Modelling Human Risk of West Nile Virus using Surveillance and Environmental Data

Shruti Mallya

A thesis submitted to the
Faculty of Graduate and Postdoctoral Studies
in partial fulfillment of the requirements for the
Master of Science degree in Epidemiology

School of Epidemiology, Public Health and Preventive Medicine
Faculty of Medicine
University of Ottawa

Contents

Abstract

Limited research has been performed in Ontario to ascertain risk factors for West Nile Virus (WNV) and to develop a unified risk prediction strategy. The aim of the current body of work was to use spatio-temporal modelling in conjunction with surveillance and environmental data to determine which pre-WNV season factors could forecast a high risk season and to explore how well mosquito surveillance data could predict human cases in space and time during the WNV season. Generalized linear mixed modelling found that mean minimum monthly temperature variables and annual WNV-positive mosquito pools were most significantly predictive of number of human WNV cases ($p<0.001$). Spatio-temporal cluster analysis found that positive mosquito pool clusters could predict human case clusters up to one month in advance. These results demonstrate the usefulness of mosquito surveillance data as well as publicly available climate data for assessing risk and informing public health practice.

Acknowledgements

I would like to extend my sincerest thanks to everyone who was involved in the conception, implementation and presentation of the work embodied in this thesis whether directly or indirectly. First and foremost, I must express my gratitude to my supervisor and mentor in this journey that is research, Dr. Manisha Kulkarni. Her encouragement of independent exploration of ideas balanced with her careful guidance allowed me to flourish and discover the joy of performing research. I offer my thanks also to Dr. Marie-Helene Roy-Gagnon whose constant willingness to offer a hand with any problems encountered with SAS coding provided much needed peace of mind after struggling for hours and sometimes days with stubborn coding problems. I credit her entirely with my proficiency and newfound enjoyment of SAS coding. I thank Dr. Curtis Russell who was always quick to provide insight into mosquito testing practices in Ontario. I thank Dr. Dan Mckenney and Pia Papadopoul who so enthusiastically processed the climate data I needed for my analysis. I thank Dr. Monica Taaljard who provided guidance in performing generalized linear mixed models in my time of dire need. I thank Dr. Ann Jolly and Dr. Beate Sander who so graciously provided valuable feedback on this work from conception to conclusion. I would like to thank my family for supportively listening to me go on and on about my research. I would also like to thank all my classmates in the M.Sc. Epidemiology program who created a sense of community and provided for comic relief in times of high stress. I couldn't have done it without you!

Chapter 1: Introduction

West Nile Virus (WNV) is a mosquito-borne virus of the family Flaviviridae similar to Japanese encephalitis virus, St. Louis encephalitis virus and Murray valley encephalitis virus. Existing on every inhabited continent,[1] WNV is one of the most widespread mosquito-borne pathogens in the world. The virus is transmitted to humans via the bite of an infected mosquito but lives primarily in an avian reservoir, with various mammals including humans and horses serving as incidental hosts.[2] An infected mosquito may transmit the virus to a bird or mammal, but only birds are capable of infecting a mosquito as the virus does not replicate to sufficient levels within mammals.[3] Competent vectors of WNV transmission to humans are mosquitoes that can feed on mammals as well as birds, allowing for transmission between the two groups. These are known as bridge vectors. In North America, the primary vectors of human WNV are of the genus *Culex*, with differences in main species across geography. In western North America for example, *Culex tarsalis* is the dominant vector species while in the eastern regions, *Culex pipiens* prevails.[4]

West Nile Virus was first identified in Uganda in 1937 and did not appear in North America until 1999 when it caused an outbreak in New York City.[5] Shortly thereafter, it spread across the continent and has since then become endemic across the United States, Mexico and the Southern parts of Canada including Ontario, Quebec and Manitoba.[6] The virus was believed to be disseminated by the migratory patterns of birds, which correlated well with epidemics in newly emergent areas.[7] With climate change, there are concerns of further range expansion of the disease. In Canada, WNV is expected to migrate further north,[8] raising concerns for wildlife and human health.

When initially identified, WNV was not of particular public health interest since infection was generally asymptomatic, and continues to be in tropical regions of the world;[7] but, with the advent and rapid spread of the virus in North America and the appearance of serious symptoms, public health interest in the disease has increased substantially. In North America, it has been observed that infection with the virus is usually asymptomatic, but in about 20% of cases it can cause flu-like symptoms.[9] Among symptomatic cases, the rate of neurological disease is approximately 35%.[10] In addition to concerns regarding population health, WNV may pose a significant burden on public health and healthcare systems. The economic burden of the disease in Canada is as of yet unknown but in the United States, it is estimated that WNV costs the American healthcare system $56 million every year.[11] All things considered, WNV has emerged as a disease of public health interest in North America with many public health agencies undertaking substantial measures to prevent WNV infection. Public awareness campaigns, mosquito control initiatives and surveillance form the core of the public health response.

1.1 Surveillance of WNV in Canada

Bird, horse, mosquito and human surveillance systems are currently in place in Canada to inform as to potential risk of WNV. To address the avian contribution to the disease cycle, passive bird surveillance is undertaken by the Canadian Wildlife Health Cooperative (CWHC).[12] This organization is responsible for dead bird collection, characterization and testing for WNV. However, the passive nature of this program necessitates public awareness and interest as the public is requested to report dead bird sightings. Unfortunately, dead bird submissions have been declining steadily over the years in Canada due to lack of public participation,[13] and potentially to a decrease in dead

birds, making this data of limited value. Avian surveillance may be of greater use in newly emergent areas where the resident bird populations are immunologically naive.[14] In addition to birds, horses can also suffer WNV infection, however they are dead-end hosts and cannot amplify the virus. Horse surveillance is undertaken by the Canadian Food Inspection Agency (CFIA). Veterinary laboratories are required to report to CFIA every confirmed and suspected case of WNV in all domestic animals, primarily horses[14]. With WNV being a mosquito-borne pathogen, mosquito surveillance plays a critical role in understanding risk. Mosquito surveillance is mandated by Ontario and undertaken independently by public health units (PHUs). This surveillance involves the collection, identification and testing of mosquitoes for WNV during the transmission season[15] which usually ranges from May-October. Human surveillance requires reporting of confirmed and probable cases by laboratories to healthcare authorities[16]. All these surveillance methods provide data to inform public health agencies regarding risk in a given season and potential outbreaks. This is of particular importance as human incidence of WNV can vary greatly year to year and at different times throughout the transmission season[16]. As such, it is necessary for public health agencies to be able to predict where and when there is elevated risk for WNV, allowing for disease prevention strategies to be employed optimally.

1.2 Prevention and surveillance of WNV in Ontario

Historically, Ontario has been amongst the provinces with the highest rate of human WNV infection. As such, 'Ontario Regulation 199/03: Control of West Nile Virus' was passed in 2003 by Ontario's Ministry of Health (MOH) pertaining to control efforts for West Nile virus.[17] This legislature requires that PHUs undertake appropriate source reduction, control, surveillance and public education efforts.[17] More details as to

the suggested actions are provided within the MOH's 'West Nile Virus Preparedness and Prevention Plan'.[18] This plan recommends that PHUs adopt an 'Integrated Vector Management' program consisting of surveillance and control activities, public education campaigns and the determination of action thresholds among others.[18] In practice, control activities are limited to source reduction and larviciding. Adulticiding is not entirely beyond the scope of control efforts, however it has yet to be employed in the context of WNV risk by any PHU in Ontario and it is recommended that there should be significant human case counts before adulticiding is considered.[15] Source reduction measures include reducing larval habitat by drainage of standing water, prevention of human contact with mosquitoes by the use of physical barriers and the use of screens on catch basins.

Public education campaigns aim to ensure awareness of source reduction habits, personal protective measures and behaviour pertinent to WNV transmission. Many of these activities are contingent on action thresholds that are unique to each PHU and consist of a set of parameters that are determined prior to the WNV season. Based on the fulfilment of the parameters outlined in an action threshold, it is decided at what point in the season intervention is necessary. These thresholds allow for public health decision-making regarding the magnitude of management strategies as well as where and when they should be implemented for maximum effectiveness. These pre-season action thresholds along with in-season risk assessments based on human and mosquito vulnerability factors[18] allow for a well rounded, and locally optimized, public health response. There are currently no explicit recommendations for setting action thresholds, however. Risk assessment decision-making is suggested to follow a decision tree

developed by the U.S. Centers for Disease Control and Prevention. However, this too does not provide any explicit instruction regarding how risk assessments should be undertaken. The development of unified guidelines pertaining to how individual PHUs should develop action thresholds and risk assessments would likely aid PHUs in optimizing their own WNV preparedness and prevention plans. One way to accomplish this may be to determine early season risk factors to predict upcoming high WNV risk seasons that would form the core of action thresholds as well as in-season risk prediction measures.

1.3 Predicting occurrence and magnitude of human West Nile Virus infection

To date, many studies have been published that aim to predict WNV occurrence, whether several months or several weeks in advance. Many of these models include a component focussed on mapping hotspots within various regions across North America to aid public health agencies in targeting their educational program.[19 - 26] Tachiiri *et al* for example performed a spatial risk assessment in British Columbia (BC). They created a predictive model for WNV risk based on abundance of *C. taraslis* and then applied parameter values including temperature and geographic variables specific to various areas of BC to determine which areas would be at greatest risk for future range expansion of WNV.[19] Rochlin *et al* performed logistic regression modelling to ascertain areas at higher risk of WNV in Suffolk County, New York based on environmental and socioeconomic factors.[20] Both of these studies made sure to investigate the unique risk factors for exposure and infection in their target areas, since these can differ substantially across various regions of North America and in different populations.[22] Other studies have aimed to create human WNV risk prediction models and measures independent of mapping. These models aim to predict either when WNV will begin to appear in humans

[27,28] or the location[29], number [30] or presence of a single case.[31] For example, Kilpatrick *et al*, used mosquito surveillance data to calculate the vector index (estimate of infective mosquitoes found by multiplying the number of mosquitoes per trap-night by the infection prevalence) to predict occurrence of human WNV 1-3 weeks in advance while Patnaik *et al*, employed animal surveillance data to predict where cases would occur. Several prediction models have been published in recent years that have outcomes on various timescales ranging from forecasting a high risk WNV season months in advance [31,32] to predicting WNV cases several weeks in advance.[27-28] These models have used past WNV cases as predictors [23-25, 27,30-33] or for evaluation of the models.[27–29,34] Most of these models make use of a combination of environmental, avian and demographic data to predict risk of infection. These are discussed below.

1.4 Climatic effects on mosquitoes and WNV transmission

Temperature is known to have a substantial effect on the transmission of the mosquito-borne viruses as well as mosquito population dynamics. Mosquito reproduction is contingent on temperature; increased temperatures will generally result in an increased abundance of mosquito populations due to shortened larval development time.[35] Increased temperatures also decrease the time interval between blood meals.[36] Thus in years with increased summer temperatures, there are a greater number of mosquitoes and higher biting frequency, indicating an increased risk of being bitten. Additionally, higher temperatures decrease the incubation time of the virus in mosquitoes (i.e. the extrinsic incubation period) making it transmissible faster.[37-39] Extremely high temperatures, however, may have the opposite effect due to impacts on mosquito survival.[40]

There are various measures of temperature that have been found to be useful for WNV prediction. The use of degree days, for example is prevalent in many public health agencies. This measure makes use of the number of days that the average temperature is 1^{o}C above a reference threshold that reflects when WNV is able to grow and replicate in mosquitoes (this differs based on species and area, for example in Ontario this value is 18.3^{o}C).[15] Accumulated degree days (ADDs) have been used to predict at which point during the season WNV can become transmissible which can be relevant for in-season risk assessments. One study that made use of ADDs to predict WNV transmission dates found that this data could at best predict 15 out of 25 first-detection dates and 15 out of 24 last-transmission dates[41], indicating that though useful, using ADD does not provide a full picture. Other papers have looked at climate earlier in the year to predict a generally high or low activity WNV season, which may be more relevant for establishing action thresholds. Wimberly *et al* found that higher than average winter temperatures were a significant predictor of increased number of human WNV disease cases in the following season[22] while Manore *et al* found that mean minimum January temperatures specifically were significant predictors for WNV. [31]

Although the effect of temperature on WNV activity is fairly established, the pattern of association between precipitation and WNV risk is less clear. The effect of precipitation has been shown to vary considerably across areas with changes in the predominant vector species. For example, drought can be an aggravating factor for WNV transmission since this reduces pools of water, increasing the contact between birds and mosquitoes.[40] Increased rainfall may also augment risk as water is critical for mosquito larval development and as such, this may have a positive impact on adult mosquito

abundance. Conversely, heavy rains may decrease the rate of larval development by diluting the nutrient content of standing water.[40] Epstein *et al* hypothesized in 2000 that risk of WNV was higher in years that had warmer winters, spring droughts, summer rains and summer heat;[42] however various studies since then have shown that these criteria are not universal. Wimberley *et al* found that less than average precipitation in January and July and higher than average in March and August were significantly associated with WNV.[32] Regardless of differences in predictors across space, many studies have found that using climate data can be very effective at predicting WNV occurrence and have used climate variables with this aim.

Land cover is also an important predictor for WNV. A 2014 study by DeGroote *et* al found that various land cover classes including cropland, forests, water and urban land cover were significantly correlated with WNV risk[26], however this varied greatly by region. Other studies have found grassland [22,33,43], urban [33,43] , wetland areas, [20,33] cropland and forest [33] to be significant predictors of WNV. The discrepancies between significant land cover predictors demonstrate the value of performing smaller scale studies to ascertain risk factors specific to different regions.

1.5 Avian associations with human West Nile virus infection

Avian testing data can be useful in forecasting immediate risk of WNV mainly because birds are the primary hosts of the virus. This kind of data has been used in previous models[24] as part of a multivariable prediction system. Liu *et al* for example found that a dead bird sighting in the last 30 days and a WNV-positive bird in the last 30 days were significantly associated with increased risk of human infection. [22] The utility of including dead bird surveillance in the Canadian context is unclear, however, since data

from dead bird surveillance may be unreliable given declining public interest/awareness.[13] However, bird species distribution and abundance may still prove useful as a predictor since host competency differs between species as does susceptibility[44]. Some bird species die almost immediately after contracting the disease, whereas others are able to carry the virus for longer, allowing for greater dissemination. As such, the relative abundance of these species may provide an estimate of the potential for amplification of the virus[45,46] and has been used effectively in the past for predicting annual human WNV cases.[47]

1.6 Demographic risk factors

Few papers have been published on determining demographic and/or socioeconomic risk factors for human WNV in North America. Further work is warranted to assess these risk factors for WNV infection in different parts of North America and in specific communities because, as noted by Liu *et al*, risk factors may differ by population.[22] The authors noted that studies in some areas of the United States found an association between urban environments and human infection, while others found the opposite. Although this may be due to environmental differences, regional differences in behaviour pertaining to mosquito exposure may contribute as well.[22] Rochlin *et al*, also noted the discrepancies between studies attributing risk to income. Some studies found that risk was higher in suburban areas and others found risk higher in low income areas. This was deemed to be due to a mix of differences in vector ecology and human behaviour.[20] Incorporating demographic variables into prediction models might greatly improve them and account for complexities in human-vector interactions.

Few studies have thoroughly studied demographic risk factors for WNV infection. These studies have found that number of senior households (aged 65+), median household income and percent of vacant housing were significant predictors of WNV, [20] as well as population density. [22] However, in order to have accurate information regarding the effect of these demographic factors, it would be necessary to undertake mapping studies at the local level to first find WNV hotspots (i.e. areas with elevated rates of WNV transmission) and then determine what demographic factors make these areas more vulnerable to infection. This would allow for better public health response during the WNV season.

1.7 Understanding the spatial dependence of WNV

The dependence of WNV on climate and vector distribution makes it a disease that is naturally spatially dependent. As such, spatial analysis of WNV forms a vital component of understanding the spread of the disease. Several studies have employed a combination of Geographic Information Systems (GIS)-based mapping and various modelling methods to visualize risk in space.[43,48,49,19,24,50] A commonly used method in GIS-based assessments of WNV risk is Moran's I analysis. Anselin's local Moran's I works with area data and measures spatial autocorrelation. As such, this serves to identify areas with similarly higher or lower incidence compared to the overall. These are identified by a z-score, which indicates a spatial outlier and a Moran's I score, which if close to +1 indicates clustering. [51] Though this does provide a fair spatially explicit estimate of risk in certain areas, mapping risk over different scales of time is difficult using this method. As such, the temporal aspect of clustering may be missed if only Moran's I is used. A frequently employed method of finding spatio-temporal clusters is Kulldorf's spatial scan statistic, which can be employed using the SaTScan software.[52]

SaTScan works using data on point incidence of a disease; it employs a scanning window of variable size that moves across geographic units with the X and Y axes corresponding to space and the Z axis corresponding to time. Whenever a case is detected, the software performs a likelihood ratio test comparing expected and observed counts inside the window and outputs a p-value based on 999 Monte Carlo replications[52] which is the default. The Monte Carlo simulations simulate potential distributions of cases across space and time. These are taken to be the expected number of cases inside the scanning window and represent the null hypothesis (that no clustering is present). The Besag-Newell test is also very popular for cluster analysis. This test determines whether disease in small units (counties for example) clusters in space based on a pre-specified threshold of cases, k, required in an area to be considered a cluster. Starting from the centroid of an area, the scanning circle will expand until k cases are located within it. The p value for the cluster is calculated based on the probability of the observed number of units needed to reach k is greater than or equal to the theoretical value. Though this test readily detects spatial clusters, it is unable to detect spatio-temporal clusters and it has been noted that this test tends to detect larger clusters resulting in less spatial precision. As such, past studies have recommended the use of SaTScan.[53] Anselin has suggested the use of SaTScan and Anselin's local Moran's I in conjunction with one another to obtain the most accurate results. Sugumaran *et al* employed both these methods for finding clusters of human WNV disease in the United States with the addition of a binary scoring method as suggested by Chen *et al.*[54] They found that though the methods were complimentary, they did not always agree.

Knowing where and when clusters of human cases tend to occur is a great asset for public health practice. This information, in conjunction with knowledge of predictors, can help to plan public health interventions in advance of human disease. It is well known that mosquito pools tend to test positive for WNV weeks in advance of human case reports. [15] As such it might prove useful to evaluate the predictive ability of WNV positive mosquito clusters using routinely collected mosquito surveillance data. Few studies thus far have aimed to find clusters of WNV positive mosquitoes and none have assessed how these clusters may correspond with human case clusters. The predictive value of mosquito surveillance is known, as mosquitoes tend to test positive earlier in the year in high incidence years,[15] however the use of this data could still be optimized to allow for a more complete picture of WNV in a given area. Real-time cluster analysis of mosquito positive pools, if predictive, could form the cornerstone of in-season public health decision-making with regards to intervention strategies.

1.8 Importance of studying WNV in Ontario

Though various articles have been published studying WNV in North America, it is important to perform studies that are specific to a given region. Risk factors that have a significant effect in some areas may not have the same effect in others due to differences in primary host, vector species and human behaviour across the continent. By knowing which areas are more prone to WNV transmission and understanding which factors are to blame, we can focus our public health efforts on education and prevention. This will hopefully help to lower the burden of disease in terms of both morbidity/mortality and the economic ramifications on Ontario's healthcare system. The following studies aim to provide methods by which public health units may estimate risk in advance of and during the WNV season.

1.9 References

1. WHO | West Nile virus. *WHO* (2014). Available at: http://www.who.int/mediacentre/factsheets/fs354/en/. (Accessed: 27th June 2016)

2. Rossi, S. L., Ross, T. M. & Evans, J. D. West Nile virus. *Clin. Lab. Med.* **30,** 47–65 (2010).

3. Colpitts, T. M., Conway, M. J., Montgomery, R. R. & Fikrig, E. West Nile Virus: biology, transmission, and human infection. *Clin. Microbiol. Rev.* **25,** 635–48 (2012).

4. Conly, J. & Johnston, B. Why the West in West Nile virus infections? *Can. J. Infect. Dis. Med. Microbiol. = J. Can. des Mal. Infect. la Microbiol. médicale* **18,** 285–8 (2007).

5. Nash, D. *et al.* The Outbreak of West Nile Virus Infection in the New York City Area in 1999. *N. Engl. J. Med.* **344,** 1807–1814 (2001).

6. Artsob, H. *et al.* West Nile Virus in the New World: trends in the spread and proliferation of West Nile Virus in the Western Hemisphere. *Zoonoses Public Health* **56,** 357–69 (2009).

7. Gubler, D. J. The continuing spread of West Nile virus in the western hemisphere. *Clin. Infect. Dis.* **45,** 1039–46 (2007).

8. West Nile virus in the context of climate change. *Paediatr. Child Health* **13,** 399–404 (2008).

9. Nash, D. *et al.* The Outbreak of West Nile Virus Infection in the New York City Area in 1999. *N. Engl. J. Med.* **344,** 1807–1814 (2001).

10. West Nile Virus Activity—United States, 2006. *JAMA* **298,** 619 (2007).

11. Barrett, A. D. T. Economic burden of West Nile virus in the United States. *Am. J. Trop. Med. Hyg.* **90,** 389–90 (2014).

12. PHAC. Surveillance of West Nile virus. (2015).

13. Zimmer, P. West Nile Virus National Report on Dead Bird Surveillance, Canadian Cooperative Wildlife Health Centre. *Canadian Cooperative Wildlife Health Centre: Newsletters & Publications* (2005).

14. CFIA. Surveillance - Animals - Canadian Food Inspection Agency. (2012).

15. Ontario Agency for Health Protection and Promotion (Public Health Ontario). Guide for public health units: Considerations for adult mosquito control. *Queen's Printer for Ontario* (2013). Available at: https://www.publichealthontario.ca/en/eRepository/Guide_Considerations_Mosquito_Control_2013.pdf. (Accessed: 29th February 2016)

16. PHAC. Surveillance of West Nile virus. (2015).

17. Ontario Ministry of Health. Ontario Regulation 199/03: Control of West Nile Virus. (2009). Available at: https://www.ontario.ca/laws/regulation/030199. (Accessed: 31st May 2016)

18. Division, P. H. *West Nile Virus Preparedness and Prevention Plan*. *Public Health Division* (2008).

19. Tachiiri, K., Klinkenberg, B., Mak, S. & Kazmi, J. Predicting outbreaks: a spatial risk assessment of West Nile virus in British Columbia. *Int. J. Health Geogr.* **5,** 21 (2006).

20. Degroote, J. P., Sugumaran, R. & Ecker, M. Landscape , demographic and climatic associations with human West Nile virus occurrence regionally in 2012 in the United States of America. **9,** 153–168 (2014).

21. Rochlin, I., Turbow, D., Gomez, F., Ninivaggi, D. V. & Campbell, S. R. Predictive mapping of human risk for west nile virus (WNV) based on environmental and socioeconomic factors. *PLoS One* **6,** (2011).

22. Liu, A. *et al.* Risk factors for human infection with West Nile Virus in Connecticut: a multi-year analysis. *Int. J. Health Geogr.* **8,** 67 (2009).

23. Kilpatrick, a. M. & Pape, W. J. Predicting human west nile virus infections with mosquito surveillance data. *Am. J. Epidemiol.* **178,** 829–835 (2013).

24. Bolling, B. G., Barker, C. M., Moore, C. G., Pape, W. J. & Eisen, L. Seasonal patterns for entomological measures of risk for exposure to Culex vectors and West Nile virus in relation to human disease cases in northeastern Colorado. *J. Med. Entomol.* **46,** 1519–1531 (2009).

25. Kwan, J. L. *et al.* Comparison of Enzootic Risk Measures for Predicting West Nile. *Emerg. Infect. Dis.* **18,** 2004–2010 (2012).

26. Patnaik, J. L., Juliusson, L. & Vogt, R. L. Environmental Predictors of Human West Nile Infections, Colorado. *Emerg. Infect. Dis.* **13,** 1788–1790 (2007).

27. Manore, C. a *et al.* Towards an early warning system for forecasting human West Nile virus incidence. *PLoS Curr. Outbreaks* **EARLY VIEW,** 1–18 (2014).

28. Wimberly, M. C., Lamsal, A., Giacomo, P. & Chuang, T.-W. Regional Variation of Climatic Influences on West Nile Virus Outbreaks in the United States. *Am. J. Trop. Med. Hyg.* **91,** 677–684 (2014).

29. Chuang, T. W. & Wimberly, M. C. Remote Sensing of Climatic Anomalies and West Nile Virus Incidence in the Northern Great Plains of the United States. *PLoS One* **7,** 1–10 (2012).

30. Winters, A. M. *et al.* Combining Mosquito Vector and Human Disease Data for

Improved Assessment of Spatial West Nile Virus Disease Risk. *Am. J. Trop. Med. Hyg.* **78,** 654–665 (2008).

31. Chuang, T.-W., Hockett, C. W., Kightlinger, L. & Wimberly, M. C. Landscape-Level Spatial Patterns of West Nile Virus Risk in the Northern Great Plains. *Am. J. Trop. Med. Hyg.* **86,** 724–731 (2012).

32. Tackett, J., Charnigo, R. & Caldwell, G. Relating West Nile Virus Case Fatality Rates to Demographic and Surveillance Variables. *Public Health Rep.* **121,** 666–673 (2006).

33. Rowley, W. a. & Graham, C. L. The effect of temperature and relative humidity on the flight performance of female Aedes aegypti. *J. Insect Physiol.* **14,** 1251–1257 (1968).

34. Paaijmans, K. P. *et al.* Temperature-Dependent Pre-Bloodmeal Period and Temperature-Driven Asynchrony between Parasite Development and Mosquito Biting Rate Reduce Malaria Transmission Intensity. *PLoS One* **8,** e55777 (2013).

35. Reisen, W. K. *et al.* Effects of temperature on the transmission of west nile virus by Culex tarsalis (Diptera: Culicidae). *J. Med. Entomol.* **43,** 309–17 (2006).

36. Dohm, D. J. *et al.* Effect of environmental temperature on the ability of Culex pipiens (Diptera: Culicidae) to transmit West Nile virus. *J. Med. Entomol.* **39,** 221–5 (2002).

37. Paz, S. Climate change impacts on West Nile virus transmission in a global context. *Philos. Trans. Roal Soc. Biol. Sci.* **370,** (2015).

38. Chen, S. *et al.* Estimating West Nile virus transmission period in Pennsylvania using an optimized degree-day model. *Vector Borne Zoonotic Dis.* **13,** 489–97 (2013).

39. Epstein, P. R. & Defilippo, C. West Nile Virus and drought. *Glob. Chang. Hum. Heal.* **2,** 105–107 (2001).

40. Yiannakoulias, N. W., Schopflocher, D. P. & Svenson, L. W. Modelling geographic variations in West Nile virus. *Can. J. public Heal. = Rev. Can. santé publique* **97,** 374–8

41. Cooke, W. H., Grala, K. & Wallis, R. C. Avian GIS models signal human risk for West Nile virus in Mississippi. *Int. J. Health Geogr.* **5,** 36 (2006).

42. Hamer, G. L. *et al.* Host selection by Culex pipiens mosquitoes and west nile virus amplification. *Am. J. Trop. Med. Hyg.* **80,** 268–278 (2009).

43. Loss, S. R. *et al.* Avian host community structure and prevalence of West Nile virus in Chicago, Illinois. *Oecologia* **159,** 415–424 (2009).

44. Fenech, A. & Chiotti, Q. in *Integrated mapping assessment* (Meteorological

Service of Canada, Environment Canada, 2005).

45. McKenzie, V. J. & Goulet, N. E. Bird community composition linked to human West Nile virus cases along the Colorado front range. *Ecohealth* **7,** 439–47 (2010).

