Web Appendix for

Predicting human West Nile virus infections with mosquito surveillance data

A. Marm Kilpatrick[1] and W. John Pape[2]

1 University of California, Santa Cruz, CA, USA

2 Colorado Dept. Public Health & Environment, Denver, CO, USA

Methods

Study Region

The fifteen counties in Colorado for which we had data for the analyses are shown in Web Figure 1, and the number of human cases in each county in each year is given in Web Table 1.

Mosquito feeding patterns

Data on the fraction of blood meals from mammals were available from Colorado (Weld and Larimer counties) for *Cx. tarsalis*, and *Cx. pipiens* (1), from California and Utah for *Cx. erythrothorax* (2-4), and from CT, MD/DC, NY/NJ, and TN for *Cx. pipiens and Cx. restuans* (5-8). Due to infrequent feedings on humans for *Cx. pipiens*, and *Cx. tarsalis* in the Colorado data (making it difficult to precisely estimate the fraction of blood meals from humans, F_h), we examined additional data on the fraction of mammalian blood meals that came from humans for these mosquito species from studies in California, Texas, and Utah for *Cx. tarsalis* (4, 9-13) and Kansas, California, Connecticut, New York, New Jersey, Tennessee, Maryland and Washington DC for *Cx. pipiens* (2, 5-8, 14, 15). Surprisingly, there was no significant difference in fraction of blood meals from humans or partial vector competence (Web Figure 2).

Estimating prevalence

One challenge in using the Vector Index is that the small number of mosquito traps used by most health departments at each site make it difficult to collect enough mosquitoes in a single week to accurately estimate WNV infection prevalence for each species. In addition, one of our key aims was to determine the predictive power of mosquito surveillance data when testing data was entirely missing, as may be the case under reduced budget conditions. Thus, we considered six ways of estimating prevalence P_i (estimated by maximum likelihood (16)) for each mosquito species *i* that we believe span potential strategies that might be used by local or state public health officials:

1. Statewide Static Prevalence: A statewide estimate, combining all the mosquitoes trapped over all weeks, all years, and across all counties. This essentially generates a risk index in which temporal variation is determined solely by mosquito abundance, with abundances of different mosquito species weighted by their statewide prevalence.
2. Statewide Weekly Prevalence: A statewide estimate from the week of trapping, resulting in different estimates for prevalence for each week of each year. This index ignores local (county) spatial variation in prevalence but explicitly includes temporal and between species differences in prevalence. It enables pooling of testing results across counties to decrease errors due to small numbers of mosquitoes tested, but, in doing so, obscures spatial variation in prevalence.
3. County Prevalence: An estimate for each county across all weeks. This index ignores temporal variation in prevalence, but incorporates among county variability.
4. Local Weekly Prevalence: An estimate from only the week of trapping in that county. Risk was calculated if the number of mosquitoes trapped was >40; otherwise the week was excluded from the analysis due to insufficient mosquitoes to estimate prevalence.

5. Local Two Week Prevalence: An estimate from the current week and previous week. As with Local Weekly Prevalence, risk was calculated if the number of mosquitoes trapped was >40; otherwise the week was excluded from the analysis.
6. Past Weekly Prevalence: An estimate from the calendar week of trapping in that county across all years except the "current" year. This estimate is a potential candidate for what counties might be forced to use in future years if no funds were available to test mosquitoes that year.

Statistical Analyses

We fit local data of the square root of counts of WNV cases vs. risk indices with generalized linear models with a quasi-poisson distribution and a square root link to equalize leverage and linearize relationships. We quantified the explanatory power of correlations with

pseudo-R^2 = 1 – deviance/null deviance

Pseudo-R^2 are approximations of the conventional R^2 but are more appropriate for non-Gaussian generalized linear models.

We fit statewide spatio-temporal numbers of WNV cases with a generalized linear mixed effect models with a poisson distribution and a square root link and with county as a random effect.

Results

Difference in key vectors

Although *Cx. tarsalis* was 3.65 times as abundant as *Cx. pipiens* overall, in two urban counties (Denver and Jefferson) *Cx. pipiens* was both more abundant and more frequently infected with WNV (Web Table 2), and in two other counties (Mesa, Pueblo), *Cx. pipiens* made up more than 39% of the mosquitoes (with *Cx. tarsalis* making up most of the remainder). Finally, in one county (Delta), *Cx. erythrothorax* was nearly as abundant as *Cx. tarsalis*. Thus, while *Cx. tarsalis* is likely the most important WNV vector in Colorado for bird-to-bird, and bird-to-mammal transmission overall, *Cx. pipiens* may be more important in transmitting WNV to birds and mammals in some counties (Web Table 2), depending on the local feeding patterns of *Cx. pipiens* and *Cx. tarsalis* (see Discussion and Supplemental Material).

Comparison of prevalence methods

The correlation of mosquito surveillance data with the number of human WNV cases was much lower when using prevalence estimates that averaged across years within a county (Table 1: County) or across all years and counties (Table: Statewide), because substantial year to year variation was present both in calculated risk, and the number of human cases (Figure 1). Using a two-week running average prevalence measure outperformed using a single week estimate based on the average and case-weighted pseudo-R^2 values (Table 1). This is likely because using two weeks of trapped mosquitoes to estimate prevalence gave a more stable and accurate estimate of prevalence, while still capturing local variation in space and time.

