2021

The Power of Exogenous Variables in Predicting West Nile Virus in South Carolina

Christopher Lee Glaze
Walden University

Walden University

College of Management and Technology

This is to certify that the doctoral dissertation by

Christopher L. Glaze

has been found to be complete and satisfactory in all respects,
and that any and all revisions required by
the review committee have been made.

Review Committee
Dr. Branford McAllister, Committee Chairperson, Management Faculty
Dr. Robert Kilmer, Committee Member, Management Faculty
Dr. Robert Haussmann, University Reviewer, Management Faculty

Chief Academic Officer and Provost
Sue Subocz, Ph.D.

Walden University
2021

Abstract

The Power of Exogenous Variables in Predicting West Nile Virus in South Carolina

by

Christopher L. Glaze

MA, Naval War College, 1998

MPA, Troy State University, 1987

BS, The Citadel, 1979

Dissertation Submitted in Partial Fulfillment

of the Requirements for the Degree of

Doctor of Philosophy

Management

Walden University

October 2021

Abstract

Despite the availability of medical data, environmental surveillance tools, and heightened public awareness, West Nile Virus (WNv) remains a global health hazard. Reliable methods for predicting WNv outbreaks remain elusive, and environmental health managers must take preventive actions without the benefit of simple predictive tools. The purpose of this ex post facto research was to examine the accuracy and timeliness of exogenous data in predicting outbreaks of WNv in South Carolina. Decision theory, the CYNEFIN construct, and systems theory provided the theoretical framework for this study, allowing the researcher to broaden traditional decision theory concepts with powerful system-level precepts. Using WNv as an example of decision making in complex environments, a statistical model for predicting the likelihood of the presence of WNv was developed through the exclusive use of exogenous explanatory variables (EEVs). The key research questions were focused on whether EEVs alone can predict the likelihood of WNv presence with the statistical confidence to make timely preventive resource decisions. Results indicated that publicly accessible EEVs such as average temperature, average wind speed, and average population can be used to predict the presence of WNv in a South Carolina locality 30 days prior to an incident, although they did not accurately predict incident counts higher than four. The social implications of this research can be far-reaching. The ability to predict emerging infectious diseases (EID) for which there are no vaccines or cure can provide decision makers with the ability to take pro-active measures to mitigate EID outbreaks.

The Power of Exogenous Variables in Predicting West Nile Virus in South Carolina

by

Christopher L. Glaze

MA, Naval War College, 1998

MPA, Troy State University, 1987

BS, The Citadel, 1979

Dissertation Submitted in Partial Fulfillment

of the Requirements for the Degree of

Doctor of Philosophy

Management

Walden University

October 2021

Dedication

I dedicate this work to Rebecca, my wife of 42 years. Without her constant love, companionship, and support, I would not have completed this effort. Having her own brush with West Nile Virus in 2010, her experience became the genesis and driving factor behind this study.

Acknowledgments

The mechanics of this research would never have occurred without the persistent guidance and help of my Committee Chair and Committee Member, Dr. Branford McAllister and Dr. Robert Kilmer. I am grateful for their mentorship and professionalism.

Table of Contents

List of Tables

List of Figures

Chapter 1: Introduction to the Study

Despite the availability of serological sampling, environmental surveillance tools, and heightened public awareness, consistently reliable methods for predicting West Nile virus (WNv) outbreaks remain elusive (Manore et al., 2014). Current predictive models of local WNv outbreaks are reliant on robust epidemiological (EPS) and environmental surveillance programs (EVS) that produce actionable data. Hadler et al. (2015) showed that environmental health programs across the United States are executed with differing levels of resources and funding. With emerging infectious diseases (EIDs) like WNv and Zika, a simple, reliable predictive tool is required to ensure public health measures can be taken before an outbreak occurs. In this research, I examined the accuracy and timeliness of using web-based, publicly accessible ecological and environmental data in predicting outbreaks of WNv in South Carolina.

This chapter contains the background, problem statement, and purpose of the research; sets the theoretical structure for the study; and defines the terms of reference. It concludes with a series of assumptions and highlights the significance of the study to the field of management.

Background of the Study

First detected in Africa in 1937, the initial cases of WNv in the United States occurred in New York City in the summer of 1999. Today, WNv has been diagnosed in all 48 states within the contiguous United States, with two nationwide epidemics occurring in 2003 and 2012 (Centers for Disease Control and Prevention, 2013; Kwan et al., 2012). At the end of 2016, the U.S. Department of Health & Human Services, Centers for Disease Control and Prevention (CDC) reported 46,086 cases of WNv, with 2,017 human deaths over the 1999 to 2016 timeframe. Although it has become the “principal cause of viral encephalitis in the United States” (Austin & Dowd, 2014, p. 1015), no vaccine or specific therapy for WNv currently exists (CDC, 2013; Gubler, 2007).

Due to a universal lack of a vaccine and approved therapy, public officials around the world have been actively seeking tools that will predict human outbreaks of WNv and aid decision making in the timely application of preventive measures when transmission cycles are high. Predictive modeling of WNv is an important decision support tool in this effort, but it remains problematic due to the dynamic temporal and spatial interdependencies of the pathogenic, ecological, and anthropological components of the virus (Pirofski & Casadevall, 2012). These interdependencies present a complex decision-space for environmental health managers (EHMs).

The presence of WNv in a locality is dependent on numerous interactive biological, environmental, and ecological factors. To quantify WNv risks, researchers and EHMs have developed EPS and EVS tools to monitor arbovirus infections in humans, to understand WNv mosquito transmission activity (vector control), and to execute preventive measures (CDC, 2013).

EPS involves testing humans for the presence of WNv in blood or cerebrospinal fluid, detecting anti-WNv immunoglobulin antibodies or through nucleic acid amplification testing. These types of surveillance data are conclusive as to the presence of

WNv and are critical to understanding the extent of human incidents of WNv in a locality. However, utilizing EPS data by itself is usually insufficient for predicting outbreaks as the data can lag several weeks behind the actual infection timeframe (CDC, 2013, p. 11). This deficiency in EPS data has led to the development of EVS activities to gather additional explanatory factors and to strengthen the indices supporting predictive capabilities.

EVS monitors WNv transmission in mosquitoes, birds, equines, chickens, and other mammals. Using several surveillance activities, such as mosquito pools, sentinel animals, and birds, EHMs gather additional explanatory data to compare against historical EVS and EPS data to better understand and quantify the antecedent conditions necessary to address WNv. Using a combination of EVS and EPS data, these types of mixed predictive models provide detection timeframes of 2 to 4 weeks prior to the onset of human symptoms. In a locality that has a robust EVS and EPS program, integrated risk management (IRM) decisions can occur within a sufficient time-period to take preventive actions. However, according to Hadler et al. (2015), “arboviral surveillance is inadequate in many states to rapidly detect and control outbreaks and to give the public critical information it needs for prevention” (p. 1165).

Following the outbreak of WNv in the United States in 1999, the federal government implemented programs and funding to support state or local arbovirus surveillance programs. The National Association of County & City Health Officials (NACCHO, 2014) reported that by 2004, approximately $45M in federal funding supported these surveillance programs. By 2010, reports began to surface of funding declines in these programs in the states of California, South Carolina, Florida, Nevada, Wyoming, and Oregon (DeLong, 2010). In 2012, federal funding for arbovirus surveillance programs had declined by 61% or $17.5M (NACCHO, 2014). This decrease sparked a shift in the capacity for IRM and required local health officials to prioritize mosquito surveillance at the expense of other EVS (equine, avian, sentinel animals) methods (Hadler et al., 2015). By 2012, Hadler et al. (2015) reported that “57% of states reported eliminating avian death surveillance, 58% decreased mosquito testing, and 46% decreased the number of human specimens tested for WNv” (p. 1161).

Although local EHMs reprioritized limited funding, the downward trend in operational capability continues. In 2016, NACCHO found that 84% of national mosquito vector control programs needed improvement due to a failure of one or more vector control program core competencies (routine mosquito surveillance, treatment decisions using surveillance data, larviciding, adulticing, or both, routine vector control activities, pesticide resistance training (NACCHO, 2017). This decrease in capability directly affects current WNv predictive tools such as California Mosquito-Virus Risk Assessment and Dynamic Continuous-Area Space-Time systems (Kwan et al., 2012). These predictive modeling tools rely on robust surveillance programs with regular sampling/reporting to provide EHMs with indices that can accurately support WNv outbreak prediction in an accurate and timely manner.

In the absence of these robust, data rich WNv surveillance programs, there is evidence that other explanatory environmental factors can provide interactive context to the presence of WNv in a locality. Explanatory factors such as meteorological data, topology, land use, and population density have been used in past studies to enrich predictive models when univariate analysis has shown some correlation to incidents of WNv (Ahmadnejad et al., 2016; Soverow et al., 2009). Most of these factors, referred to as exogenous explanatory variables (EEVs) in this research, are collected in web-based, publicly accessible data bases allowing real-time access to historical and trending sets of data. When viewed from a management science perspective, the use of these exogenous factors reflects a more holistic, systems-level approach to decision making that may provide statistically significant predictability using contextual data in the absence of primary EPS and EVS surveillance data.

In combination with EPS and EVS factors, several studies have proven these types of exogenous factors to be important to understanding WNv-vector-host interaction, virus maturity, and vector abundance. However, there are no studies that have addressed the power of these types of explanatory factors in predicting WNv when robust data from EPS and EVS are not available.

Problem Statement

Despite the availability of medical data, environmental surveillance tool sets, and heightened public awareness, reliable methods for predicting WNv outbreaks remain elusive (Manore et al., 2014). According to Manore et al. (2014) and Chevalier et al.

(2014), statistical predictions of local WNv outbreaks with the reliability and timeliness required for EHMs to make pro-active public health decisions is increasingly problematic. The existing research has identified shortfalls in current EID predictive models due to a lack of robust empirical data, longitudinal analysis, and research exploring the exclusive use of EEVs (see, for example, Liu et al., 2009; Manore et al., 2014; Rochlin et al., 2011).

