of WNV viremic blood donors transmit infection to recipients with 100-percent efficiency.

All computations were performed with software (S-Plus 6 Professional for Windows, Insightful Corp., Seattle, WA) using existing routines or routines written by the authors.

RESULTS

Figure 3 illustrates the construction of the EVC for the 31 cases in the Queens epidemic. The values shown are one realization of the many simulations used in computing the EVC. The dots in the graph represent the smoothed times of symptom onset for each case (using the smoothing parameter h = 10.4). The endpoints of the line segments are the simulated viremia time spans associated with each case. To demonstrate the computation, the numeric value used to compute the EVC at time t = 20 is obtained from the graph by drawing a vertical line at time t = 20 and counting the number of segments it intersects, which in this example is 10. Thus, for this single realization, we would estimate that 20 days after the onset of symptoms of the first case in this outbreak, 10 of the 31 known cases were viremic.

One thousand such realizations combined to produce the EVC, shown as the dark, solid line in Fig. 4. The dashed lines are the 95th percentile-t confidence bands. Also shown in Fig. 4 are 100 randomly selected realizations of the EVC from the 1000 generated. The scale on the left axis is the number of the observed cases with viremia.

As seen in Fig. 4, the maximum of the EVC is 6.2 (95% CI, 2.0-13.2), occurring at 17.8 days, roughly on August 19 to 20, 1999. Given the estimate that for every WNV-infected individual who develops severe neurologic disease, there are 140 who do not, we estimated that the maximum number of people who were viremic at any one time in Queens during the outbreak was 865, and the proportion of those who were asymptomatic was 0.68 times this, or 588. Finally, dividing this number by the size of the population of Queens (2,229,379[15]) provides an estimate of the proportion of residents of Queens who were asymptomatic and viremic during the outbreak. Given the assumptions outlined in the methods, we estimated the maximum risk of transmission from transfusion of a single unit as 2.7 per 10,000 (95% CI, 0.9-5.6/10,000) and the mean risk of transmission over the course of the outbreak as 1.8 per 10,000 (95% CI, 1.4-2.2/10,000). The risk scale is shown on the right side of Fig. 4.

DISCUSSION

Our results indicate a small but nonzero risk of WNV transmission from transfusion of blood components. We calculated that during an epidemic of WNV neurologic disease in Queens, New York, in 1999, the risk peaked at approximately 2.7 per 10,000 donations in mid- to late August, with a mean risk over the course of the outbreak of 1.8 per 10,000 donations. The calculated risk was highly limited in time, with the risk approaching zero among donations before August and after September.

By way of comparison, recent estimates[16] of the risk for transfusion transmission of HBV, HCV, and HIV are 1 per 30,000 to 250,000, 1 per 30,000 to 150,000, and 1 per 200,000 to 2,000,000, respectively, one or two orders of