46. Brownstein, J. S. *et al.* Spatial analysis of West Nile virus: rapid risk assessment of an introduced vector-borne zoonosis. *Vector Borne Zoonotic Dis.* **2,** 157–64 (2002).

47. Epp, T. Y., Waldner, C. L. & Berke, O. Predicting geographical human risk of West Nile virus--Saskatchewan, 2003 and 2007. *Can. J. public Heal. /Revue Can. santé publique* **100,** 344–8

48. Jones, R. C. *et al.* Use of the Vector Index and Geographic Information System to Prospectively Inform West Nile Virus Interventions. *J. Am. Mosq. Control Assoc.* **27,** 315–319 (2011).

49. Anselin, L. Local Indicators of Spatial Association-LISA. *Geogr. Anal.* **27,** 93–115 (1995).

50. Kulldorff, M. A spatial scan statistic. *Commun. Stat. - Theory Methods* **26,** 1481–1496 (1997).

51. Costa, M. A. A fair comparison between the spatial scan and the Besag–Newell Disease clustering tests. *Environ. Ecol. Stat.* **12,**

52. Chen, J., Roth, R. E., Naito, A. T., Lengerich, E. J. & Maceachren, A. M. Geovisual analytics to enhance spatial scan statistic interpretation: an analysis of U.S. cervical cancer mortality. *Int. J. Health Geogr.* **7,** 57 (2008).

Chapter 2 (Manuscript 1): Use of Generalized Linear Mixed Modeling to Ascertain Predictors for Human West Nile Virus Infection in Ontario

2.1 Preface to Manuscript 1

The work embodied in this paper has gone through ethical approval by the Ottawa Health Sciences network research ethics board for the use of secondary human data pertaining to cases of WNV. Other sources of data used in this analysis come from sources that are open to the public including Statistics Canada and Natural Resources Canada (NRCAN). Climate data provided by NRCAN was generated for the study locations by Ms. Pia Papadopoul. Descriptive statistics of human and mosquito surveillance data are provided in appendix 1. Full results of modelling are provided in appendix 2. The conceptualization of the project was jointly performed between Dr. Manisha Kulkarni and Shruti Mallya. Technical advice on preparing the datasets for analysis was provided by Dr. Marie-Helene Roy-Gagnon. Expertise regarding the development of generalized linear mixed modelling was provided by Dr. Monica Taljaard. All data preparation and analysis was performed by Shruti Mallya.

2.2 Manuscript 1

Use of Generalized Linear Mixed Modeling to Ascertain Predictors for Human West Nile Virus Infection in Ontario

Shruti Mallya[1], Beate Sander[2,3], Marie-Hélène Roy-Gagnon[1], Monica Taljaard[1], Ann Jolly[1], Manisha Kulkarni[1]

[1]School of Epidemiology, Public Health and Preventive Medicine, University of Ottawa

[2]Public Health Ontario

[3]University of Toronto

Abstract

In recent years, West Nile virus (WNV) has become a disease of public health concern in North America with many agencies undertaking preventive measures. The current study aims to combine environmental, census and historical mosquito and human surveillance data for Ontario to create a WNV case prediction model to identify risk factors at the public health unit (PHU) level and provide insight for public health intervention. To this aim, a generalized linear mixed model with a Poisson distribution was developed using backwards selection to predict yearly human WNV case counts based on monthly historical climate, landcover, census populations, yearly historical human cases and yearly mosquito WNV testing results from 2002-2013. It was found that only climate variables were significantly predictive of human WNV with the most significant variables being February, March and April mean minimum temperature ($P<0.001$) and annual positive mosquito pools ($p<0.001$). These results demonstrate the usefulness of environmental and surveillance data for predicting human WNV and informing public health decision-making.

Introduction

West Nile virus (WNV) is an emerging infectious disease that was initially characterized in Uganda in 1937 and made its first appearance in North America in 1999 in New York.[1] The lifecycle of this flavivirus is sustained by the feeding of mosquitoes on birds, the main reservoir for the disease. Although they are unable to amplify the virus, mammals can also acquire the infection.[2] Humans are typically asymptomatic but

nevertheless display febrile illness in 20%[1] of cases and of these, 35% will suffer from neuroinvasive disease. [3] In Ontario, high incidence seasons of human WNV infection occur very sporadically. The highest incidence year to date occurred in 2002, following the initial introduction of the virus to the province. Case counts decreased dramatically over the next few years until 2012, when there was a sudden increase in number of cases[4] . The erratic nature of the disease poses a challenge for public health agencies in planning intervention strategies in advance of high incidence WNV seasons. However, the use of several known environmental and demographic predictors may help in this endeavour.

Among environmental predictors, temperature is known to have a substantial effect on WNV incidence given its role in mosquito population dynamics and feeding behaviour as well as viral replication and transmissibility. Higher temperatures often result in increased mosquito abundance, biting rate, viral replication and rate of transmission. [5- 6,7] Land cover is also expected to be an important predictor given that mosquito distribution is dictated by habitat. Past studies have found that warmer winter temperatures[8,9] as well as various land cover classes including grassland, wetland and urban cover are significant predictors of WNV activity.[10,11,12] However, the importance of these predictors can vary greatly across large geographic areas due to differences in primary vector species and ecology. In Western Canada for example, *Culex tarsalis* is the predominant vector species while in eastern Canada, *Culex pipiens/restuans* dominates.[13] These two species have different lifecycles and habitat preferences, which can cause differences between environmental predictors of WNV in these areas. The same may hold true for demographic risk factors since human behaviour, hence interaction with mosquitoes, can change greatly across space.[14] These changes in behaviour can be

challenging to measure, however. Variance in population structure between areas may also impact the incidence of WNV. Past studies for example have found that the male/female ratio of a region[15]as well as the number of senior households in a given area[10] are predictive of WNV incidence. This spatial dependence in predictor significance necessitates the ascertainment of geographically specific risk factors. In Ontario, limited studies have been performed that investigate local predictors of high WNV incidence seasons. As such, the goals of this study were to determine environmental and demographic risk factors for human WNV infection specific to Ontario, based on the development of a model to predict the annual number of human WNV cases. Anticipating the level of risk for an upcoming WNV season early in the year might assist public health agencies in tailoring their surveillance and response efforts.

Past studies have made use of various modeling methods such as logistic regression[10] and partial least squares regression[16] among others for predicting WNV. Here, we use a multilevel generalized linear mixed model (GLMM) with environmental and demographic data to predict WNV occurrence in southern Ontario.

Methods

Data Sources

Epidemiological data

Confirmed and probable human WNV cases in Ontario from 2002-2014 were obtained from the Integrated Public Health Information System (iPHIS) of Public Health Ontario (PHO), which is the system used for documenting cases of reportable communicable disease. Cases are reported to the medical officer of health by labs and physicians and this data is updated in iPHIS on a weekly basis. Cases were aggregated by public health unit (PHU) on a yearly basis. Only data for the southern portion of Ontario (Table 1, Figure 1) was included since northern PHUs tend to have very few cases of WNV possibly due to low climatic and environmental suitability for the virus and vector.

Figure 1: Map of Ontario Public Health units and the area under study.

Table 1: Public Health Unit names with associated abbreviations.

Abbreviation	Full Public Health Unit Name
BRN	Brant County Health Unit
CHK	Chatham-Kent Health Unit
DUR	Durham Regional Health Unit
ELG	Elgin-St. Thomas Health Unit
EOH	The Eastern Ontario Health Unit
GBO	Grey Bruce Health Unit
HAL	Halton Regional Health Unit
HAM	City of Hamilton Health Unit
HDN	Halimand-Norfolk Health Unit
HKP	Haliburton, Kawartha, Pine Ridge District Health Unit
HPE	Hastings and Prince Edward Counties Health Unit
HUR	Huron County Health Unit
KFL	Kingston, Frontenac and Lennox and Addington Health Unit
LAM	Lambton Health Unit
LGL	Leeds, Grenville and Lanark District Health Unit
MSL	Middlesex-London Health Unit
NIA	Niagara Regional Area Health Unit
OTT	City of Ottawa Health Unit
OXF	Oxford County Health Unit
PDH	Perth District Health Unit
PEE	Peel Regional Health Unit
PTC	Peterborough County-City Health Unit
SMD	Simcoe Muskoka District Health Unit
TOR	City of Toronto Health Unit
WAT	Waterloo Health Unit
WDG	Wellington-Dufferin-Guelph Health Unit
WEC	Windsor-Essex County Health Unit
YRK	York Regional Health Unit

Population data

The 2011 census population for each PHU was obtained from the Statistics Canada Health Profile, December 2013.[17] Variables were created for the population structure of each PHU: percentage of population in each age category (corresponding to

CDC age groups), percentage male population and population density.[15] The population structure was assumed to remain the same throughout the years of study.

Climate data

Estimates for mean monthly minimum and maximum temperature and mean monthly precipitation were obtained from Natural Resources Canada for each PHU centroid (geometric centre) for the years 2001-2013.

Environmental data

A 27 class raster file for land cover (year 2000) was obtained from the Ontario Ministry of Natural Resources.[18] The percentage of land falling in each land cover class was calculated for each PHU using the 'Tabulate area' tool in ArcMap 10.2 (ESRI, Redlands, USA). This tool tabulates the number of pixels in an area that have a certain value - in this case each land cover class was assigned a value and the number of pixels having that value were tabulated - the percentage of land in each PHU falling under each land cover class was then calculated. Classes were then aggregated in SAS 9.4 © 2012 (SAS Institute Inc., Cary, NC, USA) based on spearman correlation values above 0.7 (Table 2) in order to reduce collinearity between variables. This is because many of the landcover classes were very similar in nature. Minute differences between them were not expected to have a differential impact on mosquito habitat.

Table 2: Land cover class aggregations performed based on high correlation values

Aggregate variable	**Original variables**
Swamp	Coniferous swamp, deciduous swamp
Coniferous	Mainly coniferous, dense coniferous forest, sparse coniferous forest
Cuts and Burns	Old cuts and burs, recent burns, recent cutovers
Bog and fen	Open bog, treed bog, treed fen

Deciduous	Sparse deciduous forest, mainly deciduous forest

Mosquito surveillance data

WNV testing and species identification data were obtained from the PHO Vector Surveillance dataset maintained by Public Health Ontario. This dataset documents the frequency, location and results of mosquito testing for WNV. During this process, mosquito traps are placed at pre-determined locations. Mosquitoes are then collected, pooled together and tested for WNV. Mosquito pools that test positive for WNV are referred to as WNV-positive mosquito pools. Mosquito testing data was aggregated at the level of PHU and year. The annual percentage of WNV-positive mosquito pools for each PHU was calculated by totalling the number of WNV-positive mosquito pools collected in a year divided by the total number of pools collected for that year and PHU.

Data Analysis

Spatial Autocorrelation

Spatial autocorrelation is important to assess because it results in dependence between observations, violating the assumption of independence embedded in most statistical analyses. If detected, however, steps can be taken to adjust for spatial autocorrelation. With this goal in mind, a generalized linear mixed model with Poisson distribution was developed with all potential variables (Appendix 2) using proc Glimmix in SAS 9.4 © 2012 (SAS Institute Inc., Cary, NC, USA) following a methodology similar to Yiannakoulias *et al*.[19] This modelling method has various advantages over others used in similar studies. Firstly, spatial autocorrelation is easily adjusted for by modelling the spatial distribution of variance as part of a random effect. This is necessary since certain

areas are prone to WNV based simply on their geography and this association can confound the relationship between WNV and other variables. Another important advantage is that the effect of repeat measures can be taken into account. For studies that span multiple years and make use of aggregate data, observations are not independent and treating them as such can invalidate results. Neither of these important factors can be adjusted for appropriately in other model types suited to the type of data used in this study other than fitting separate models for separate years or locations. Some may seek to simply include location as a variable in the model but this ignores underlying spatial associations between locations. For example, counties that are beside each other may naturally have similar characteristics that are not already adjusted for in the model. Conditional autoregressive models work well with spatial data since they account for neighbourhood effects, yet they do not easily adjust for repeat measurements and as such, multiple models would be required for different years. GLMMs also allow for continuous outcomes, whereas logistic regression only allows for a dichotomous outcome. Given these advantages, a GLMM was chosen for this study. The residuals from this model were calculated and a variogram was plotted. A variogram measures the difference of variance in values between two locations, in this case, the pairwise difference in variance between PHUs. By plotting this against distance, it is possible to observe the spatial dependence of the residuals. If there is spatial autocorrelation, a pattern of increasing difference in variance will be visible between locations that are increasingly far apart. Distance between locations is called the lag. The lags of all the pairs can be grouped into a number of classes which each correspond to a range of distance. In order to adjust for this spatial autocorrelation, we can approximate the shape of the curve using an

appropriate distribution and use this to model the G-side covariance matrix, which is the covariance matrix of the random effects, in this case the public health units. This was done by comparing exponential, spherical and Gaussian distributions to the curve of the variogram plotted against the average distance for lag class.

Model development

To adjust for spatial autocorrelation in the final model, PHU was included as a random effect with a spherical spatial covariance structure. A random slope for time with PHU as the subject was also included with the 'residual' statement to model PHUs as a repeated subject based on year. Since the data were aggregated at the level of PHU and observations corresponded to human WNV incidence in each PHU for a given year, it was necessary to treat these as multiple samplings of the same subject, and thus dependent. Since the list of potentially important variables was too large for the size of the dataset, multiple smaller models were first built using a stepwise selection procedure with inclusion and exclusion p-values of 0.1. Separate models were fit for the land cover variables, mean minimum monthly temperature, mean maximum monthly temperature, mean monthly precipitation and age and sex distributions for each PHU. Significant variables from each of these models were then added in a stepwise fashion with inclusion and exclusion p-values of 0.1 to develop the final model. Studentized and conditional studentized residuals were examined as part of the model diagnostics.

Results

Testing and adjusting for spatial autocorrelation

Figure 2 shows a plot of the variogram against lag distance. The plot indicates increasing variance between PHU residuals with increasing distance, typical of spatial autocorrelation. There is a stark increase from 3.3 lag distances to 4, after which point the variogram levels off. Based on visual inspection, it appears as though a spherical distribution is closest to the actual data and this was further used to model the G-side covariance matrix of the model.

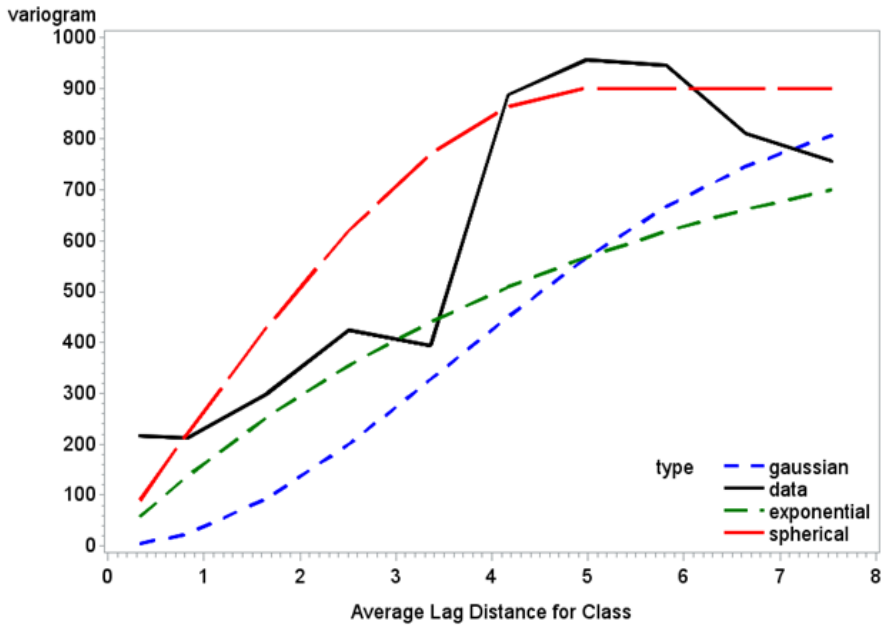

Figure 2: Variance plotted by lag distance. The 'data' line corresponds to the variogram calculated from the residuals of the full model. Gaussian, exponential and spherical lines were generated using a range value of 8 (maximum lag distance) and a scale value of 1000 (maximum variogram value).

Significant predictors of WNV

Table 3: Significant predictors of human WNV infection in southern Ontario Public Health Units, 2002-2013, based on Poisson modelling.

Variable category	Month	Parameter estimate	Standard Error	P-value
Mean monthly minimum temperature	January	0.047	0.024	0.0496
	February	0.457	0.053	<.0001
	March	-0.265	0.038	<.0001
	April	-0.662	0.105	<.0001
	July	0.192	0.072	0.0085
	August	0.230	0.107	0.0323
Mean monthly precipitation	February	-0.0154	0.004	0.0002
Mean monthly maximum temperature	April	-0.148	0.072	0.0412
	August	0.2266	0.1052	0.0321
Annual percent positive mosquito pools		0.0706	0.017	<.0001

Only climatic variables and percent annual WNV-positive mosquito pools were significantly predictive of human WNV cases (Table 4). Scatter plots indicate that with increases in mean minimum January, February, March, April, July and temperatures, there are increases in the number of WNV cases. Mid-range April and higher August maximum temperatures showed an increase in WNV case counts. Lower February and mean precipitation showed increased numbers of WNV case counts (Figure 3). Plots of residuals showed that there was a fair amount of overdispersion indicated by a chi-

square/ degrees of freedom value of 1.58, particularly for the Windsor-Essex region (appendix 2). Models with values close to 1 are considered to display good fit.

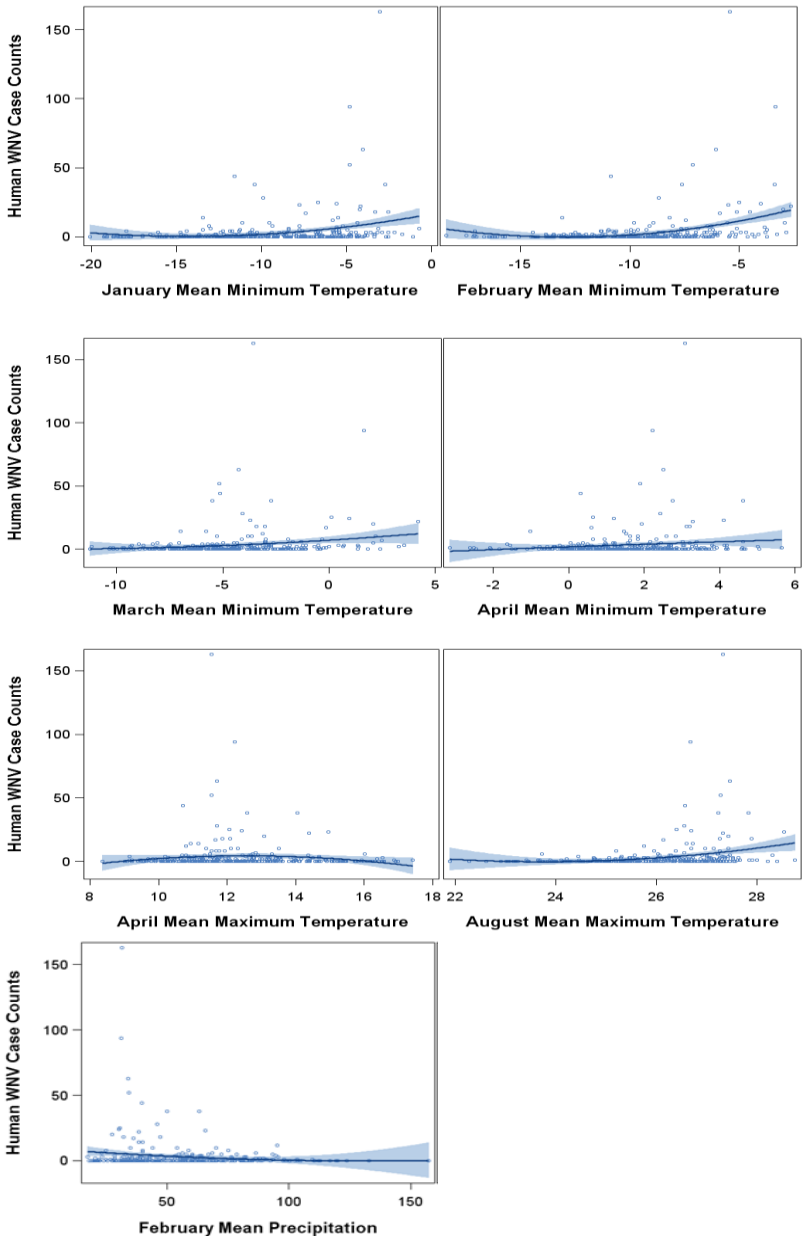

Figure 3: Number of annual human WNV cases in a public health unit (PHU) versus climate predictor values. The regression line is shown with the standard errors (in blue). Data points are indicated by hollow blue circles.

Discussion and Conclusions

To address the gap in knowledge regarding predictors of human WNV infection in Ontario, we applied a multilevel modeling approach that incorporated key environmental and population factors. The model showed that climate variables, particularly February, March and April minimum temperature along with February precipitation, and annual percentage of WNV-positive mosquito pools in a given PHU, were significantly predictive of human WNV cases across PHUs of the southern portion of Ontario. We found the effect of monthly temperature to be similar to previous studies, which also found winter temperatures to be significantly predictive of WNV transmission in an upcoming season.[16] Wimberly *et al,* for example, found in 2014 that winter temperature variables had the greatest influence on rate of West Nile human infection. December and January temperatures, in particular, were most significant. Similarly, Manore *et al*, found mean minimum January temperature to be a highly significant predictor of human WNV. [9] While these studies from the United States found January temperature to be most significant, our analysis found February to be most significant. This may be due in part to climatic differences between the study areas. Both the Wimberly and Manore papers looked at WNV across the United States whereas our study concentrated on the southern portion of Ontario. Temperatures during the winter months have a considerable impact on the WNV's ability to survive into the spring. In colder years, effective overwintering of mosquitoes is lessened.[20] In Ontario, the coldest days of the year tend to occur in February. Hence, it is possible that even if January temperatures are warmer than average, cold weather in February will adversely affect the virus' ability to overwinter successfully. March and April mean minimum temperatures may be similarly important for overwintering, but it is not clear why January mean minimum

temperature was less significantly predictive of human WNV than these months. A study by Winters *et al* in the American Journal of Tropical Medicine and Hygiene found that February, March and April mean temperatures have a larger effect on *Culex tarsalis* abundance than January mean temperatures. [21] As such, it is possible that our findings regarding mean minimum temperature may also reflect factors that affect mosquito abundance in addition to virus overwintering, however the findings of Winters *et al* might not necessarily be applicable for *Culex pipiens,* the main vector species in Ontario.

Maximum temperatures also had a significant effect on WNV activity. It was found that moderate April mean maximum temperatures and higher August mean maximum temperatures were significantly related to human WNV cases. The significance of the April mean maximum temperature may be related to virus amplification in the avian host, which is believed to occur in the early spring.[8] Why moderate temperatures are associated with human WNV infection, however, is unclear. Colder temperatures may be unfavourable since these result in later migration of birds. Spring temperatures that are too warm might result in faster melting of snow which could dilute nutrients in standing water, impeding larval proliferation.[22]

In addition to temperature effects, precipitation was found to contribute significantly to increased human WNV risk, however more subtly. Finally, we found that the annual percentage of WNV-positive mosquito pools was significantly predictive of human cases, confirming our expectation that higher mosquito infection rates should result in more transmission events. This has been previously remarked by Brownstein *et al,* who noted that mosquito surveillance data is the most sensitive marker for human risk,[23] and suggested that this data should be included in any surveillance system. They

found that positive mosquito pools could account for 38% of human risk.[23] The association between WNV-positive mosquito pools and human WNV case counts has also been remarked by Rochlin *et al*, who found a strong association between human risk and proximity to a single WNV-positive mosquito pool[10] and Liu *et al* who found that presence of a WNV-positive pool in the last 30 days was significantly predictive of human infection risk. The results of the current study indicate that a 1% increase in annual positive pools will result in a 0.06772 increase in the log of the incidence rate ratio when the other variables are held constant. However, these results should be interpreted with caution since the model suffers from overdispersion.

This study identifies several predictors of human WNV infection in southern Ontario that can be of public health use, however it has several limitations. The proposed model was successful in identifying key predictors of human WNV cases; but the model was overdispersed even after adjustments for spatial autocorrelation and repeated measures. Thus, the parameter estimates may not be reliable, necessitating the use of scatter plots to confirm the effects of the individual variables. This also means that the model may not be very good at predicting the number of cases in a season. The model was particularly overdispersed for the Windsor-Essex county PHU at the southern-most point of Ontario, indicating that fit was especially poor for this region. Overdispersion may have various causes including an imperfect fit of the covariance matrix (Figure 2) chosen to model the spatial dependence of observations, and potentially the absence of an important variable. In this model, avian data was not included due to paucity of data. Since birds are the main reservoir of WNV, bird dynamics may substantially influence seasonal risk, and previous studies have found links between bird community

composition and WNV incidence.[24] Future studies may seek to include this type of data to produce a more comprehensive model. Finally, population characteristics were assumed to stay constant across time, however this may not be the case. This assumption may have masked potential associations within the data.

Given the aforementioned limitations, it is unlikely that the presented model would be exploitable for predictive purposes. However, the variables identified as significant in the model may still be applicable for public health planning, particularly the early and previous year climate variables. Climate data similar to that used in this study are often routinely collected at the municipal level. By monitoring the monthly average temperature and precipitation in a given area and comparing to normal or historical values, it may be possible to estimate risk for an upcoming WNV season at the PHU level. The pre-WNV season climate variables identified in this study could serve as a checklist for PHUs to estimate risk and tailor their WNV strategies accordingly. It is important to note, however, the contribution of small scale weather events that can have a large impact on WNV risk, such as sudden heavy rains or short cold periods which affect mosquito survival.[8] For this reason, it is also necessary to monitor WNV activity during the season. In order for PHUs to have a well informed WNV campaign, it may be useful for PHUs to first estimate overall risk for an upcoming season and then monitor level of risk during the season. Early year estimates of virus activity could then be used to inform decisions regarding quantity of resources that could be put towards in-season measures.

Overall, although WNV does not consistently pose a public health risk to the majority of Ontarians, studying this disease is important both for improving the health and safety of the population but also for strengthening the capacity of our health system

for dealing with vector-borne illness. Our results indicate that by using measures of risk that are detectable early in the year it may be possible to estimate the level of WNV activity for the upcoming season, allowing PHUs to enforce appropriate preventive measures and decrease risk.

2.3 References

1. Nash, D. *et al.* The Outbreak of West Nile Virus Infection in the New York City Area in 1999. *N. Engl. J. Med.* **344,** 1807–1814 (2001).

2. Artsob, H. *et al.* West Nile Virus in the New World: trends in the spread and proliferation of West Nile Virus in the Western Hemisphere. *Zoonoses Public Health* **56,** 357–69 (2009).

3. West Nile Virus Activity—United States, 2006. *JAMA* **298,** 619 (2007).

4. Public Health Ontario. Vector-Borne Disease Surveillance Reports. Available at: https://www.publichealthontario.ca/en/ServicesAndTools/SurveillanceServices/Pages/Vector-Borne-Disease-Surveillance-Reports.aspx#.V1BzByHys6w.

5. Rowley, W. a. & Graham, C. L. The effect of temperature and relative humidity on the flight performance of female Aedes aegypti. *J. Insect Physiol.* **14,** 1251–1257 (1968).

6. Reisen, W. K. *et al.* Effects of temperature on the transmission of west nile virus by Culex tarsalis (Diptera: Culicidae). *J. Med. Entomol.* **43,** 309–17 (2006).

7. Dohm, D. J. *et al.* Effect of environmental temperature on the ability of Culex pipiens (Diptera: Culicidae) to transmit West Nile virus. *J. Med. Entomol.* **39,** 221–5 (2002).