Discussion

In collecting data for this analysis we were surprised by the variability observed in previous studies of mosquito feeding behavior, and discrepancies with conventional wisdom. Conventional wisdom suggested that *Cx. tarsalis* was a more mammalophilic vector and would thus feed more on humans than both *Cx. pipiens* and *Cx. restuans* which were thought to feed primarily on birds (17, 18). Instead, the average fraction of blood meals coming from humans across 24 studies, 550 – 13,600 blood meals/species, and 6-11 regions/species (Web Table 3), was highest for *Cx. restuans* (7.7% ±1 SE 3.5%; range 0-21%), followed by *Cx. pipiens* (4.4% ±

1.4%; range 0-18%), *Cx. erythrothorax* (2.7% ± 1.4%; range 0-5.4%), and *Cx. tarsalis* (0.64% ± 2.7%; range 0-3.9%). What is clearly missing are studies determining the factors the influence mosquito feeding on humans and other mammals, or more generally, on all hosts broadly. Previous efforts have only considered temporal variation in feeding and have been either inconclusive (19), or have identified changes in the abundance of over-utilized hosts (American robins, *Turdus migratorius*) as predictors (20). Future studies should aim to identify causes of spatial variation in feeding patterns, especially with regard to humans. This would enable a more accurate estimating of the vector index that includes the likelihood of mosquitoes feeding on humans. This might help explain why we found no correlations of the number of human WNV cases with the population or population density within a county.

Conventional wisdom had also suggested that *Cx. tarsalis* was a more efficient vector in terms of vector competence than *Cx. pipiens* (17). However, recent results suggest this may not be the case when considering the standard measure of vector competence (the fraction of mosquitoes feeding on an infected blood meal that subsequently transmit) (21), and we found no difference in the part of vector competence relevant for estimating risk indices using WNV testing data (the fraction of infected mosquitoes that can transmit WNV) between the four species in this study (Web Figure 3).

Web Table 1. Numbers of reported human WNV cases and total population for 15 counties in Colorado from 2003 to 2007.

County	2003	2004	2005	2006	2007	Total	Population
Adams	238	15	4	12	32	301	363,857
Arapahoe	140	0	2	2	19	163	487,967
Boulder	457	14	5	76	98	650	291,288
Delta	10	27	1	34	6	78	27,834
Denver	173	3	5	5	29	215	554,636
El Paso	114	2	1	5	4	126	516,929
Fremont	77	4	4	1	15	101	46,145
Jefferson	160	8	6	8	34	216	527,056
Larimer	563	17	13	42	96	731	251,494
Mesa	20	127	10	38	37	232	116,255
Otero	28	0	1	6	10	45	20,311
Prowers	42	3	4	7	7	63	14,483
Pueblo	185	4	5	7	20	221	141,472
Weld	414	8	17	68	98	605	180,936
Morgan	64	0	0	2	14	80	28,109
All counties	2,685	232	78	313	519	3,827	3,568,772

Web Table 2. Number of mosquitoes trapped (WNV+ pools/pools tested) for each species for each county, across all weeks of 2004-2007. Prevalence (bottom row) for each species was estimated by maximum likelihood with bias correction.

Species County	*Cx.* *erythrothorax*	*Cx.* *Pipiens*	*Cx.* *restuans*	*Cx.* *spp.*	*Cx.* *Tarsalis*
Adams	0	687(1/42)	149(0/15)	0	471(0/26)
Alamosa	0	0	0	0	5208(1/111)
Arapahoe	0	203(1/29)	0	0	864(0/50)
Boulder	0	369(0/53)	0	0	2123(0/74)
Chaffee	0	8(0/1)	0	0	243(0/15)
Delta	2967(2/64)	108(0/10)	0	0	1302(4/45)
Denver	0	2814(0/101)	0	1(0/1)	575(1/39)
El Paso	0	26(0.2)	0	59(0/5)	145(0/9)
Fremont	0	0	0	0	196(0/14)
Jefferson	0	475(0/42)	0	0	68(0/17)
La Plata	0	0	0	0	45(0/5)
Larimer	0	455(0/47)	0	39(0/39)	2184(2/79)
Las Animas	0	137(2/13)	0	0	869(4/29)
Mesa	30(0/9)	207(7/50)	0	0	2717(23/101)
Otero	0	514(1/30)	0	0	3068(6/88)
Prowers	0	2(0/1)	0	0	1021(5/28)
Pueblo	0	457(1/35)	0	0	558(1/23)
Weld	0	389(1/42)	0	367(2/33)	3417(9/174)

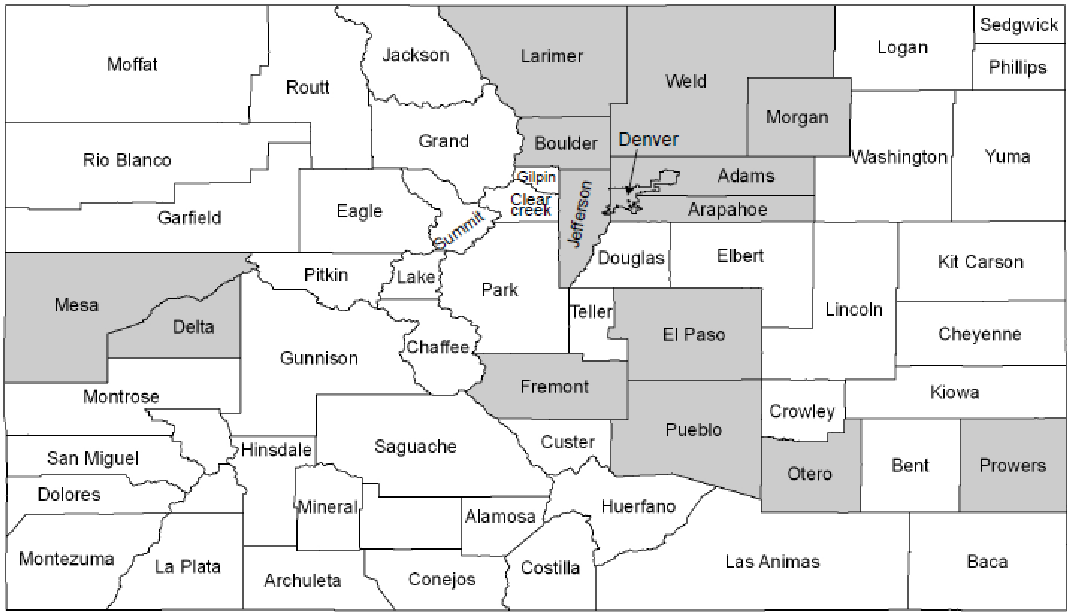
All Counties	2997(2/73)	6851(14/498)	149(0/15)	466(2/51)	25074(56/927)
Prevalence	0.00067	0.0020	0	0.0035	0.0021