The research problem was that no current mechanism, model, or algorithm exists for the accurate and timely prediction of WNv outbreaks without robust EPS and EVS data (Chevalier et al., 2014; Manore et al., 2014). Specifically, there was a lack of understanding of the predictive power of EEVs when used by themselves or in combination when robust EPS and EVS data are unavailable.

Purpose of the Study

The purpose of this ex post facto quantitative, correlational research was to examine the use of EEV data in predicting outbreaks of WNv in SC when robust EPS and EVS data are unavailable. To address the scholarly gap of accurate and timely predictive modeling of WNv, I examined 10 EEVs listed in Table 1. These 10 EEVs were proposed based on a systems-level review of the WNv decision-space within the literature review and are readily available from publicly accessible data sets.

Table 1

Study Variables

Variable type	Description	Label	Scale	Measure
Dependent	Presence of WNv	$DV_{PRESENCE}$	Categorical	Presence of WNv in a time-lagged period 0 (no) 1 (yes)
Dependent	Number of WNv incidents	DV_{COUNT}	Discrete	Count of WNv incidents in a time-lagged period
Independent	Average temperature	*EV1ATM*	Ratio	Average 30-day county temperature in degrees Fahrenheit (°F)
Independent	Average rainfall	*EV2ARN*	Ratio	Average 30-day county rainfall in (inches)
Independent	Average dewpoint	*EV3ADP*	Ratio	Average 30-day county dew point in degrees Fahrenheit (°F)
Independent	Average snow depth	*EV4ASD*	Ratio	Average 30-day county snow depth in (inches)
Independent	Average barometric pressure	*EV5ABP*	Ratio	Average 30-day county barometric pressure in inches of mercury (HG)
Independent	Average wind speed	*EV6AWS*	Ratio	Average 30-day county wind speed in miles per hour (MPH)
Independent	Topology	*EV7ELV*	Interval	County seat elevation in (feet)
Independent	Land use	*EV8USE*	Categorical	0 (agricultural use) 1 (industrial use)
Independent	Urbanization	*EV9POP*	Ratio	Population density per county square miles
Independent	Dew point deficit	*EV10ADD*	Interval	The difference between *EV1ATM* and *EV3ADP*

The dependent variables (DVs) were various measures of the presence of WNv and were suitable for analysis using Statistical Package for the Social Sciences (SPSS) binary logistic regression (BLR) and generalized linear model (GZLM) regression. There were five DVs, with varying degrees of time lag (0, 30, 60, 90-days; and a 90-day moving average). These varying degrees of time lag directly addressed EEV predictive capability in the research questions. The issue of time lag is addressed further in Chapter 2.

Research Questions and Hypotheses

The following research questions examined the utility of EEVs in predicting outbreaks of WNv. Research question two was developed with multiple hypothesis pairs. In this research, a WNv incident was the positive identification of the virus in a locality in either a human, mosquito, bird, equine, or sentinel animal.

Research Question 1 (RQ1): In the absence of robust EPS and EVS data, which EEVs are predictors of incidents of WNv in South Carolina (SC) in a current month?

$H_0 1$: When used alone or in combination, EEVs do not accurately predict incidents of WNv in SC in the same month.

$H_a 1$: At least one EEV accurately predicts incidents of WNv in SC in the same month.

Research Question 2 (RQ2): In the absence of robust EPS and EVS data, which EEVs accurately predict incidents of WNv in SC in the future?

$H_0$2: When used alone or in combination, EEVs do not predict incidents of WNv in SC 30 days later.

H_a2: At least one EEV predicts incidents of WNv in SC 30 days later.

$H_0$3: When used alone or in combination, EEVs do not predict incidents of WNv in SC 60 days later.

H_a3: At least one EEV predicts incidents of WNv in SC 60 days later.

$H_0$4: When used alone or in combination, EEVs do not predict incidents of WNv in SC 90 days later.

H_a4: At least one EEV predicts incidents of WNv in SC 90 days later.

In this research, I examined 10 EEVs and a numerical DV indicating the presence of WNv (see Table 1). I used BLR and GZLM regression to test the hypotheses, then developed and validated predictive models of the effects and associations of the EEVs with the relevant DVs.

Theoretical Foundation

This research leveraged Simon's (1960) decision-making process, decision theory (DT), and Snowden and Boone's (2007) contextual decision-making framework called CYNEFIN. These theoretical constructs aligned with this research because EHMs are facing complex decision-spaces associated with WNv and require a broader, systems-level approach to explanatory factor selection.

The epidemiological cycle associated with WNv incubation, transport, and transmission reflect components of both deterministic biological processes and stochastic ecological/environmental conditions. The theoretical concepts within DT and CYNEFIN allow the manager to frame WNv decision-space complexity within Simon's (1960) iterative three-phase decision process, allowing for a more holistic systems approach to EID prediction and prevention.

DT and CYNEFIN provide the underlying theoretical construct for this research and provided the means for addressing the complex decision-space associated with WNv. The identification and use of contextually based EEVs in predicting EID outbreaks provided a unique contribution to the field of management science and disease prevention.

Nature of the Study

The study was quantitative and retrospective in nature. To answer the research questions, I initially selected nine contextually derived EEVs based on the literature review. These nine EEVs were (a) average 30-day temperature (*EV1ATM*), (b) average 30-day rainfall (*EV2ARN*), (c) average 30-day dew point (*EV3ADP*), (d) average 30-day snow depth (*EV4ASD*), (e) average 30-day barometric pressure (*EV5ABP)*, (f) average 30-day wind speed (*EV6AWS)*, (g) topology (*EV7ELV*), (h) land use (*EV8USE)*, and (i) urbanization (*EV9POP*). The DV was the presence of WNv (see Table 1) computed as a number of WNv incidents within a SC county.

The EEVs acted as predictors in this research. Using longitudinal South Carolina Department of Health and Environmental Control (SC DHEC) data from 1999 to 2016, each EEV was subjected to a purposeful selection process, as detailed by Field (2013).

The EEV selection process was guided by the acknowledgement that the decision-space surrounding the WNv challenge consists of a diverse set of biological, ecological, environmental, topological, and demographic factors that interact to create a dynamic, complex system. For my analysis, these explanatory factors must have met two criteria: They must have exhibited an interactive contextual relationship to the genesis, maturation, and vectorization of WNv; and they must have been drawn from publicly available, web-based sources that provide historical and real-time data. My challenge was to determine which factors correlated strongly with the likelihood of WNv outcomes. I further constrained the study to exogenous factors alone.

Initially, this culling involved exploratory data analysis (EDA), planned use of multiple linear regression (MLR; models of multiple EEVs), and univariate analysis of the influence of each EEV on the DVs. If the analysis showed significance ($p < .20$) for any EEV, the variable would be carried forward into the final MLR model. The temporal dimension was examined (five DVs measuring no lag; 30-, 60-, 90-day lags; and a 90-day moving average time lag) to support the examination of IRM timeframes and predictive accuracy.

The EEV data were extracted from original and publicly available sources (e.g., SC DHEC, United States Geological Survey, South Carolina Department of Natural Resources). Because of the type of EEV and their standard measures, the reliability of the source data were high. The predictive validity of the EEVs' impact on model outcomes was reliant on the fit of those variables within the statistical model. The final model was developed using data from 2002 to 2016.

The original DVs were measures of the number of WNv incidents present in SC. The ability to predict each DV was directly dependent on the fidelity of the data used to construct the statistical model. Predictive models were developed by both BLR and regression using GZLMs in SPSS. GZLMs refer to a broad family of regression models that follow an exponential family distribution (Javaras & Vos, 2002). The empirical validity of the predictive models was assessed using historical results of WNv in SC.

Definitions

Abiotic: "Not biotic" (Abiotic, n.d., para. 1).

Arbovirus: "Any of various RNA viruses (as an arenavirus, bunyavirus, or flavivirus) transmitted principally by arthropods and including the causative agents of encephalitis, yellow fever, and dengue" (Arbovirus, n.d., para. 1).

Arthropod: "Any of a phylum (Arthropoda) of invertebrate animals (such as insects, arachnids, and crustaceans) that have a segmented body and jointed appendages, a usually chitinous exoskeleton molted at intervals, and a dorsal anterior brain connected to a ventral chain of ganglia" (Arthropod, 2019, para. 1).

Biotic: "Of, relating to, or caused by living organisms" (Biotic, n.d., para. 1).

Biological systems theory: "Combines experimental . . . techniques and mathematical modeling and analysis, with the ultimate goal of understanding the emergence of biological function on the basis of interdependencies among molecular components" (Radde & Hutt, 2016, p. 1).

Causation: An awareness of what causes what in the world and why it matters. There are generally two types of causality, direct and indirect (Pearl, 2009).

Complex adaptive system: A system that lightly constrains agent behavior and in turn the agents through their interactions constantly modify the nature of the system (Snowden & Boone, 2007).

Complexity theory: "A scientific theory which asserts that some systems display behavioral phenomena that are completely inexplicable by any conventional analysis of the systems' constituent parts. These phenomena, commonly referred to as emergent behavior, seem to occur in many complex, robust systems involving living organisms, such as a stock market or the human brain" (Casti, 2017, para. 1).

Complex dynamical systems theory: "Complex dynamical systems theory and its related disciplines and tools -- network theory, agent-based modeling -- provide the appropriate prism through which interdependent systems such as social groups can be understood, and coherent, integrated policy recommended" (Juarrero, 2010, p. 1).

Context: Decision making is context dependent; context influences the decision analysis process (Riabacke, 2006).

CYNEFIN: "The CYNEFIN framework is derived from several years of action research into the use of narrative and complexity theory in organizational knowledge exchange, decision making, strategy, and policy-making" (Kurtz & Snowden, 2003, p. 463).