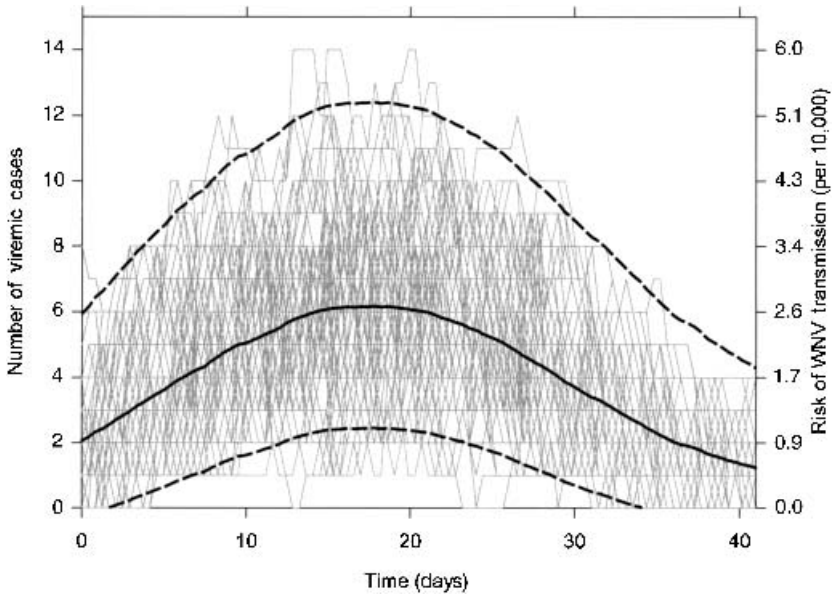

Fig. 4. The EVC (solid, dark line) from 1000 simulations. Simultaneous 95th percentile-*t* confidence bands are shown as dashed lines. The left axis gives the scale of the EVC as the number viremic among the 31 cases. The right axis gives the scale as the risk of WNV transmission from transfusion of a single unit of blood, after inferring to the entire Queens population and accounting for symptomatic individuals who would not donate or be deferred from donating. The light lines depict 100 sample realizations of the 1000 used to compute the EVC.

magnitude lower than our maximal estimates for WNV. As we saw, however, the risk for transfusion transmission of WNV is highly time limited, and the estimates we give are in a sense the worst-case estimates because they were generated during a recognized epidemic.

Our risk estimates for WNV transfusion transmission are in line with results computed using the window-period method.[17,18] Assuming that 5 percent of the population donates blood in the outbreak year and that these donations are evenly distributed over the year, the window-period approach yields estimates of the number of infectious donors as follows. Using an R value of 140:1, one expects 4340 infections in Queens based on the 31 cases we used. Then, 5 percent of these, or 217, would be donors, and 41 ÷ 365 = 11 percent of these, or 25, would be donors who became infected during the 41 days of the outbreak. Using a window period of 3 days (the mean of our assumed distribution for the time from onset of viremia to symptom onset) and noting the outbreak was 41 days long, 3 ÷ 41 = 7.3 percent, or roughly 2 of these 25 would donate during their window period and so would be infectious. If the window period is increased to 5 days (the upper limit of our assumed distribution), the estimated number of infectious donors who would donate during their window period is 3. Using our estimated mean risk of 1.8 per 10,000 donations and making the same assumptions (i.e., that 5% of the population would be donors and that they would donate uniformly over the year), we compute the expected number of infectious donors over the outbreak as 2,229,379 × 0.05 × 0.11 × (1.8/10,000) = 2.2. Although the window-period method and our method give similar estimates for the expected number of infectious donors, our approach provides more information in the form of the full risk curve (Fig. 4) and appropriate confidence bands.

Despite the theoretic risk of transfusion-related transmission from WNV and related flavivirus infections such as yellow fever, dengue, and Japanese encephalitis, infections of these agents from transfusion of blood or blood components have not been reported. One explanation is that the prevalence of WNV viremia among humans even during an outbreak of this magnitude may be low, as is suggested by our findings. Another is that in endemic regions, infections often occur among nonimmune youth, who would be unlikely to donate blood. The fact that most flavivirus infections result in asymptomatic or mildly symptomatic infections would cause most transfusion-related infections to go unrecognized, though the potentially higher dose and IV inoculation from a transfusion-induced infection would likely result in a higher rate of clinical symptoms than for natural, mosquito-borne infection.[10] Diagnostic tests are also unavailable in many areas. Finally, it may be difficult to distinguish transfusion-related infection from infection from mosquito vectors in endemic areas.

Our estimated peak transmission risk of 2.7 per 10,000 for the Queens epidemic may be too high for several reasons. We assumed a rate of 100-percent transmission from viremic donors; the true transmission rate is probably lower. Adjustments to our risk estimates may be made directly to account for this by multiplying our estimates by the assumed transmission rate. There are also uncertainties about the duration of viremia. We used data from experimental infections of cancer patients. The duration of viremia in cancer patients after WNV infection may be longer than that of previously healthy individuals who would be donating blood. Conversely, our estimate of the transmission risk over time might be too low if case ascertainment by the surveillance systems was imperfect. Further, some viremic individuals who do develop symptoms may donate before exhibiting these symptoms.

