Modelling Geographic Variations in West Nile Virus

Nikolaos W. Yiannakoulias, MA[1,2]
Donald P. Schopflocher, PhD[2-4]
Lawrence W. Svenson, BSc[2-4]

ABSTRACT

Background: This paper applies a method for modelling the spatial variation of West Nile virus (WNv) in humans using bird, environmental and human testing data.

Methods: We used data collected from 503 Alberta municipalities. In order to manage the effects of residual spatial autocorrelation, we used generalized linear mixed models (GLMM) to model the incidence of infection.

Results: There were 275 confirmed cases of WNv in the 2003 calendar year in Alberta. Our spatial model indicates that living in the grasslands natural region and levels of human testing are significant positive predictors of WNv; living in an urban area is a significant negative predictor.

Conclusion: Infected bird data contribute little to our model. The variability of West Nile virus incidence in Alberta may be partly confounded by the variations in the rate of testing in different parts of the province. However, variation in infection is also associated with known environmental risk factors. Our findings are consistent with existing knowledge of WNv in North America.

MeSH terms: West Nile virus; regression analysis; decision support techniques

La traduction du résumé se trouve à la fin de l'article.

1. Department of Earth and Atmospheric Sciences, University of Alberta, Edmonton, AB
2. Public Health Surveillance and Environmental Health, Alberta Health and Wellness, Edmonton
3. Department of Public Health Sciences, University of Alberta
4. Department of Community Health Sciences, University of Calgary, Calgary, AB

Correspondence and reprint requests: Mr. Nikolaos Yiannakoulias, Department of Earth and Atmospheric Sciences, 1-26 Earth Sciences Building, University of Alberta, Edmonton, AB T6G 2E3, E-mail: nwy@ualberta.ca

Acknowledgements: The authors thank Debra Mooney for her advice and support in the preparation of this manuscript.

The spread of West Nile virus (WNv) to new parts of the world has necessitated a renewed interest in the disease.[1-3] The 1999 outbreak of WNv in New York City was soon followed by a spread of infection across North America. Since the initial outbreak, human cases have been found throughout the United States and Canada, and evidence of WNv has been found as far south as the Caribbean and Mexico.[4]

A number of North American mosquito species are suitable vectors of WNv.[5,6] Of these, one of the most likely vectors in Western Canada is *Culex tarsalis*.[7] This species is present across the Southern non-alpine regions of the prairies, as well as the British Columbia interior.[8] Although other prairie-dwelling mosquito species are competent vectors of WNv, *Cx. tarsalis* feeds on avian and mammalian hosts, particularly in the summer and fall months.[7] These feeding habits makes it well suited for disease transmission between infected birds and humans.

With a few recent exceptions,[9,10] little research has considered the geographical (spatial) dimensions of WNv in North America. To aid in the allocation of resources that mitigate future outbreaks (such as funding for mosquito fogging/spraying, or local awareness campaigns), we developed a spatial model to predict municipal-level incidence of human WNv infections. We believe that an explicitly spatial approach is necessary since mosquito habitat and population density vary considerably across the province.

METHODS

Data

Our analysis includes 503 Alberta cities, towns and villages (which we refer to as 'municipalities') that have a population greater than 100 people. Municipal populations are defined based on mailing address; residents residing near rural areas are still assigned to a municipality, even if they do not live within the city/town limits. Municipalities aggregate well into larger administrative regions, and are of significant administrative and political importance.[11]

Several data sources were available at the municipal level. Human infection and testing data were acquired through a commu-