Web Table 3. Explanatory power of the vector index using six different methods for estimating prevalence for fifteen counties. The first three columns give the number of counties (out of 15 in the analysis) where the risk index was a significant predictor (P<0.05) of the number of reported human cases, for all three lags (one, two, or three weeks in advance of the date of onset of illness), two of the three lags, or just one lag. The percentage in parentheses in the first column gives the percent of all human WNV cases that occurred in counties where the risk index was a significant predictor for all three lags. The next three columns give the average pseudo-R^2 for all fifteen counties for that lag (see Web Figure 3 for detailed results), and last column gives the weighted average pseudo-R^2 across all three lags for the Risk index for that prevalence estimate where the weights are the number of human cases in that county.

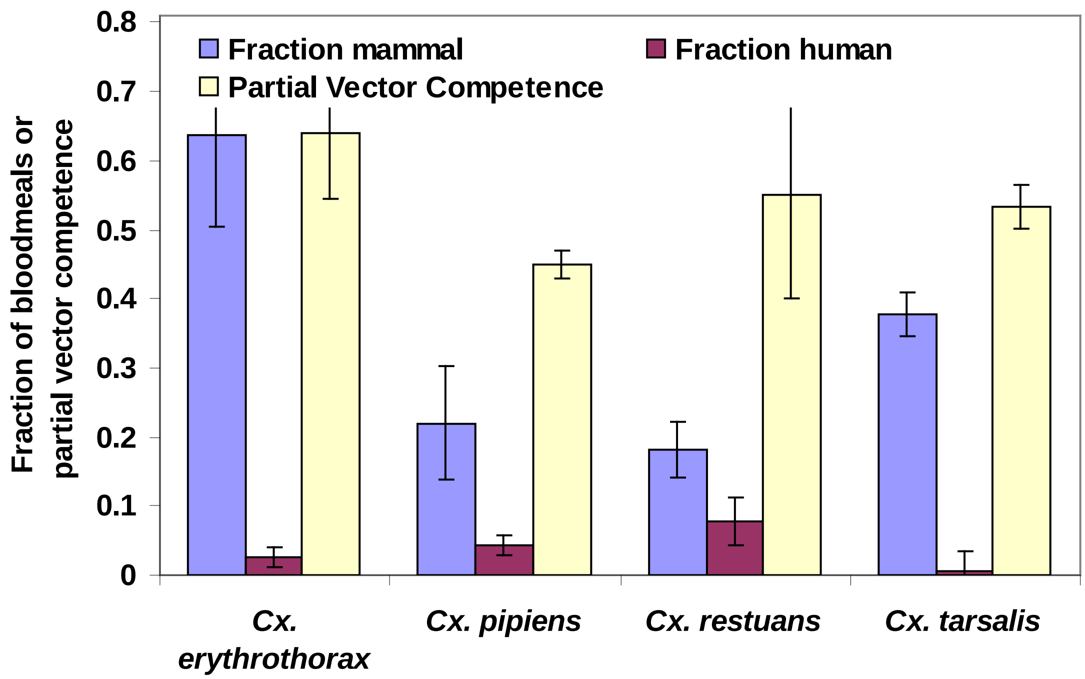
Prevalence Estimate	Counties all 3 lags (% of human cases)	Counties 2 lags	Counties 1 lag	Average pseudo-R^2, all counties 1 week lag	2 week lag	3 week lag	Case-weighted pseudo-R^2
Statewide	10 (90%)	2	0	0.30	0.29	0.32	0.33
Statewide Weekly	14(98%)	0	0	0.54	0.47	0.40	0.53
County	7(78%)	4	1	0.31	0.30	0.32	0.35
Local week	12(90%)	1	0	0.46	0.39	0.29	0.44
Local two-week	11(88%)	3	0	0.50	0.43	0.30	0.45
Local weekly (no	5(52%)	3	2	0.14	0.10	0.06	0.13

current year data)*

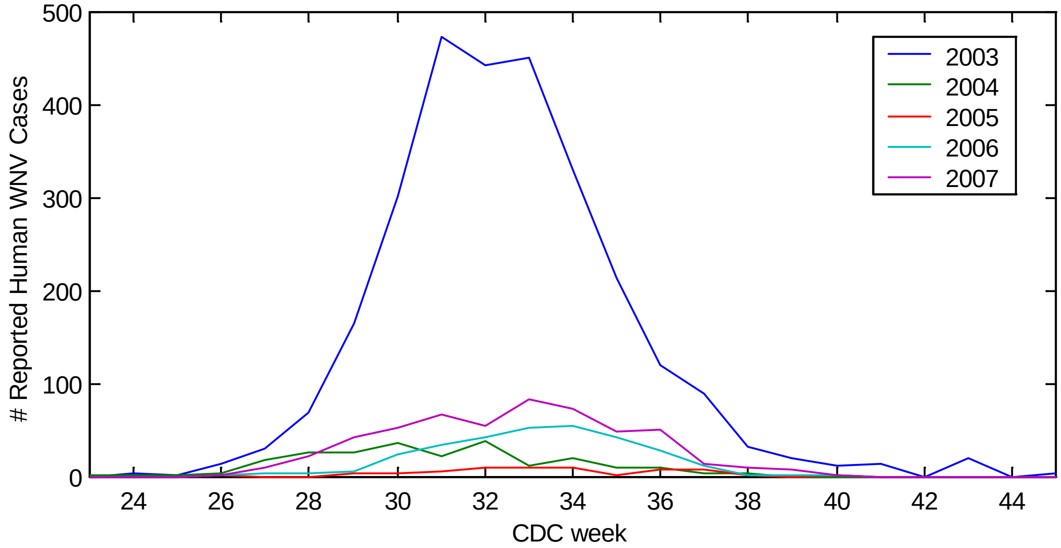
*One county only had sufficient mosquito surveillance data for analysis in one year (Morgan), so this risk measure could not be calculated and assessed. Thus total in 2nd column is out of 14 counties instead of 15.