Other issues with respect to the possibility of transmission of WNV through transfusion are the stability of the virus during refrigeration of blood and blood compo-

nents and whether WNV has any cell tropisms. We are unaware of any studies concerning WNV stability in refrigerated blood or blood components. However, flaviviruses have a long survival in fluids containing a high protein content that stabilizes the virus, and HCV, also a flavivirus, survives well in blood components.

The statistical approach we took was one of several we might have taken. As detailed now, our considerations of alternative methods indicate that the nonparametric nature of the direct resampling approach we used mimics a reasonable but more complicated parametric modeling framework, that the choice of the form for the density of the duration from inoculation to symptom onset is reasonably robust, and that smoothing the case series before estimation of the EVC does affect the shape of the viremia curve. A sensible parametric augmentation to our analyses would be to assume the case onset series is the realization of an inhomogeneous Poisson process. Estimation of the intensity function for this process could then be performed nonparametrically,[19,20] and incorporation of viremia distribution information could be made following a (Bayesian) conditioning argument via imputation.[21] Results from this approach would actually be similar to those we presented because the smoothing we used for the case onset series is akin to nonparametric estimation of the intensity function, and the way we incorporated the viremia distribution information is operationally what one would do using imputation. Indeed, we analyzed the Queens data using this approach and the results differed little; we therefore chose to present the simpler method. The choice of the density for the duration from initial viremia to symptom onset was reasonable but arbitrary. Another parametric class of densities with larger variances, and thus less informative, was considered, but there was no appreciable difference in the results. The results did differ noticeably when smoothing of the original series (Y_i) was not done. In the nonsmoothed case, the EVC ($\widehat{E[V(t)]}$ in Appendix 1) was notably influenced by the discreteness of the data, being much less smooth than the curve we present and following the bumps apparent in the case series shown in Fig. 1. We favored the smoothed version because of the theoretical continuity of the underlying biologic process and the general agreement between the smoothed method and the Poisson process approach, which assumes continuous variables, noted above. One shortcoming of our inference to the general Queens population is that we did not incorporate uncertainty in the exogenous estimation of R or A; the effect of this would be to widen the confidence bands and the 95-percent CIs.

Despite the lack of data about blood transfusion-related flavivirus transmission, the fact that dengue transmission has occurred from marrow transplantation[22] and the fact that dengue has been transmitted by needlestick injury[23] provide evidence that transfusion-related flavivirus transmission is plausible. WNV should be considered in the differential diagnosis of unexplained fever, meningitis, or encephalitis among recent recipients of blood or blood components, particularly if the donation occurred when WNV activity is highest—in late summer in the northeast US—and in areas where WNV infections have recently been documented in humans. If acute WNV infection is detected in a transfusion recipient, the donors should be tested for WNV antibody. The presence of WNV antibody in the donor and in other recipients of blood components from that same donation would provide corroborative evidence that transfusion-related transmission has occurred.

ACKNOWLEDGMENTS

The authors thank Grant L. Campbell, MD, PhD, of the Division of Vector-Borne Infectious Diseases, National Center for Infectious Diseases, CDC, Fort Collins, CO, for providing the case onset data.