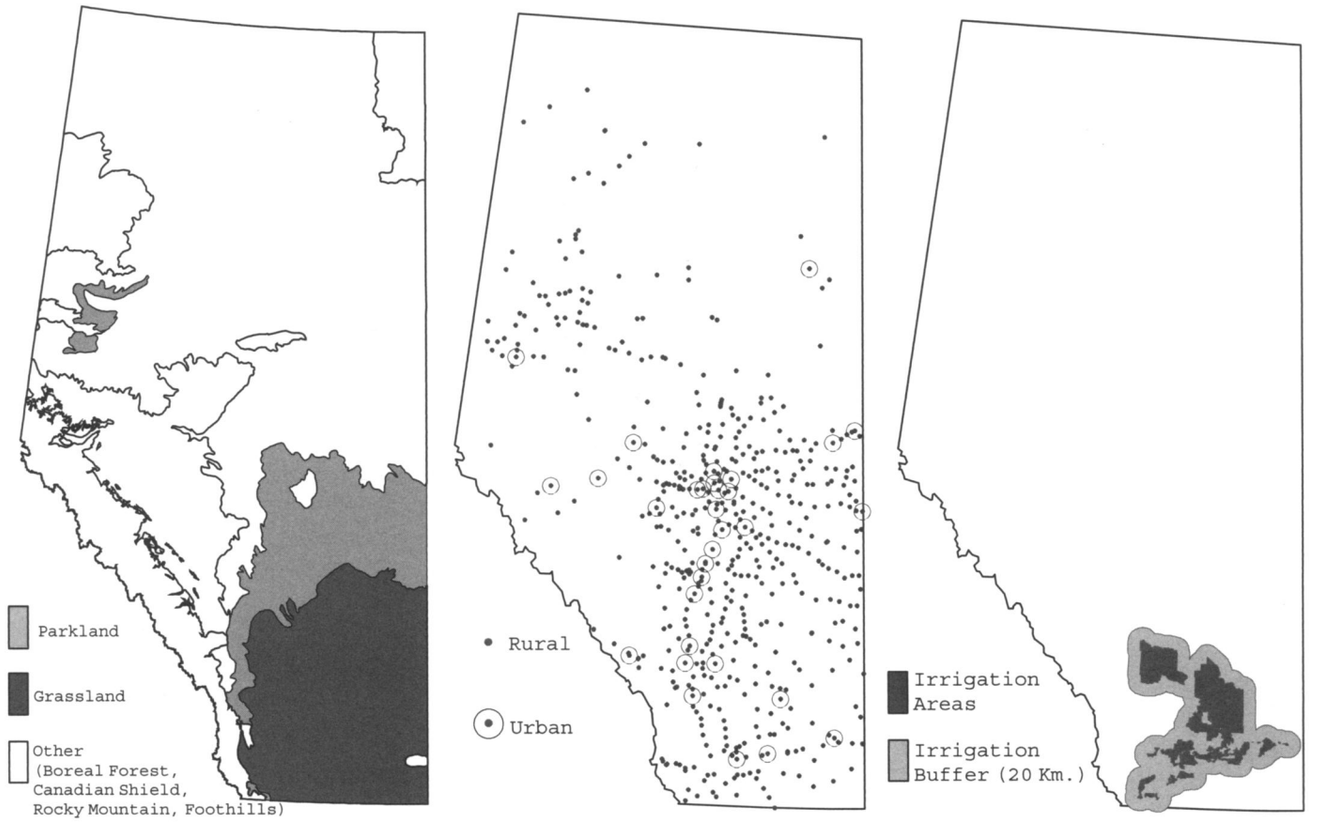

Figure 1a-c. Maps of Alberta with ecological and municipal features
a. Natural regions of Alberta
b. Urban and rural municipalities
c. Irrigation areas

nicable disease reporting system administered by Alberta Health and Wellness. North American research suggests that 79% of persons infected with WNv are asymptomatic.[12] As a result, our reported cases are a proxy of the real presence of disease in the population. Infected bird data were acquired through a passive surveillance system administered by Alberta Sustainable Resource Development in which the public turned in dead birds. We also incorporated irrigation, ecological and urban/rural information into the model (Figure 1a-c). These variables were meant to capture the environmental suitability of a municipality for sustaining a population of *Cx. tarsalis*. Irrigation data were obtained from Alberta Agriculture, and were included in the model as a dichotomous indicator variable. Municipalities within a 20-km distance of an irrigation district were considered within an area of irrigation. This 20-km buffer is meant to take into account the short-range dispersion characteristics of *Cx. tarsalis*.[7,13] Municipalities with a population greater than or equal to 10,000 persons were considered urban, the remaining municipalities were considered rural. This corresponds to the definition of city as established by the Alberta *Municipal Government Act*.[11] The six natural regions represent a number of different ecological characteristics – such as physical landscape, temperature and precipitation.[14] We put the Boreal Forest, Rocky Mountain, Foothills and Canadian Shield regions into a single index variable. These regions are not known to support *Cx. tarsalis* habitat.[8] The Grasslands and Parkland regions were entered into the model as separate dummy variables.