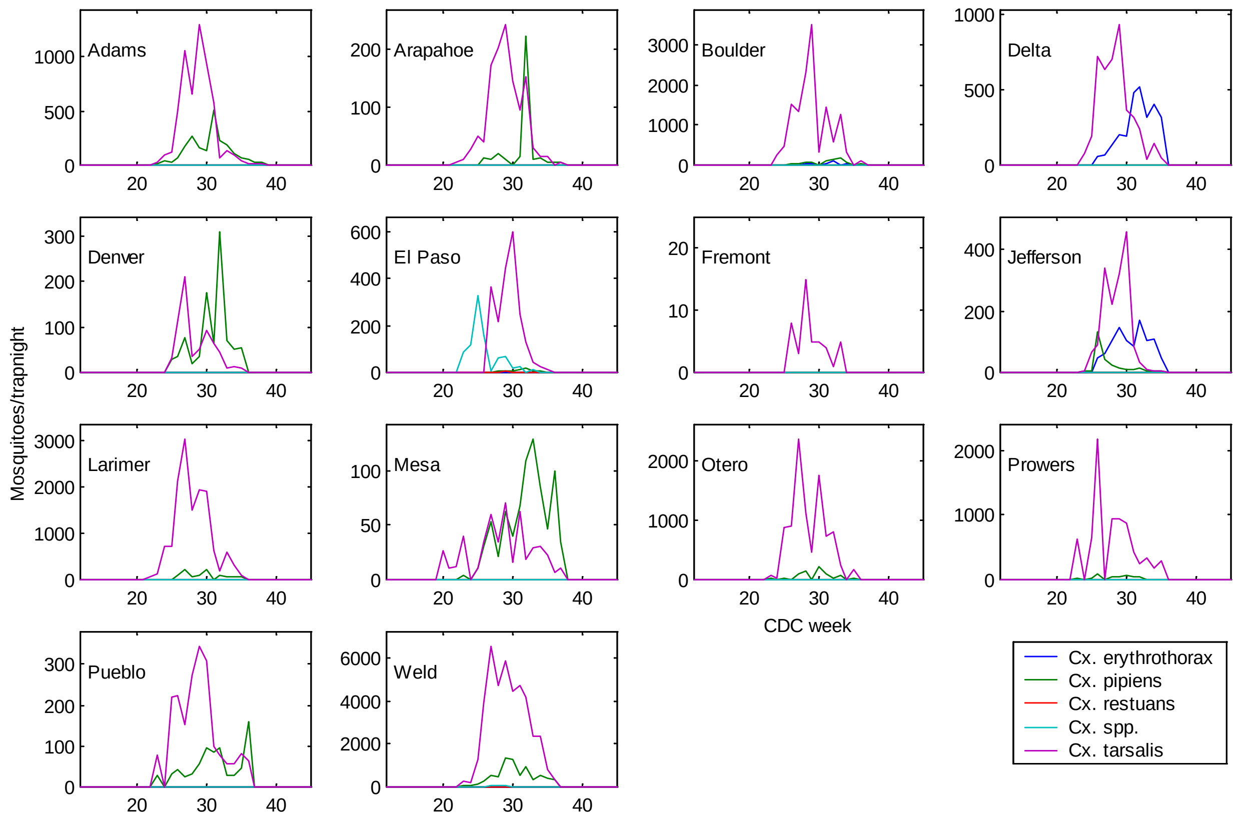
Web Figure 1. Map of Colorado counties. Shaded counties are those used in this study.



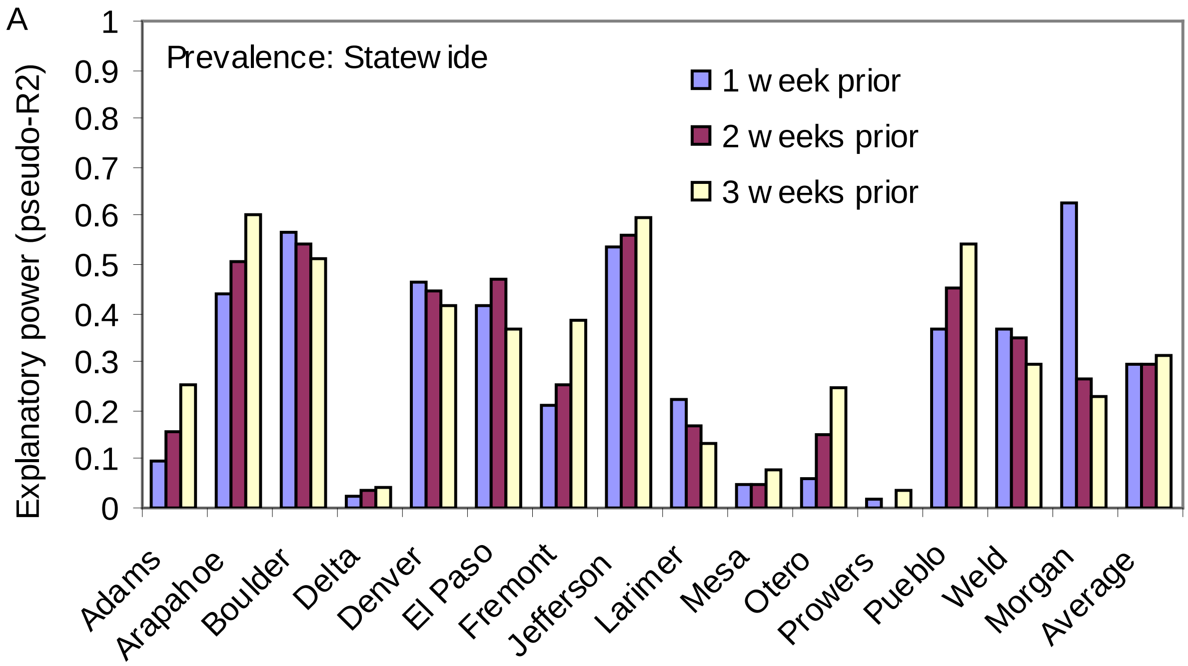
Web Figure 2. Feeding patterns and partial vector competence for four *Culex* mosquito species. Error bars show 1 standard error. The fraction of blood meals from mammals, F_m, differs significantly between mosquito species (ANOVA with arc-sin square-root transformed data: $F_{3,29}=4.47$; P = 0.011, but the fraction from humans, F_h, does not ($F_{3,25}=1.71$; P = 0.19). Partial vector competence, C_v, also did not differ significantly between species ($F_{3,11}$; P = 0.93).