REFERENCES

1. Tsai TF, Popovici F, Cernescu C, and for the Investigative Team. West Nile encephalitis epidemic in southeastern Romania. Lancet 1998;352:767-71.
2. Weinberger M, Pitlik SD, Gandacu D, et al. West Nile fever outbreak, Israel, 2000: epidemiologic aspects. Emerg Infect Dis 2001;7:686-91.
3. Platonov AE, Shipulin GA, Shipulina OY, et al. Outbreak of West Nile virus infection. Volgograd Region, Russia, 1999. Emerg Infect Dis 2001;7:128-32.
4. Nash D, Mostashari F, Fine A, Miller J, et al. The outbreak of West Nile virus infection in the New York City area in 1999. N Engl J Med 2001;344:1807-14.
5. Centers for Disease Control and Prevention Weekly update: West Nile virus activity–United States, November 14-20, 2001. MMWR Morb Mortal Wkly Rep 2001;50:1061-2.
6. Mostashari F, Bunning ML, Kitsutani PT, et al. Epidemic West Nile encephalitis, New York, 1999: results of a household-based seroepidemiological survey. Lancet 2001;358:261-4.
7. Serosurveys for West Nile virus infection, New York and Connecticut counties, 2000. MMWR Morb Mortal Wkly Rep 2001;50:37-9.
8. Goldblum N, Sterk VV, Jasinska-Klingberg W. The natural history of West Nile fever: II. Virological findings and the development of homologous and heterologous antibodies in West Nile infection in man. Am J Trop Med Hyg 1957; 66:363-80.
9. Southam CM, Moore AE. West Nile, ilheus, and bunyamwera virus infections in man. Am J Trop Med Hyg 1951; 31:724-41.
10. Southam CM, Moore AE. Induced virus infections in man by the Egypt isolates of West Nile virus. Am J Trop Med Hyg 1954;3:19-50.

11. Charrel RN, Lamballerie X, Durand JP, et al. Prevalence of antibody against West Nile virus in volunteer blood donors living in southeastern France (letter). Transfusion 2001;41:1320-1.
12. Goldblum N, Sterk VM, Paderski B. The clinical features of the disease and the isolation of West Nile virus from the blood of nine human cases. Am J Hygiene 1954;59:89-103.
13. Kokernot RH, McIntosh BM. Isolation of West Nile virus from a naturally infected human and from a bird, *Sylvitta rufescens (Vieillot)*. S Afr Med J 1959;33:987-9.
14. Hannoun C, Corniou B, Causse G, Panthier R. Development of serum antibodies in 4 cases of West Nile virus infection. Ann Inst Pasteur (Paris) 1967;113:29-36.
15. US Census Bureau. 2000 US Census, http://www.census.gov.
16. Goodnough LT, Brecher ME, Kanter MH, AuBuchon JP. Transfusion medicine. First of two parts–Blood transfusion. New England J Med 1999;340:438-47.
17. Petersen LR, Satten GA, Dodd R, et al. Duration of time from onset of human immunodeficiency virus type 1 infectiousness to development of detectable antibody. The HIV Seroconversion Study Group. Transfusion 1994;34:283-9.
18. Kleinman S, Busch MP, Korelitz JJ, Schreiber GB. The incidence/window period model and its use to assess the risk of transfusion-transmitted human immunodeficiency virus and hepatitis C virus infection. Transfusion Med Rev 1997;11:155-72.
19. Diggle P. A kernel method for smoothing point process data. Appl Stat 1985;34:138-47.
20. Cowling A, Hall P, Phillips MJ. Bootstrap confidence regions for the intensity of a Poisson point process. J Am Stat Assoc 1996;91:1516-24.
21. Shafer JL. Analysis of incomplete multivariate data. London: Chapman & Hall, 1997.
22. Rigau-Perez JG, Vorndam AV, Clark GG. The dengue and dengue hemorrhagic fever epidemic in Puerto Rico, 1994-1995. Am J Trop Med Hyg 2001;64:67-74.
23. de Wazieres B, Gil H, Vuitton DA, Dupond JL. Nosocomial transmission of dengue from a needstick injury. Lancet 1998;351:498.
24. Silverman BW. Density estimation for statistics and data analysis. London: Chapman & Hall, 1986.
25. Diggle P, Marron JS. Equivalence of smoothing parameter selectors in density and intensity estimation. J Am Stat Assoc 1988;83:793-800.
26. Sheather SJ, Jones MC. A reliable data-based bandwidth selection method for kernel density estimation. J Royal Stat Soc, Series B 1991;53:683-90.
27. Jones MC, Marron JS, Sheather SJ. A brief survey of bandwidth selection for density estimation. J Am Stat Assoc 1996;91:401-7.
28. Davison AC, Hinkley DV. Bootstrap methods and their application. Cambridge, UK: Cambridge University Press, 1997.
29. Lange K. Numerical analysis for statisticians. New York: Springer-Verlag, 1999.

