Appendix

Human Surveillance Model

Because the human West Nile virus (WNV) case number is low relative to the base population, it was assumed to have a Poisson distribution. Under general conditions, the Poisson provides a good description for the distribution of the numerator for an incidence rate (1). However, our model also allowed for the estimation of "extra-Poisson" variation in case it is also needed to provide an accurate description of these data.

The log linear model used for spatial smoothing assumed that the number of disease cases in the *i*-th county, n_i, has a mean, $P_i\lambda_i$, where P_i, is the denominator for the rates and

$$\lambda_i = \exp\left(\alpha_o + b_i + h_i\right)$$

where α_o is the intercept, b_i is the spatially correlated random variation with mean 0 and variance σ_b^2, and h_i the unstructured extra-Poisson variability with mean 0 and variance σ_h^2. In addition, we assume that both the spatial and the unstructured variability have Gaussian distributions, which are independent in the latter case. On the other hand, the mean for the spatial component, conditional on the means for the contiguous neighbors, is

$$\mathrm{E}\left[\mu_i \middle| \text{mean for all regions}\right] = \frac{\sum_{j \text{ neighbors of region } i} \mu_j}{r_i}$$

where r_i is the number of neighbors for region *i*. The adjacent neighbors for each county were determined by using a geographic information system (GIS, ArcView 3.2, ESRI, Redlands, CA). Thus, the overall log linear model for the number of cases in the *i* county that incorporates both spatial correlation and unstructured variability is

$$\log n_i = \log P_i + b_i + h_i + \alpha_0$$

The population size for county *i* (P_i) was determined from the Census 2000 data.

Markov Chain Monte Carlo (MCMC) simulation methods were used to find Bayesian estimates of the model parameters as implemented in WinBUGS v1.4 (Imperial College and Medical Research Council) (2,3). Gamma prior distribution parameters were assumed for the variances of the Gaussian distributions, and a plot of the history of the simulation was used to determine the number of iterations required for the process to equilibrate. The approach provides improved estimates of county-specific rates that have been spatially smoothed.

In the MCMC method, parameters estimated from each step are used in turn to determine values for the next step; therefore, a good set of initial values is essential before gleaning the values that will be used in the estimation. To accomplish robust parameter estimates, an arbitrary set of values was chosen, and the number of successive steps taken to stabilize the simulations was noted, which is known as the burn-in. The burn-in period was determined through the use of two chains and the modified Gelman-Rubin convergence statistic. This statistic indicates the point at which the process stabilized by describing how well the chains overlap. Final estimates were obtained by using 1,000 iterations as the burn-in

period, and the next 9,000 were used as the sample for deriving the Bayes estimates of the smoothed WNV incidence rates.

Nonhuman Surveillance Model

The quantitative predictive ability of the nonhuman surveillance systems was assessed by once again fitting a log-linear model to the rate of WNV human cases. For this analysis, we instead used a maximum likelihood approach, in which we assumed a Poisson distribution for the number of cases, allowing for extra-Poisson variation by estimating the scale factor. In this model,

$$\log n_i = \log P_i + \beta_A M_i + \beta_M A_i + \alpha_0$$

where P_i is the population offset, A_i is avian mortality attributable to WNV, and M_i is the number of virus-positive mosquito pools. The model was implemented by using GENMOD in SAS (SAS Institute Inc., Cary, NC). Only counties that submitted both mosquito and bird samples were included in the analysis (N = 382) (Appendix Figure).

Appendix References

1. Brillinger DT. The natural variability of vital rates and associated statistics. Biometrics. 1986;42:693–734.
2. Breslow N, Clayton D. Approximate inference in generalized linear mixed models. J Am Stat Assoc. 1993;88:9–25.
3. Clayton D, Kaldor J. Empirical Bayes estimates of age-standardized relative risks for use in disease mapping. Biometrics. 1987;43:671–81.