Analysis

We combined these data into a regression model formulated to predict cases of West Nile virus infection at the municipal level. Counts of disease at the municipal level were incorporated into a generalized linear model. Since municipal West Nile virus counts are relatively rare events, we assume a Poisson-distributed dependent variable. The log of the municipal population was used as an offset variable in these Poisson regression models. By log transforming the population, we scaled the population variation between larger municipalities (such as Edmonton and Calgary) and smaller municipalities.

Spatial phenomena exhibit two general classes of patterns: first-order effects (trend) and second-order effects (dependence/autocorrelation).[15,16] Since West Nile virus is not normally communicable between people, we would not expect any direct second-order effects at the municipal level; i.e., the frequency of infection in one municipality should not have much direct impact on the frequency of infection in another municipality. Nonetheless, spatial dependence may also be an issue when the features that influence frequency of WNv are spatially dependent – such as the interaction between vectors and other host species. When spatial dependence is not accounted for in a model, standard errors of model

TABLE I

Ten Municipalities with Highest Number of Positive WNv Cases

Community	Positive Human Cases (obs)	Expected Human Cases (exp)	Incidence Ratio (obs/exp)	1 - (p≤obs)	Persons Tested	WNv Testing Rate/100,000 (tests/population)
Medicine Hat	62	4.940	12.551	0.000	230	26.758
Calgary	34	81.614	0.417	1.000	504	2.402
Bow Island	18	0.256	70.224	0.000	44	224.305
Edmonton	15	61.110	0.245	1.000	249	2.255
Taber	14	0.909	15.399	0.000	30	52.539
Brooks	13	1.456	8.927	0.000	23	28.742
Lethbridge	9	6.217	1.448	0.101	110	14.723
Foremost	8	0.111	72.015	0.000	10	247.628
Hanna	7	0.379	18.451	0.000	22	107.701
Redcliff	5	0.434	11.526	0.000	12	69.658

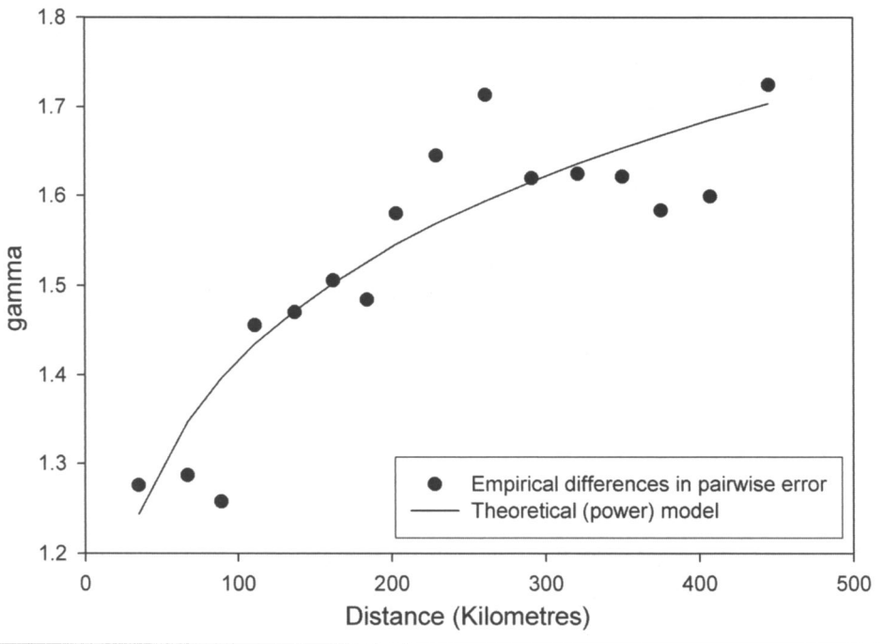

Figure 2. Variogram of empirical and theoretical spatial dependence in residual error

estimates are too large.[17] In the presence of these effects, there are methods available that allow for dependence between observations of the dependent variable (in this study, the count of human WNv cases). We used the SAS GLIMMIX macro to solve both a 'spatial' and 'non-spatial' model. For the spatial model, we conditioned variance on a theoretical model of spatial dependence based on an empirical variogram of the residuals of a non-spatial model.[18] Less technically, the pattern of spatial dependence in the error from the non-spatial model was used to inform the spatial model. This approach produces model estimates that take spatial-dependent error into account by including spatially-dependent random effects. A detailed overview of the spatial modelling procedure we employed can be found elsewhere.[17-19]

RESULTS

In 2003, 275 positive cases of WNv were identified in Alberta. Seroprevalence estimates suggest an asymptomatic infection rate of 3 per 1000 in the province but a rate of 46 per 1000 in rural Southeastern Alberta.[20] The resolution and sample size of the seroprevalence survey was not sufficient to estimate the presence of disease in individuals at the municipal level.