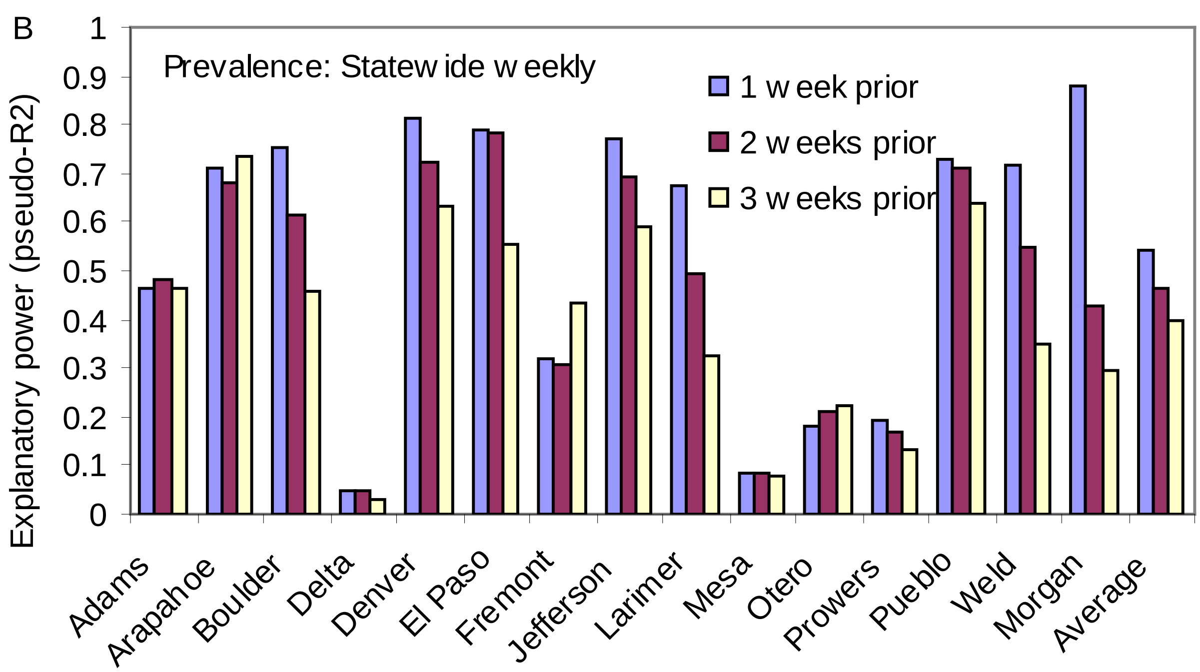


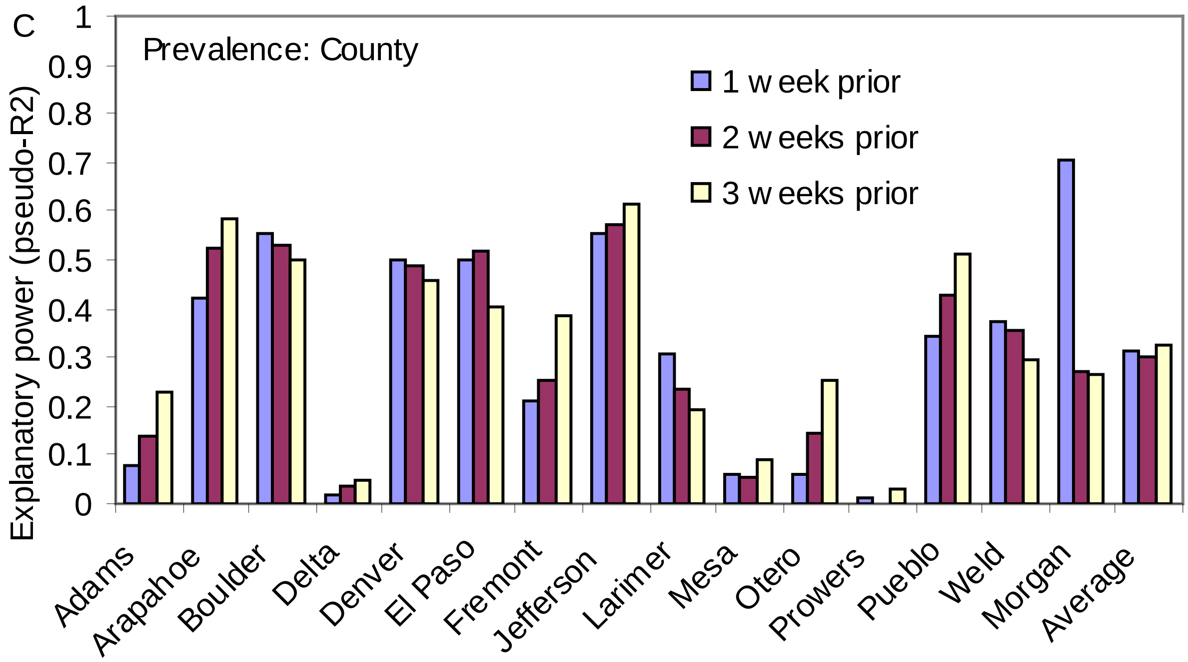
Web Figure 3. Number of human WNV cases reported each week in Colorado, from 2003 to 2007.