Decision-support tools: "Computer-based information systems that help support decision-making activities" (Decision-support tools, n.d., para. 1).

Decision theory: "Decision theory is concerned with the reasoning underlying an agent's choices" (Zalta, 2016, para 1).

Emerging infectious disease (*EID*): An emerging infectious disease is one "whose incidence in humans has increased within the past two decades or threatens to increase in the near future" (van Doorn, 2014, p. 1).

Endogenous: "Caused by factors inside the organism or system" (Endogenous, n.d., definition 2a).

Environmental surveillance program (*EVS*): EVS monitors WNv transmission in mosquitoes, birds, equines, chickens, and other mammals.

Epidemiological surveillance program (*EPS*): EPS involves the testing of humans for the presence of WNv in blood or cerebrospinal fluid, detecting anti-WNv immunoglobulin antibodies, or through nucleic acid amplification testing.

Exogenous: "Caused by factors or an agent from outside the organism or system" (Exogenous, n.d., definition 2b).

Exogenous explanatory variable (*EEV*): Variables that have the ability to produce interactive context with a WNv eco-system. (e.g., average rainfall, average temperature, average windspeed).

Explanatory variable (*EV*): An independent variable postulated to influence or predict the dependent variable in a real-world process or that explains the behavior of the dependent variable (Montgomery, 2019). In this study, EVs were the set of independent variables that included EEVs, their two-factor interactions (2FI), and months.

Flavivirus: *"*Are responsible for a number of important mosquito-borne diseases of man and animals globally" (Hobson-Peters, 2012, p. 1).

Pathogen: "A microorganism that causes, or can cause, disease" (Pirofski & Casadevall, 2012, p. 1).

Predictive analytics: "A collective term for techniques with the aim of predicting the future based on static or historical data" (Geerdink, 2013, p. 1).

Real-time business intelligence: The real-time capture, access, understanding, and analysis of raw data into actionable intelligence to improve business performance (Azvine et al., 2006).

Systems biology: "The science that studies how biological function emerges from the interactions between the components of living systems and how these emergent properties enable and constrain the behavior of these components" (Wolkenhauer, 2014b, p. 1).

West Nile virus (*WNv*): A mosquito-borne zoonotic arbovirus in the family of Japanese encephalitis serocomplex (Hobson-Peters, 2012).

Zoonotic: Infectious diseases of animals (usually vertebrates) that can naturally be transmitted to humans (Rosenberg, 2015).

Assumptions

An assumption is a belief that cannot be proven but is critical to the success of the study (Simon & Goes, 2013). For this research, I assumed the following:

- Contextually based EEVs can provide sufficient statistical power for use in WNv prediction.
- These explanatory data were available from publicly accessible sources and could be readily used in predictive models.
- Publicly accessible data were generally available to decision makers in their work environments and did not require permission for use.
- Publicly accessible data could be collected with standardized measurement tools.

These assumptions were necessary in the context of the study as the research findings were based on the availability of publicly accessible, timely, relevant contextual data to a decision maker.

Scope and Delimitations

In this research, I examined the power of exogenous explanatory data in WNv predictive models when robust EPS and EVS data are lacking. To examine the use of these explanatory data in complex decision-spaces, the research scope was confined to a single EID in the southeastern United States. Research data and analysis were focused on the ability to predict WNv in SC localities alone.

According to Simon and Goes (2013), "The delimitations of a study are those characteristics that arise from limitations in the scope of the study (defining the boundaries) and by the conscious exclusionary and inclusionary decisions made during the development of the study plan" (p. 4). The problem of timely predictive modeling of EIDs is not limited to WNv, but because of the complex epidemiological and environmental cycles associated with different EIDs (e.g., Zika) and disparate health resources across local environmental health organizations, it was not feasible to develop a single model suitable for all geographic regions and EIDs. An additional delimitation within this research was the use of near real-time, web-based publicly accessible data. Using readily available data from public sources allowed me to determine if publicly accessible exogenous data alone could reliably and accurately predict outbreaks of WNv.

Limitations

Although the climate and topography of the region is similar to other temperate areas, the ability to generalize the study was limited by the use of regionally focused publicly accessible data (Liu et al., 2009; Ozdenerol et al., 2013). The scope of this study was limited to the state of SC, and the data collected were primarily historical, raising issues of external validity about the numerous means of collection and accuracy. Hence, any attempt to infer results beyond the scope of this study should be done with caution.

Beyond the scope of this study, further research should consider placing contextually based, publicly available EEVs in more advanced statistical toolsets examining complex decision spaces. Also, the application of a priori beliefs to data collection strategies is an area of interest. The field of combinatorial probability is an interesting alternative means of research in this area.

Significance of the Study

The application of DT and CYNEFIN to predictive analytics was a unique aspect of this research. The results of the study may provide insight into the robustness of systems-level, contextually based EEVs in predictive models. According to Gatherer (2010), Motta and Pappalardo (2013), and Wolkenhauer (2013), a system-level research approach aligns with the rise of systems-level biology and its attempts to characterize biological complexity in more holistic terms. An EID case study provided the complexity required to potentially generalize any findings to other predictive frameworks such as decision support tools, analytical decision management models, and real-time business intelligence applications. However, these frameworks were not addressed in this study. A secondary benefit of the research could be the examination of a predictive contextual framework for the proactive management of EIDs.

Significance to Theory

In this research, I examined the relationship among DT, decision-making context (CYNEFIN), and systems theory to understand the impact of exogenous data on complex decision making. The tenets of these theories were combined to provide a theoretical framework that challenges traditional linear-causal approaches to decision making, to expand the manager’s perception of the decision-space, and to provide greater fidelity to the design and choice phases of the decision-making process. This was accomplished through an emphasis on data intelligence and specifically the use of systems-level exogenous data to provide context and reduce uncertainty in the decision-making process.

Significance to Practice

Managers of all professions are required to make decisions daily. These decisions are made within decision-spaces that range from simple to chaotic (Snowden & Boone, 2007). The ability for a manager to make decisions when dealing with complicated and complex decision-spaces is dependent on their ontological understanding, breadth of intelligence, and analytic support.

In this research, I used a contextually based theoretical foundation that leveraged the dynamic presence of publicly accessible data in forming intelligence collection strategies for decision making. The theoretical foundation approached the decision-space in a way that allows practitioners to make decisions when empirical data are not available.

Significance to Social Change

The significance of this study was two-fold. First, I leveraged the tenets of DT, CYNEFIN, and systems-thinking to more robustly account for EEV data use in WNv predictive modeling. The premise was that if contextually based explanatory factors within a linear regression model can predict WNv presence in a locality prior to an outbreak, then the research and the predictive models may provide practitioners and decision makers in other like professions with an alternative theoretical framework for decision making in data-poor environments.

The second goal of the research was to develop a predictive model that would allow proactive EID decision making in a number of different localities based on the use of exogenous exploratory factors unique to the region. I used SC as a case study, as NACCHO rated the state as needing improvement across all core competencies of their mosquito vector control programs. This means that local detection and other preventive measures like mosquito abatement procedures are less robust than in other states.

According to Hadler et al. (2015), many public health care organizations are resource constrained, and the mosquito abatement programs in SC are no exception. Funding for these types of programs have decreased in the state since 2008, and this was exemplified in the SC DHEC's request for five additional EHMs in a recurring operating request in the Fiscal Year 2018-1019 Agency Budget Plan (SC DHEC, 2018). A lack of resources also requires local EHMs to examine and use a variety of predictive models of WNv.

A predictive model developed using SC county EEV data could reduce the latency of the current process and positively influence abatement and preventive measures, reduce WNv outbreaks, and provide a cost-effective means for disease control. It may also prove generalizable to other communities and mosquito-borne viruses around the United States and the world. This is a critical point when considering the effects of environmental factors like temperature on the WNv ecosystem and in the face of rising temperatures associated with global climate change (Soverow et al., 2009).

Summary and Transition

Historically, the timely and accurate prediction of WNv in a locality requires robust EPS and EVS programs. These programs produce surveillance data that populate predictive models, which allow EHMs to make timely decisions on preventive measures (Manore et al., 2014). With EIDs like WNv and more recently Zika, simple, reliable predictive tools are required to ensure public health measures can be taken before outbreaks occur. To address this scholarly gap, I examined the accuracy and timeliness of contextually based exogenous explanatory data in predicting outbreaks of WNv in SC. In doing so, I also examined the importance of context and system-level thinking in decision making.

Management tools that predict trends and services need to adapt to the complexity of today's information environment and to the systems-level data it produces. A systems-level, context driven approach could offer an answer to these data challenges. When required, this practical approach could allow the manager to place a decision within a broader systems-level context, using exogenous data to enrich and define a less ordered decision-space. This is particularly relevant for decision makers and managers who work within the complex field of EIDs. In Chapter 2, I examine relevant literature in decision-making theory and explore the complexities associated with emerging infectious disease.

Chapter 2: Literature Review

A scholarly gap exists in the accurate and timely predictive modeling of WNv. As revealed in the literature review, this shortfall is broadly true within the field of epidemiology and EID, where predictive models have been developed to assist managers in making resource decisions associated with EID preventive actions (e.g., public education and mosquito abatement; Reiner et al., 2013). In this research, I addressed this gap through the exclusive use of contextually based, system-level EEVs to enhance decision-space context and to predict outcomes in complex environments. The purpose of this ex post facto research was to examine the accuracy and timeliness of publicly accessible exogenous explanatory data in predicting outbreaks of WNv in SC.

The literature review accomplished two research objectives: the identification of a gap in the scholarly research relating to the power of contextual data in EID predictive modeling and the development of a theoretical foundation for addressing that gap. To properly examine the theoretical foundations for this research, a cross-disciplinary review was conducted, incorporating the core elements of DT, the CYNEFIN construct, and systems-thinking. A contextually rich, decision-making approach emerged from this review and establishes a potential process of inquiry for decision making in complex environments.