8. Chuang, T. W. & Wimberly, M. C. Remote Sensing of Climatic Anomalies and West Nile Virus Incidence in the Northern Great Plains of the United States. *PLoS One* **7,** 1–10 (2012).

9. Manore, C. a *et al.* Towards an early warning system for forecasting human West Nile virus incidence. *PLoS Curr. Outbreaks* **EARLY VIEW,** 1–18 (2014).

10. Rochlin, I., Turbow, D., Gomez, F., Ninivaggi, D. V. & Campbell, S. R. Predictive mapping of human risk for west nile virus (WNV) based on environmental and socioeconomic factors. *PLoS One* **6,** (2011).

11. Chuang, T.-W., Hockett, C. W., Kightlinger, L. & Wimberly, M. C. Landscape-Level Spatial Patterns of West Nile Virus Risk in the Northern Great Plains. *Am. J. Trop. Med. Hyg.* **86,** 724–731 (2012).

12. Yiannakoulias, N. W., Schopflocher, D. P. & Svenson, L. W. Modelling

geographic variations in West Nile virus. *Can. J. public Heal. = Rev. Can. santé publique* **97,** 374–8

13. Conly, J. & Johnston, B. Why the West in West Nile virus infections? *Can. J. Infect. Dis. Med. Microbiol. = J. Can. des Mal. Infect. la Microbiol. médicale* **18,** 285–8 (2007).

14. Liu, A. *et al.* Risk factors for human infection with West Nile Virus in Connecticut: a multi-year analysis. *Int. J. Health Geogr.* **8,** 67 (2009).

15. Degroote, J. P., Sugumaran, R. & Ecker, M. Landscape , demographic and climatic associations with human West Nile virus occurrence regionally in 2012 in the United States of America. **9,** 153–168 (2014).

16. Wimberly, M. C., Lamsal, A., Giacomo, P. & Chuang, T.-W. Regional Variation of Climatic Influences on West Nile Virus Outbreaks in the United States. *Am. J. Trop. Med. Hyg.* **91,** 677–684 (2014).

17. Health Profile, December 2013. (2013). Available at: http://www12.statcan.gc.ca/health-sante/82-228/ind. (Accessed: 1st February 2016)

18. Ontario Ministry of Natural Resources. Provincial Landcover 2000 - 27 Classes. (2002). Available at: http://geo.scholarsportal.info/#r/details/_uri@=538735837.

19. Yiannakoulias, N. W., Schopflocher, D. P. & Svenson, L. W. Modelling Geographic Variations in West Nile Virus. *Can. J. Public Heal.* **97,** 374–378 (2006).

20. Nasci, R. S. *et al.* West Nile virus in overwintering culex mosquitoes, New York City, 2000. *Emerg. Infect. Dis.* **7,** 742–744 (2001).

21. Winters, A. M. *et al.* Combining Mosquito Vector and Human Disease Data for Improved Assessment of Spatial West Nile Virus Disease Risk. *Am. J. Trop. Med. Hyg.* **78,** 654–665 (2008).

22. Paz, S. Climate change impacts on West Nile virus transmission in a global context. *Philos. Trans. Roal Soc. Biol. Sci.* **370,** (2015).

23. Brownstein, J. S., Holford, T. R. & Fish, D. Enhancing West Nile virus surveillance, United States. *Emerg. Infect. Dis.* **10,** 1129–33 (2004).

24. McKenzie, V. J. & Goulet, N. E. Bird community composition linked to human West Nile virus cases along the Colorado front range. *Ecohealth* **7,** 439–47 (2010).

Bridge to Chapter 3

The previous paper identified environmental risk factors that could be used to predict risk of human WNV several months in advance. However as discussed, there are many short term events that can have a substantial impact on human risk of WNV necessitating the use of short term measures of risk. Importantly, the previous paper identified the significance of WNV-positive mosquito pools as a predictor of human WNV cases in a given health unit and year. The following paper explores the possibility of using mosquito positive pool clusters as a means of determining immediate risk of human WNV infection as a complement to the use of early year risk factors. The availability of mosquito surveillance data at high spatial and temporal resolution permits a detailed investigation of the patterns in WNV transmission within Ontario public health units, and how this relates to human WNV risk.

Chapter 3 (Manuscript 2): Spatio-temporal associations of West Nile Virus Human Infection and Mosquito Positive Pool Clusters in Ontario

3.1 Preface to Manuscript 2

The work embodied in this paper has gone through ethical approval by the Ottawa Health Sciences network research ethics board for the use of secondary human data pertaining to cases of WNV. Full results from yearly human cluster detection is provided in appendix 3. Full results from yearly mosquito cluster detection is provided in appendix 4. The conceptualization of the project was jointly performed between Dr. Manisha Kulkarni and Shruti Mallya. Technical advice on preparing the datasets for analysis was provided by Dr. Marie-Helene Roy-Gagnon. All data preparation and analysis was performed by Shruti Mallya.

3.2 Manuscript 2

Spatio-temporal associations of West Nile Virus Human Infection and Mosquito Positive Pool Clusters in Ontario

Shruti Mallya[1], Beate Sander[2,3], Marie-Hélène Roy-Gagnon[1], Ann Jolly[1], Manisha Kulkarni[1]

[1]School of Epidemiology, Public Health and Preventive Medicine, University of Ottawa

[2]Public Health Ontario

[3]University of Toronto

Abstract

This study aims to assess epidemiological risk factors associated with human WNV infection in Ontario, Canada, using spatiotemporal analysis and Geographic Information Systems (GIS). Data from public health surveillance of human cases of WNV and mosquito surveillance programs in Ontario (2002-2014) were analyzed for space-time clusters of human WNV and mosquito WNV-positive pools using SaTScan and Moran's I. Recurrent human clusters were noted in Windsor-Essex county, Toronto and Halton area. Mosquito positive pool clusters corresponding in space to the human case clusters were also identified, however several weeks to a month in advance. These results demonstrate the potential of space-time cluster analyses for predicting human WNV incidence and informing public health practice. By knowing where and when WNV

cases are likely to occur, it may be possible for public health units to take action in order to prevent disease and mitigate risk.

Introduction

Since its introduction to New York State in 1999[1] and subsequent spread across North America, West Nile Virus (WNV) has become a disease of considerable public health concern across the continent. WNV is a virus of the family Flaviviridae and is similar to Japanese encephalitis, St. Louis encephalitis and Murray valley encephalitis, all mosquito-borne pathogens. WNV is contracted from the bite of an infected mosquito, and is maintained in nature by mosquito-avian transmission cycles[2] which vary geographically. For example, in western North America, the primary vector species is *Culex tarsalis* while in eastern North America it is the *Culex pipiens/restuans* complex.[3] The main reservoir species is fairly consistent, however, in the eastern and western United States with the American robin, *Turdus migratorius,* being the predominant reservoir species,[4,5] although secondary reservoir species may vary.[6] Humans, horses and other animals[7] are considered dead-end hosts of WNV since they cannot amplify the virus, with infection resulting in varying degrees of illness or death.[8] Though the disease is usually asymptomatic in humans, febrile illness can occur in about 20%[1] of cases and of these individuals, 35% will develop neuroinvasive disease[9]. The rapid dissemination of the disease across the continent and the emergence of neuroinasive disease as a result necessitated spatial investigations of the disease in order to understand its spread and to identify areas of high risk. However, it is still difficult in Ontario to predict where and when the risk of WNV is highest to humans, since high incidence years are very sporadic.

Since 2001, there have been 1135 cases of WNV reported in the province and a total of 5230 in Canada as of 2015[10], with the majority of Ontario infections arising in 2002 and 2012 (appendix 4). By identifying areas historically prone to high WNV incidence, and understanding where and when human infections arise in these areas, it may be possible to intervene to decrease incidence and the associated costs. One very promising source of predictive data comes from routine surveillance of mosquitoes for WNV in Ontario.

Surveillance via routine trapping and testing of mosquitoes for WNV is a required practice in all Ontario public health units (PHUs) to serve as an early warning system for WNV risk to humans,[11] however the magnitude of the program in terms of location of traps and frequency of testing is at the discretion of the PHUs. Mosquito species are identified and individual mosquitoes are pooled before testing, with approximately 50 mosquitoes per pool in most instances, and thus known as mosquito pools. West Nile virus mosquito positive pools are those that tested positive for infection with WNV based on a standardized molecular assay. In Ontario, the provincial public health authority, Public Health Ontario (PHO), analyzes this data in conjunction with human disease surveillance data and climatic variables to produce a weekly WNV surveillance report.[12] There are some suggested methods for analyzing this data to determine risk, however there are currently no unified guidelines for how PHUs should be analyzing this data to optimize prevention strategies.[11]

The use of mosquito surveillance data for alerting public health officials to a high-risk WNV season is well documented in various WNV-endemic areas across the continent.[13] A popularly used method is the vector index, which provides an estimate of infective mosquitoes for a given period by multiplying the number of mosquitoes per

trap-night by the infection prevalence. Estimates of human disease risk are based on this value. In Colorado, for example, a value of 0.75 is understood to mean increasing risk and intervention efforts are recommended.[14] The calculation of accumulated degree-days (ADD) is a commonly used method for determining when human cases may appear, though not the level of risk. A degree-day is defined as a day when the average temperature is one degree above a reference temperature. In Ontario, the reference value is 18.3ºC for *Culex pipiens/restuans*, the main vector species for WNV. It has been noted by Public Health Ontario that human cases tend to occur between 100-125 ADD, with higher numbers at 180-220 ADD. When the threshold value for ADD is reached earlier in the year, this is indicative of elevated risk. [15] Though this method may provide a fair estimate of when risk is highest and the magnitude of the risk, discrete spatial risk cannot be deduced since ADD will be similar across relatively large areas owing to the inherent spatial autocorrelation of temperature indices. The vector index may be more useful in this regard, since it may be possible to aggregate vector data to acquire an estimate of regional risk, though aggregation may overshadow smaller hotspots within an area. By using point occurrence data in conjunction with area data, we may be able to avoid these issues.

In the past, various cluster detection methods have been used to identify areas at high risk for human infection based on past human incidence records. Commonly used methods include Kulldorf's spatial scan statistic, which detects clustering of cases, and Anselin's local Moran's I which is used to uncover areas of high spatial autocorrelation. Relating WNV-positive mosquito pool clusters to human WNV case clusters has been employed previously to aid in planning mosquito surveillance activities.[16] Yet no studies

have assessed the predictive potential of WNV-positive mosquito pool clusters. With this in mind, this study aimed to identify PHUs in Ontario at high risk of human WNV infection using spatio-temporal cluster detection, compare two frequently used methods of cluster identification for identifying human WNV infections, and relate clusters of WNV-positive mosquito pools to human WNV case clusters in space and time.

Methods

Data Sources

Human surveillance data

Confirmed (laboratory positive) and probable (based on symptoms) human WNV cases in Ontario from 2002-2014 were obtained from the Integrated Public Health Information System (iPHIS) of Public Health Ontario (PHO). Cases were aggregated at the level of public health unit (PHU) and date. The 2011 census population for each PHU was obtained from the Statistics Canada Health Profile, December 2013,[17] and used as the background population (population file). The PHU centroids in latitude and longitude were calculated in ArcMap 10.2 using the 'Calculate Feature Centroid' tool and used as the coordinates for the cases. For Moran's I analysis, yearly incidence rates of WNV were calculated based on the number of cases and the 2011 populations for each PHU and joined to a shapefile of Ontario PHUs in ArcMap10.2.

Mosquito surveillance data

WNV testing and species identification data were obtained from the PHO Vector Surveillance dataset maintained by Public Health Ontario. Three data files were generated for subsequent cluster analysis. First, a case file, representing all WNV-positive mosquito pools, was created by aggregating the number of positive pools by trap

site and date. Aggregation at the level of PHU was not performed as this would result in loss of spatial resolution. A control file was created by aggregating the number of WNV-negative mosquito pools by test site and date. Finally, latitude and longitude coordinates of each site were stored in the coordinates file.

Data Analysis

Cluster analysis

To detect spatio-temporal clusters of human WNV cases and WNV-positive mosquito pools, two methods were used similar to Sugumaran *et al*, with some modifications.[18] To increase the sensitivity and specificity of cluster detection, Kulldorf's spatial scan statistic was used in conjunction with Anselin's Local Moran's I to detect spatio-temporal clusters.[19] The addition of a binary scoring method to assess for homogeneity of incidence rates within clusters produced by SaTScan has been previously suggested [20] and was implemented in this study.

Yearly cluster detection – Spatial scan statistics

Initially, a cluster detection approach using Kulldorf's spatial scan statistic was applied using SaTScan software, which works by scanning across an area in space and/or time using cylindrical windows of varying sizes. The maximum size of the window is user-specified by inputting values for a maximum spatial scanning window size and maximum temporal scanning window size. The spatial scanning window specifies the percentage of the population that can be included in a single cluster. The temporal scanning window size specifies the length of time that can be included in a single cluster. Monte Carlo replications are used to simulate potential case distributions across space and time. The software then performs a likelihood function test to determine which

clusters in the real data are significant based on the observed and expected counts within the window. A relative risk is also outputted based on the number of expected and observed cases within the cluster[21] The advantage of the spatial scan statistic is that analysis can be performed directly on point location data, permitting retention of spatial resolution.

To detect clusters of human WNV cases, a retrospective space-time scan with a Poisson probability model was employed in SaTScan with the default 999 Monte Carlo replications using the human surveillance and population data across all years, 2002 to 2014. Maximum spatial scanning window sizes of 5, 7, 10, 20 and 30% were tested and the 30% window was retained to best reflect the population distribution of the study area. The population distribution of Ontario is highly skewed with the majority of the population living in Southern Ontario. Further, 2011 census data indicates that the City of Toronto PHU accounts for 20.3% of the population of Ontario. Thus, a maximum scanning window size of less than 20.3% would fail to include the Toronto PHU in or as a cluster since its population would be too large. However, when the scanning window is too large, it may fail to detect clusters in areas with smaller populations or detect a group of many PHU with low rates as a cluster. For this reason, it was necessary to employ multiple spatial scanning windows and to use multiple cluster detection methods to ensure that significant clusters were appropriately identified. Chen *et al* have prescribed the use of a binary variable to screen out large clusters containing low incidence rates, [20] this method was applied to the cluster results. Single year models were also performed to facilitate better comparison with clustering detected using Moran's I.

Yearly mosquito positive pool (WNV-positive mosquito pool) clusters were detected using a retrospective Bernoulli model in SaTScan. A Bernoulli model was chosen since data for case (WNV positive) and control (WNV negative) mosquito pools were available for each trap location. Maximum spatial scanning window sizes of 5, 7, 10, 20 and 30% were tested but only results from the 10% window were retained as these produced the most discrete clusters.

Yearly cluster detection – Local Moran's I

Cluster analysis using Anselin's Local Moran's I was done using the Cluster and Outliers Analysis tools' in ArcMap 10.2, which works by assessing whether or not values for rate or prevalence for an area are significantly similar to the values of its neighbours. This is determined based on the values of a z-score and p-value.

As per Sugumaran *et al*, human WNV incidence rates pooled for all years were analyzed using Euclidean distance bands of 200-1000 km. Various distance bands provided similar results so yearly analyses were then performed for 2002-2014 using the optimal distance band calculated by ArcMap.

Cluster comparison

Age and sex distribution of human WNV cases in outbreak years and similar areas were compared using a chi-square test for sex and Fisher test for age as expected counts for some age categories were too low.

Weekly Cluster Detection – Spatial scan statistics

To detect and relate spatio-temporal clusters of high WNV incidence (human and mosquito) within two high incidence years (2002 and 2012), a daily retrospective space-

time scan with a Poisson probability model was employed with 999 Monte Carlo replications and with a maximum time window of 7 days. For comparison, similar analyses were conducted for 2006, a low incidence year with similar levels of mosquito testing to 2002. Maximum spatial scanning window sizes of 5, 10, 20 and 30% were employed for human WNV case data and the 5% and 30% windows were retained as the final. For mosquito surveillance data, a Bernoulli distribution with 5, 10 and 20% spatial scanning windows were tested and the 10% window was retained. The time period used for human cases was June-October to reflect the actual period of human WNV risk. The mosquito time window was specified from May to October to reflect timing of WNV surveillance and the mosquito trapping season. Spatio-temporal scans using only data prior to August were also performed for mosquito data using the same parameters for these same years to detect early season clusters, since these may be obscured by larger late season clusters.

A sensitivity analysis was performed by removing the populations and cases from more northerly PHUs in order to determine whether the inclusion of these PHUs was inflating the relative risk, since these areas are inherently at low risk for WNV due to their climate. This was confirmed, and as such, weekly cluster analysis was performed excluding data from northern health units.

Results

Rates of Human WNV incidence in Ontario

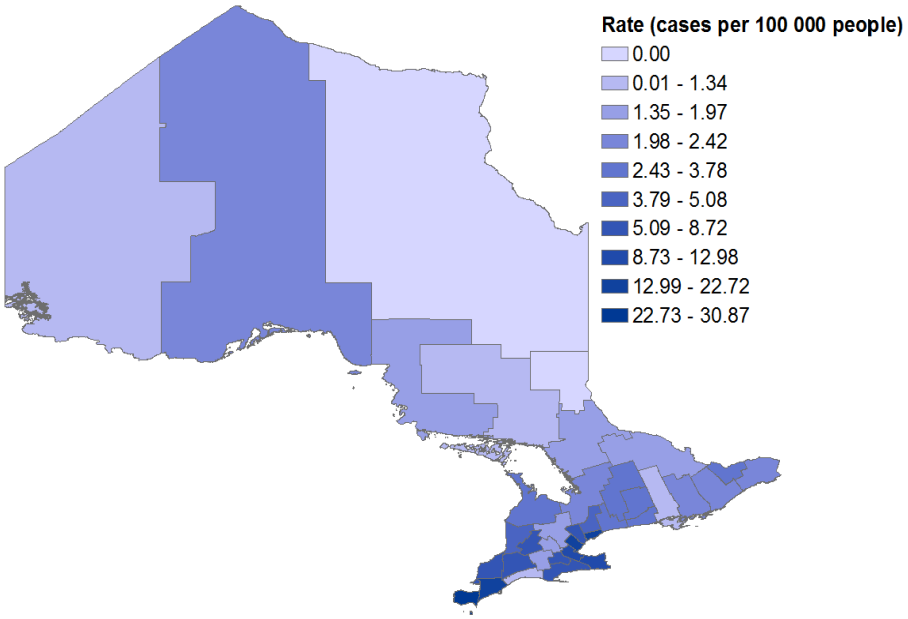

Figure 1: Cumulative incidence of human WNV in Ontario from 2002-2014 based on data from the integrated Public Health Information System (iPHIS) and Statistics Canada. Rates are given in cases per 100 000. Map generated using ArcMap 10.2 (Esri, Redlands, California).

Highest incidence rates generally occur in and around urbanized areas of Southern Ontario (Figure 1). Human WNV case counts were highest in 2002 and 2012 and lowest between 2008-2010.

Yearly Scan

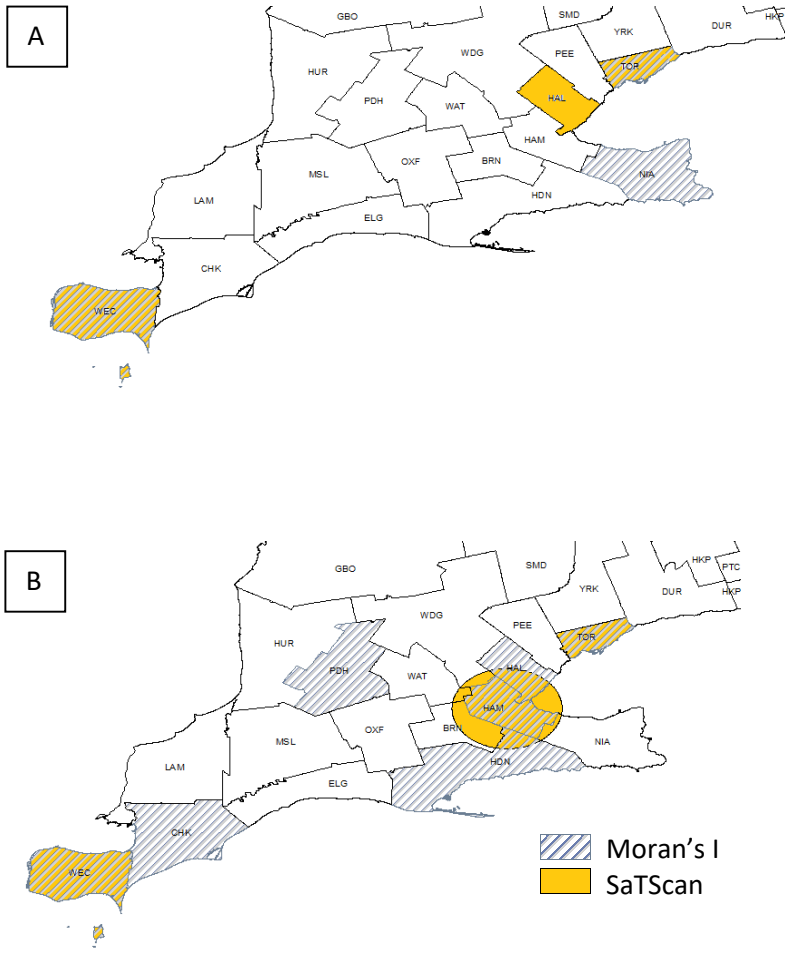

Figure 2: Clusters identified by SaTScan and Moran's I analysis for high WNV incidence years in southern Ontario public health units. Panel A shows locations of significant clusters for 2002 and panel B, clusters for 2012.

Significant human WNV case clusters were found by SaTScan analysis in 2002-2006, and 2012-2013. Moran's I analysis found clusters in 2002, 2004-2005 and 2011-2013. Between the two methods, Windsor-Essex county was most consistently found to have clustering of WNV human cases; clusters of cases in this PHU were found for every year that clustering was detected. Toronto was also found to have frequent clustering of human WNV cases. Clustering between the two highest years (2002 and 2012) were

fairly consistent in space, both years had clusters identified in Windsor-Essex county, Toronto and the Golden horseshoe area; however, Moran's I analysis identified a greater number of clusters in 2012 (Figure 2). Toronto, Windsor-Essex and the Hamilton/ Halton clusters were consistent between the two methods, however the Moran's I analysis identified Halimand-Norfolk (HDN), Perth District (PDH) and Chatham-Kent (CHK) as additional clusters. Fisher and chi-square analysis of demographic characteristics for these years showed no significant differences in age or sex distribution (Table 1). Clustering of WNV-positive mosquito pools generally occurred in proximity to human WNV case clusters except for in 2012 (not shown); this was explored further in the weekly cluster analysis. SaTScan cluster analysis found the relative risks for all three clustered areas were higher in 2002 versus 2012. All p-values were significant. Moran's I analysis showed a slightly different pattern; I values were generally higher and clusters more significant in 2012 versus 2002 (Table 2). The binary method suggested by Chen *et al* was originally planned for implementation in this project, however this method is better applied to areas with smaller areal units such as counties in the United States where there is better spatial distinction of incidence and a larger overall area. For Ontario, no clusters of more than 3 or 4 PHUs were uncovered. As such, though the method was implemented, no useful information was gleaned.

Table 1: Fisher's Exact and Chi-square p-values for age and sex distribution of cases in 2002 versus 2012.

Public Health Unit	Fisher's exact test p-value for Age	Chi-square test p-value for Sex
Toronto	0.6808	0.7952
Windsor-Essex County	0.5802	1.00

Halton	0.5874	0.1279

Table 2: P-values and test statistics obtained from SaTScan and Moran's I analysis by year and public health unit.

Year	Public Health Unit	SaTScan		Moran's I	
		Relative Risk	P-Value	I value	P-value
2002	TOR	2.51	< 0. 001	4.50	0.03
	HAL	4.34	< 0. 001	NA*	NA
	NIA	NA	NA	4.19	0.04
	WEC	3.17	< 0. 001	5.35	<0.01
2012	TOR	2.08	< 0. 001	5.31	0.01
	HAL/ HAM	2.24	< 0. 001	NA	NA
	HAL	NA	NA	8.53	<0.01
	HDN	NA	NA	8.90	<0.01
	PDH	NA	NA	5.54	0.01
	CHK	NA	NA	4.06	0.03
	WEC	2.83	0.0013	6.94	<0.01

*Indicates that this PHU was not detected as being part of a cluster by the specified technique.

Weekly Scan

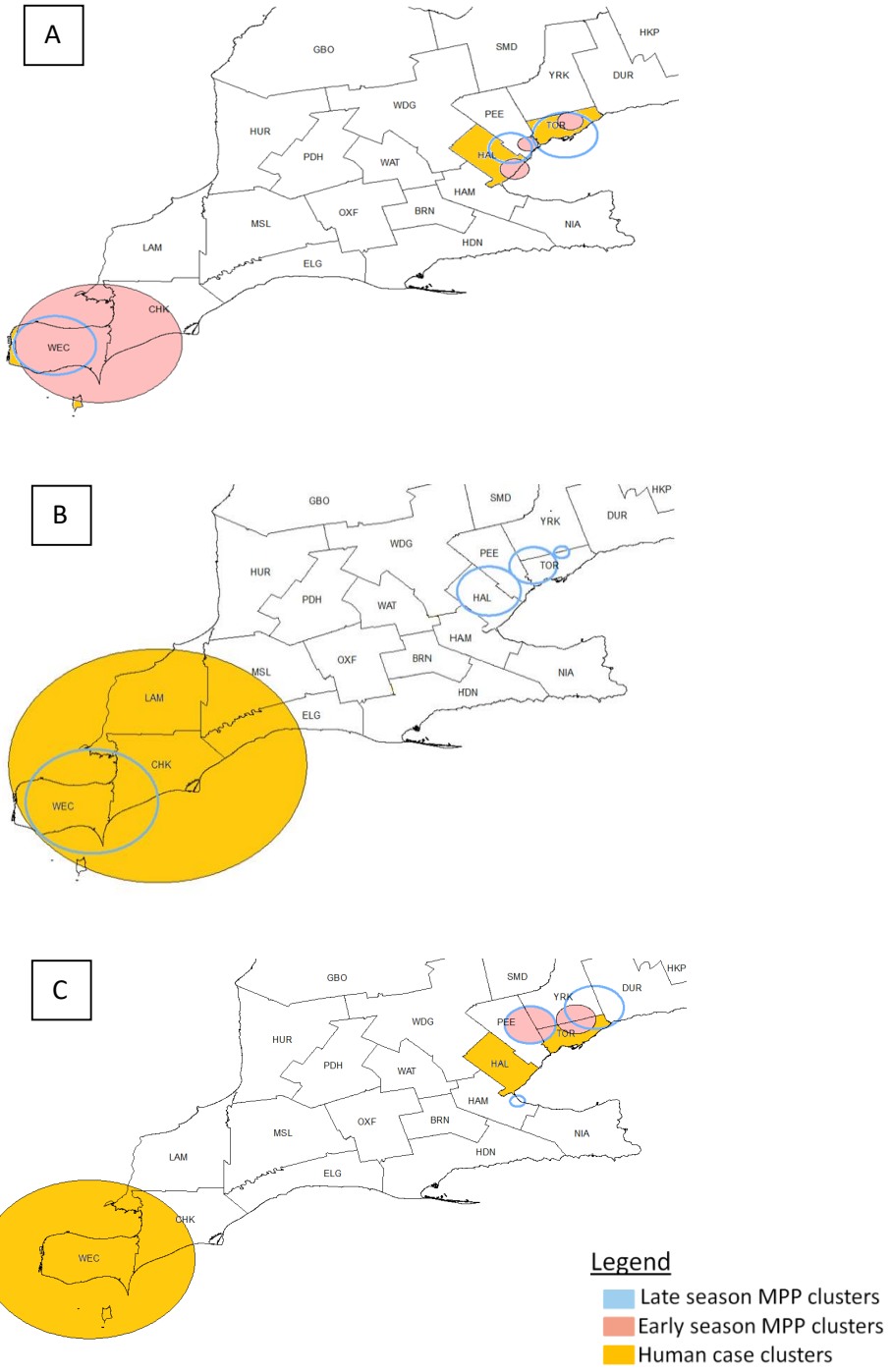

Figure 3: Weekly Human case and WNV-positive mosquito pool (MPP) clusters identified by retrospective SaTScan analysis using Poisson and Bernoulli models, respectively. Panel A shows clusters for 2002, panel B clusters for 2006 and panel C clusters for 2012.