The ten municipalities with the largest number of positive human cases are shown in Table I. Incidence of disease and the rate of testing vary considerably between municipalities. The municipalities with the highest testing rates also appear to have the highest disease incidence.

We use a variogram to visualize the magnitude of spatial dependence in the residuals of the non-spatial model (Figure 2). The vertical axis describes the variation of residual error between municipalities – values near the origin indicate greater similarity of residuals. The horizontal axis measures pairwise geographic distance (i.e., distance between municipalities), increasing from left to right. The points on the graph represent the pairwise differences in residual error with increasing distance. Based on the variogram, we see that there is spatial dependence of residual error among municipalities within 200 km of each other. The flattening out of the curve at greater distances indicates the point at which dependence in error dissipates. Based on the appearance of this graph, we chose a power variogram model to represent this structure in the spatial modelling process.

The results from the non-spatial and spatial Poisson regression models are compiled into Table II. All variables but one are significant in both non-spatial models. In the spatial model, residence in the grasslands region and testing rates have a significant positive association with disease incidence, and urban areas have a significant negative association with disease incidence.

DISCUSSION

Since 2003, WNv has not been found in significant numbers in the human population of Alberta. With only one year of data, it was not possible to use time series approaches to predict future incidence of WNv for each Alberta municipality. Instead, we used a spatial model to illustrate how rates of human infection vary across Alberta municipalities. Admittedly, the dynamics of WNv are probably changing, and the exact character of future outbreaks (if they occur) may be hard to antic-

TABLE II

Non-spatial and Spatial Model Results

Effect	Poisson Regression, Non-spatial		Poisson Regression, Non-spatial*		Poisson Regression, Spatial		Poisson Regression, Spatial*	
	Estimate	p	Estimate	p	Estimate	p	Estimate	p
Intercept	-1.6369	0.0005	-1.4571	0.0079	-1.6059	<0.0001	-1.5167	0.0002
% positive birds	0.0148	0.0001	0.0144	0.0002	0.0058	0.2058	0.0088	0.0615
Irrigation (1/0)	0.6728	0.0021	1.2128	0.0001	-0.4929	0.1280	-0.0673	0.8474
Natural areas	–	–	–	–	–	–	–	–
Parkland	0.3819	0.4615	0.7434	0.2166	0.7329	0.1024	1.0727	0.0220
Grassland	1.5975	0.0015	1.8559	0.0017	2.5438	0.0001	2.8205	0.0001
Rate of human testing	228.2400	0.0001	–	–	304.1700	<0.0001	–	–
Urban	-0.4490	0.0295	0.2148	0.0002	-0.6729	0.0271	-1.0088	0.0041

* Human testing data not included in the model

ipate; indeed, the 2004 and 2005 seasons of West Nile produced very few human cases. Nonetheless, the findings of our model are consistent with information available on the distribution of disease, and we have shown that municipal-level spatial models provide a reasonable framework for analysis, and resource allocation.

Positive bird information does not appear to offer much predictive value in our spatial model. Birds are indicators, not vectors of disease, and data collected were based on voluntary public surveillance. The majority of birds identified as positive for WNv were found in the Edmonton area, where the risk of infection was comparatively low. Several factors unrelated to human exposure could explain patterns of bird infection in some municipalities, such as localized abundance of competent avian-feeding mosquito species which are less likely to transmit infection to humans. The distribution of the bird population, sampling bias and other factors could all complicate any meaningful associations between WNv infection in humans and birds. Irrigation data may have also been too crude to measure the local features of a municipality that could increase its propensity for WNv. Our measure of irrigation areas only gives a general indication of the irrigation activity, and does not describe the local factors which are more likely to explain the variation in mosquito habitats. We tested the model with different irrigation buffers (0, 50 and 100 km) and found similar results.