To examine the statistical power of contextual data, I chose WNv as an example of a complex decision-space and as such, a detailed review of literature associated with this EID was accomplished. To predict outbreaks of WNv, EEVs were selected from exogenous data that were contextually related to WNv. The DVs were $DV_{PRESENCE}$ and DV_{COUNT}. The exogenous variables were employed within a statistical model to determine its utility in predicting outbreaks of WNv.

This chapter consists of three sections: (a) The search strategy establishes the boundaries of the literature review; (b) the theoretical foundation section examines Simon's contributions to contemporary DT, the CYNEFIN construct, and systems-thinking; and (c) a review of the history of WNv provides current modeling approaches for predicting WNv and key predictor variables traditionally used within the EID community. Insights gained from the literature review highlight a practical and theoretically based foundation for the use of contextual variables in complex decision spaces.

Literature Search Strategy

The search strategy for the review was to gather literature that contributed to the identification of a scholarly gap and provided the necessary rigor for addressing the formulation of research questions and hypothesis. The strategy was executed in two phases: first, a focused foundational review of Simon's contributions to DT the impact of complexity and systems-thinking on that theory (concept formulation) using Snowden and Boone's (2007) CYNEFIN construct; second, a broadly defined search associated with WNv predictive modeling (problem specificity).

The search strategy focused on keywords associated with *biological systems theory*, *bounded rationality*, *complexity*, *complexity theory*, *complex dynamical systems* *theory*, *decision-making context*, *decision-making process*, *causality* (direct and indirect), *systems theory*, *reductionism*, *West Nile virus*, and *West Nile virus modeling*. The years of 2010 to 2018 were used as search filters to ensure literature currency, but a more targeted search on the theoretical foundations of the study required a relaxation of literature currency to adopt the grounded theoretical concepts introduced by Simon's mid-19th century contributions to DT and the decision-making process. The review spanned a wide range of publications to include published books, peer-reviewed journal articles, conference proceedings, corporate studies and reports, academic studies, federal and state government publications, and articles and papers retrieved from the internet. I used research database structures available through the Walden University library, the Defense Advanced Research Projects Agency research portal, and Google Scholar for this literature search.

Theoretical Foundation

This research leveraged DT, Simon's (1960) decision-making process, and Snowden and Boone's (2007) framework of contextual decision making called CYNEFIN. These theoretical constructs aligned with this research because EHMs are facing complex decision-spaces associated with WNv and require a broader, systems-level approach to explanatory variable selection.

The epidemiological cycle associated with WNv incubation, transport, and transmission reflects components of both deterministic biological processes and stochastic ecological/environmental conditions. The theoretical concepts within DT and

CYNEFIN allow the manager to frame WNv decision-space complexity within Simon's (1960) iterative three-phase decision process, allowing for a more holistic systems approach to EID prediction and prevention.

DT and CYNEFIN acted as the underlying theoretical construct for this research and provided the means for addressing the complex decision-space associated with WNv. The theoretical foundation in this study offered a new approach to the challenges of EID predictive modeling through the identification and use of contextually based EEVs in predicting EID outbreaks. This approach provided a unique contribution to the field of management science and disease prevention.

Decision Theory and Simon's Concept of Bounded Rationality

According to Moreno-Jimenez and Vargas (2018), "The ability to make decisions is an inherent and essential characteristic of human beings that reflects their degree of evolution, knowledge, and freedom" (p. 68). The study of decision making has been a part of scholarly research for centuries, but despite this fact, there is no definitive, universally accepted description of DT. North (1968), a mid-20th century decision theorist, stated that "decision theory provides a rational framework for choosing between alternative courses of action when the consequences resulting from this choice are imperfectly known" (p. 220). According to Devinney and Siegel (2012), DT is a mature and broad field sharing deep practical connectivity within the social science disciplines of economics, management, psychology, sociology, anthropology, and political science. Steele and Stefansson (2016) provided a more contemporary definition: "Decision theory is concerned with the reasoning underlying an agent's choices" (para. 1). This definition resonates with current technological thrusts in artificial intelligence and machine learning, where software agents are placed within data streams to extract and process intelligence for decision making.

According to Buchanan and O'Connell (2006), classical DT has its roots in philosophy and psychology. It is in these roots that one finds the enduring concept of the *rational man*. The concept of the rational man in decision making spans nearly 70 years of contemporary DT and has sparked the theories of normative, descriptive, and prescriptive DT.

Based on his life's work in the field of decision making and his challenge to DT's classical concept of the rational man, I elected to use Simon's work in DT and the decision-making process as the theoretical underpinning for my research. Simon's interest in political science and economics began with his undergraduate studies at the University of Chicago in the early 1930s. In the process of earning his B.A. (1936) and PhD (1943), Simon developed an academic interest in reasoning and decision making. By 1949, amid postwar thought on human rationality and the rise of computational models, Simon began to challenge the traditional economic theories of maximizing expected utility and the precepts behind normative DT. Social scientist, economist, and mathematician, Simon was awarded the ACM Turing Award in 1975 and then went on to win the Nobel Prize for Economics in 1978.

Gorzen-Mitka and Okreglicka (2014) highlighted that over the last 70 years, DT has primarily adhered to the unitary foundation of rationality. In other words, when confronted with a decision, the decision maker will make rationale choices with respect to the expected utility of the outcome and within moral boundaries. According to Moreno-Jimenez and Vargas (2018), scientific decision making well into the 1970s was executed using mechanistic processes that ignored the subjective aspects of human cognitive processes. Rationally driven decision making can be described by three theoretical models: normative, descriptive, and prescriptive**.**

Normative DT

Normative DT sits at the core of classic decision-making theory and is dependent on the rationality of human behavior. The normative model of DT states that an individual makes decisions that best satisfy a desired outcome or objective. According to Gigerenzer and Selten (2001), the concept of reasonableness or rationality emerged out of the 1950s and 1960s, where psychology, probability, and optimization combined to yield new models of statistical inference and cognitive processes. Within a normative decision process, managers exercise a theory of how humans will rationally address problems and decide on courses of action. In normative decision making, individuals strive to understand the impact of sociocultural biases, experience, training, time, and personal well-being on the decision process. Proponents of normative theory believe individuals can accurately assess the risk, probability, and utility of their decision-making options (Simon, 1955).

Challenging the rational man theories of the time, Simon (1955) believed the decision maker or economic man was not cognizant of all aspects of their decision-making environment. He argued that the normative definition of rational choice was modified by the levels of informational access and computational capacities available to the decision maker. According to Buchanan and O'Connell (2006), Simon argued that decision makers were constrained or bounded by the costs of acquiring information for decision making. These bounds on rationality constitute Simon's (1956) main argument, stating that decision makers essentially adapt their choices well enough to *satisfice* rather than to optimize or maximize. Simon's principle of satisficing or bounded rationality disrupted the classical normative model of DT, which strove for optimal solutions.

Descriptive DT

Descriptive DT is more heuristically oriented than its normative kin. Dillon (1998) claimed that descriptive DT is focused on what the decision maker actually does in the decision process. In other words, descriptive DT describes what happened in a particular situation and why.

Following Simon's theory, in descriptive DT, the decision maker is bound by organic and informational constraints that force them to work to a decision that is satisfactory vice optimal. This reveals the promise of complexity in decision making, where organic and informational elements of a decision exist in some state of equilibrium within the decision maker. Simon (1956) believed a decision maker in constant interaction with their environment will *satisfice* to reduce the complexity of that interaction. In Dillon's (1998) words, "Humans have limits that they cannot exceed" and will seek some form of satisfactory equilibrium (p. 101). Koopmans (2014) illustrated these limits in research when a researcher seeks to apply linear causality to dynamic states of equilibrium in randomized controlled trial or controlled tests. In today's complex environments, this approach is questionable and reflective of linear vice recursive thinking.

In descriptive DT, individuals use highly tailored and rational approaches to balancing decisions based on available time, resources, and commitment in descriptive theory. Because the practice of descriptive decision making is more heuristic than its normative cousin, it is possible to identify variables within a decision-space that can provide acceptable statistical results when a rudimentary understanding of correlation or causality is required. However, descriptive theory and its supporting heuristic processes and statistical tools have limits. According to Koopmans (2014), the ability to quantify a decision-space defined by complexity, feedback loops, and emerging behavior is extremely difficult in a process that seeks to reduce the problem to a set of hierarchical variables. This is particularly true when addressing decision making within the field of epidemiology.

Building on Simon's theory of bounded rationality, Tversky and Kahneman (1981) introduced an additional aspect of uncertainty to DT related to transitivity. While descriptive DT explains how and why decisions are made, it does not address why certain decisions, when framed differently, result in shifts of choice preference. Tversky and

Kahneman called this phenomenon prospect theory, arguing that decision makers prefer prospects that offer the highest expected utility. Divekar et al. (2012) differentiated descriptive DT and prescriptive DT through the inclusion of prospect theory.

Prescriptive Decision Theory

French and Rios Insua (2000) described the prescriptive theory of decision making "as the application of normative theories, mindful of the descriptive realities, to guide real decision-making" (p. 5). In this theoretical framework, a prescriptive means is available to the decision maker to make a specific decision with the full knowledge of the normative and descriptive elements. The decision maker can pursue an optimized solution understanding the constraints of Simon's bounded rationality for making a specific decision. The prescriptive approach is more malleable to the uncertainty of data and information used within the decision process because it accounts for its presence.

Uncertainty

A review of DT and the concept of bounded rationality would be incomplete without broaching the topic of uncertainty. According to Russell et al. (2017), uncertainty can limit course of action development. Uncertainty can also delay or even corrupt the decision-making process making front-end intelligence a critical factor. In the information gathering phase of the decision process, incomplete or missing data reduce an individual's ability to identify the problem, frame the decision-space, and conduct the type of analysis (codification and idealization of data) necessary for decisions of utility.