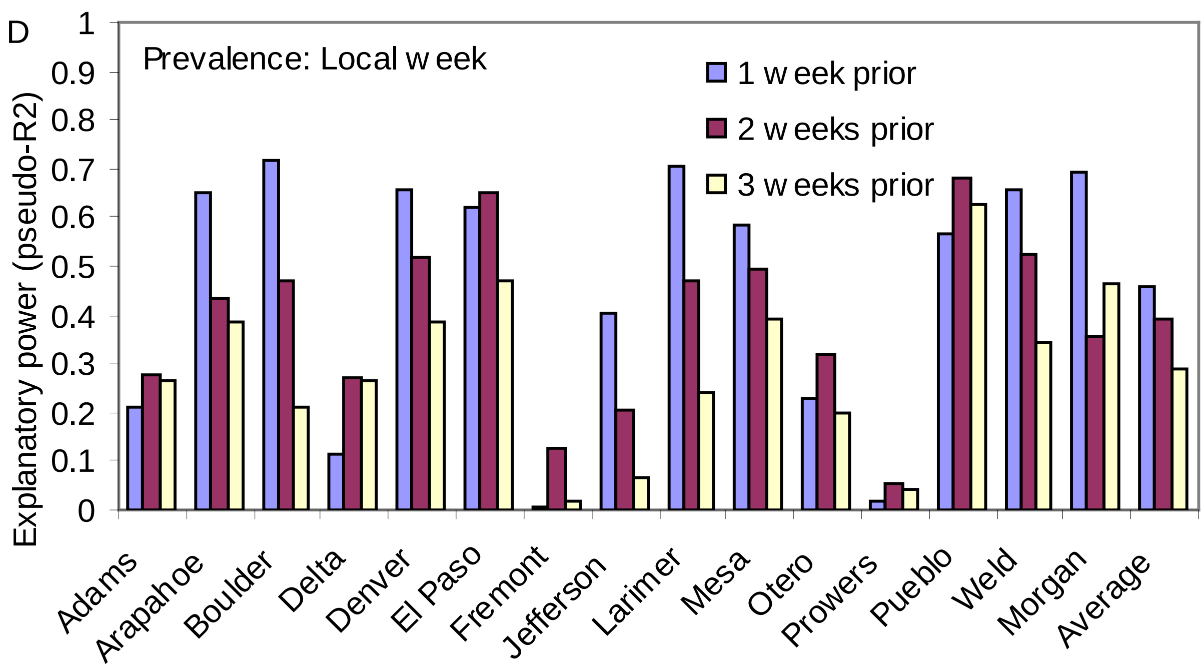


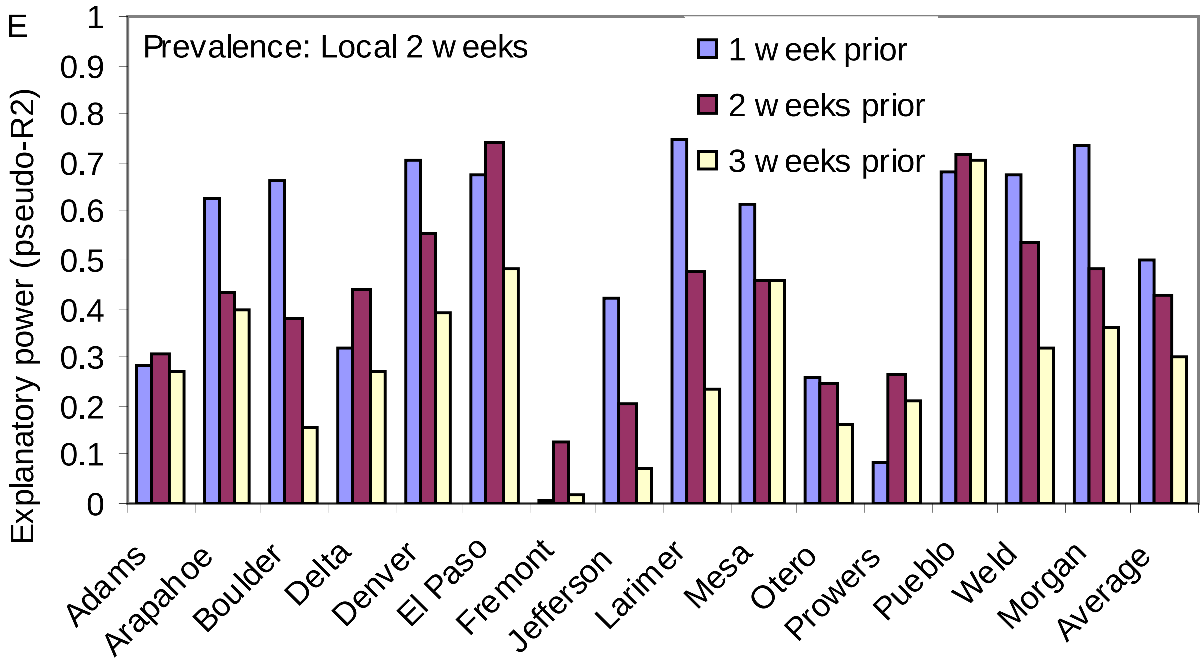
Web Figure 4. Mosquito abundance, by species, in fourteen counties in Colorado in 2007 vs. CDC week. Note different y-axis scales.