Municipal testing rate appears to be a methodological confounder. Increased awareness of WNv may have led to higher testing in some areas, which may have led to a higher sensitivity of detection. Geographic inconsistency in disease diagnosis has been observed in the United States[21] and internationally[22] and may be related to physician behaviour, availability of resources, or even media/cultural influences.[23] The relative predictive power of testing rates in all our models is consistent with the suggestion that variation in testing practices may be partly responsible for the observed variation in rates of WNv. The variation in WNv infection we observed exceeds the range of variation observed in a seroprevalence survey over the same year. For example, the incidence-ratio in Bow Island (70.2) exceeds the seroprevalence incidence-ratio of rural Southeastern Alberta (15.3). It is also possible that higher real infection in some municipalities led to higher presentation of symptoms to physicians, and in turn, higher testing. In this case, the higher testing rates may simply be the result of higher disease incidence. Two expected risk factors remain significant in the model when testing frequency was added, and seroprevalence information confirmed that WNv was more common in Southeastern Alberta in 2003. Therefore, we also caution against any conclusion that the variation in infection rates is *only* the result of variations in testing.

Based on currently available data, the most parsimonious model for predicting future patterns of West Nile virus in Alberta would include three variables: natural region of residence (as a series of dummy variables), urban/rural residence and rate of testing for WNv in humans. The first variable describes the regional habitat of the vector; the grasslands natural region supports *Cx. tarsalis*. The second variable may be a more local indicator of habitat. *Cx. tarsalis* is a 'foul water'-preferring species found in rural and suburban rather than urban areas.[24] Both of these variables may be related to human exposure as well. The grasslands natural region supports a large number of farms, and perhaps a large number of outdoor workers who are likely to spend more time outdoors than urban residents. Both of these characteristics could result in increased exposure to mosquitoes.

Awareness of disease, and a higher testing rate in some municipalities, may have influenced the observed rates of illness, but known risk factors remain important in our spatial model of WNv. Predicting future outbreaks may be confounded by testing behaviour, public awareness, physician practice style and temporal interactions between these factors. For example, a single positive case may result in public awareness that spawns a surge in testing, which further increases the number of confirmed cases. It may be difficult to separate true local spikes of infection from spikes that are due to increased testing frequency until enough years of WNv infection have occurred to accumulate robust temporal data. In the future, a spatial-temporal approach to investigating outbreaks may provide information that helps untangle these issues.

REFERENCES

1. Sayao A, Suchowersky O, Al-Khathaami A, Klassen B, Katz NR, Sevick R, et al. Calgary experience with West Nile virus neurological syndrome during the late summer of 2003. *Can J Neurol Sci* 2004;31:194-203.
2. Weir E, Shapiro H. West Nile virus: Round five. *CMAJ* 2004;170:1669-70.
3. Power C, van Marle G. The emergence of West Nile virus in Canada. *Can J Neurol Sci* 2004;31:135-37.
4. Hayes EB, Komar N, Nasci RX, Montgomery SP, O'Leary DR, Campbell GL. Epidemiology and transmission dynamics of West Nile virus disease. *Emerg Infect Dis* 2005;11:1167-73.
5. Turell MJ, O'Guinn ML, Dohm DJ, Jones JW. Vector competence of North American mosquitoes (Diptera: Culicidae) for West Nile virus. *J Med Entomol* 2001;38:130-34.
6. Sardellis MR, Turell MJ, Dohm DJ, O'Guinn ML. Vector competence of selected North