Data analysts typically deal with incomplete or missing data by imputing values. Hewett (2004) stated that this can occur in two ways: “Eliminating cases with missing data” (p. 182) or by estimating values for substitute data. Empty set, fuzzy set, and rough set theories are all mathematical approaches for estimating missing or partial data, and these have been used in numerous automated information systems (Rissino & Lambert-Torres, 2009; Walczak & Massart, 1999). For EHMs, imputing values to mitigate uncertainty may not be viable. Motta and Pappalardo (2013) applied this same viability to the science of biology, positing that the development of numerical equivalents for biological parameters may result in large uncertainty.

In complex professions such as finance, economics, and medicine, the application of value approximations for data can lead to incorrect inferences and poor decisions. This may require managers within these professions to either delay the decision process until all data are in, approximate missing or partial data, or make decisions based on the information at hand. The risk associated with decisions made without broader level contextual data can have severe consequences. This is particularly true when one addresses preventive measure decisions associated with EIDs such as WNv. The failure to detect WNv before an outbreak can result in a lack of action to educate the public on its presence or to take effective mosquito abatement actions.

Lichtenstein et al. (2006) stated that a new paradigm is rising within the art of organizational leadership. The reality of today’s complex, interrelated, and adaptive business environments are causing decision makers to view complex decision-spaces from a more holistic perspective, willing to incorporate heterogeneous factors into traditional decision processes. This is particularly true in the biological sciences, where systems biology has embraced the concepts of emergence and multilevel systems-thinking. This approach paves the way for the wider use of contextual variables in decision-making tools and models.

Simon's Decision-Making Process

Central to Simon's views on bounded rationality and "satisficing" are his thoughts on the decision-making process. As the Ford Distinguished Visiting Professor at New York University, Simon gave three lectures that were then synthesized and published in the book, *The New Science of Management*. Within this book, Simon (1960) introduced the idea of programmed and nonprogrammed decisions. He described them as bookends of a decision continuum. He believed programmed decisions are transactional by nature; they recur often enough that decision makers can quickly recognize and implement experientially informed decisions. Nonprogrammed decisions lie at the other end of the continuum. Simon described these types of decisions as "novel, unstructured, and consequential" (p. 6).

Simon (1960) described a manager's decision-making process across this continuum in three phases: intelligence, design, and choice. This concise decision-making framework provides the foundation for my research, providing a grounded theoretical starting point for comparison to more contemporary decision-making tools such as Snowden and Boone's (2007) CYNEFIN construct.

To examine the impact of complexity, uncertainty, and context on decision making, I framed DT through lens of Simon's (1960) decision-making process. I began with an overview of Simon's three-phased process of intelligence, design, and choice. I then examined each phase separately using Snowden and Boone's (2007) CYNEFIN construct to address the impact of complexity, uncertainty, and increasing importance of context in Simon's decision-making construct.

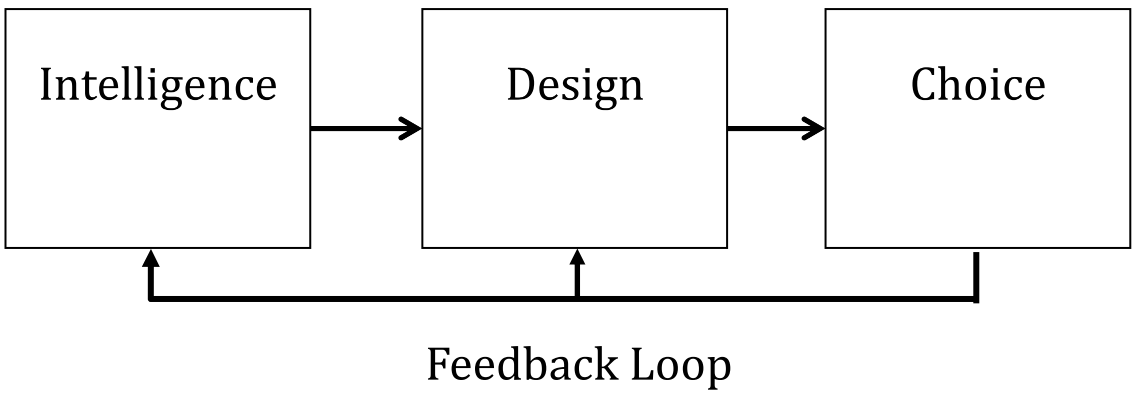
According to Simon (1960), the decision-making process occurs in three phases: intelligence, design, and choice (Figure 1). Presented in a simple form, the complex nature of decision-making lies completely within his model. Simon described the phases as follows:

> The first phase of the decision-making process – searching the environment for conditions calling for decision – I shall call *intelligence* (borrowing the military meaning of intelligence). The second phase – inventing, developing, and analyzing possible courses of action – I shall call *design* activity. The third phase – selecting a particular course of action from those available – I shall call *choice* activity. (p. 2)

Simon (1960) declared that decision makers spend large amounts of time in the gathering of intelligence and the development of courses of action. If the intelligence and design phases are done correctly, choice becomes less time consuming as the decision-space context will have been adequately defined and COA alternatives mentally prioritized. In Simon's construct, intelligence and design work in a recursive fashion to establish the context behind a decision and to set the stage for an informed choice. While this construct is theoretically complete and incorporates feedback, its simplicity masks the complex issues facing decision makers in the information age.

Figure 1

Simon's Decision-Making Cycle



Note. The decision-making cycle diagramed above reflects Simon's (1960) thoughts described in "*The New Science of Management Decision*,*"* Harper & Brothers Publishers, N.Y. (1960). I modified Simon's decision-making construct with a feedback loop.

Azvine et al. (2006), Babu and Sastry (2014), Gorzen-Mitka and Okreglicka (2014), Hewett (2004), and Osman et al. (2013) all posited that the ability to rapidly stratify, discriminate, and synthesize decision quality information from the midst of complex dynamical systems is one of the great challenges associated with information age decision making. Artikis et al. (2012) highlighted several trends driving this challenge: the virtual instrumentation of the world through the *internet of things*; expanding sources of cheap storage; the pervasiveness of sensor technology resident in personal devices such as smart phones; and the spread of broadband connectivity.

According to Boisot (1999), Geerdink (2013), and Osman et al. (2013), while these drivers create a data rich environment, the ongoing convergence of personal computing, mobile communications, and web-enabled technologies has increased the complexity and uncertainty of the decision-making process due to the sheer volume and heterogeneity of data. In many cases, the uncertainty lies in how data were produced, recorded, or disseminated to decision makers. Trust in the completeness, accuracy, and relevance of outsourced data and the latency associated with that data are a concern to enterprise-level decision makers who, according to Azvine et al. (2006), “are no longer satisfied with scheduled analytics reports, pre-configured key performance indicators, or fixed dashboards” (p. 1).

To provide similar contextual fabric to complex decision spaces such as EID, management tools predicting trends and services need to adapt to the complexity of today’s information environment and to the system-level data it produces (Babu & Sastry, 2014; Koehler, 2014, Raia, 2008). Adopting a more holistic strategy to data collection and analysis is required to maintain contextual robustness in a complex decision-space that is shaped by interacting contextual factors (Gorzen-Mitka & Okreglicka, 2014).

Intelligence

Dillon (1998) explained in Phase 1 of Simon's decision-making construct that the term intelligence is used to describe the function of collecting data to design and make intelligent choices about the decision at hand. Seventy years ago, business intelligence and data collection were made without the ubiquitous data streams of the 21st century; data sources were created and aligned to record observable factors reflecting mechanistic and causal relationships. Meadows (2008) believed that while Western society benefited from these past mechanistic approaches, simple causal observations introduced a form of reductionism into intelligence gathering and decision-space formulation. Riabacke (2006) described how centralized management can develop organizational norms that constrict decision-making context. Koopmans (2014) and Straub (2013) warned these constricted, linear approaches in today's complex business environments will not be sufficient. Intelligence derived from constrained data streams are more likely suited to the type of transactional decision making experienced in simplified problem sets with repetitive histories and do not incorporate Meadow's (2008) principle of "equifinality" in its data strategy (p. 41).

According to Sargut and McGrath (2011), access to, collection, and analysis of data in today's business environments can be constrained by locally derived and business-oriented data mining strategies that have evolved to obsolescence. For example, a decision maker may only have the authority to access and use data applicable to his/her product, service, or division. This could unnecessarily constrain the broader intelligence required to appropriately abstract and design the decision-space and to make quality choices in the decision process. These constraints appear to be present in the intelligence being captured for EHMs in general. While EHMs recognize the dynamics associated with the inherent biological and environmental complexities of WNv, the traditional EPS and EVS surveillance process currently in use provides non-real time, incomplete intelligence.

The EID intelligence phase (i.e., surveillance) tends to ignore the interactive, hard and soft effects of feedback, equilibrium, and emergence within complex systems. Motta and Pappalardo (2013) also addressed this reductive approach to intelligence collection in the choosing of mathematical models to explain biological systems. They posit that the mathematical representations of biological systems need to consider multiple sources of data, essentially using more integrative strategies. Consequently, data collection strategies and intelligence gathering become critical elements in creating context in EID decision making.

For Simon's (1960) intelligence to be actionable and constructive in today's environments, the decision maker needs to adapt data collection and analysis strategies that align with their current decision-space. This is an iterative (lower feedback loop in Figure 1) but necessary process to ensure the ongoing mental model constructed within a decision-space is rich enough to support quality decisions.

For these reasons, Simon's (1960) Design and Choice phases become highly dependent on the following:

1. The fidelity, relevance, and wholeness of collected intelligence data.

2. The completeness of a decision-space; constructed to support either ordered, unordered, or disordered decision-spaces.