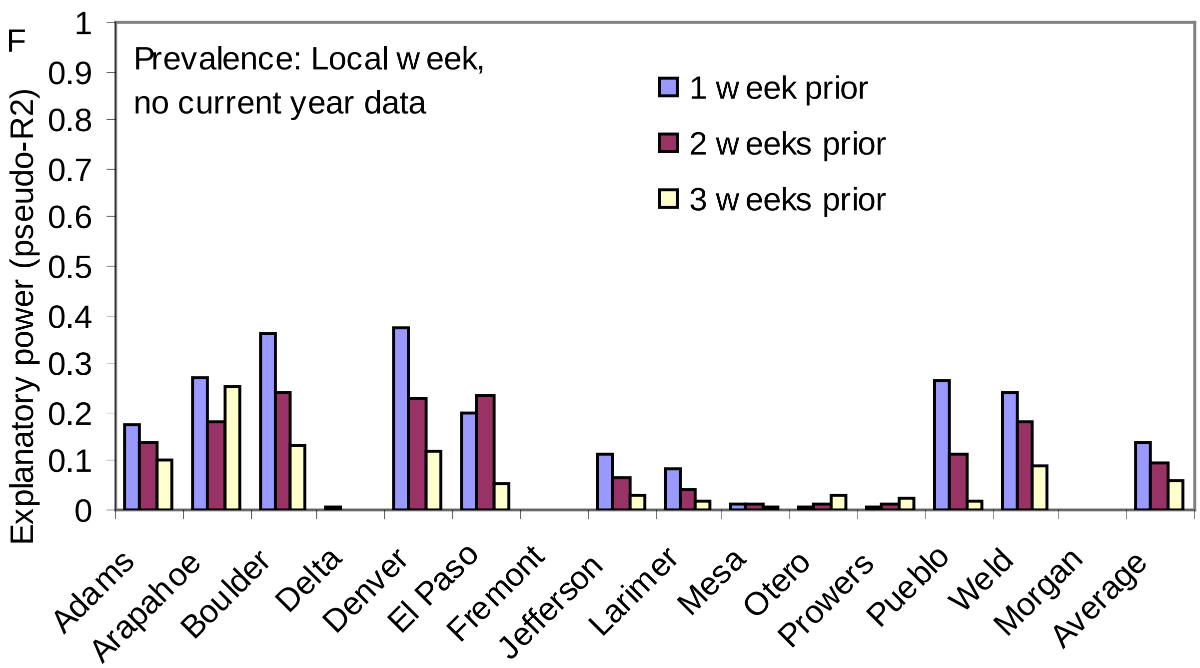

A
Prevalence: Statewide
1 week prior
2 weeks prior
3 weeks prior
Explanatory power (pseudo-R2)
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
Adams
Arapahoe
Boulder
Delta
Denver
El Paso
Fremont
Jefferson
Larimer
Mesa
Otero
Prowers
Pueblo
Weld
Morgan
Average


B
Prevalence: Statewide weekly
1 week prior
2 weeks prior
3 weeks prior
Explanatory power (pseudo-R2)
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
Adams
Arapahoe
Boulder
Delta
Denver
El Paso
Fremont
Jefferson
Larimer
Mesa
Otero
Prowers
Pueblo
Weld
Morgan
Average


C
Prevalence: County
1 w eek prior
2 w eeks prior
3 w eeks prior
Explanatory power (pseudo-R2)
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
Adams
Arapahoe
Boulder
Delta
Denver
El Paso
Fremont
Jefferson
Larimer
Mesa
Otero
Prowers
Pueblo
Weld
Morgan
Average


D
Prevalence: Local w eek
1 w eek prior
2 w eeks prior
3 w eeks prior
Explanatory power (pseudo-R2)
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
Adams
Arapahoe
Boulder
Delta
Denver
El Paso
Fremont
Jefferson
Larimer
Mesa
Otero
Prowers
Pueblo
Weld
Morgan
Average


E
Prevalence: Local 2 w eeks
1 w eek prior
2 w eeks prior
3 w eeks prior
Explanatory power (pseudo-R2)
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
Adams
Arapahoe
Boulder
Delta
Denver
El Paso
Fremont
Jefferson
Larimer
Mesa
Otero
Prowers
Pueblo
Weld
Morgan
Average


F
Prevalence: Local w eek,
no current year data
1 w eek prior
2 w eeks prior
3 w eeks prior
Explanatory power (pseudo-R2)
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
Adams
Arapahoe
Boulder
Delta
Denver
El Paso
Fremont
Jefferson
Larimer
Mesa
Otero
Prowers
Pueblo
Weld
Morgan
Average

Web Figure 5. Predictive power (measured using the pseudo-R^2) of the square root of risk indices for predicting the square root of the number of human WNV cases using six different methods to estimate prevalence for fifteen counties in Colorado, 2003-2007. A) Statewide: prevalence estimate combines all mosquitoes of each species trapped over all years in all counties. It is essentially a sum of mosquito abundance weighted by the average prevalence for each species across the five years. Correlations using this risk index were significant ($P<0.05$) for all counties for all lags, except: Freemont and Otero (which were significant for 2 and 3 week lags), Morgan (which was significant for 1 week lag only), and Delta and Prowers (which were non-significant for all lags). B) Statewide weekly: prevalence estimate combines all mosquitoes of each species trapped over all years in all counties for each week. This risk index was a significant predictor ($P<0.05$) for all counties for all lags, except Delta, which was non-significant for all lags. C) County: prevalence was estimated using all mosquitoes trapped across all weeks in that county. This index was a significant predictor ($P<0.05$) for eight of the fifteen counties for all lags, for one and two week lags in Arapahoe, for two and three week lags in Adams, Fremont and Otero, at one week lag in Morgan, and was non-significant for all lags in Prowers and Delta. D) Local week: prevalence was estimated using mosquitoes trapped in that county in that week, and risk is estimated if number of mosquitoes trapped is >40. This index was a significant predictor ($P<0.05$) for twelve of the fifteen counties for all lags and for Jefferson for 1 and 2 week lags, but was non-significant for all lags for Fremont and Prowers. E) Local two weeks: prevalence was estimated using mosquitoes trapped in the current and previous week in that county. This index was a significant predictor ($P<0.05$) for twelve of the fifteen counties for all 3 lags, for 1 and 2 week lags in Jefferson and Morgan, for 2 and 3