Table 3: Results from SaTScan weekly cluster detection.

Year	Public Health Unit	Cluster type	Time frame	Duration (days)	Observed	Expected	Relative risk	P-value	time interval preceding human WNV case clusters
2002	TOR	Human WNV Case	Aug 30-Sep 4	5	35	3.79	10.06	<0.001	NA
		Late season WNV-positive mosquito pool	Aug 27-29	3	46	7.46	6.61	<0.001	3 days
		Early Season WNV-positive mosquito pool	July 30-31	2	6	0.42	15.31	0.013	1 month
	HAL	Human Case	Sep 9-15	7	19	0.61	32.79	<0.001	NA
		Late season mosquito pool	Aug 13	1	19	2.36	9.97	<0.001	2 weeks
		Early Season Mosquito	July 24-30	7	11	1.07	11.74	<0.001	>1 month
	WEC	Human Case	Aug 26 to Sep 1	5	14	0.47	30.76	<0.001	NA
		Late season mosquito pool	Aug 13	1	19	2.36	8.30	<0.001	3 weeks
		Early Season Mosquito	July 23-29	7	11	1.44	8.70	0.0012	1 month
	HAM	Human Case	Sept 2-8	7	7	0.63	11.28	0.020	NA
	NIA	Human Case	Sept 9-13	5	6	0.37	16.29	0.019	NA
2006	TOR	Late season mosquito pool 1	Aug 22	1	10	0.77	13.62	<0.001	NA

		Late season mosquito pool 2	Aug 1	1	9	0.14	67.67	<0.001	NA
		Early Season Mosquito	None	NA	NA	NA	NA	NA	NA
	HAL	Late season mosquito pool	Aug 9	1	10	0.56	18.87	<0.001	NA
		Early Season Mosquito	None	NA	NA	NA	NA	NA	NA
	WEC, CHK LAM, ELG , MSL	Human Case	Aug 28- Sep 2	5	5	0.15	37.51	0.0055	NA
		Late season mosquito pool	Aug15- Aug18	4	8	0.45	18.65	<0.001	2 weeks
		Early Season Mosquito	None	NA	NA	NA	NA	NA	NA
2012	**TOR**	Human Case	Aug 10- 16	7	24	2.45	10.69	<0.001	NA
		Late season mosquito pool 1	July 25-31	7	17	3.5	5.01	0.0033	3 days
		Late season mosquito pool 2	Aug 8- 14	7	30	4.82	6.59	<0.001	2 weeks
		Early Season Mosquito 1	July 25-31	7	17	2.77	6.81	<0.001	2 weeks
		Early Season Mosquito 2	July 31	1	11	1.72	6.83	0.028	2 weeks
	HAL	Human Case	Aug 9- 14	6	7	0.36	20.07	0.0012	NA
		Late season mosquito pool	None	NA	NA	NA	NA	NA	NA

		Early Season Mosquito	None	NA	NA	NA	NA	NA	NA
	WEC/ CHK	Human Case	Aug 14-20	7	7	0.41	17.51	0.0026	NA
		Late season mosquito pool	None	NA	NA	NA	NA	NA	NA
		Early Season Mosquito	None	NA	NA	NA	NA	NA	NA
	HAM	Late Season mosquito	Aug 8-14	7	9	0.64	14.43	<0.001	NA
	OTT	Late Season mosquito	Jul 30-Aug 5	7	9	0.69	13.32	<0.001	NA

Results from the weekly models showed that WNV-positive mosquito pool clusters overlapped human WNV case clusters in space (Figure 3) and generally appeared up to 2-3 weeks in advance of human case clusters (Table 3). Exclusion of post-July data showed smaller WNV-positive mosquito pool clusters occurring in the same health units as human cases (Figure 3) and 1 month in advance for 2002 and 2012 (Table 3). Early and late season WNV-positive mosquito pool clusters corresponded in space though the overlap was not exact (Figure 3). There were no late season WNV-positive mosquito pool clusters detected for 2006. Relative risks appeared to be of a fairly similar magnitude between 2002 and 2012, though again, those of 2002 were marginally higher. Relative risks for 2006 were comparatively high ($p<0.05$).

Discussion

Through spatiotemporal cluster analysis of entomological and epidemiological surveillance data, this study identified Ontario public health units at high risk of human

WNV infection, and demonstrated a close association between the presence of WNV-positive mosquito pools and human WNV cases in space and time. Clustering of human WNV cases was most apparent in urbanized areas of southern Ontario, after accounting for population size, indicating that factors other than climate may contribute to higher rates of WNV. Human WNV case clusters in both years of high WNV incidence (2002 and 2012) occurred in similar areas, suggesting that these PHUs may be more prone to WNV generally. Demographic factors including socioeconomic characteristics may play a role in the risk of infection. Previous studies in the United States have found various demographic risk factors such as human population density[22] and number of seniors[23] to play a significant role in elevating human risk for WNV infection. Performing such studies in Ontario may help to increase our understanding of why WNV incidence is elevated in these areas. It is also interesting to note that there were no differences in age or sex distribution of cases between 2002 and 2012. This may indicate that the sudden increase of cases in 2012 may not have resulted from a change in target demographic of the virus.

WNV-positive mosquito pool clusters were consistently found within PHUs that had clustering of human WNV infections and occurred up to a month in advance. These results show us the potential of applying cluster detection to mosquito surveillance data to predict human disease incidence. The 2012 data, however, displayed no WNV-positive mosquito pool clustering in PHUs that had human WNV case clustering, other than Toronto. This is likely due to scaling back of mosquito surveillance efforts in the Windsor-Essex county and Halton PHUs in 2012 resulting in low levels of mosquito trapping and sampling, which would decrease the power to detect clusters.

Recognizing the potential of mosquito surveillance to inform public health decision-making, to optimize mosquito surveillance efforts it may prove useful to:

1) Scale mosquito sampling efforts based on early season indicators. Without undertaking sufficient mosquito surveillance efforts, the mosquito surveillance data loses its ability to serve as an early warning for high human disease risk. With budgetary restrictions it is necessary to optimize expenditures, and mosquito surveillance may not be prioritized by individual PHUs. Many studies have found that there are various environmental indicators for a high risk WNV season that occur earlier in the year including higher than average January temperatures.[24] This data is routinely collected and publicly available. If used, this could allow PHUs to make decisions regarding sampling efforts early in the year without spending money to collect data and without training employees to use a complex model.

2) In years with expected high case counts, perform real time cluster detection for WNV-positive mosquito pool clusters. These data would provide useful information regarding where and when human cases are predicted to occur. Based on this data, PHUs may decide to employ targeted prevention strategies such as mosquito control and public awareness campaigns during the season. These would be particularly important in areas that have a greater number of residents at higher risk for complications. If clusters of WNV-positive mosquito pools were to be found in a neighbourhood with a nursing home for example, it would be useful for Public Health authorities to inform residents of the increased risk in order to safeguard against infection (i.e. by using mosquito repellent, covering up and checking the environment for standing water, etc.) and improve care-seeking in the event of febrile illness which may improve reporting of WNV cases.

This study is subject to several limitations. Since the analyses in this study are based on reporting of disease it is important to note the potential for reporting bias. This is particularly relevant due to the spatial nature of the analysis. It is quite possible that in some regions, physicians are more or less likely to order a test[25] for WNV. In areas of historically higher incidence, physicians may be more likely to recognize febrile illness as indicative of a potential case of WNV. In areas historically considered to be at lower risk, physicians may not immediately consider WNV to be the culprit of a patient's symptoms. Demographic factors such as age, gender and education level may also affect care-seeking behaviour. The differing spatial resolutions of the entomological and epidemiological surveillance data (point locations versus PHU-level) also prevented fine-scale assessment of correlation between WNV-positive mosquito pool clusters and human WNV case clusters. Though this was a limitation in our analysis, PHUs will have access to this data for their own population and should be able to implement this analysis more effectively.

Conclusions

Though routinely conducted in many PHUs in Ontario, mosquito surveillance may not as yet be used optimally for alerting local public health officials to WNV risk. The results of this paper show that it may be possible to effectively use cluster analysis of WNV positive mosquito pools to determine approximately where and when human case clusters are likely to occur. However, the usefulness of this analysis and any other analysis based on mosquito surveillance data is dependent on the optimal placement and number of traps as well as the frequency of trapping. In order for mosquito surveillance data to alert us to epidemiological risk, it is necessary that sufficient resources be

allocated to establish a robust surveillance program, which may be resource intensive. By using a combination of early and in-season risk assessments, it may be possible for PHUs to cater their planned responses to high WNV risk in such a way to optimize time, effort and finances, thereby decreasing the economic burden of public health measures while increasing safety to the public.

3.3 References

1. Nash, D. *et al.* The Outbreak of West Nile Virus Infection in the New York City Area in 1999. *N. Engl. J. Med.* **344,** 1807–1814 (2001).
2. Artsob, H. *et al.* West Nile Virus in the New World: trends in the spread and proliferation of West Nile Virus in the Western Hemisphere. *Zoonoses Public Health* **56,** 357–69 (2009).
3. Conly, J. & Johnston, B. Why the West in West Nile virus infections? *Can. J. Infect. Dis. Med. Microbiol. = J. Can. des Mal. Infect. la Microbiol. médicale* **18,** 285–8 (2007).
4. Kilpatrick, A. M., Daszak, P., Jones, M. J., Marra, P. P. & Kramer, L. D. Host heterogeneity dominates West Nile virus transmission. *Proc. Biol. Sci.* **273,** 2327–33 (2006).
5. McKenzie, V. J. & Goulet, N. E. Bird Community Composition Linked to Human West Nile Virus Cases Along the Colorado Front Range. *Ecohealth* **7,** 439–447 (2010).
6. Nemeth, N. M. & Oesterle, P. T. West Nile virus from an avian conservation perspective. *Int. Zoo Yearb.* **48,** 101–115 (2014).
7. Gubler, D. J. The continuing spread of West Nile virus in the western hemisphere. *Clin. Infect. Dis.* **45,** 1039–46 (2007).
8. Colpitts, T. M., Conway, M. J., Montgomery, R. R. & Fikrig, E. West Nile Virus: biology, transmission, and human infection. *Clin. Microbiol. Rev.* **25,** 635–48 (2012).
9. West Nile Virus Activity—United States, 2006. *JAMA* **298,** 619 (2007).
10. Government of Canada. Surveillance of West Nile Virus. *Jan 5* (2016). Available at: http://healthycanadians.gc.ca/diseases-conditions-maladies-affections/disease-maladie/west-nile-nil-occidental/surveillance-eng.php. (Accessed: 2nd June 2016)
11. Division, P. H. *West Nile Virus Preparedness and Prevention Plan*. *Public Health Division* (2008).

12. Public Health Ontario. Vector-Borne Disease Surveillance Reports. Available at: https://www.publichealthontario.ca/en/ServicesAndTools/SurveillanceServices/Pages/Vector-Borne-Disease-Surveillance-Reports.aspx#.V1BzByHys6w.

13. Kwan, J. L. *et al.* Comparison of Enzootic Risk Measures for Predicting West Nile. *Emerg. Infect. Dis.* **18,** 2004–2010 (2012).

14. Microsoft Word - WNV report FINAL.docx - DC_CD-Zoo-WNV-Report-2014.pdf. Available at: https://www.colorado.gov/pacific/sites/default/files/DC_CD-Zoo-WNV-Report-2014.pdf. (Accessed: 2nd June 2016)

15. Ontario Agency for Health Protection and Promotion (Public Health Ontario). Guide for public health units: Considerations for adult mosquito control. *Queen's Printer for Ontario* (2013). Available at: https://www.publichealthontario.ca/en/eRepository/Guide_Considerations_Mosquito_Control_2013.pdf. (Accessed: 29th February 2016)

16. Mulatti, P. *et al.* Retrospective space-time analysis methods to support West Nile virus surveillance activities. *Epidemiol. Infect.* **143,** 202–13 (2015).

17. Health Profile, December 2013. (2013). Available at: http://www12.statcan.gc.ca/health-sante/82-228/ind. (Accessed: 1st February 2016)

18. Sugumaran, R., Larson, S. R. & Degroote, J. P. Spatio-temporal cluster analysis of county-based human West Nile virus incidence in the continental United States. *Int. J. Health Geogr.* **8,** 43 (2009).

19. Anselin, L. Review of Cluster Analysis Software. *Report in fulfill- ment of consultant agreement #2003-04-01 with the North American Association of Central Cancer Registries, Inc* (2004). Available at: https://www.naaccr.org/LinkClick.aspx?fileticket=pScBROXkeTg%3D&tabid=239&mid=699. (Accessed: 6th June 2016)

20. Chen, J., Roth, R. E., Naito, A. T., Lengerich, E. J. & Maceachren, A. M. Geovisual analytics to enhance spatial scan statistic interpretation: an analysis of U.S. cervical cancer mortality. *Int. J. Health Geogr.* **7,** 57 (2008).

21. Kulldorff, M. A spatial scan statistic. *Commun. Stat. - Theory Methods* **26,** 1481–1496 (1997).

22. Liu, A. *et al.* Risk factors for human infection with West Nile Virus in Connecticut: a multi-year analysis. *Int. J. Health Geogr.* **8,** 67 (2009).

23. Rochlin, I., Turbow, D., Gomez, F., Ninivaggi, D. V. & Campbell, S. R. Predictive mapping of human risk for west nile virus (WNV) based on environmental and socioeconomic factors. *PLoS One* **6,** (2011).

24. Manore, C. a *et al.* Towards an early warning system for forecasting human West

Nile virus incidence. *PLoS Curr. Outbreaks* **EARLY VIEW,** 1–18 (2014).

25. Yiannakoulias, N., Svenson, L. W. & Schopflocher, D. P. Diagnostic uncertainty and medical geography: what are we mapping? *Can. Geogr. / Le Géographe Can.* **49,** 291–300 (2005).

Chapter 4: Discussion

West Nile virus, though it does not present an overwhelming danger to the majority of Ontario's population, nevertheless poses a significant public health challenge. It has been mandated by law that PHUs conduct routine collection and testing of mosquitoes during the WNV season,[1] however conducting these activities is laborious and often resource intensive. With budget cuts and seemingly declining WNV activity, funding to these surveillance programs was reduced in several PHUs, leaving these PHUs unprepared for years of high WNV incidence. In order for PHUs to make the most of occasionally scarce resources while maintaining practices to guard the health and safety of the public, we have proposed a 2-stage method of WNV risk assessment.

In the first paper presented in this work, a number of environmental risk factors signalling a potentially high risk WNV season were identified. We suggested that by following these risk factors throughout the year, it could be possible to infer potential risk of WNV for an upcoming WNV season. This information could inform decisions regarding the magnitude of mosquito surveillance activities for WNV, allowing PHUs to scale the program up or down depending on the anticipated level of risk. Monitoring of mosquitoes in these predicted low risk seasons would still be important, nonetheless, as the literature shows that short term, small-scale weather events can also have a substantial impact on WNV risk.[2] Most PHUs perform analyses of mosquito data as per their own chosen methodologies and, as of yet, no uniform guidelines on analyzing these data have been published.

As discussed in the second paper, the different methodologies used by PHUs accurately assess different aspects of risk, but none effectively capture spatio-temporal

risk in a unified measure. To address this, we have proposed spatio-temporal cluster detection of WNV-positive mosquito pools to accurately predict where and when risk of WNV is highest. This analysis could be performed in real time during the WNV season, and was shown to predict human case clusters up to one month in advance. It is important to note that this may be contingent on the level of mosquito collection and testing. We demonstrated that decreasing the number of mosquito pool collections has an adverse impact on the ability to detect clusters of positive pools, and thus the predictive power of the mosquito data in years of high WNV incidence. Similar results have been found in other studies, such as a 2013 paper by Kilpatick *et al.,* which found that the power to predict WNV was greatly reduced without current year testing data.[3] The reduction in mosquito collection and testing efforts hinders surveillance programs' effectiveness as early warning systems, rendering them close to redundancy. In order to preserve the value of WNV mosquito surveillance, we suggest a dynamic, evidence-based approach by which scale and locality of surveillance efforts is decided based on risk factors for infection as well as in season testing results. Currently, location of traps and testing frequency are determined based on a combination of past year and in season testing results in many PHUs. Employing a dynamic approach to mosquito surveillance could aid in ensuring that there is sufficient capacity of the system to alert to high risk while allowing efforts to be decreased in times of low risk, saving time and effort. Nevertheless, a method such as this should be employed in a conscientious manner, specifically in times that low risk is predicted. A prediction of low risk does not guarantee that probability of infection is in fact low. Mosquito surveillance efforts should persist in a manner that would allow for rapid scale up of efforts should there be

sufficient human case counts in a given PHU. In this manner, mosquito surveillance would be harmonized with human surveillance data.

Studies such as those presented in this work improve our understanding of when risk of vector-borne diseases is highest and lowest. However, in addition to these studies, further investigation of the impact of various socio-demographic factors on human WNV infection may be helpful to understand the behavioural aspect of risk. Especially when related to decision making at the PHU level, an understanding of socioeconomic risk may help to guide placement of traps and frequency of testing. It is also necessary to note that studies relying on reported disease incidence are at risk of bias as disease reporting may be spatially dependent.[4] Further studies assessing the spatial dependence of possible human WNV case testing may help to tease apart the true spatial dependence of incidence versus spatial dependence brought on by reporting bias.

By strengthening our surveillance and forecasting systems for vector-borne illness, we prepare the public health system for future epidemics, whether WNV or other pathogens. The current emergency state of global health caused by Zika virus is a testament to the importance of strengthening our local public health systems so that when faced with global health threats, Canada may keep its citizens safe while also remaining a public health leader.

4.1 References

1. Ontario Ministry of Health. Ontario Regulation 199/03: Control of West Nile Virus. (2009). Available at: https://www.ontario.ca/laws/regulation/030199. (Accessed: 31st May 2016)

2. Chuang, T. W. & Wimberly, M. C. Remote Sensing of Climatic Anomalies and West Nile Virus Incidence in the Northern Great Plains of the United States. *PLoS One* **7,** 1–10 (2012).
3. Kilpatrick, a. M. & Pape, W. J. Predicting human west nile virus infections with mosquito surveillance data. *Am. J. Epidemiol.* **178,** 829–835 (2013).

4. Yiannakoulias, N., Svenson, L. W. & Schopflocher, D. P. Diagnostic uncertainty and medical geography: what are we mapping? *Can. Geogr. / Le Géographe Can.* **49,** 291–300 (2005).

Appendix 1: Characteristics of human and mosquito WNV surveillance data

Table 1: Age and Sex Distribution of Human WNV Cases

	Overall		Gender			
			FEMALE		MALE	
Age	**N**	**(%)**	**N**	**(%)**	**N**	**(%)**
0-4	2	0.18	0	0	2	0.38
5-9	6	0.54	2	0.35	4	0.75
10-14	6	0.54	2	0.35	4	0.75
15-19	15	1.36	9	1.56	6	1.13
20-24	24	2.17	18	3.13	6	1.13
25-29	42	3.79	22	3.82	20	3.77
30-34	48	4.34	30	5.21	18	3.39
35-39	71	6.41	41	7.12	30	5.65
40-44	102	9.21	59	10.24	43	8.1
45-49	122	11.02	76	13.19	46	8.66
50-54	145	13.1	88	15.28	57	10.73
55-59	140	12.65	59	10.24	81	15.25

60-64	120	10.84	66	11.46	54	10.17
65-69	73	6.59	27	4.69	46	8.66
70-74	79	7.14	29	5.03	50	9.42
75-79	63	5.69	27	4.69	36	6.78
80-84	32	2.89	13	2.26	19	3.58
85-89	11	0.99	6	1.04	5	0.94
90+	6	0.54	2	0.35	4	0.75
Total	1107	100	576	100	531	100

Table 2: Mosquito WNV test results by Public Health Unit.

	Overall		**Result**			
	1		**Negative**		**Positive**	
Health unit	**N**	**(%)**	**N**	**(%)**	**N**	**(%)**
ALG	2809	1	2803	1	6	0
BRN	4635	2	4623	2	12	1
CHK	4319	2	4285	2	34	1
DUR	5853	3	5786	3	67	3

ELG	1146	1	1141	1	5	0
EOH	3405	2	3393	2	12	1
GBO	828	0	828	0	0	0
HAL	10000	5	10000	5	236	10
HAM	14000	7	14000	7	128	5
HDN	4897	2	4885	2	12	1
HKP	5196	3	5188	3	8	0
HPE	4390	2	4367	2	23	1
HUR	2975	1	2975	1	0	0
KFL	1205	1	1201	1	4	0
LAM	4175	2	4152	2	23	1
LGL	3034	1	3033	1	1	0
MSL	10000	5	10000	5	79	3
NIA	6680	3	6611	3	69	3
NPS	2794	1	2793	1	1	0
NWR	1487	1	1481	1	6	0
OTT	9831	5	9750	5	81	3
OXF	4361	2	4329	2	32	1

PDH	3072	1	3058	2	14	1
PEE	18000	9	17000	9	431	18
PQP	2326	1	2324	1	2	0
PTC	1893	1	1871	1	22	1
REN	1711	1	1711	1	0	0
SMD	6485	3	6473	3	12	1
SUD	5386	3	5370	3	16	1
THB	1239	1	1239	1	0	0
TOR	23000	11	22000	11	708	29
TSK	1442	1	1441	1	1	0
WAT	6956	3	6928	3	28	1
WDG	4185	2	4176	2	9	0
WEC	9336	5	9117	4	219	9
YRK	13000	6	13000	6	112	5
total	206051	100	203332	100	2413	100

Table 3: WNV test result by mosquito species.

	Overall		Result			
			Negative		Positive	
Species	**N**	**(%)**	**N**	**(%)**	**N**	**(%)**
Ae. cinereus	717	0	717	0	0	0
Ae. vexans	24	0	24	0	0	0
Ae. vexans nipponi	11	0	11	0	0	0
Ae. vexans vexans	55000	27	55000	27	142	6
Ae. vexans/cantator	1406	1	1399	1	7	0
Aedes vexans nipponi	56	0	56	0	0	0
Aedes/Ochlerotatus species	449	0	447	0	2	0
An quadrimaculatus	3	0	3	0	0	0
An. Punctipennis	13000	6	13000	6	23	1
An. quadrimaculatus	2495	1	2491	1	4	0
An. quadrimaculatus/walkeri	81	0	80	0	1	0
An. walkeri	2154	1	2153	1	1	0
Anopheles species	403	0	401	0	2	0
Cq. perturbans	12000	6	12000	6	27	1

Cq. perturbans (pale legs)	113	0	113	0	0	0
Cs. melanura	109	0	109	0	0	0
Culex species	4386	2	4216	2	170	7
Culiseta morsitans	314	0	314	0	0	0
Cx . salinarius	1	0	1	0	0	0
Cx pipiens/restuans	52	0	46	0	6	0
Cx. Tarsalis	2	0	2	0	0	0
Cx. erraticus	21	0	21	0	0	0
Cx. pipiens	11000	5	10000	5	316	13
Cx. pipiens/restuans	50000	24	48000	24	1530	63
Cx. quinquefasciatus	18	0	17	0	1	0
Cx. restuans	6680	3	6597	3	83	3
Cx. salinarius	853	0	825	0	28	1
Cx. spp.	263	0	262	0	1	0
Cx. tarsalis	140	0	140	0	0	0
Males (unidentified)	3	0	3	0	0	0
Oc canadensis	4	0	4	0	0	0
Oc triseriatus	11	0	11	0	0	0

Oc trivittatus	7	0	7	0	0	0
Oc. black legged	381	0	381	0	0	0
Oc. broad-banded	698	0	697	0	1	0
Oc. canadensis	3818	2	3818	2	0	0
Oc. excrucians	738	0	736	0	2	0
Oc. hendersoni	22	0	22	0	0	0
Oc. japonicus	12000	6	12000	6	19	1
Oc. provocans	102	0	102	0	0	0
Oc. sollicitans	462	0	461	0	1	0
Oc. stimulans	7507	4	7505	4	2	0
Oc. triseriatus	6517	3	6494	3	23	1
Oc. triseriatus/hendersoni	2314	1	2313	1	1	0
Oc. trivittatus	11000	5	11000	5	20	1
Other species	83	0	83	0	0	0
Stegomyia albopicta (Aedes albopictus)	3	0	3	0	0	0
Ur. species	1	0	1	0	0	0

Appendix 2: Full Modelling Results

Modelling code:

```
proc glimmix data= dataset
plots= (studentpanel (type=noblup) studentpanel (type= blup))
plots=boxplot(random marginal conditional observed);
class Health_unit year;
model cases=  Jan_meanmin_temp Feb_meanmin_temp Mar_meanmin_temp
Apr_meanmin_temp
Jul_meanmin_temp Aug_meanmin_temp Feb_mean_prec Mar_mean_prec
 Apr_meamax_temp Aug_meamax_temp percent_positive_pools
/dist=poisson offset= log_total link=log solution;
random Health_unit/ subject=intercept type= sp(sph)(latitude
longitude);
random year/ subject=Health_unit residual;
run;
```

The GLIMMIX Procedure

Model Information	
Data Set	WORK.DATASET
Response Variable	cases
Response Distribution	Poisson
Link Function	Log
Variance Function	Default
Offset Variable	log_total
Variance Matrix	Not blocked
Estimation Technique	Residual PL
Degrees of Freedom Method	Containment

Class Level Information		
Class	**Levels**	**Values**
Health_Unit	28	BRN CHK DUR ELG EOH GBO HAL HAM HDN HKP HPE HUR KFL LAM LGL MSL NIA OTT OXF PDH PEE PTC SMD TOR WAT WDG WEC YRK
Year	12	2002 2003 2004 2005 2006 2007 2008 2009 2010 2011 2012 2013

Number of Observations Read	336
Number of Observations Used	329

Dimensions	
G-side Cov. Parameters	2

R-side Cov. Parameters	1
Columns in X	12
Columns in Z	28
Subjects (Blocks in V)	1
Max Obs per Subject	329

Optimization Information	
Optimization Technique	Dual Quasi-Newton
Parameters in Optimization	2
Lower Boundaries	2
Upper Boundaries	0
Fixed Effects	Profiled
Residual Variance	Profiled
Starting From	Data

Iteration History

Iteration	Restarts	Subiterations	Objective Function	Change	Max Gradient
0	**0**	**3**	1016.932359	0.55436512	6.833E-7
1	**0**	**5**	1265.1394155	0.43128003	1.131E-6
2	**0**	**4**	1338.9260334	0.16555881	3.674E-7
3	**0**	**2**	1344.2864204	0.01668493	6.557E-7
4	**0**	**1**	1344.3400676	0.00061632	8.673E-8
5	**0**	**0**	1344.3389608	0.00000000	2.214E-6

Convergence criterion (PCONV=1.11022E-8) satisfied.

Fit Statistics	
-2 Res Log Pseudo-Likelihood	1344.34
Generalized Chi-Square	500.05
Gener. Chi-Square / DF	1.58

Covariance Parameter Estimates

Cov Parm	Subject	Estimate	Standard Error
Variance	**Intercept**	79.2531	42.1827
SP(SPH)	**Intercept**	661.81	.