3. The managerial belief functions (experience, trust, commitment) associated with both the intelligence data and model used.

Design and Context

The design phase determines how the decision maker abstracts their decision-space from experiential and collected intelligence data. It is here that decision makers encounter the challenges of modern-day complexity in nature and in business. Schafer (1976) described a decision-space as a frame of discernment or dynamic mental model that organizes the variables within a decision. Groen and Mosleh (2005) defined a decision-space as the idealized sum of a decision makers understanding of reality combined with their socio-cultural biases and life experiences as applied to a current decision.

According to Bunge (1973), the art of abstracting reality begins with idealization. Groen and Mosleh (2005) and Motta and Pappalardo (2013) believed decision makers accomplish this idealization through a fluid cognitive synthesis of reality, experience, and understanding. These elements of idealization help the decision maker to refine their frame of discernment in a complex but contextually rich decision-space.

In more complex situations, a decision-space becomes a dynamic mental mosaic of internal and external factor relationships and pathways. The multi-dimensional process of decision making becomes more complex by a decision-space affected by modifying factors such as time (e.g., schedule and speed of decision), resources (e.g., funding and capital assets), and commitment (e.g., stakeholder interest and utility; Dietrich, 2010). If present, these factors may shape the decision-space in a way that modifies the final decision.

In decision-space construction, the decision maker ultimately builds context about the problem set. As stated by Riabacke (2006), "No decision takes place in vacuo: there is always a context" (p. 1). Decision-making context builds and gains fidelity as intelligence gleaned from data are made available and are examined from different perspectives. This sets up a recursive process (Figure 1 feedback loop between intelligence and design) of collection and idealization for the decision maker that is driven by the complexity of the problem set and the completeness of data.

Choice

In Simon's (1960) construct, intelligence and design work in a recursive fashion to establish the context behind a decision and to set the stage for an informed choice. Choice lies at the end of Simon's decision-making construct and represents the pathway taken by a decision maker based on the clarity of the decision-space once intelligence has been applied to design. The decision maker acts with the understanding they are executing a decision within a quadrant of the CYNEFIN construct.

There is a necessary feedback loop from choice to intelligence that needs to follow the decision. The real-time observation of a decision in action can inform the decision maker of any adjustments required and will update the previous decision-space with new information (Snowden & Boone, 2007).

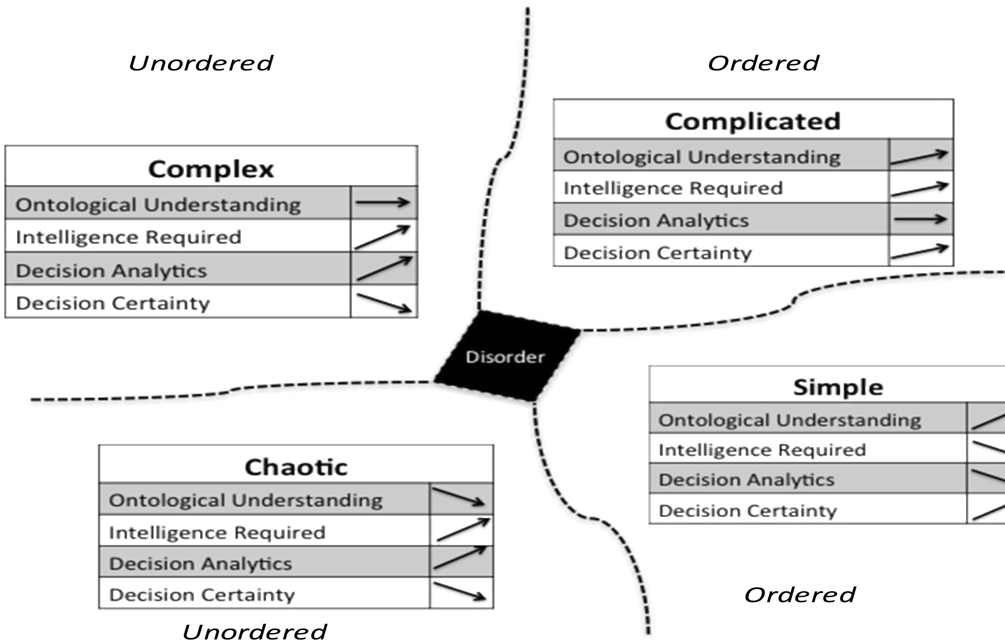
The CYNEFIN Construct in the Design Phase

Kurtz and Snowden's (2003) decision-making framework called CYNEFIN (pronounced ku-ne-vin) provided an interesting construct to explain the different levels of cause-and-effect relationships the decision maker faces when designing their decision-space. The CYNEFIN framework was constructed around three types of relational contexts; ordered, unordered, and disordered.

Within the ordered category, Snowden and Boone (2007) identified simple and complicated contexts. The unordered category includes complex and chaotic contexts. Disorder is situated in the center of the diagram signifying the ease at which a decision-space may slip from one category to another. In Figure 2, I expand on the original CYNEFIN framework to highlight the decision maker's initial system-level (ontological) understanding, level of intelligence data required to support decision-space construction (contextual understanding), the level of decision analytics required to support decisions, and finally the level of decision certainty; that is the level of certainty that the decision maker has in their decision process.

Figure 2

CYNEFIN Construct



Note. The CYNEFIN construct is organized around three types of cause-and-effect relationships, ordered, unordered, and disordered with simple and complicated being categorized as ordered. Adapted from CYNEFIN construct at https://en.wikipedia.org/wiki/Cynefin_framework#/media/File:Cynefin_as_of_1st_June_2014.png. Content freely available from Creative Commons (CC BY-SA 3.0).

The CYNEFIN construct is complementary to Simon's (1960) design phase. Here the decision maker identifies where their decision-space context lies and thus what level of ontological understanding, intelligence, decision analytics, and decision certainty they are addressing. Each of the quadrants are now addressed separately.

Ordered-Simple Context

According to Meadows (2008), ontological understanding of systems range from simple to complex. Simple systems consist of simple feedback mechanisms that allow repetitive, transactional activities to occur, responding to internal and external stimuli in bounded ways. For instance, the fuel gauge within an automobile relies on a feedback system that consists of a fuel sender unit (or float) and a lever arm attached to a resistor that is powered by the car's battery (Nice, 2019). In a fully fueled vehicle, the float will sit at the level of the fuel and the lever arm will cause the resistor to send a current to the fuel gauge that will result in a full tank reading. As fuel is used, the float will fall with the fuel level and the lever arm will cause the resistor to send more current to the fuel gauge indicating levels less than full. In this example, the system accepts external stimulus from the filling or draining of the tank. However, the system is bounded by several factors such as the size of the tank and the instrumentation of the fuel gauge. While many simple systems are more complicated than the example, it is consistent with the characteristics of systems of this type.

The decision maker can conceptualize this type of system as being in the lower right quadrant (Figure 2) of the CYNEFIN framework. In this quadrant, the decision maker is dealing with simple systems and thus the context surrounding that system is well understood. Intelligence data are collected, categorized, and a response/decision can be quickly made. Directly causal effects are observable and repetitive, allowing the decision maker to make decisions aligned with best practices (Snowden & Boone, 2007).

Decisions with simple systems can be transactional by nature and fit within Simon's (1960) programmed definition.

Ordered-Complicated Context

In the complicated systems quadrant, Snowden and Boone (2007) moved away from transactional decisions to those that require expert knowledge and good practice. Kempermann (2017) stated that "contexts, in which causalities can at least in theory (or retrospect) be known but are non-linear and difficult to untangle are called complicated" (p. 3). Expert knowledge and experiential rule-based approaches are used to idealize the decision-space and construct courses of action. In this category of system understanding, decision making relies on subject matter experts who are supported by knowledge-based systems with historical precedents.

Using the CYNEFIN construct, Smit and Derksen (2017) looked at complex and complicated problems in primary mental healthcare. Of 113 primary care vignettes, 35% were classified as complicated. Although these complicated vignettes contained one or more relationships, the outcomes were predictable. The professions of medicine and engineering are prime examples of the types of decision-spaces seen in the complicated quadrant. The decision-space in these systems can be very complicated but there are historical precedents, expert-driven decision support tools, and established standards of practice available to the decision maker.

Based on the WNv literature review, it appeared that EHMs have been reliant on traditional surveillance tools and historical best practices. It is likely that many EHMs are gathering intelligence data and constructing decision-spaces appropriate to the complicated quadrant vice the complex quadrant. If this is the case, their frames of discernment may be skewed.

Unordered-Complex Context

According to Udelll (2017), the study of complexity has been in the literature for over 25 years. Walton (2016) wrote that in those 25 years, complexity theory (CT) precepts have expanded from the mathematical and physical sciences into organizational and social sciences such as management, health, public policy, and evaluation. The sciences of biology and ecology have embraced the complexity of nature for decades and are rapidly adopting system-thinking approaches to their fields. For example, Wolkenhauer and Green (2013) described “systems biology as the science that studies how biological function emerges from interactions between the components of living systems, and how these emergent properties constrain the behavior of these components” (p. 5939).

While Morcol (2001) and Walton (2016) believed that a specific definition for complexity remains elusive, Bar-Yam (1997) developed a common set of central properties or commonalities that steers adherents away from the deterministic and reductionist approaches of Newtonian thinking. Bar-Yam championed a post-positivist system view that recognizes the concept of emergence and embraces context in the form of (a) elements and their number; (b) interactions and their strengths; (c) operating structure and their time slots; (d) diversity and variability; (e) the environment and its external demands; and (f) activities and objectives.

Meadows (2008) aligned these central properties in the terms of a system; "a system is an interconnected set of elements that is coherently organized in a way that achieves something" (p. 11). While their findings are separated by 11 years, Bar-Yam (1997) and Meadows (2008) defined complexity and systems as inextricably linked through the properties of interconnectivity, diversity, and objective. Morgan (2007) urged decision makers to apply these properties in a systems-thinking approach. An approach that is more inclusive than the mechanistic decision processes of the past.