APPENDIX 1. METHODS USED IN ESTIMATION OF WNV TRANSFUSION RISK

Let Y_i, for $i = 1, 2, \ldots, n$, be the case symptom onset times, with $Y_1 = 0$ and $Y_n = T$. Let V_{0i} be the duration from the initial time of viremia to the onset of symptoms, and let V_{1i} be the duration of viremia. Note that the onset of symptoms may occur after the time of viremia has ended.[8] Assume that the timing and duration of viremia for a case is independent of the symptom onset time, that is, Y_i is independent of V_{0i} and V_{1i}; the V_{0i} are independent and identically distributed; the V_{1i} are independent and identically distributed; and the V_{0i} is independent of the V_{1i}.

Y_i is theoretically continuous, though recorded as discrete and rounded to the nearest day. To account for this, rather than using the recorded Y_i, we used a smoothed version X_i derived by setting $X_i = Y_i + h\varepsilon_i$, where $h > 0$ is a smoothing parameter and ε_i is an independent standard normal deviate. The value of h is important and is related to the problem of density estimation or Poisson intensity estimation;[19,24,25] we used the value given by Sheather and Jones[26] as recommended by Jones et al.[27] The values for V_{li} are also rounded to the nearest day, but smoothing of these values did not appreciably affect the results, so to simplify computation we retained the original values of V_{li} in the computations below.

For each case $i = 1, 2, \ldots, n$, $X_i - V_{0i}$ represents the time viremia starts, and $X_i - V_{0i} + V_{1i}$ represents the time viremia ends. For a given time t, case i is thus viremic at t if $X_i - V_{0i} < t < X_i - V_{0i} + V_{1i}$. We used this to count the number of viremic cases at t with the viremia count function

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

where $I_A(x) = 1$ when $x \in A$ and $I_A(x) = 0$ when $x \notin A$. $V(t)$ is a random function, and we assumed that a particular realization of $V(t)$ represents a population of such realizations, the random mechanisms being the infection and subsequent advancement to viremia of an individual in the human population at risk. The expected value of $V(t)$, $E[V(t)]$, is the mean of such a population and is the quantity we want to estimate.

We estimated $E[V(t)]$ using resampling or Monte Carlo simulation, in a method combining ideas from the bootstrap[28] and imputation methods.[21] We evaluated our estimate of $E[(V(t)]$ on an equally spaced grid of points t_r

$\in (0,T)$, for $r = 1, 2, \ldots, 100$. Our approach is to generate many simulated values $V^{(j)}(t_r)$, for $j = 1, 2, \ldots, B$, with B some large number. The mean of these resample values $V^{(j)}(t_r)$ for each fixed t_r provides an estimate of $\mathrm{E}[V(t_r)]$.