Residual (VC)	1.5774	0.1297

Solutions for Fixed Effects

Effect	Estimate	Standard Error	DF	t Value	Pr > |t|	Alpha	Lower	Upper
Intercept	-17.6075	9.1821	27	-1.92	0.0658	0.05	-36.4475	1.2325
Jan_meanmin_temp	0.04722	0.02395	290	1.97	0.0496	0.05	0.000086	0.09435
Feb_meanmin_temp	0.4572	0.05334	290	8.57	<.0001	0.05	0.3522	0.5621
Mar_meanmin_temp	-0.2650	0.03773	290	-7.03	<.0001	0.05	-0.3393	-0.1908
Apr_meanmin_temp	-0.6616	0.1047	290	-6.32	<.0001	0.05	-0.8677	-0.4554
Jul_meanmin_temp	0.1917	0.07239	290	2.65	0.0085	0.05	0.04927	0.3342
Aug_meanmin_temp	0.2300	0.1069	290	2.15	0.0323	0.05	0.01959	0.4403
Feb_mean_prec	-0.01535	0.004021	290	-3.82	0.0002	0.05	-0.02327	-0.00744
Mar_mean_prec	0.003096	0.002928	290	1.06	0.2913	0.05	-0.00267	0.008859
Apr_meamax_temp	-0.1482	0.07226	290	-2.05	0.0412	0.05	-0.2904	-0.00594
Aug_meamax_temp	0.2266	0.1052	290	2.15	0.0321	0.05	0.01950	0.4338
percent_positive_poo	0.07061	0.01702	290	4.15	<.0001	0.05	0.03711	0.1041

Type III Tests of Fixed Effects

Effect	Num DF	Den DF	F Value	Pr > F
Jan_meanmin_temp	1	290	3.89	0.0496
Feb_meanmin_temp	1	290	73.46	<.0001

Mar_meanmin_temp	1	290	49.36	<.0001
Apr_meanmin_temp	1	290	39.90	<.0001
Jul_meanmin_temp	1	290	7.02	0.0085
Aug_meanmin_temp	1	290	4.63	0.0323
Feb_mean_prec	1	290	14.58	0.0002
Mar_mean_prec	1	290	1.12	0.2913
Apr_meamax_temp	1	290	4.20	0.0412
Aug_meamax_temp	1	290	4.64	0.0321
percent_positive_poo	1	290	17.21	<.0001

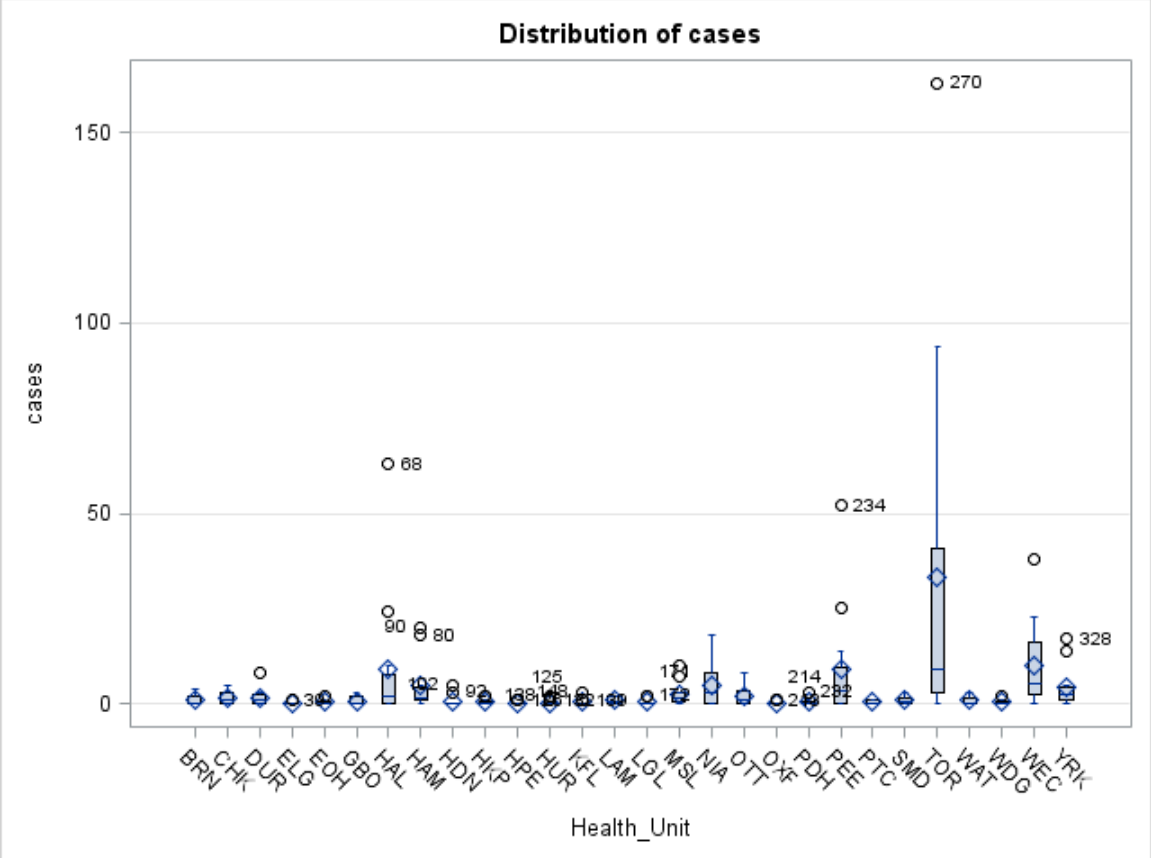

Figure 1: Box plot of distribution of WNV cases by Health unit. Blue diamond indicates the mean with the length of the box indicating the interquartile range, the horizontal line indicating the median and the vertical blue lines indicating the minimum and maximum values.

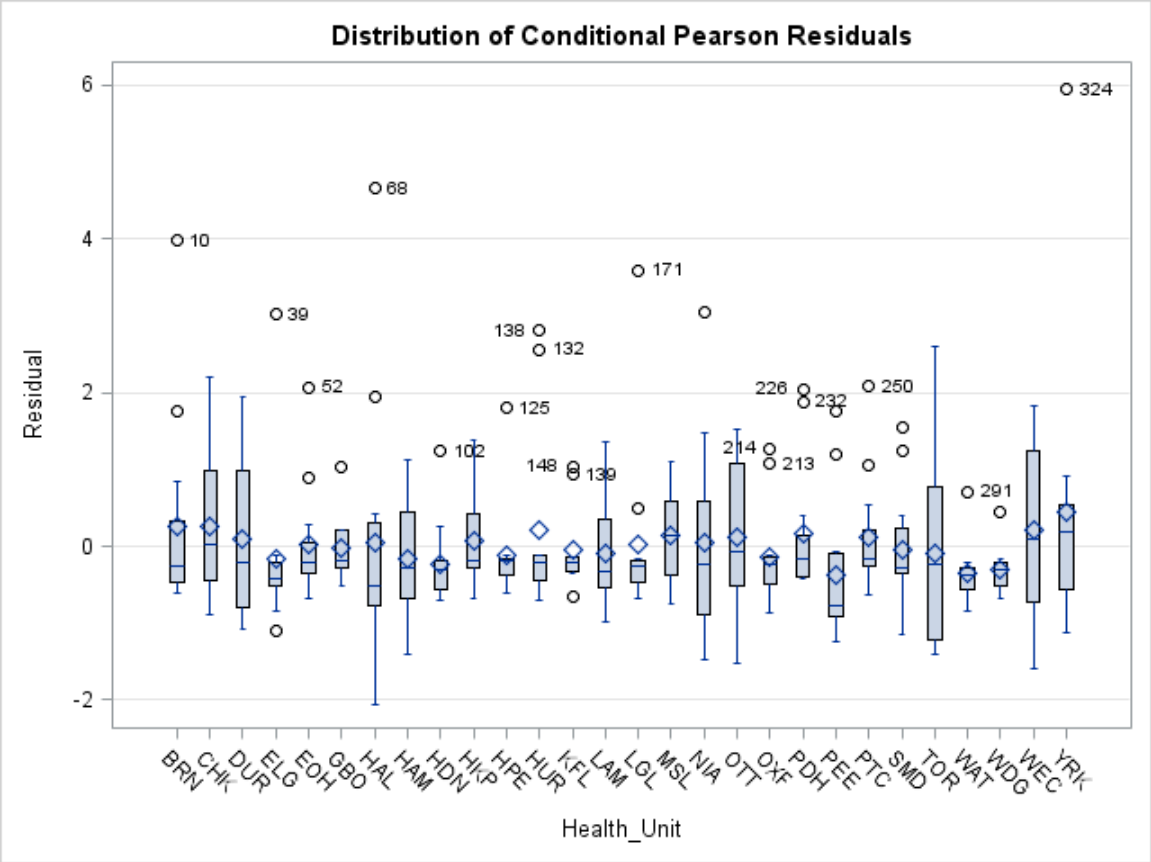

Figure 2: Box plot of distribution of conditional pearson residuals. Blue diamond indicates the mean with the length of the box indicating the interquartile range, the horizontal line indicating the median and the vertical blue lines indicating the minimum and maximum values.

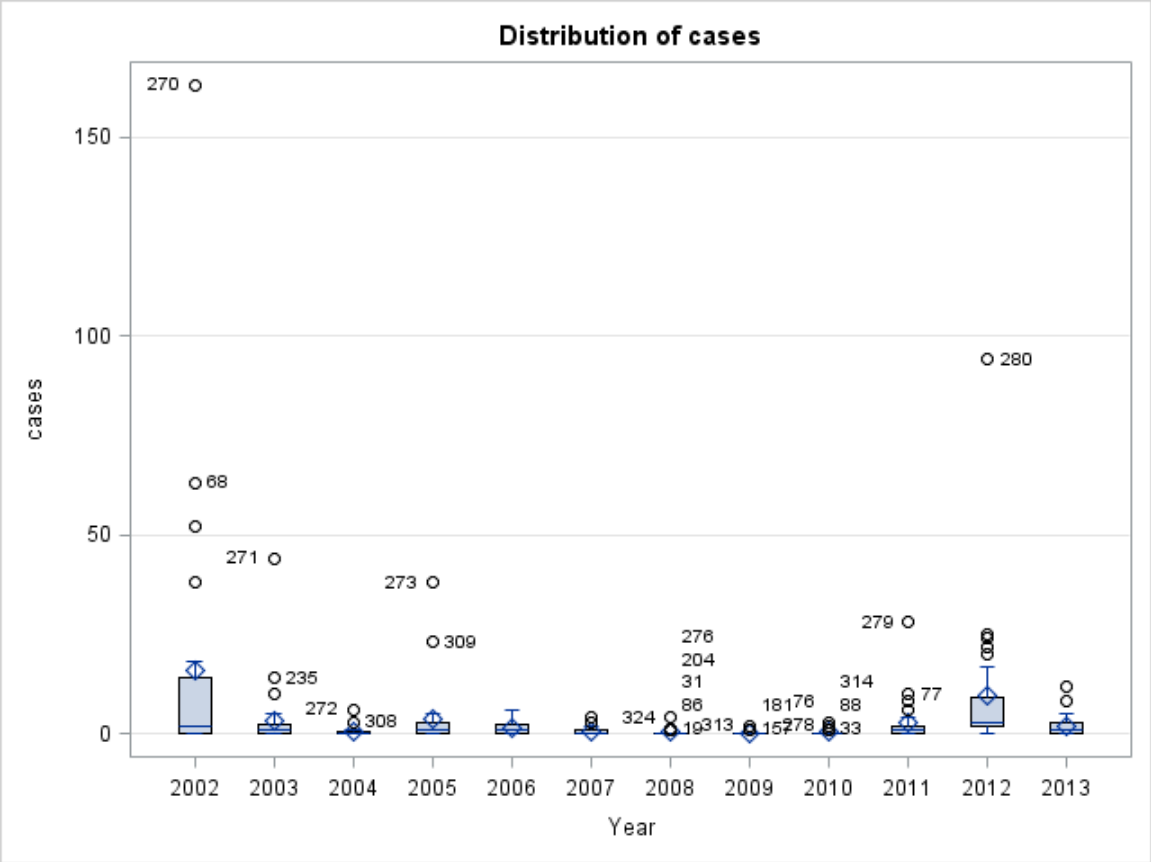

Figure 3: Box plot of distribution of WNV cases in a health unit by year Blue diamond indicates the mean with the length of the box indicating the interquartile range, the horizontal line indicating the median and the vertical blue lines indicating the minimum and maximum values.

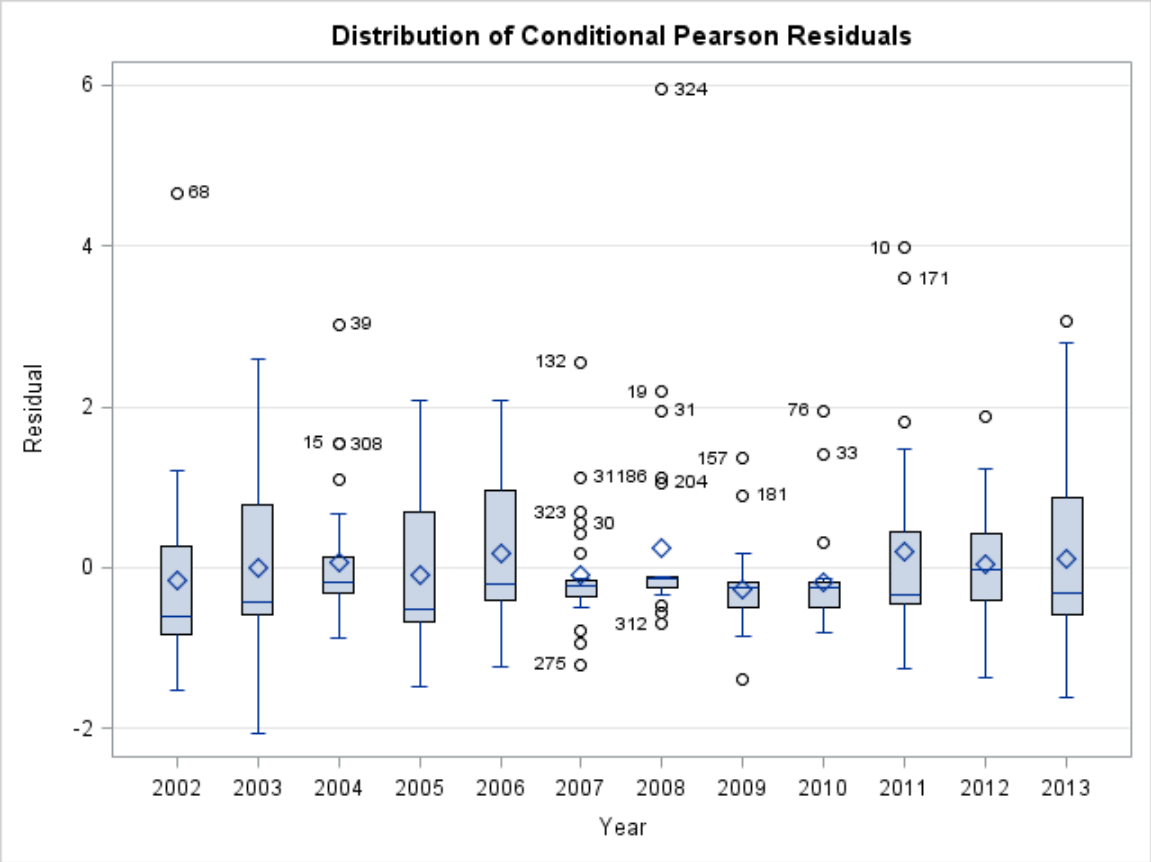

Figure 4: Box plot of distribution of conditional Pearson residuals by year. Blue diamond indicates the mean with the length of the box indicating the interquartile range, the horizontal line indicating the median and the vertical blue lines indicating the minimum and maximum values.

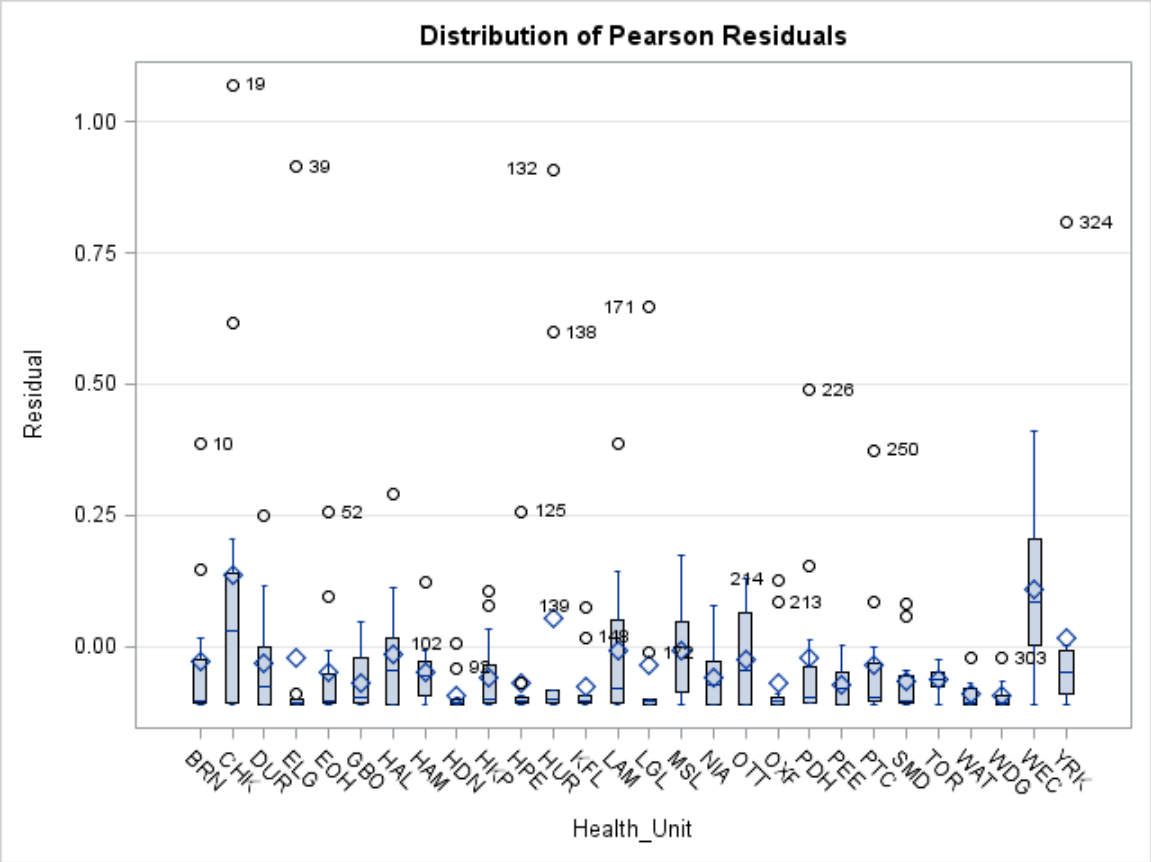

Figure 5: Box plot of distribution of Pearson residuals by health unit. Blue diamond indicates the mean with the length of the box indicating the interquartile range, the horizontal line indicating the median and the vertical blue lines indicating the minimum and maximum values.

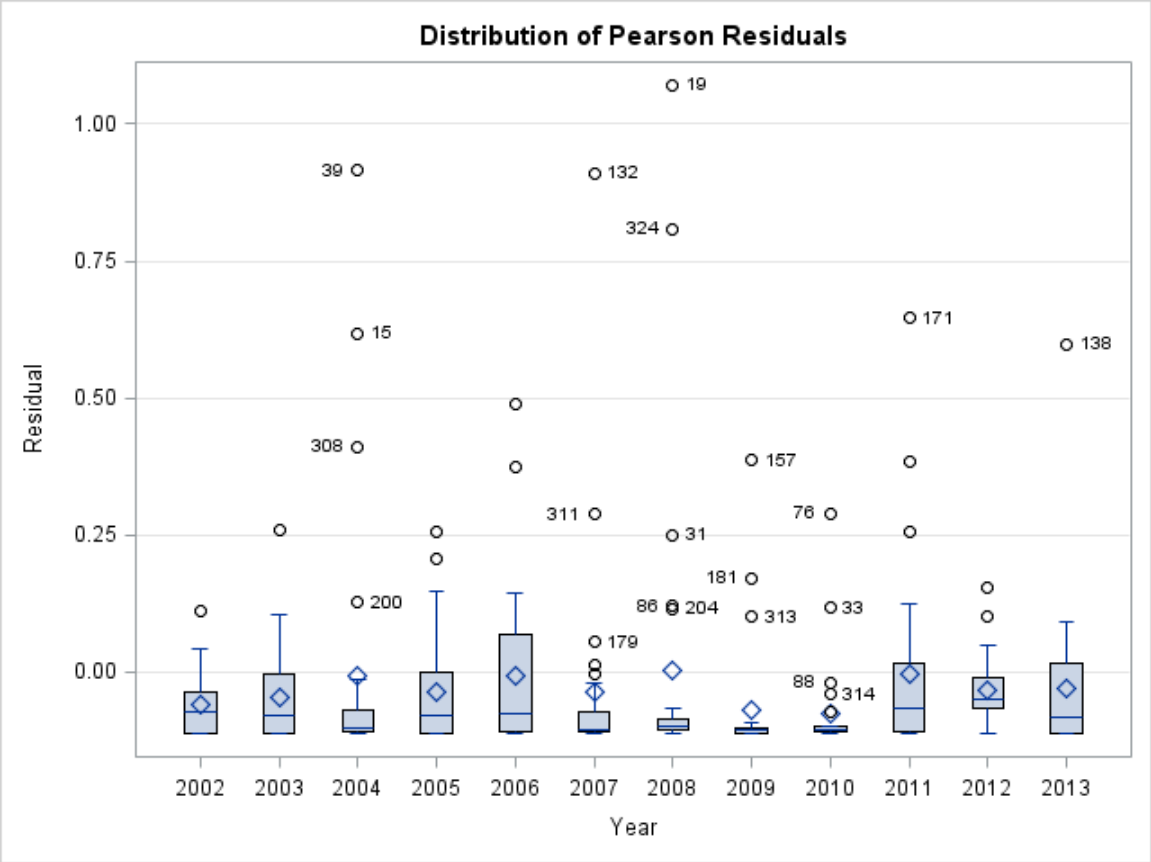

Figure 6: Box plot of distribution of Pearson residuals by year. Blue diamond indicates the mean with the length of the box indicating the interquartile range, the horizontal line indicating the median and the vertical blue lines indicating the minimum and maximum values.

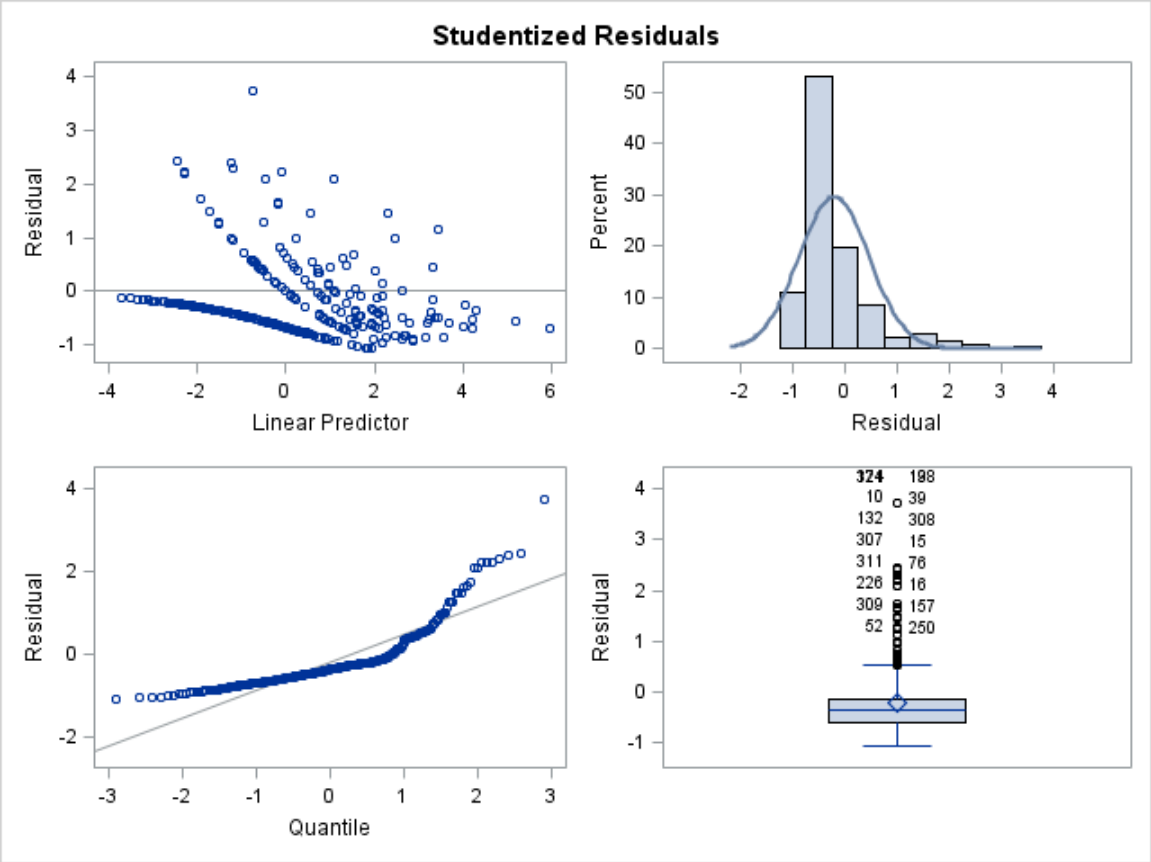

Figure 7: Studentized residuals of full model. Clockwise from right: Distribution of residuals, boxplot of residuals, Residuals by quantile and residuals by linear predictor.