A complex systems decision-space is more dynamic than simple or complicated decision-spaces. It is characterized by symbiotic relationships formed by component interdependencies and on internal and external interactions. According to Gatherer (2010), these dynamic relationships allow system components and their associated structures to adapt, evolve, and to behave in new ways.

As an example, in the top-hand pane of Figure 3, two independent systems operate for a function or purpose (Meadows, 2008, p. 11). While each contains internal elements that interact (endogenous), the two systems may also act upon each other dynamically. In the bottom pane, the external interactions (exogenous) between the two systems results in Outcome Y. Therefore, Outcome Y is dependent on some level of exogenous interaction between System 1 and 2.

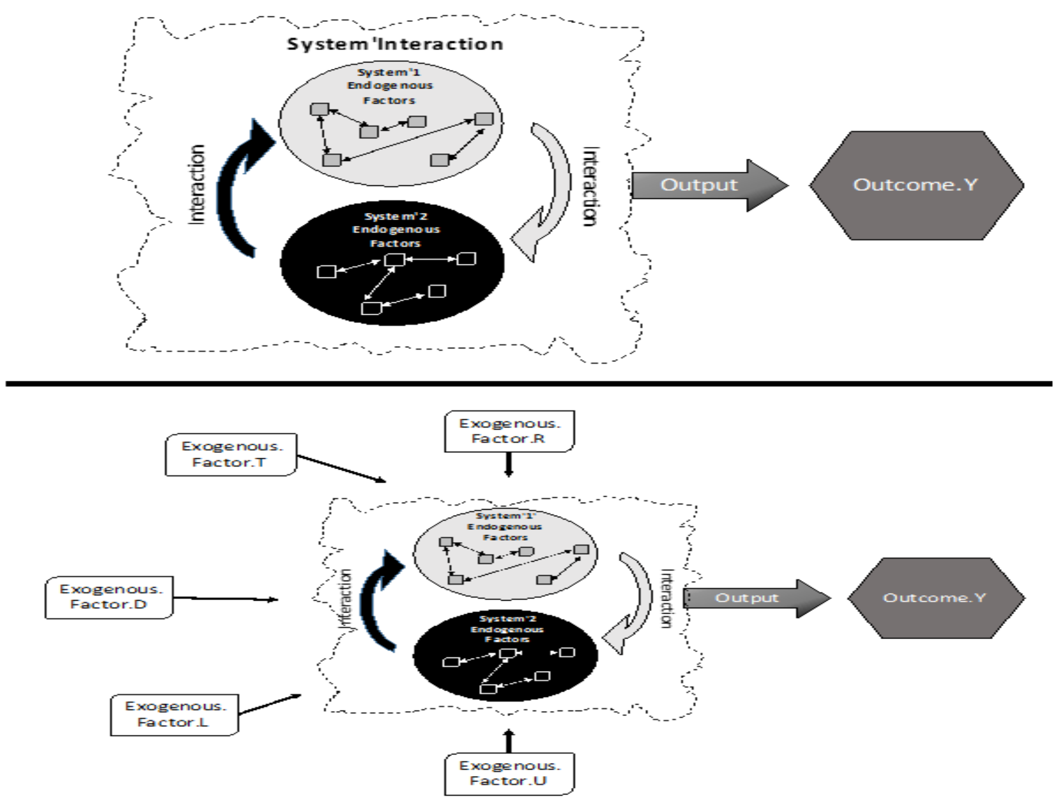
Meadows (p. 17) posited that the changing relationship between systems will also change the system behavior or outcome. This means that Outcome Y can change based on some dynamic change in System 1 or 2. If Outcome Y involved financial, health, defense, or safety related outcomes, a decision maker may wish to predict the outcome of this system-level interaction.

To fully understand the dynamic relationship between System 1, System 2, and Outcome Y, the decision maker requires not only an understanding of the endogenous elements of the interacting systems but *also* the exogenous factors that shape these interactions as depicted in the bottom pane in Figure 3. These exogenous interactions are depicted as input arrows to the system-level interaction.

Senge (2006) defined the composite of these inputs as the “invisible fabrics of interrelated actions” (p. 7). Wolkenhauer (2014a) addressed these interactions stating, “biological systems are complex, not only as a consequence of non-linear dynamics but also as a consequence of multi-levelness; the functioning of tissues is determined by interactions taking place across multiple levels of structural and functional organization” (p. 247). As a part of the functional organization of the system depicted in Figure 3, it makes sense that these exogenous factors are included in any understanding of Outcome Y because they are an extended part of system interactions. They add context to the understanding of an outcome. Examination of EPS and EVS explanatory factors in WNv predictive modeling found that many localities do not have the resources to produce the primary surveillance data necessary to robustly populate their models (Hadler et al., 2015). In the absence of these data, understanding the power of contextually derived exogenous factors in predicting WNv outbreaks is required.

Figure 3

Endogenous and Exogenous Factors in System Interactions and Outcomes



Note. Endogenous and exogenous factors in system interactions and outcomes. The top pane of the figure depicts two interacting systems that produce Outcome Y. The bottom pane shows that there can be EEVs that act as catalyzing agents to the interactions between Systems 1 and 2.

According to Wolkenhauer (2014a), when addressing complex decisions within areas such as systems biology and EID, integrating system-level exogenous data to define and enrich a less ordered decision-space, may invigorate traditional statistical tools and enhance cognitive thresholds for decision making. Based on the non-linear, spatio-temporal dynamics associated with WNv, EHMs will find themselves in the unordered-complex quadrant, attempting to understand the interactions and relationships between key biological parameters such as pathogenicity, environmental conditions, a myriad of virus hosts, and transmission pathways (Russell et al., 2017; Wolkenhauer & Green, 2013). Additional intelligence is required to fully understand system interactions and to build context for the WNv decision-space. Effective decisions in this quadrant require a more holistic characterization of the decision-space to provide EHM decision makers with the proper frame of discernment. A holistic characterization of WNv may benefit from a time-lagged examination of contextual variables and the temporal aspect of their influence on WNv outbreaks.

In the design phase of Simon's (1960) construct, the study of EID and systems biology in general, may be enhanced by viewing the decision-space within a complex quadrant that is idealized at multiple levels (Russell et al., 2017; Wolkenhauer & Green, 2013). To develop this contextually rich decision-space, EHM decision makers may expand their intelligence strategy to include both endogenous and exogenous data. Decisions taken in this quadrant are more likely to reflect new or emergent practices (Snowdon & Boone, 2007).

Unordered-Chaotic Context

In the Chaotic decision quadrant, systems that appear to have no patterns of discernment. Simple causal, rule-based, or pattern recognition approaches to problem solving and decision making cannot provide the decision quality information required to make decisions of high utility. Gatherer (2010) defined this level as irreducible complexity. This type of system and their effects tend to drive the decision maker to decisions that spark practices without precedent.

Kempermann (2017) described biological systems in advanced stages of disease as chaotic. He goes on to stress that responses to chaotic decision-spaces require some form of immediate action. There are no precedents or standard operating procedures available to the decision maker and unlike complex decisions, probing for causal effect is inappropriate. Here the decision maker must perturbate the system into a new, perhaps ordered equilibrium if possible.

Disorder

Finally, the CYNEFIN framework recognizes the fact that disorder may exist in systems, and this is shown in the middle of the diagram. In this area, an immediate categorization of the decision-space may not be possible. Kurtz and Snowden (2003) and Snowden and Boone (2007) suggested that decision makers should attempt to break down the elements into actionable parts. Driving those elements that they can, into the unordered or ordered categories of context.

In summary, CYNEFIN is a conceptual framework that recognizes and places decision and systems theory at its core. It categorizes the types of systems that decision makers must deal with and aligns the type of decision approach (best practice, good practice, emergent practice, and novel practice) that is suited to the decision-space (Snowdon & Boone, 2007). While the CYNEFIN construct allows the decision maker to recognize and posture their supporting intelligence collection activities, they will still face challenges related to resources, commitment, and uncertainty. These may constrain, reduce, or provide additional freedom of design and choice.

A contextually rich, systems-level approach to decision making using DT and CYNEFIN could provide EHMs with a more inclusive framework to gather EID intelligence, construct complex decision-spaces, and choose EEVs for predictive modeling. The following literature review establishes shortfalls in WNv predictive modeling and provides insight into the types of contextually derived EEVs necessary for more dynamic EID modeling.

Literature Review

In the previous section, Simon's (1960) decision-making process was integrated with the CYNEFIN construct to provide a more contemporary, contextually rich approach to decision making for EHMs. Snowden and Boone's (2007) CYNEFIN construct was interwoven into the Simon's (1960) framework to show how intelligence and design worked iteratively to identify and develop ordered, unordered, and disorders contexts.

As the EHM considers their WNv decision-space, they need to construct a mental model that is more inclusive and holistic than traditional mechanistic and reductionist approaches. An expansion of the EHM decision-making process to include both endogenous and exogenous data may provide a broader frame of discernment in EID decision-spaces. Theoretically, this type of approach drives the decision maker to a system-level orientation; an approach better suited to the systems biology of EIDs (Gatherer, 2010). This systems-level orientation guided my literature review of the predictive modeling of WNv.

Presence and Prediction of West Nile Virus

First detected in Africa in 1937, the initial cases of WNv in the United States occurred in New York City in the summer of 1999. Today, WNv has been diagnosed in all 48 states within the contiguous United States, with two nationwide epidemics occurring in 2003 and 2012 (CDC, 2013; Kwan et al., 2012). At the end of 2016, the U.S. Department of Health & Human Services, CDC reported 46,086 cases of WNv, with 2,017 human deaths over the 1999 to 2016 timeframe. Although it has become the "principal cause of viral encephalitis in the United States" (Austin & Dowd, 2014, p. 1015), no vaccine or specific therapy for WNv currently exists (CDC, 2013; Gubler, 2007).