For each $j = 1, 2, \ldots, B$, compute a resample value $V^{(j)}(t_r)$ by first creating a resample set $X_1^{(j)}, X_2^{(j)}, \ldots, X_n^{(j)}$ by sampling with replacement from the series $X_1, X_2, \ldots, X_n$, where a new smoothing component ε_i in X_i is generated for each resample. For each resample observation $X_k^{(j)}$, for $k = 1, 2, \ldots, n$, sample independently a value $V_{0k}^{(j)}$ from the density

$$q(v_0) = \frac{15}{8(b-a)}\left(1 - \left[\frac{2x-(a+b)}{b-a}\right]^2\right)^2 I_{[a,b]}(v_0),$$

where and a and b are chosen to correspond to the distribution of V_0s. We adopted this $q(v_0)$ because the only information we have on this variable, as noted in the body of the paper, is that the duration of time from initial viremia to onset, V_{0i}, is roughly 1 to 5 days. The density $q(v_0)$ is symmetric with mean $(a + b)/2$ and variance $(b - a)/20$. We give the general form using the parameters a and b to facilitate future application of these methods; for our application, $a = 1$ and $b = 5$. Sampling from $q(v_0)$ may be performed using, for example, acceptance-rejection sampling.[29] Finally, for each $k = 1, 2, \ldots, B$, resample independently with replacement a value $V_{1k}^{(j)}$ from the observed data shown in Fig. 2. Thus a resample value $V^{(j)}(t_r)$ is obtained by inserting the values $X_k^{(j)}$, $V_{0k}^{(j)}$, and $V_{1k}^{(j)}$ into Equation 1. Repeat this procedure B times to produce values $V^{(j)}(t_r)$, which are then averaged to estimate $\mathrm{E}[V(t)]$ over the grid t_r:

$$\widehat{\mathrm{E}[V(t_r)]} = \frac{1}{B}\sum_{j=1}^{B} V^{(j)}(t_r) = \frac{1}{B}\sum_{j=1}^{B}\sum_{k=1}^{n} I_{(X_k^{(j)} - V_{0k}^{(j)},\, X_k^{(j)} - V_{0k}^{(j)} + V_{1k}^{(j)})}(t_r) \qquad (2)$$

for $r = 1, 2, \ldots, R$. The quantity $\widehat{\mathrm{E}[V(t)]}$ is called the EVC (estimated viremia curve) in the text.

Confidence bands for $\mathrm{E}[V(t)]$ can be computed from $\widehat{\mathrm{E}[V(t)]}$ in various ways. Pointwise 95th percentile bands at each t_r are easily computed, for example, by using the 95th percentile points of $V^{(j)}(t_r)$ for each r as the limits. In the bootstrap literature for density estimation and Poisson process intensity estimation, simultaneous percentile-t bands have been shown to have superior properties as noted by Davison and Hinkley[28] and Cowling et al.[20] in the context of intensity estimation. Similar to interval B_3 of Cowling et al., we used a simultaneous percentile-t approach. Suppressing the subscript r, define

$$T^{(j)}(t) = \frac{V^{(j)}(t) - \widehat{\mathrm{E}[V(t)]}}{\sqrt{\dfrac{V^{(j)}(t)[n - V^{(j)}(t)]}{n}}}$$

for each resample $j = 1, 2, \ldots, B$. For fixed t, $V^{(j)}(t)$ is binomially distributed, and the denominator of $T^{(j)}(t)$ estimates its SD. To compute a $100(1 - \alpha)\%$ simultaneous confidence band $C_{1-\alpha}$ for $\mathrm{E}[V(t)]$ over $t \in (0,\mathrm{T})$, find q_1 and q_2 such that

$$P[q_1 \le T^{(j)}(t) \le q_2, \forall t \in (0,T) \mid X] = \alpha$$

and

$$P[T^{(j)}(t) \le q_1, \forall t \in (0,T) \mid X] = P[T^{(j)}(t) \ge q_2, \forall t \in (0,T) \mid X],$$

where $X = (X_1, \ldots, X_n, V_{01}^{(j)}, \ldots, V_{0n}^{(j)}, V_{11}^{(j)}, \ldots, V_{1n}^{(j)})$ is the data. Then set

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

where $s = \sqrt{\dfrac{\widehat{\mathrm{E}[V(t)]}(n - \widehat{\mathrm{E}[V(t)]})}{n}}$.