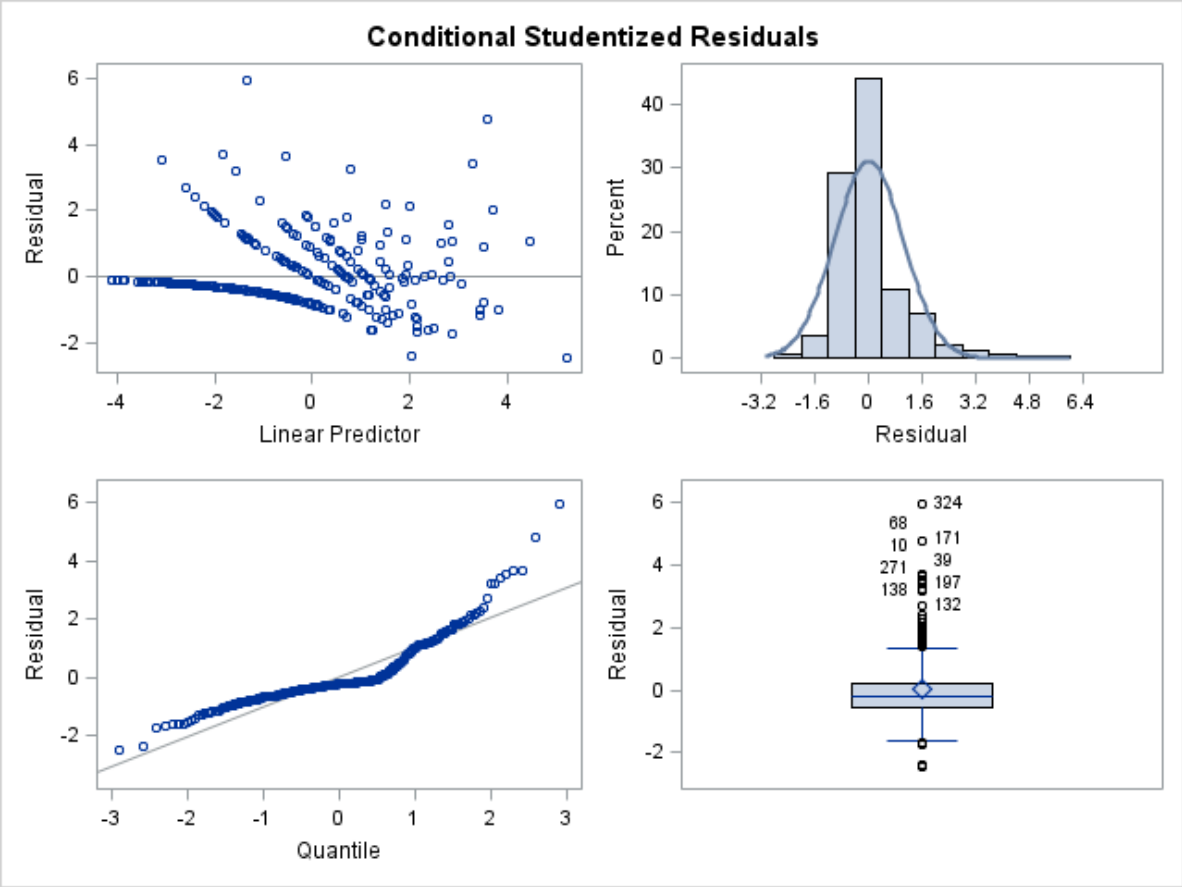

Figure 8: Conditional studentized residuals of full model. Clockwise from right: Distribution of residuals, boxplot of residuals, Residuals by quantile and residuals by linear predictor.

Appendix 3: Univariate Analysis Results

univariate results

The GLIMMIX Procedure

Model Information	
Data Set	WORK.DATASET
Response Variable	cases
Response Distribution	Poisson
Link Function	Log
Variance Function	Default
Offset Variable	log_total
Variance Matrix	Not blocked
Estimation Technique	Residual PL
Degrees of Freedom Method	Containment

Class Level Information

Class	Levels	Values
Health_Unit	28	BRN CHK DUR ELG EOH GBO HAL HAM HDN HKP HPE HUR KFL LAM LGL MSL NIA OTT OXF PDH PEE PTC SMD TOR WAT WDG WEC YRK
Year	12	2002 2003 2004 2005 2006 2007 2008 2009 2010 2011 2012 2013

Number of Observations Read	336
Number of Observations Used	336

Dimensions	
G-side Cov. Parameters	2
R-side Cov. Parameters	1
Columns in X	2
Columns in Z	28
Subjects (Blocks in V)	1
Max Obs per Subject	336

Optimization Information	
Optimization Technique	Dual Quasi-Newton
Parameters in Optimization	2

Optimization Information	
Lower Boundaries	2
Upper Boundaries	0
Fixed Effects	Profiled
Residual Variance	Profiled
Starting From	Data

Iteration History

Iteration	Restarts	Subiterations	Objective Function	Change	Max Gradient
0	**0**	**4**	1168.9952504	0.16582318	2.516E-6
1	**0**	**4**	1442.3825075	0.57515474	3.102E-6
2	**0**	**4**	1513.7200749	0.23366152	3.556E-7
3	**0**	**2**	1517.0137018	0.03956952	0.000023
4	**0**	**1**	1516.7226617	0.00456698	5.789E-6
5	**0**	**1**	1516.6543021	0.00044246	1.44E-8
6	**0**	**1**	1516.6470899	0.00004466	2.433E-7
7	**0**	**0**	1516.6463536	0.00000000	1.2E-6

Convergence criterion (PCONV=1.11022E-8) satisfied.

Fit Statistics	
-2 Res Log Pseudo-Likelihood	1516.65
Generalized Chi-Square	1741.21
Gener. Chi-Square / DF	5.21

Covariance Parameter Estimates

Cov Parm	Subject	Estimate	Standard Error
Variance	**Intercept**	88.0037	52.0877
SP(SPH)	**Intercept**	673.32	.
Residual (VC)		5.2132	0.4125

Solutions for Fixed Effects								
Effect	**Estimate**	**Standard Error**	**DF**	**t Value**	**Pr > |t|**	**Alpha**	**Lower**	**Upper**
Intercept	-10.2975	9.3436	27	-1.10	0.2802	0.05	-29.4689	8.8740
Jan_meanmin_temp	0.1853	0.02333	307	7.94	<.0001	0.05	0.1394	0.2312

Type III Tests of Fixed Effects				
Effect	**Num DF**	**Den DF**	**F Value**	**Pr > F**
Jan_meanmin_temp	1	307	63.11	<.0001

univariate results

The GLIMMIX Procedure

Model Information	
Data Set	WORK.DATASET
Response Variable	cases
Response Distribution	Poisson
Link Function	Log
Variance Function	Default
Offset Variable	log_total
Variance Matrix	Not blocked
Estimation Technique	Residual PL
Degrees of Freedom Method	Containment

Class Level Information

Class	Levels	Values
Health_Unit	28	BRN CHK DUR ELG EOH GBO HAL HAM HDN HKP HPE HUR KFL LAM LGL MSL NIA OTT OXF PDH PEE PTC SMD TOR WAT WDG WEC YRK
Year	12	2002 2003 2004 2005 2006 2007 2008 2009 2010 2011 2012 2013

Number of Observations Read	336
Number of Observations Used	336

Dimensions	
G-side Cov. Parameters	2
R-side Cov. Parameters	1
Columns in X	2
Columns in Z	28
Subjects (Blocks in V)	1
Max Obs per Subject	336

Optimization Information	
Optimization Technique	Dual Quasi-Newton
Parameters in Optimization	2
Lower Boundaries	2

Optimization Information	
Upper Boundaries	0
Fixed Effects	Profiled
Residual Variance	Profiled
Starting From	Data

Iteration History

Iteration	Restarts	Subiterations	Objective Function	Change	Max Gradient
0	**0**	**5**	1172.9559932	0.33319203	7.494E-6
1	**0**	**5**	1454.1417647	0.90802759	6.144E-6
2	**0**	**4**	1545.390896	0.67509732	2.167E-7
3	**0**	**3**	1554.0560825	0.26060692	0.000011
4	**0**	**2**	1554.0917611	0.07868964	9.608E-6
5	**0**	**2**	1553.8064671	0.02258031	6.975E-7
6	**0**	**2**	1553.6989199	0.00642610	1.04E-7
7	**0**	**1**	1553.6662272	0.00188122	0.000011
8	**0**	**1**	1553.6564811	0.00049557	3.045E-6
9	**0**	**1**	1553.6539021	0.00013011	8.058E-7
10	**0**	**0**	1553.6532239	0.00000000	6.881E-6

Convergence criterion (PCONV=1.11022E-8) satisfied.

Fit Statistics	
-2 Res Log Pseudo-Likelihood	1553.65
Generalized Chi-Square	1814.02
Gener. Chi-Square / DF	5.43

Covariance Parameter Estimates

Cov Parm	Subject	Estimate	Standard Error
Variance	**Intercept**	21.9016	33.4112
SP(SPH)	**Intercept**	673.45	.
Residual (VC)		5.4312	0.4358

Solutions for Fixed Effects								
Effect	**Estimate**	**Standard Error**	**DF**	**t Value**	**Pr > |t|**	**Alpha**	**Lower**	**Upper**
Intercept	-8.9198	4.6693	27	-1.91	0.0668	0.05	-18.5005	0.6608
Feb_meanmin_temp	0.3563	0.03050	307	11.68	<.0001	0.05	0.2963	0.4163

Type III Tests of Fixed Effects				
Effect	**Num DF**	**Den DF**	**F Value**	**Pr > F**
Feb_meanmin_temp	1	307	136.43	<.0001

univariate results

The GLIMMIX Procedure

Model Information	
Data Set	WORK.DATASET
Response Variable	cases
Response Distribution	Poisson
Link Function	Log
Variance Function	Default
Offset Variable	log_total
Variance Matrix	Not blocked
Estimation Technique	Residual PL
Degrees of Freedom Method	Containment

Class Level Information

Class	Levels	Values
Health_Unit	28	BRN CHK DUR ELG EOH GBO HAL HAM HDN HKP HPE HUR KFL LAM LGL MSL NIA OTT OXF PDH PEE PTC SMD TOR WAT WDG WEC YRK
Year	12	2002 2003 2004 2005 2006 2007 2008 2009 2010 2011 2012 2013

Number of Observations Read	336
Number of Observations Used	336

Dimensions	
G-side Cov. Parameters	2
R-side Cov. Parameters	1
Columns in X	2
Columns in Z	28
Subjects (Blocks in V)	1
Max Obs per Subject	336

Optimization Information	
Optimization Technique	Dual Quasi-Newton
Parameters in Optimization	2
Lower Boundaries	2

Optimization Information	
Upper Boundaries	0
Fixed Effects	Profiled
Residual Variance	Profiled
Starting From	Data

Iteration History

Iteration	Restarts	Subiterations	Objective Function	Change	Max Gradient
0	**0**	**4**	1209.8850217	0.21045359	2.07E-6
1	**0**	**6**	1492.8419169	0.73581751	0.000013
2	**0**	**6**	1580.7136247	0.40008815	1.786E-7
3	**0**	**3**	1591.8830881	0.11222255	7.682E-6
4	**0**	**2**	1592.2648213	0.02216631	3.58E-6
5	**0**	**1**	1592.1173871	0.00398053	4.049E-6
6	**0**	**1**	1592.0811137	0.00069385	1.35E-7
7	**0**	**1**	1592.074435	0.00012253	1.648E-7
8	**0**	**0**	1592.0732435	0.00000000	8.804E-6

Convergence criterion (PCONV=1.11022E-8) satisfied.

Fit Statistics	
-2 Res Log Pseudo-Likelihood	1592.07
Generalized Chi-Square	2671.12
Gener. Chi-Square / DF	8.00

Covariance Parameter Estimates

Cov Parm	Subject	Estimate	Standard Error
Variance	**Intercept**	97.2256	10187
SP(SPH)	**Intercept**	674.08	70623
Residual (VC)		7.9974	0.6307

Solutions for Fixed Effects								
Effect	**Estimate**	**Standard Error**	**DF**	**t Value**	**Pr > |t|**	**Alpha**	**Lower**	**Upper**
Intercept	-11.4761	9.8205	27	-1.17	0.2528	0.05	-31.6261	8.6739
Mar_meanmin_temp	0.1340	0.03388	307	3.95	<.0001	0.05	0.06732	0.2006

Type III Tests of Fixed Effects				
Effect	**Num DF**	**Den DF**	**F Value**	**Pr > F**
Mar_meanmin_temp	1	307	15.64	<.0001

univariate results

The GLIMMIX Procedure

Model Information	
Data Set	WORK.DATASET
Response Variable	cases
Response Distribution	Poisson
Link Function	Log
Variance Function	Default
Offset Variable	log_total
Variance Matrix	Not blocked
Estimation Technique	Residual PL
Degrees of Freedom Method	Containment

Class Level Information

Class	Levels	Values
Health_Unit	28	BRN CHK DUR ELG EOH GBO HAL HAM HDN HKP HPE HUR KFL LAM LGL MSL NIA OTT OXF PDH PEE PTC SMD TOR WAT WDG WEC YRK
Year	12	2002 2003 2004 2005 2006 2007 2008 2009 2010 2011 2012 2013

Number of Observations Read	336
Number of Observations Used	336

Dimensions	
G-side Cov. Parameters	2
R-side Cov. Parameters	1
Columns in X	2
Columns in Z	28
Subjects (Blocks in V)	1
Max Obs per Subject	336

Optimization Information	
Optimization Technique	Dual Quasi-Newton
Parameters in Optimization	2
Lower Boundaries	2

Optimization Information	
Upper Boundaries	0
Fixed Effects	Profiled
Residual Variance	Profiled
Starting From	Data

Iteration History

Iteration	Restarts	Subiterations	Objective Function	Change	Max Gradient
0	**0**	**4**	1164.2696539	0.83030713	1.809E-7
1	**0**	**5**	1458.4888691	2.00000000	3.532E-6
2	**0**	**5**	1566.8224553	0.73378332	1.11E-6
3	**0**	**3**	1582.1384679	0.11809573	3.806E-6
4	**0**	**2**	1581.1371614	0.01917756	5.203E-7
5	**0**	**1**	1580.9137373	0.00183570	3.586E-7
6	**0**	**1**	1580.8885464	0.00016845	2.422E-7
7	**0**	**0**	1580.8861799	0.00000000	2.642E-6

Convergence criterion (PCONV=1.11022E-8) satisfied.

Fit Statistics	
-2 Res Log Pseudo-Likelihood	1580.89
Generalized Chi-Square	2620.05
Gener. Chi-Square / DF	7.84

Covariance Parameter Estimates

Cov Parm	Subject	Estimate	Standard Error
Variance	**Intercept**	170.59	16133
SP(SPH)	**Intercept**	672.92	63638
Residual (VC)		7.8445	0.6197

Solutions for Fixed Effects

Effect	Estimate	Standard Error	DF	t Value	Pr > |t|	Alpha	Lower	Upper
Intercept	-11.9186	13.0061	27	-0.92	0.3676	0.05	-38.6049	14.7678

Solutions for Fixed Effects								
Effect	**Estimate**	**Standard Error**	**DF**	**t Value**	**Pr > |t|**	**Alpha**	**Lower**	**Upper**
Apr_meanmin_temp	-0.02795	0.07481	307	-0.37	0.7089	0.05	-0.1752	0.1193

Type III Tests of Fixed Effects				
Effect	**Num DF**	**Den DF**	**F Value**	**Pr > F**
Apr_meanmin_temp	1	307	0.14	0.7089

univariate results

The GLIMMIX Procedure

Model Information	
Data Set	WORK.DATASET
Response Variable	cases
Response Distribution	Poisson
Link Function	Log
Variance Function	Default
Offset Variable	log_total
Variance Matrix	Not blocked
Estimation Technique	Residual PL
Degrees of Freedom Method	Containment

Class Level Information

Class	Levels	Values
Health_Unit	28	BRN CHK DUR ELG EOH GBO HAL HAM HDN HKP HPE HUR KFL LAM LGL MSL NIA OTT OXF PDH PEE PTC SMD TOR WAT WDG WEC YRK
Year	12	2002 2003 2004 2005 2006 2007 2008 2009 2010 2011 2012 2013

Number of Observations Read	336
Number of Observations Used	336

Dimensions	
G-side Cov. Parameters	2
R-side Cov. Parameters	1
Columns in X	2
Columns in Z	28
Subjects (Blocks in V)	1
Max Obs per Subject	336

Optimization Information	
Optimization Technique	Dual Quasi-Newton
Parameters in Optimization	2
Lower Boundaries	2

Optimization Information	
Upper Boundaries	0
Fixed Effects	Profiled
Residual Variance	Profiled
Starting From	Data

Iteration History

Iteration	Restarts	Subiterations	Objective Function	Change	Max Gradient
0	**0**	**5**	1214.7987299	0.28151832	2.268E-7
1	**0**	**6**	1491.9639142	0.84637085	4.065E-6
2	**0**	**3**	1573.8002847	0.58858707	0.000015
3	**0**	**4**	1584.8982222	0.22739733	4.228E-7
4	**0**	**3**	1587.365451	0.06604826	8.564E-6
5	**0**	**3**	1587.8238595	0.01729455	9.993E-8
6	**0**	**2**	1587.9168392	0.00439865	1.416E-6
7	**0**	**1**	1587.9385136	0.00084998	0.000081
8	**0**	**1**	1587.9426132	0.00036143	0.000034
9	**0**	**1**	1587.9443377	0.00015339	0.000014
10	**0**	**1**	1587.9450692	0.00006504	6.125E-6
11	**0**	**1**	1587.9453794	0.00002757	2.596E-6
12	**0**	**0**	1587.9455109	0.00000000	4.784E-6

Convergence criterion (PCONV=1.11022E-8) satisfied.

Fit Statistics	
-2 Res Log Pseudo-Likelihood	1587.95
Generalized Chi-Square	2171.52
Gener. Chi-Square / DF	6.50

Covariance Parameter Estimates

Cov Parm	Subject	Estimate	Standard Error
Variance	**Intercept**	26.6162	14511
SP(SPH)	**Intercept**	672.71	366746
Residual (VC)		6.5016	0.5143

Solutions for Fixed Effects								
Effect	**Estimate**	**Standard Error**	**DF**	**t Value**	**Pr > |t|**	**Alpha**	**Lower**	**Upper**
Intercept	-21.0103	5.2593	27	-3.99	0.0004	0.05	-31.8015	-10.2191
Jul_meanmin_temp	0.5367	0.06463	307	8.30	<.0001	0.05	0.4095	0.6639

Type III Tests of Fixed Effects				
Effect	**Num DF**	**Den DF**	**F Value**	**Pr > F**
Jul_meanmin_temp	1	307	68.95	<.0001

univariate results

The GLIMMIX Procedure

Model Information	
Data Set	WORK.DATASET
Response Variable	cases
Response Distribution	Poisson
Link Function	Log
Variance Function	Default
Offset Variable	log_total
Variance Matrix	Not blocked
Estimation Technique	Residual PL
Degrees of Freedom Method	Containment

Class Level Information

Class	Levels	Values
Health_Unit	28	BRN CHK DUR ELG EOH GBO HAL HAM HDN HKP HPE HUR KFL LAM LGL MSL NIA OTT OXF PDH PEE PTC SMD TOR WAT WDG WEC YRK
Year	12	2002 2003 2004 2005 2006 2007 2008 2009 2010 2011 2012 2013

Number of Observations Read	336
Number of Observations Used	336

Dimensions	
G-side Cov. Parameters	2
R-side Cov. Parameters	1
Columns in X	2
Columns in Z	28
Subjects (Blocks in V)	1
Max Obs per Subject	336

Optimization Information	
Optimization Technique	Dual Quasi-Newton
Parameters in Optimization	2
Lower Boundaries	2

Optimization Information	
Upper Boundaries	0
Fixed Effects	Profiled
Residual Variance	Profiled
Starting From	Data

Iteration History

Iteration	Restarts	Subiterations	Objective Function	Change	Max Gradient
0	**0**	**5**	1201.3518739	0.28073186	1.129E-6
1	**0**	**5**	1489.3244393	1.02934109	4.975E-6
2	**0**	**2**	1578.4627156	0.95505218	9.548E-8
3	**0**	**5**	1595.8856408	0.47230942	1.31E-6
4	**0**	**3**	1601.9878067	0.15221625	5.041E-6
5	**0**	**2**	1603.2955994	0.04257195	0.000024
6	**0**	**2**	1603.5744986	0.01154338	4.989E-7
7	**0**	**1**	1603.6431828	0.00312760	3.753E-6
8	**0**	**1**	1603.6612595	0.00085822	2.828E-7
9	**0**	**1**	1603.6661799	0.00023387	5.018E-8
10	**0**	**1**	1603.6675179	0.00006361	5.013E-8
11	**0**	**0**	1603.6678816	0.00000000	7.311E-6

Convergence criterion (PCONV=1.11022E-8) satisfied.

Fit Statistics	
-2 Res Log Pseudo-Likelihood	1603.67
Generalized Chi-Square	2838.44
Gener. Chi-Square / DF	8.50

Covariance Parameter Estimates

Cov Parm	Subject	Estimate	Standard Error
Variance	**Intercept**	11.6291	1635.08
SP(SPH)	**Intercept**	672.73	94578
Residual (VC)		8.4983	0.6674

Solutions for Fixed Effects								
Effect	**Estimate**	**Standard Error**	**DF**	**t Value**	**Pr > |t|**	**Alpha**	**Lower**	**Upper**
Intercept	-18.7534	3.5757	27	-5.24	<.0001	0.05	-26.0902	-11.4166
Aug_meanmin_temp	0.4390	0.06949	307	6.32	<.0001	0.05	0.3022	0.5757

Type III Tests of Fixed Effects				
Effect	**Num DF**	**Den DF**	**F Value**	**Pr > F**
Aug_meanmin_temp	1	307	39.90	<.0001

The GLIMMIX Procedure

Model Information	
Data Set	WORK.DATASET
Response Variable	cases
Response Distribution	Poisson
Link Function	Log
Variance Function	Default
Offset Variable	log_total
Variance Matrix	Not blocked
Estimation Technique	Residual PL
Degrees of Freedom Method	Containment

Class Level Information

Class	Levels	Values
Health_Unit	28	BRN CHK DUR ELG EOH GBO HAL HAM HDN HKP HPE HUR KFL LAM LGL MSL NIA OTT OXF PDH PEE PTC SMD TOR WAT WDG WEC YRK
Year	12	2002 2003 2004 2005 2006 2007 2008 2009 2010 2011 2012 2013

Number of Observations Read	336
Number of Observations Used	336

Dimensions	
G-side Cov. Parameters	2
R-side Cov. Parameters	1
Columns in X	2
Columns in Z	28
Subjects (Blocks in V)	1
Max Obs per Subject	336

Optimization Information	
Optimization Technique	Dual Quasi-Newton
Parameters in Optimization	2
Lower Boundaries	2

Optimization Information	
Upper Boundaries	0
Fixed Effects	Profiled
Residual Variance	Profiled
Starting From	Data

Iteration History

Iteration	Restarts	Subiterations	Objective Function	Change	Max Gradient
0	**0**	**3**	1189.1723664	0.17304185	1.542E-7
1	**0**	**4**	1462.8982009	0.48270709	1.388E-7
2	**0**	**4**	1541.2859254	0.20574994	2.219E-6
3	**0**	**3**	1545.711697	0.03749225	1.437E-7
4	**0**	**1**	1545.3377806	0.00384161	5.458E-6
5	**0**	**1**	1545.2711419	0.00031945	1.42E-7
6	**0**	**0**	1545.265077	0.00000000	9.238E-6

Convergence criterion (PCONV=1.11022E-8) satisfied.

Fit Statistics	
-2 Res Log Pseudo-Likelihood	1545.27
Generalized Chi-Square	2020.92
Gener. Chi-Square / DF	6.05

Covariance Parameter Estimates

Cov Parm	Subject	Estimate	Standard Error
Variance	**Intercept**	145.79	12186
SP(SPH)	**Intercept**	673.46	56287
Residual (VC)		6.0506	0.4777

Solutions for Fixed Effects

Effect	Estimate	Standard Error	DF	t Value	Pr > |t|	Alpha	Lower	Upper
Intercept	-10.9948	12.0237	27	-0.91	0.3686	0.05	-35.6654	13.6759
Feb_mean_prec	-0.02010	0.003756	307	-5.35	<.0001	0.05	-0.02749	-0.01271

Type III Tests of Fixed Effects				
Effect	**Num DF**	**Den DF**	**F Value**	**Pr > F**
Feb_mean_prec	1	307	28.65	<.0001

univariate results

The GLIMMIX Procedure

Model Information	
Data Set	WORK.DATASET
Response Variable	cases
Response Distribution	Poisson
Link Function	Log
Variance Function	Default
Offset Variable	log_total
Variance Matrix	Not blocked
Estimation Technique	Residual PL
Degrees of Freedom Method	Containment

Class Level Information		
Class	**Levels**	**Values**
Health_Unit	28	BRN CHK DUR ELG EOH GBO HAL HAM HDN HKP HPE HUR KFL LAM LGL MSL NIA OTT OXF PDH PEE PTC SMD TOR WAT WDG WEC YRK
Year	12	2002 2003 2004 2005 2006 2007 2008 2009 2010 2011 2012 2013

Number of Observations Read	336
Number of Observations Used	336

Dimensions	
G-side Cov. Parameters	2
R-side Cov. Parameters	1
Columns in X	2
Columns in Z	28
Subjects (Blocks in V)	1
Max Obs per Subject	336

Optimization Information	
Optimization Technique	Dual Quasi-Newton
Parameters in Optimization	2
Lower Boundaries	2

Optimization Information	
Upper Boundaries	0
Fixed Effects	Profiled
Residual Variance	Profiled
Starting From	Data

Iteration History

Iteration	Restarts	Subiterations	Objective Function	Change	Max Gradient
0	**0**	**3**	1208.0874094	0.49048672	2.91E-6
1	**0**	**3**	1495.4365444	0.62082640	6.863E-6
2	**0**	**5**	1589.2486196	0.30776403	8.613E-6
3	**0**	**3**	1600.2681657	0.07403935	2.113E-6
4	**0**	**2**	1600.0879884	0.01067383	6.417E-7
5	**0**	**1**	1599.9322901	0.00130879	6.271E-7
6	**0**	**1**	1599.9097481	0.00015556	1.347E-7
7	**0**	**0**	1599.9070038	0.00000000	7.5E-6

Convergence criterion (PCONV=1.11022E-8) satisfied.

Fit Statistics	
-2 Res Log Pseudo-Likelihood	1599.91
Generalized Chi-Square	2626.47
Gener. Chi-Square / DF	7.86

Covariance Parameter Estimates

Cov Parm	Subject	Estimate	Standard Error
Variance	**Intercept**	145.14	12386
SP(SPH)	**Intercept**	674.15	57527
Residual (VC)		7.8637	0.6197

Solutions for Fixed Effects

Effect	Estimate	Standard Error	DF	t Value	Pr > |t|	Alpha	Lower	Upper
Intercept	-11.2368	11.9988	27	-0.94	0.3573	0.05	-35.8562	13.3827
Mar_mean_prec	-0.01296	0.004211	307	-3.08	0.0023	0.05	-0.02125	-0.00468

Type III Tests of Fixed Effects				
Effect	**Num DF**	**Den DF**	**F Value**	**Pr > F**
Mar_mean_prec	1	307	9.47	0.0023

univariate results

The GLIMMIX Procedure

Model Information	
Data Set	WORK.DATASET
Response Variable	cases
Response Distribution	Poisson
Link Function	Log
Variance Function	Default
Offset Variable	log_total
Variance Matrix	Not blocked
Estimation Technique	Residual PL
Degrees of Freedom Method	Containment

Class Level Information

Class	Levels	Values
Health_Unit	28	BRN CHK DUR ELG EOH GBO HAL HAM HDN HKP HPE HUR KFL LAM LGL MSL NIA OTT OXF PDH PEE PTC SMD TOR WAT WDG WEC YRK
Year	12	2002 2003 2004 2005 2006 2007 2008 2009 2010 2011 2012 2013

Number of Observations Read	336
Number of Observations Used	336

Dimensions	
G-side Cov. Parameters	2
R-side Cov. Parameters	1
Columns in X	2
Columns in Z	28
Subjects (Blocks in V)	1
Max Obs per Subject	336

Optimization Information	
Optimization Technique	Dual Quasi-Newton
Parameters in Optimization	2
Lower Boundaries	2

Optimization Information	
Upper Boundaries	0
Fixed Effects	Profiled
Residual Variance	Profiled
Starting From	Data

Iteration History

Iteration	Restarts	Subiterations	Objective Function	Change	Max Gradient
0	**0**	**3**	1209.2675977	0.05420484	1.112E-7
1	**0**	**5**	1488.801195	0.52633606	1.133E-7
2	**0**	**5**	1568.2726035	0.22751124	3.33E-7
3	**0**	**3**	1573.3879052	0.04561116	2.331E-7
4	**0**	**2**	1572.8785665	0.00520986	1.227E-7
5	**0**	**1**	1572.7641196	0.00049358	2.465E-7
6	**0**	**1**	1572.7521913	0.00004451	1.486E-7
7	**0**	**0**	1572.7511026	0.00000000	1.54E-6

Convergence criterion (PCONV=1.11022E-8) satisfied.