American *Culex* and *Coquillettidia* mosquitoes for West Nile virus. *Emerg Infect Dis* 2001;7:1018-22.
7. Turell MJ, Dohm DJ, Sardelis MR, O'Guinn ML, Andreadis TG, Blow JA. An update on the potential of North American mosquitoes (Diptera: Culicidae) to transmit West Nile virus. *J Med Entomol* 2005;42:57-62.
8. Darsie RF, Ward RA. *Identification and Geographical Distribution of the Mosquitoes of North America, North of Mexico.* Gainesville FL: University Press of Florida, 2005;225-340.
9. Johnson GD, Eidson M, Schmit K, Ellis A, Kulldorff M. Geographic prediction of human onset of West Nile virus using dead crow clusters: An evaluation of year 2002 data in New York State. *Am J Epidemiol* 2005;163:171-80.
10. Gibbs SEJ, Wimberly MC, Madden M, Masour J, Yabsley MJ, Stallknecht DE. Factors affecting the geographic distribution of West Nile virus in Georgia, USA: 2002-2004. *Vector-Borne and Zoonotic Diseases* 2006;6:73-82.
11. Alberta Municipal Affairs. *Municipal Government Act.* Available online at: http://www.qp.gov.ab.ca/documents/Acts/M26.cfm?frm_isbn=0779737822&type=htm (Accessed August 3, 2005).
12. Mostashari F, Bunning ML, Kitsutani PT, Singer DA, Nash D, Cooper MJ, et al. Epidemic West Nile encephalitis, New York, 1999: Results of a household-based seroepidemiological survey. *Lancet* 2001;358:261-64.
13. Reisen WK, Lothrop HD. Population ecology and dispersal of Culex Tarsalis (Diptera, Culcidae) in the Coachella Valley of California. *J Med Entomol* 1995;32:490-502.
14. Alberta Sustainable Resource Development, Alberta Environment, Alberta Community Development and Agriculture and Agri-Food Canada. Natural Regions and Subregions of Alberta, 2005. Available online at: http://www.cd.gov.ab.ca/preserving/parks/anhic/download_data.asp (Accessed July 27, 2005).
15. Bailey TC, Gatrell AC. *Interactive Spatial Data Analysis.* Essex, England: Longman Scientific & Technical, 2005.
16. Bartlett MS. The spectral analysis of two-dimensional point processes. *Biometrika* 1964;51;299-311.
17. Littell RC, Milliken GA, Stroup WW, Wolfinger RD. *SAS System for Mixed Models.* Cary, NC: SAS Institute Inc., 1996;303-20.
18. Kleinschmidt I, Sharp BL, Clarke GPY, Curtis B, Fraser C. Use of generalized linear mixed models in the spatial analysis of small-area malaria incidence rates in KwaZulu Natal, South Africa. *Am J Epidemiol* 2001;153:1213-21.
19. Waller LA, Gotway CA. *Applied Spatial Statistics for Public Health Data.* Hoboken, NJ: John Wiley & Sons Inc., 2004;380-409.
20. Alberta Health & Wellness 2004. West Nile virus Infection Rate Study. Available online at: http://www.health.gov.ab.ca/public/wnv_bloodsurvey.pdf (Accessed July 25, 2005).
21. Forand SP, Talbot TO, Druschel C, Cross PK. Data quality and the spatial analysis of disease rates: Congenital malformations in New York State. *Health and Place* 2002;8:191-99.
22. Murray CJL, Lopez AD. Estimating causes of death: New methods and global and regional applications for 1990. In: Murray CJL, Lopez AD (Eds.), *The Global Burden of Disease 1996.* Geneva: World Health Organization, 1996; Chapter 3.
23. Yiannakoulias N, Svenson LW, Schopflocher DP. Commentary: Diagnostic uncertainty and medical geography: What are we mapping? *The Canadian Geographer* 2005;49:291-300.
24. Horsfall WR. *Mosquitoes: Their Bionomics and Relation to Disease.* New York, NY: The Ronald Press Company, 1955;547-98.

Received: August 9, 2005
Accepted: May 25, 2006

RÉSUMÉ

Contexte : Dans cet article, nous appliquons une méthode de modélisation de la variation spatiale du virus du Nil occidental (VNO) chez les humains à l'aide de données d'essais sur les oiseaux, l'environnement et les humains.

Méthode : Nous avons utilisé des données recueillies auprès de 503 municipalités de l'Alberta. Pour atténuer les effets de l'autocorrélation spatiale résiduelle, nous avons fait appel à des modèles linéaires généralisés mixtes (GLMM) pour modéliser l'incidence de l'infection.

Résultats : Il y a eu 275 cas confirmés de VNO en Alberta au cours de l'année civile 2003. Notre modèle spatial montre que le fait de vivre dans la région naturelle des prairies et l'envergure des essais sur les humains sont d'importants prédicteurs positifs du VNO; le fait de vivre en milieu urbain est quant à lui un important prédicteur négatif.

Conclusion : Les données sur les oiseaux infectés jouent peu dans notre modèle. L'envergure variable des essais à différents endroits de la province pourrait être un facteur confusionnel dans la variabilité de l'incidence du virus du Nil occidental en Alberta. Cependant, les écarts dans les taux d'infection sont aussi associés à des facteurs de risque environnementaux connus. Nos constatations sont compatibles avec les connaissances actuelles sur le VNO en Amérique du Nord.