week lags in Prowers and was non-significant for all lags in Fremont. F) Local week, no current year data: prevalence was estimated using mosquitoes trapped in all years during the current week in the local county (similar to the statewide weekly index), but excludes testing results from mosquitoes trapped from the current year (simulating the situation where no funds are available to test mosquitoes locally). Risk was only a significant predictor (P<0.05) for five of the fifteen counties for all 3 lags (Adams, Arapahoe, Boulder, Denver, Weld), for 1 and 2 week lags in El Paso, for a one week lag in Jefferson and Larimer, and was non-significant for all lags for five counties (Delta, Mesa, Otero, Prowers, Fremont). There was insufficient mosquito surveillance data from Morgan except in a single year (2003) so this prevalence method could not be used for this county.

Web Appendix References

1. Tempelis CH, Francy DB, Hayes RO, et al. Variations in feeding patterns of 7 culicine mosquitoes on vertebrate hosts in Weld and Larimer counties Colorado. *Am J Trop Med Hyg* 1967;16(1):111-119.
2. Gunstream SE, Chew RM, Hagstrum DW, et al. Feeding Patterns of 6 Species of Mosquitoes in Arid Southeastern California. *Mosquito News* 1971;31(1):99-101.
3. Tempelis CH. Estimation of vectorial capacity: mosquito host selection. *Bulletin of the Society for Vector Ecology* 1989;14(1):55-59.
4. Reisen WK, Milby MM, Presser SB, et al. Ecology of mosquitos and St. Louis Encephalitis Virus in the Los Angeles Basin of California, 1987-1990. *J Med Entomol* 1992;29(4):582-598.
5. Molaei G, Andreadis T, Armstrong P, et al. Host feeding patterns of *Culex* mosquitoes and West Nile virus transmission, northeastern United States. *Emerg Infect Dis* 2006;12(3):468-474.
6. Kilpatrick AM, Daszak P, Jones MJ, et al. Host heterogeneity dominates West Nile virus transmission. *Proceedings of the Royal Society B: Biological Sciences* 2006;273(1599):2327-2333.
7. Apperson CS, Hassan HK, Harrison BA, et al. Host feeding patterns of established and potential mosquito vectors of West Nile virus in the eastern United States. *Vector-Borne Zoonotic Dis* 2004;4(1):71-82.
8. Savage HM, Aggarwal D, Apperson CS, et al. Host choice and West Nile virus infection rates in blood fed mosquitoes, including members of the *Culex pipiens* complex, from

Memphis and Shelby County, Tennessee 2002-2003. *Vector-Borne Zoonotic Diseases* 2007;7(3):365-386.
9. Hayes RO, Tempelis CH, Hess AD, et al. Mosquito host preference studies in Hale County, Texas. *Am J Trop Med Hyg* 1973;22(2):270-277.
10. Andersen DM, Collett GC, Winget RN. Preliminary host preference studies of *Culex tarsalis* Coquillett and *Culiseta inornata* (Wiluston) in Utah. *Mosquito News* 1967;27(1):12-15.
11. Tempelis CH, Reeves WC, Bellamy RE, et al. A 3-Year study of feeding habits of *Culex tarsalis* in Kern County California. *Am J Trop Med Hyg* 1965;14(1):170-177.
12. Tempelis CH, Washino RK. Host-feeding patterns of *Culex tarsalis* in Sacramento Valley California with notes on other species. *J Med Entomol* 1967;4(3):315-318.
13. Wekesa JW, Yuval B, Washino RK, et al. Blood feeding patterns of *Anopheles freeborni* and *Culex tarsalis* (Diptera: Culicidae): effects of habitat and host abundance. *Bull Entomol Res* 1997;87(6):633-641.
14. Edman JD, Downe AER. Host-blood sources and multiple-feeding habits of mosquitoes in Kansas. *Mosq News* 1964;24(2):154-160.
15. Apperson CS, Harrison BA, Unnasch TR, et al. Host-feeding habits of *Culex* and other mosquitoes (Diptera: Culicidae) in the Borough of Queens in New York City, with characters and techniques for identification of Culex mosquitoes. *J Med Entomol* 2002;39:777-785.
16. Biggerstaff BJ. Confidence intervals for the difference of two proportions estimated from pooled samples. *Journal of Agricultural Biological and Environmental Statistics* 2008;13(4):478-496.
17. Turell MJ, Dohm DJ, Sardelis MR, et al. An update on the potential of North American mosquitoes (Diptera : Culicidae) to transmit West Nile virus. *J Med Entomol* 2005;42(1):57-62.
18. Turell MJ, Sardelis MR, O'Guinn ML, et al. Potential vectors of West Nile virus in North America. In: Mackenzie J, Barrett A, Deubel V, eds. *Japanese Encephalitis and West Nile Viruses Vol 267 Current Topics in Microbiology and Immunology*. Berlin: Springer-Verlag, 2002:241-252.
19. Edman JD, Taylor DJ. *Culex Nigripalpus* - Seasonal Shift in Bird-Mammal Feeding Ratio in a Mosquito Vector of Human Encephalitis. *Science* 1968;161(3836):67-68.
20. Kilpatrick AM, Kramer LD, Jones MJ, et al. West Nile virus epidemics in North America are driven by shifts in mosquito feeding behavior. *PLoS Biology* 2006;4(4):606-610.
21. Reisen WK, Barker CM, Fang Y, et al. Does variation in *Culex* (Diptera: Culicidae) vector competence enable outbreaks of West Nile virus in California? *J Med Entomol* 2008;45(6):1126-1138.