Fit Statistics	
-2 Res Log Pseudo-Likelihood	1572.75
Generalized Chi-Square	2374.54
Gener. Chi-Square / DF	7.11

Covariance Parameter Estimates

Cov Parm	Subject	Estimate	Standard Error
Variance	**Intercept**	198.69	18069
SP(SPH)	**Intercept**	674.51	61338
Residual (VC)		7.1094	0.5611

Solutions for Fixed Effects

Effect	Estimate	Standard Error	DF	t Value	Pr > |t|	Alpha	Lower	Upper
Intercept	-9.5037	14.0527	27	-0.68	0.5046	0.05	-38.3375	19.3301

Solutions for Fixed Effects								
Effect	**Estimate**	**Standard Error**	**DF**	**t Value**	**Pr > |t|**	**Alpha**	**Lower**	**Upper**
Apr_meamax_temp	-0.1917	0.05593	307	-3.43	0.0007	0.05	-0.3017	-0.08161

Type III Tests of Fixed Effects				
Effect	**Num DF**	**Den DF**	**F Value**	**Pr > F**
Apr_meamax_temp	1	307	11.74	0.0007

univariate results

The GLIMMIX Procedure

Model Information	
Data Set	WORK.DATASET
Response Variable	cases
Response Distribution	Poisson
Link Function	Log
Variance Function	Default
Offset Variable	log_total
Variance Matrix	Not blocked
Estimation Technique	Residual PL
Degrees of Freedom Method	Containment

Class Level Information

Class	Levels	Values
Health_Unit	28	BRN CHK DUR ELG EOH GBO HAL HAM HDN HKP HPE HUR KFL LAM LGL MSL NIA OTT OXF PDH PEE PTC SMD TOR WAT WDG WEC YRK
Year	12	2002 2003 2004 2005 2006 2007 2008 2009 2010 2011 2012 2013

Number of Observations Read	336
Number of Observations Used	336

Dimensions	
G-side Cov. Parameters	2
R-side Cov. Parameters	1
Columns in X	2
Columns in Z	28
Subjects (Blocks in V)	1
Max Obs per Subject	336

Optimization Information	
Optimization Technique	Dual Quasi-Newton
Parameters in Optimization	2
Lower Boundaries	2

Optimization Information	
Upper Boundaries	0
Fixed Effects	Profiled
Residual Variance	Profiled
Starting From	Data

Iteration History

Iteration	Restarts	Subiterations	Objective Function	Change	Max Gradient
0	**0**	**4**	1199.5620604	0.12852085	4.192E-7
1	**0**	**4**	1475.3705583	0.52100853	2.476E-6
2	**0**	**4**	1547.9843182	0.20270081	8.271E-6
3	**0**	**2**	1551.257948	0.03016917	0.000011
4	**0**	**1**	1551.1412073	0.00303581	2.627E-6
5	**0**	**1**	1551.1126633	0.00026332	7.402E-8
6	**0**	**0**	1551.1099408	0.00000000	7.652E-6

Convergence criterion (PCONV=1.11022E-8) satisfied.

Fit Statistics	
-2 Res Log Pseudo-Likelihood	1551.11
Generalized Chi-Square	1674.24
Gener. Chi-Square / DF	5.01

Covariance Parameter Estimates

Cov Parm	Subject	Estimate	Standard Error
Variance	**Intercept**	75.7775	9186.96
SP(SPH)	**Intercept**	672.64	81546
Residual (VC)		5.0127	0.3962

Solutions for Fixed Effects

Effect	Estimate	Standard Error	DF	t Value	Pr > |t|	Alpha	Lower	Upper
Intercept	-32.9968	9.0102	27	-3.66	0.0011	0.05	-51.4843	-14.5093
Aug_meamax_temp	0.7862	0.09082	307	8.66	<.0001	0.05	0.6075	0.9650

Type III Tests of Fixed Effects				
Effect	**Num DF**	**Den DF**	**F Value**	**Pr > F**
Aug_meamax_temp	1	307	74.94	<.0001

univariate results

The GLIMMIX Procedure

Model Information	
Data Set	WORK.DATASET
Response Variable	cases
Response Distribution	Poisson
Link Function	Log
Variance Function	Default
Offset Variable	log_total
Variance Matrix	Not blocked
Estimation Technique	Residual PL
Degrees of Freedom Method	Containment

Class Level Information

Class	Levels	Values
Health_Unit	28	BRN CHK DUR ELG EOH GBO HAL HAM HDN HKP HPE HUR KFL LAM LGL MSL NIA OTT OXF PDH PEE PTC SMD TOR WAT WDG WEC YRK
Year	12	2002 2003 2004 2005 2006 2007 2008 2009 2010 2011 2012 2013

Number of Observations Read	336
Number of Observations Used	329

Dimensions	
G-side Cov. Parameters	2
R-side Cov. Parameters	1
Columns in X	2
Columns in Z	28
Subjects (Blocks in V)	1
Max Obs per Subject	329

Optimization Information	
Optimization Technique	Dual Quasi-Newton
Parameters in Optimization	2
Lower Boundaries	2

Optimization Information	
Upper Boundaries	0
Fixed Effects	Profiled
Residual Variance	Profiled
Starting From	Data

Iteration History

Iteration	Restarts	Subiterations	Objective Function	Change	Max Gradient
0	**0**	**2**	1064.9295266	0.12394645	5.701E-7
1	**0**	**6**	1288.0439405	0.33858388	1.632E-7
2	**0**	**4**	1339.2535077	0.08702958	1.738E-7
3	**0**	**2**	1341.2246246	0.00636896	2.565E-6
4	**0**	**1**	1341.2057643	0.00018409	9.033E-6
5	**0**	**1**	1341.2045738	0.00003520	1.719E-6
6	**0**	**0**	1341.2043254	0.00000000	1.994E-6

Convergence criterion (PCONV=1.11022E-8) satisfied.

Fit Statistics	
-2 Res Log Pseudo-Likelihood	1341.20
Generalized Chi-Square	1002.03
Gener. Chi-Square / DF	3.06

Covariance Parameter Estimates

Cov Parm	Subject	Estimate	Standard Error
Variance	**Intercept**	91.3318	7097.77
SP(SPH)	**Intercept**	658.56	51179
Residual (VC)		3.0643	0.2454

Solutions for Fixed Effects

Effect	Estimate	Standard Error	DF	t Value	Pr > |t|	Alpha	Lower	Upper
Intercept	-12.7164	9.5149	27	-1.34	0.1926	0.05	-32.2393	6.8066
percent_positive_poo	0.2479	0.01303	300	19.02	<.0001	0.05	0.2222	0.2735

Type III Tests of Fixed Effects				
Effect	**Num DF**	**Den DF**	**F Value**	**Pr > F**
percent_positive_poo	1	300	361.64	<.0001

<u>Appendix 4</u>: Full results from SaTScan vs. Moran's I yearly cluster detection

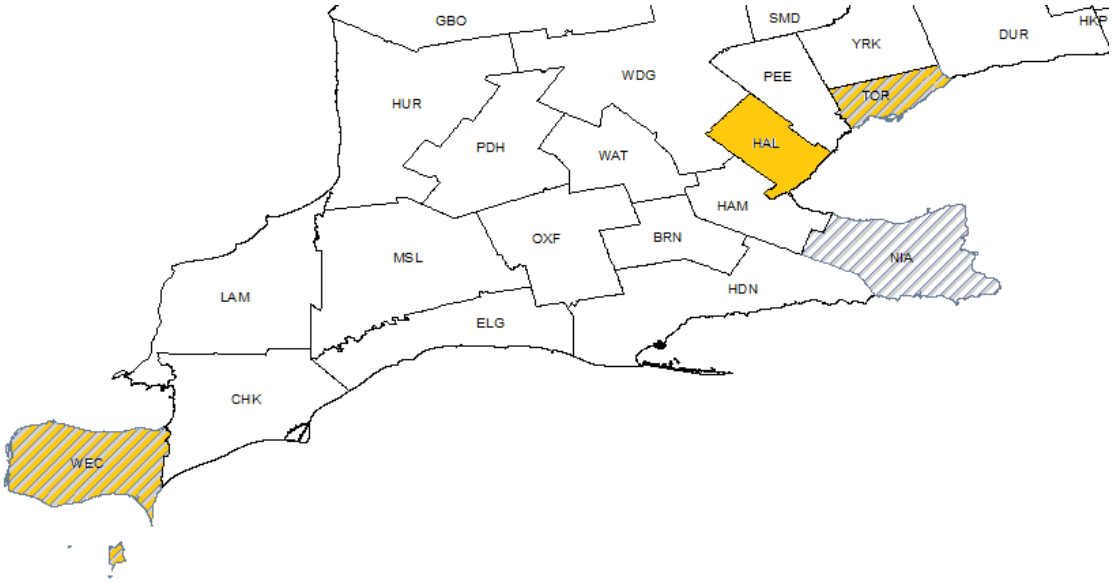

Figure 1: Clusters identified in Southern Ontario for 2002. Yellow indicates clusterd found by SaTScan analysis, hashed pattern indicates clusters found by Moran's I analysis.

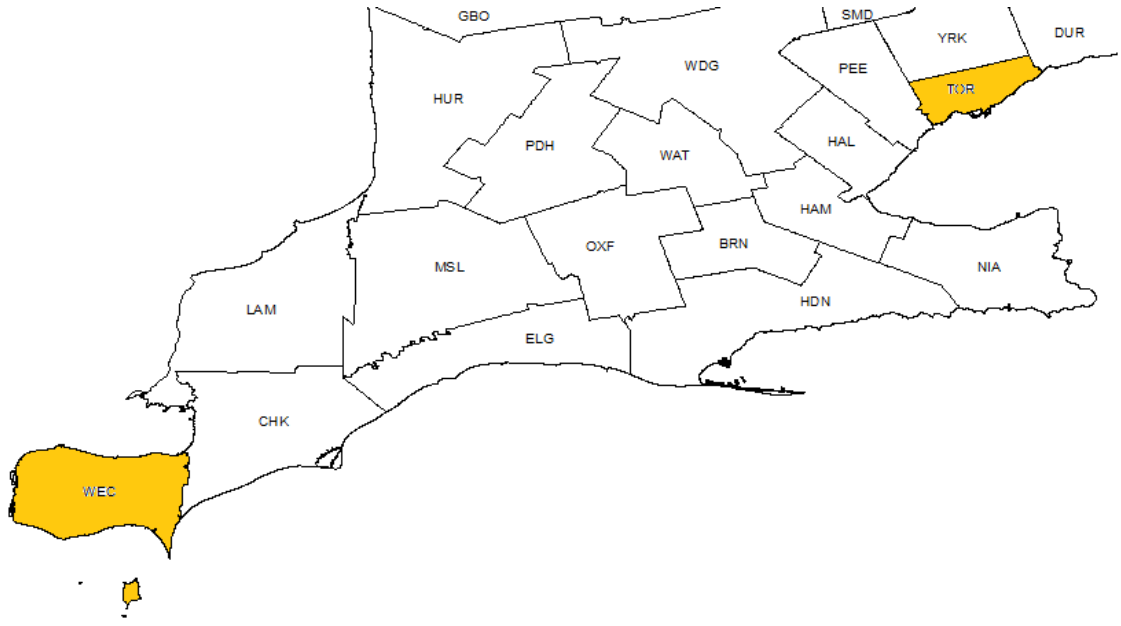

Figure 2: Clusters identified in Southern Ontario for 2003. Yellow indicates clusterd found by SaTScan analysis, hashed pattern indicates clusters found by Moran's I analysis.

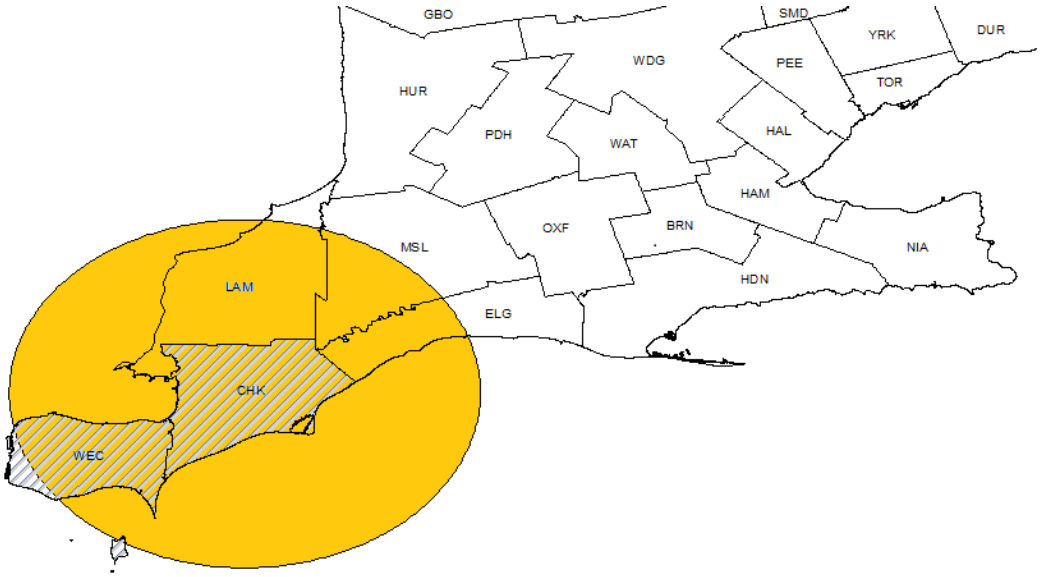

Figure 3: Clusters identified in Southern Ontario for 2004. Yellow indicates clusterd found by SaTScan analysis, hashed pattern indicates clusters found by Moran's I analysis.

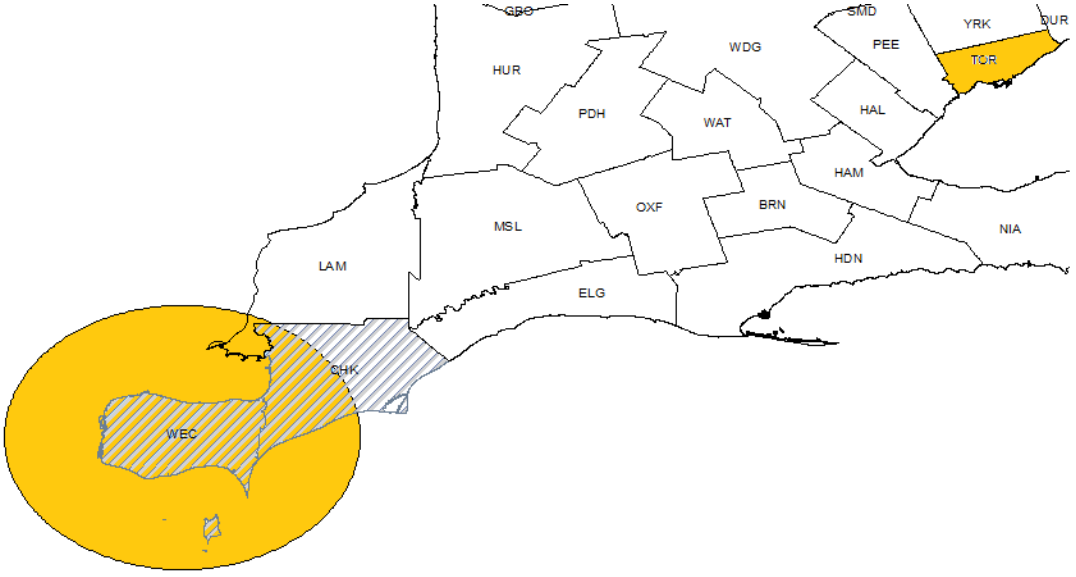

Figure 4: Clusters identified in Southern Ontario for 2005. Yellow indicates clusterd found by SaTScan analysis, hashed pattern indicates clusters found by Moran's I analysis.

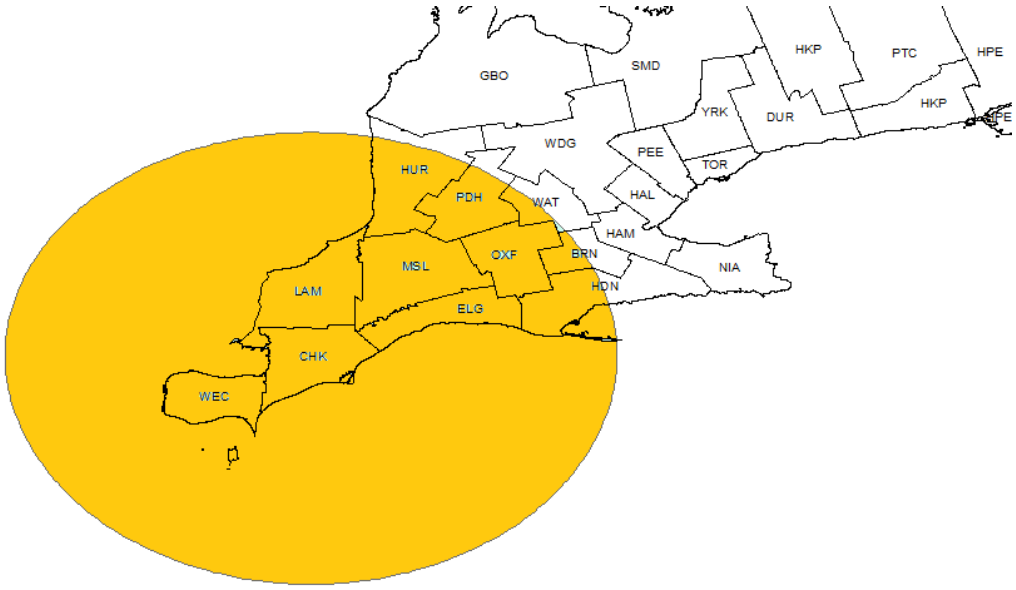

Figure 5: Clusters identified in Southern Ontario for 2006. Yellow indicates clusterd found by SaTScan analysis, hashed pattern indicates clusters found by Moran's I analysis.

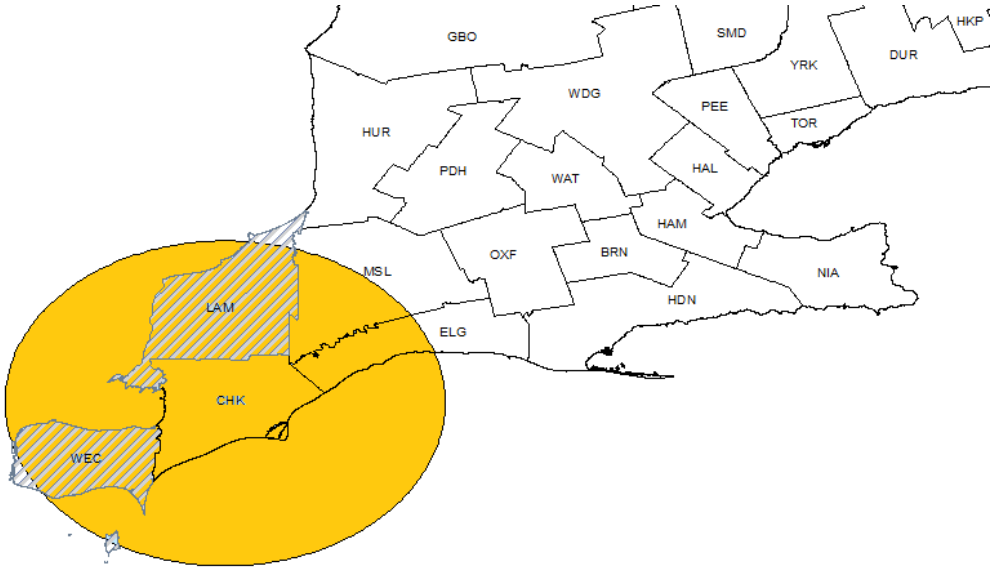

Figure 6: Clusters identified in Southern Ontario for 2009. Yellow indicates clusterd found by SaTScan analysis, hashed pattern indicates clusters found by Moran's I analysis.

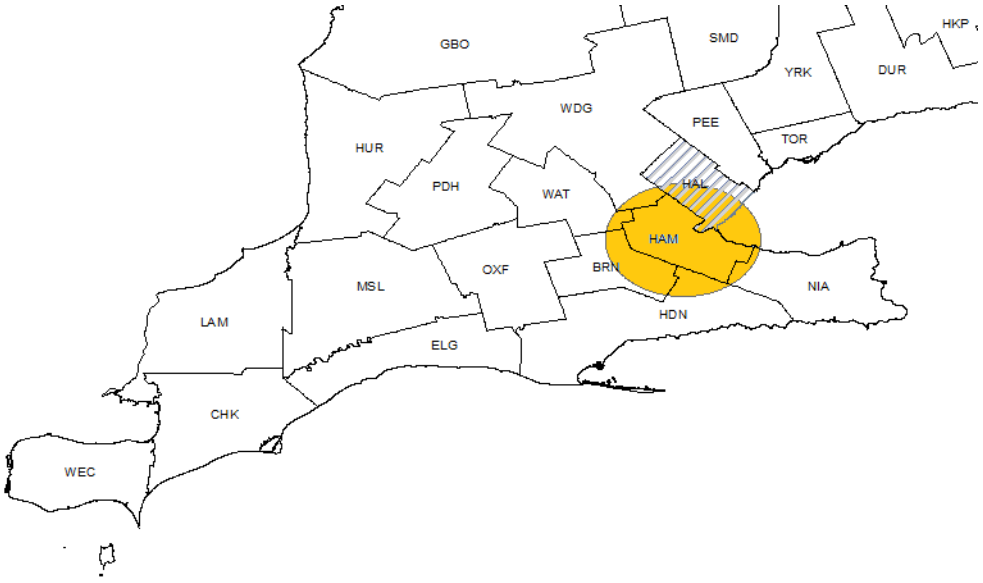

Figure 7: Clusters identified in Southern Ontario for 2010. Yellow indicates clusterd found by SaTScan analysis, hashed pattern indicates clusters found by Moran's I analysis.

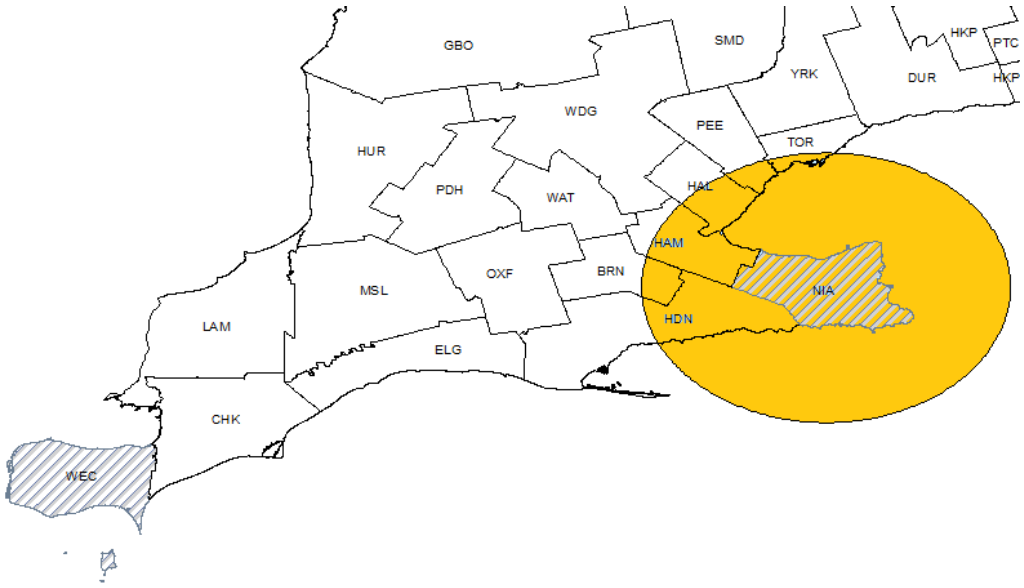

Figure 8: Clusters identified in Southern Ontario for 2011. Yellow indicates clusterd found by SaTScan analysis, hashed pattern indicates clusters found by Moran's I analysis.

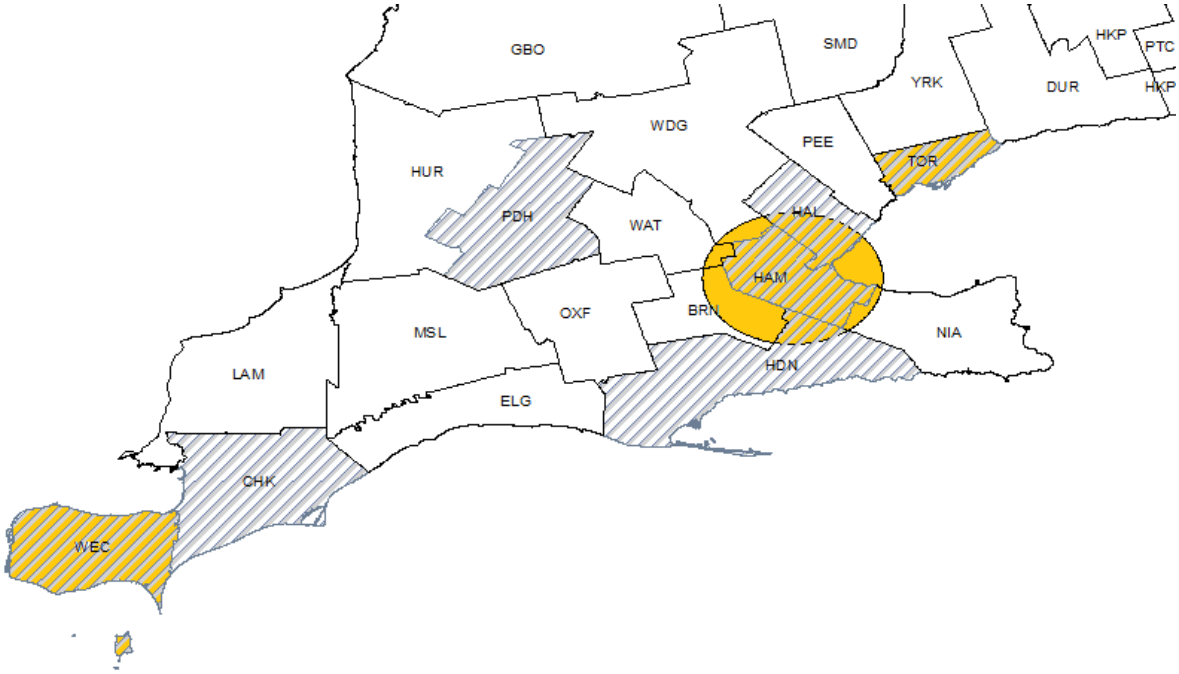

Figure 9: Clusters identified in Southern Ontario for 2012. Yellow indicates clusterd found by SaTScan analysis, hashed pattern indicates clusters found by Moran's I analysis.

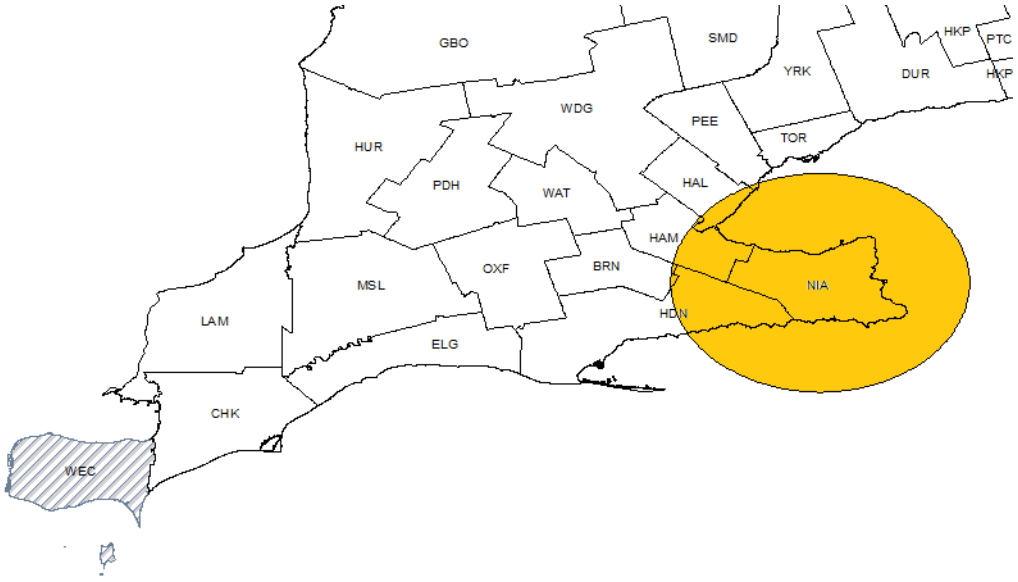

Figure 10: Clusters identified in Southern Ontario for 2013. Yellow indicates clusterd found by SaTScan analysis, hashed pattern indicates clusters found by Moran's I analysis.

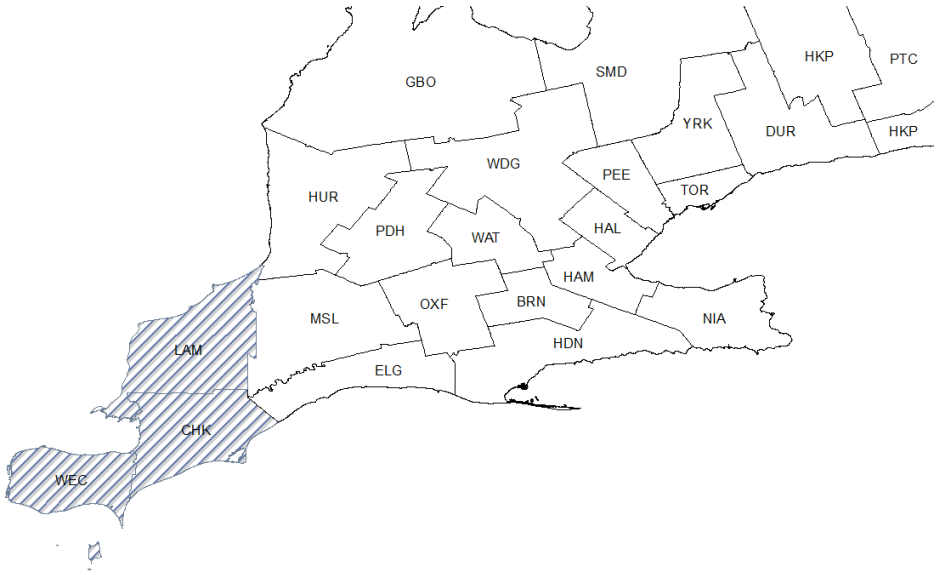

Figure 11: Clusters identified in Southern Ontario for 2014. Yellow indicates clusterd found by SaTScan analysis, hashed pattern indicates clusters found by Moran's I analysis.

Appendix 5: Full results from Yearly Mosquito SaTScan

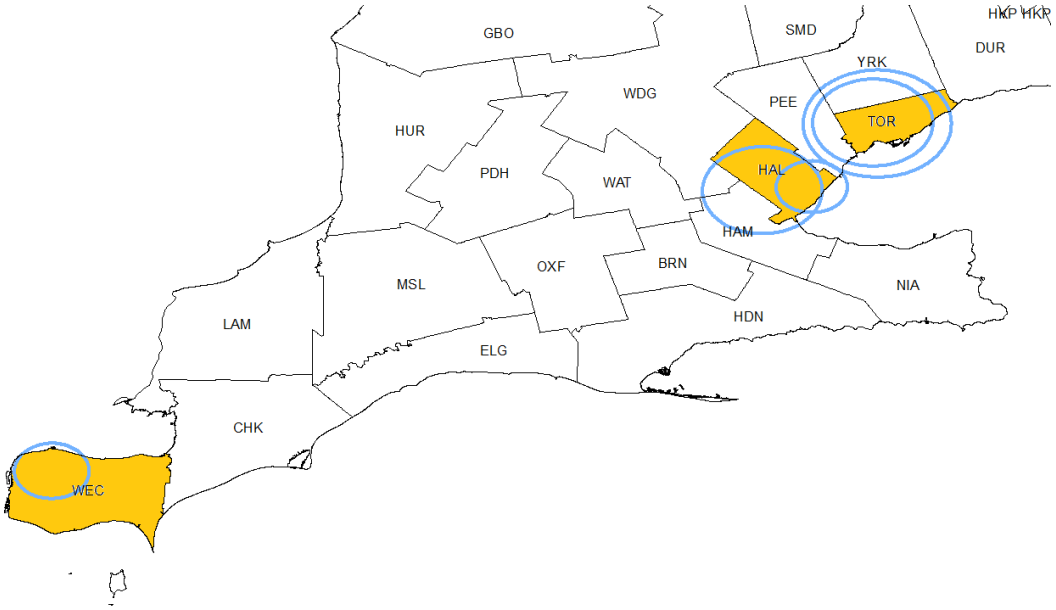

Figure 1: Mosquito positive pool clusters in Southern Ontario in 2002 as determined by SaTScan analysis. Mosquito clusters are given in blue and human clusters in yellow.

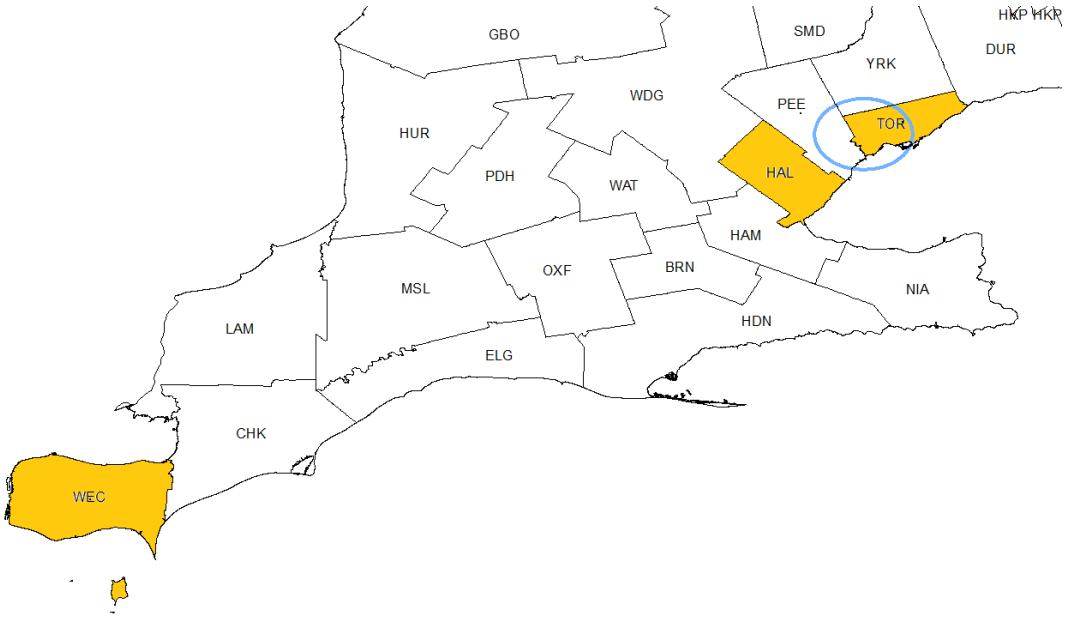

Figure 2: Mosquito positive pool clusters in Southern Ontario in 2003 as determined by SaTScan analysis. Mosquito clusters are given in blue and human clusters in yellow.

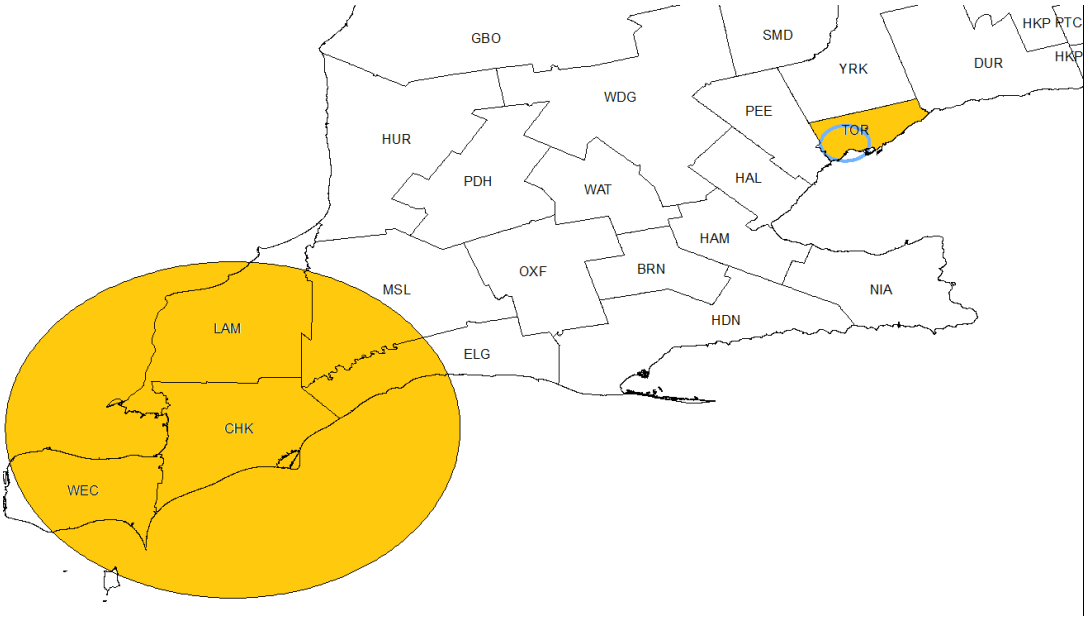

Figure 3: Mosquito positive pool clusters in Southern Ontario in 2004 as determined by SaTScan analysis. Mosquito clusters are given in blue and human clusters in yellow.

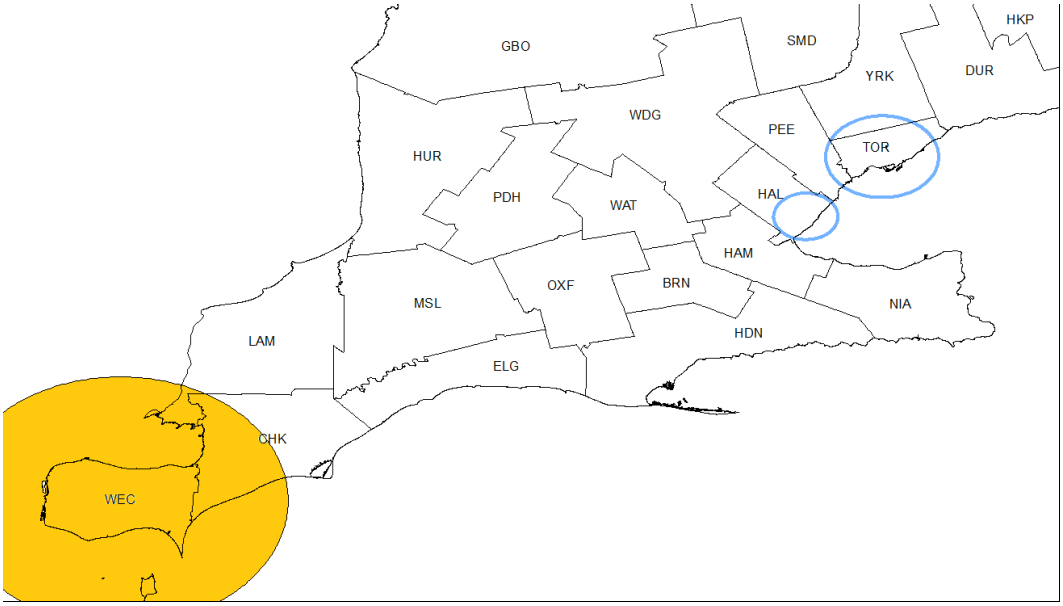

Figure 4: Mosquito positive pool clusters in Southern Ontario in 2005 as determined by SaTScan analysis. Mosquito clusters are given in blue and human clusters in yellow.

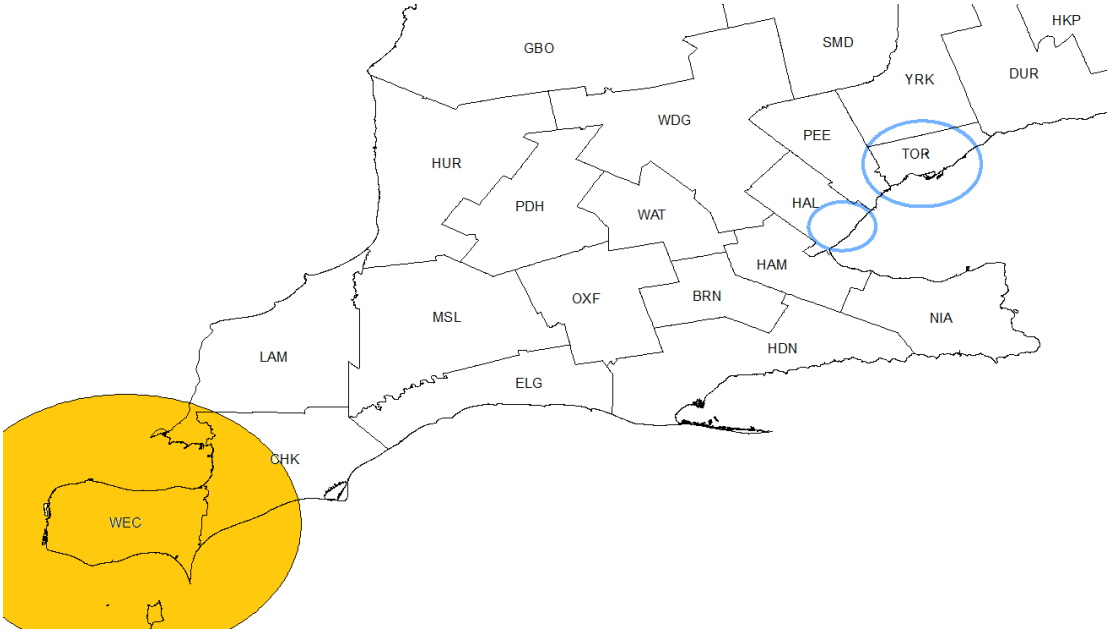

Figure 5: Mosquito positive pool clusters in Southern Ontario in 2006 as determined by SaTScan analysis. Mosquito clusters are given in blue and human clusters in yellow.

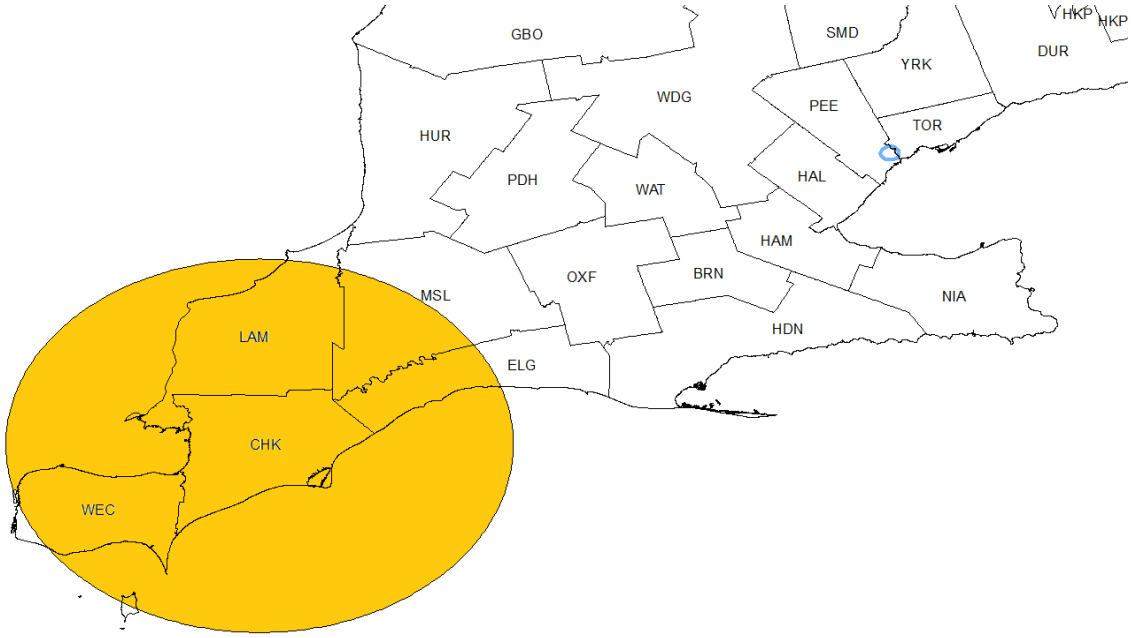

Figure 6: Mosquito positive pool clusters in Southern Ontario in 2009 as determined by SaTScan analysis. Mosquito clusters are given in blue and human clusters in yellow.

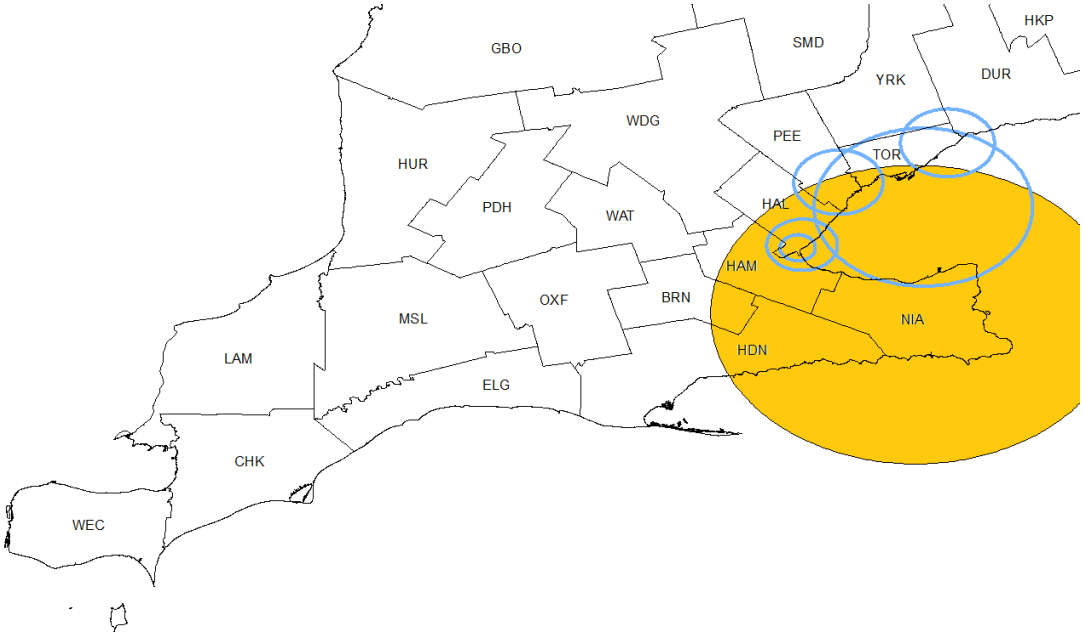

Figure 7: Mosquito positive pool clusters in Southern Ontario in 2011 as determined by SaTScan analysis. Mosquito clusters are given in blue and human clusters in yellow.

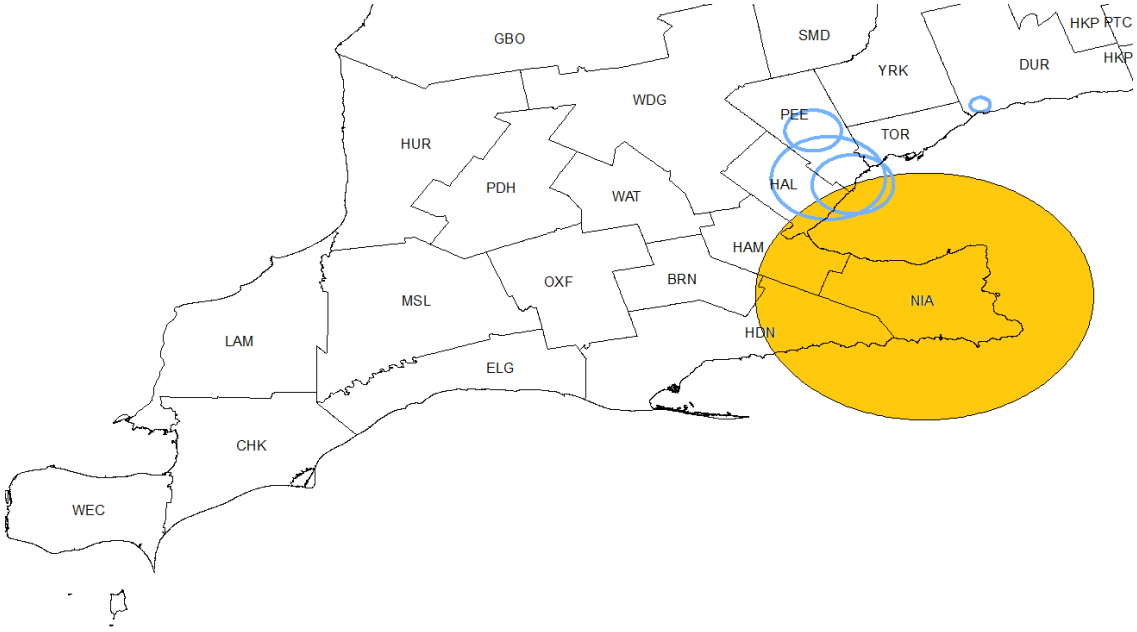

Figure 8: Mosquito positive pool clusters in Southern Ontario in 2013 as determined by SaTScan analysis. Mosquito clusters are given in blue and human clusters in yellow.

Appendix 6: Mosquito testing frequency in cluster-prone health units across time

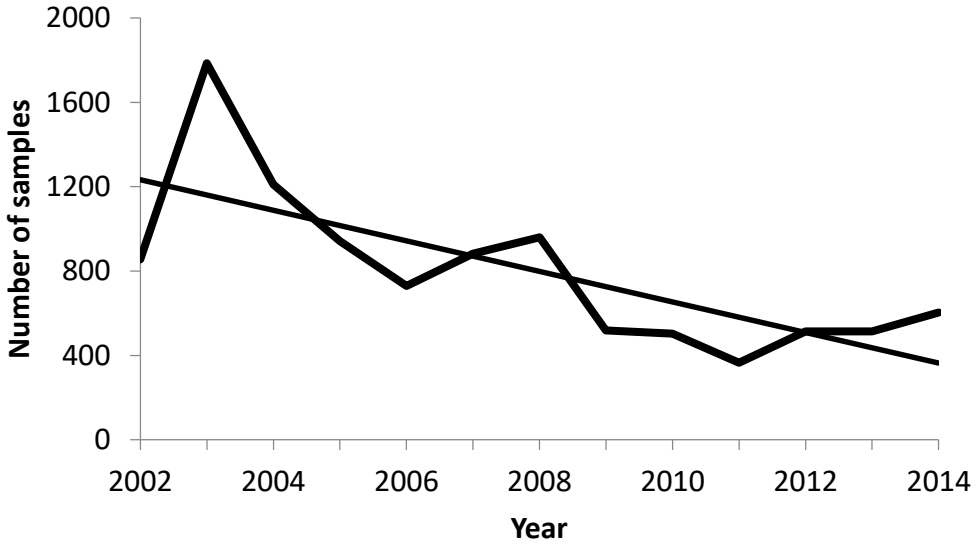

Figure 1: Number of samples (mosquito pools) collected by year in Halton PHU.

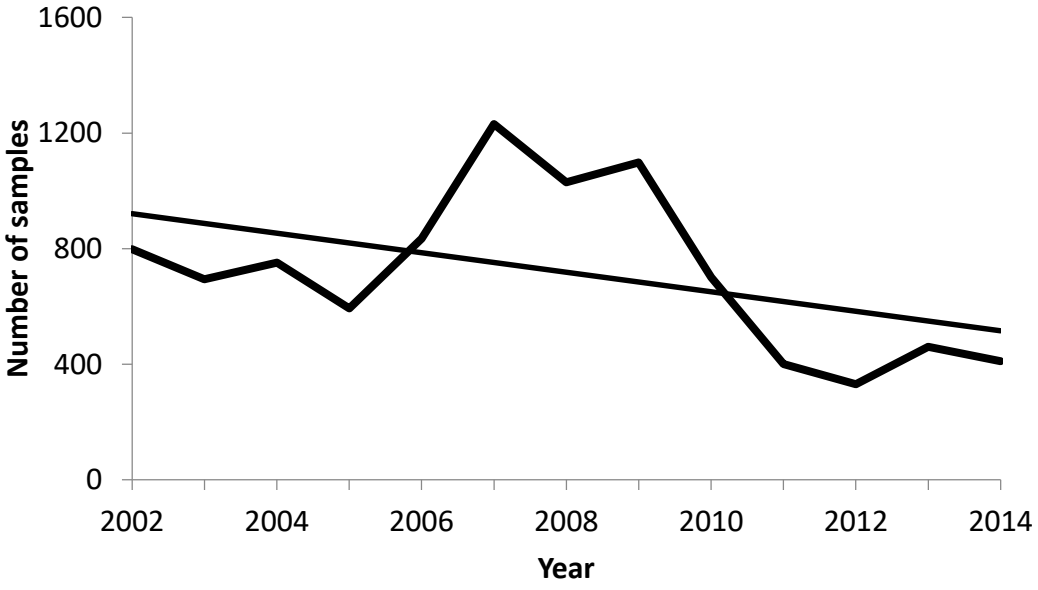

Figure 2: Number of samples (mosquito pools) collected by year in Windsor-Essex County PHU.

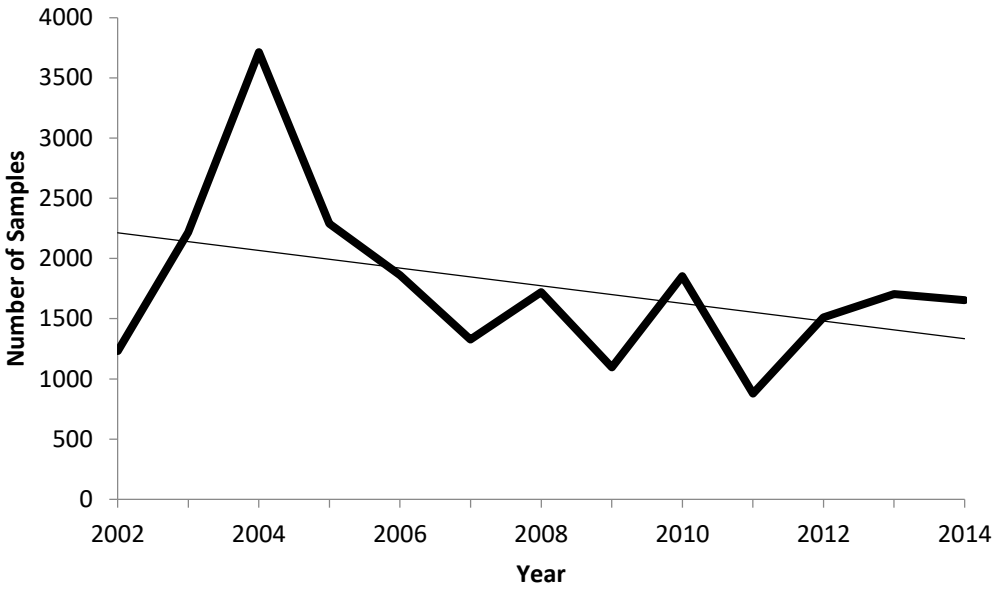

Figure 3: Number of samples (mosquito pools) collected by year in Toronto PHU.