Due to a universal lack of a vaccine and approved therapy, public officials around the world have been actively seeking tools that will predict human outbreaks of WNv and aid decision making in the timely application of preventive measures when transmission cycles are high. Predictive modeling of WNv is an important decision support tool in this effort, but it remains problematic due to the dynamic temporal and spatial interdependencies of the pathogenic, ecological, and anthropological components of the virus (Pirofski & Casadevall, 2012). These interdependencies present a complex decision-space for environmental health managers (EHMs).

The presence of WNv in a locality is dependent on numerous interactive biological, environmental, and ecological factors. To quantify WNv risks, researchers and EHMs have developed EPS and EVS tools to monitor arbovirus infections in humans, to understand WNv mosquito transmission activity (vector control), and to execute preventive measures (CDC, 2013).

EPS involves testing humans for the presence of WNv in blood or cerebrospinal fluid, detecting anti-WNv immunoglobulin antibodies or through nucleic acid amplification testing. These types of surveillance data are conclusive as to the presence of WNv and are critical to understanding the extent of human incidents of WNv in a locality. However, utilizing EPS data by itself is usually insufficient for predicting outbreaks as the data can lag several weeks behind the actual infection timeframe (CDC, 2013, p. 11). This deficiency in EPS data has led to the development of EVS activities to gather additional explanatory factors and to strengthen the indices supporting predictive capabilities.

EVS monitors WNv transmission in mosquitoes, birds, equines, chickens, and other mammals. Using a number of surveillance activities, such as mosquito pools, sentinel animals, and birds, EHMs gather additional explanatory data to compare against historical EVS and EPS data to better understand and quantify the antecedent conditions necessary to address WNv. Using a combination of EVS and EPS data, these types of mixed predictive models provide detection timeframes of 2 to 4 weeks prior to the onset of human symptoms. In a locality that has a robust EVS and EPS program, integrated risk management (IRM) decisions can occur within a sufficient time-period to take preventive actions. However, according to Hadler et al. (2015), “Arboviral surveillance is inadequate in many states to rapidly detect and control outbreaks and to give the public critical information it needs for prevention” (p. 1165).

Following the outbreak of WNv in the United States in 1999, the federal government implemented programs and funding to support state or local arbovirus surveillance programs. NACCHO (2014) reported that by 2004, approximately $45M in federal funding supported these surveillance programs. By 2010, reports began to surface of funding declines in these programs in the states of California, South Carolina, Florida, Nevada, Wyoming, and Oregon (DeLong, 2010). In 2012, federal funding for arbovirus surveillance programs had declined by 61% ($17.5M; 2014). This decrease sparked a shift in the capacity for IRM and required local health officials to prioritize mosquito surveillance at the expense of other EVS (equine, avian, sentinel animals) methods (Hadler et al., 2015). By 2012, Hadler et al. (2015) reported that “57% of states reported eliminating avian death surveillance, 58% decreased mosquito testing, and 46% decreased the number of human specimens tested for WNv” (p. 1161).

Although local EHMs reprioritized limited funding, the downward trend in operational capability continues. In 2016, NACCHO found that 84% of national mosquito vector control programs needed improvement due to a failure of one or more vector control program core competencies (routine mosquito surveillance, treatment decisions using surveillance data, larviciding, adulticing, or both, routine vector control activities, pesticide resistance training (NACCHO, 2017). This decrease in capability directly affects current WNv predictive tools such as California Mosquito-Virus Risk Assessment and Dynamic Continuous-Area Space-Time systems (Kwan et al., 2012). These predictive modeling tools rely on robust surveillance programs with regular sampling/reporting to provide EHMs with indices that can accurately support WNv outbreak prediction in an accurate and timely manner.

In the absence of these robust, data-rich WNv surveillance programs, there is evidence that other explanatory environmental factors can provide interactive context to the presence of WNv in a locality. Explanatory factors such as meteorological data, topology, land use, and population density have been used in past studies to enrich predictive models when univariate analysis has shown some correlation to incidents of WNv (Ahmadnejad et al., 2016; Soverow et al., 2009). Most of these factors, referred to as EEVs in this research, are collected in web-based, publicly accessible data bases allowing real-time access to historical and trending sets of data. When viewed from a management science perspective, the use of these exogenous factors reflects a more holistic, systems-level approach to decision making that may provide statistically significant predictability using contextual data in the absence of primary EPS and EVS surveillance data.

In combination with EPS and EVS factors, several studies have proven these types of exogenous factors to be important to understanding WNv-vector-host interaction, virus maturity, and vector abundance (see, for example, Cotar et al. (2016); Ozdenerol et al. (2013); Rochlin et al. (2011). However, there are no studies that have addressed the power of these types of explanatory factors in predicting WNv when robust data from EPS and EVS are not available.

Key Variables and Concepts

In the summer of 1999, U.S. federal and local EHMs faced a new and deadly EID. Previously undetected in North America, WNv was identified in human serology in New York City and the virus would leave seven patients dead. Over the next two years, 78 additional cases of WNv were reported within the confines of greater New York City (Campbell et al., 2002). With two nationwide epidemics in 2003 and 2012, the virus had spread to 48 states within the contiguous United States by 2016 with 46,086 reportable cases and 2,017 deaths (CDC, 2018; Kwan et al., 2012). According to Austin and Dowd (2014), CDC (2013), and Gubler (2007), although it has become the principal cause of viral encephalitis in the United States, no vaccine or specific therapy for WNv currently exists.

Due to the lack of a vaccine and approved therapy, EHMs around the world actively seek decision support tools that will predict human outbreaks of WNv and aid decision making to apply preventive measures when virus transmission cycles are high. According to Pirofski and Casadevall (2012) and Russell et al. (2017), predictive modeling of WNv is an important decision support tool in this effort but remains problematic due to the dynamic temporal and spatial interdependencies of its pathogenic, ecological, and anthropological components.

In a comprehensive literature review of 325 publications, Reiner et al. (2013) found inherent problems with current mathematical models of mosquito-borne pathogen transmission due to the complexities of accurately parameterizing biological and ecological factors. While the numerous interactions extant within the genesis, maturation, and vectorization of WNv reflect the inherent biological complexity of the virus, that same complexity relies on certain preconditions or antecedent explanatory variables (e.g., temperature, rainfall, wind, etc.). This fact has led to modeling approaches that use combinations of EPS, EVS, and EEV data.

Ozdenerol et al. (2013) reported that researchers focusing on human WNv infection risk in Connecticut in the period 2000 to 2005 identified static and dynamic EEVs associated with land use (static), “daily temperature, yearly precipitation, growing degree days (GDD), and animal sentinel data” (p. 5411). Using multiple regression, the predictive power of these risk factors was analyzed, and they found significant predictors in land use, population density, GDD, no positive mosquito pools in the last 30 days, no mosquito testing in the last 30 days, dead bird counts in last 30 days, positive WNv in birds in the last 30 days, and average temperature over the last 30 days. Liu et al. (2009)

found that while environmental and climate factors were significant predictors of WNv, these predictors appeared to be geographically dependent.

Cotar et al. (2016) examined mosquitoes in the Danube Delta between 2011-2013. They found significant positive and significant negative linkages between time-lagged temperature and precipitation data with regards to WNv infection rates. These findings point to the necessity of further time-lagged studies of WNv.

Rochlin et al. (2011) explored “the association between vector-borne WNv and habitat, landscape, virus activity, and socioeconomic variables derived from publicly accessible data” (para. 1). Using data from 2000-2004 and focusing on Suffolk County, New York, the researchers examined EEVs such as college education, distance to tidal wetland, number of senior households, road polygons, and vacant housing. The researchers highlighted the interdependence of socioeconomic and natural environments in the prediction of WNv outbreaks, finding the highest WNv human risk associated with middle class suburbia vice affluent suburbia and in the inner city. It was also noted that birds tested positive for WNv were not significant predictors, whereas habitat fragmentation by roads was an important factor. The researchers also found that “to be useful for disease surveillance and control program, a geographic human risk model should: (a) use predictors that are easily available and interpretable; (b) be accurate against independent data; and (c) generate outputs that can assist control decisions” (2011, p. 7).

Young et al. (2013) used landscape epidemiology to develop a national scale model of WNv using remotely sensed data. Manore et al. (2014) identified the importance of the combination of antecedent factors such as temperature, topology, precipitation, population density, economic status, bird migration, and positive cases of WNv in mosquito pools to the promotion of WNv outbreaks. According to Ozdenerol et al. (2013) and Petersen et al. (2013), evidence also shows that statistical models tailored to a specific locale are better suited to predict the unique interactions of EPS and EVS interdependencies. For example, the virus reflects a different biological fingerprint in Europe than it does in the United States.

Petersen et al. (2013) highlighted that within the United States, research shows that these biological fingerprints can vary within the diverse ecological conditions present in the individual states. Mancayo (2014) reported that University of Tennessee researchers looked at the use of biotic, non-pathogenic markers to identify WNv pathogenic events. However, the CDC (2013) and Elith et al. (2011) reported that while researchers have developed numerous statistical models using variations of these EEVs, the ability to generalize these models to specific cities or counties remains problematic.

Ozdenerol et al. (2013) conducted a comprehensive review of literature on WNv spatio-temporal dynamics and patterns. Covering 14 years of studies, this work categorized 47 WNv research papers into nine categories:

- Spatial analysis of human case incidence
- Spatial-temporal analysis of bird species

- Spatial analysis of horses
- Spatial modeling of mosquito pools
- Real-time geographic information systems (GIS) studies for WNv surveillance
- Habitat-based studies
- Remote sensing (RS) studies for early warning systems
- Spatial analysis of genetic variation
- Spatial uncertainty analysis

Ozdenerol et al. (2013) identified a dynamic and complex WNv decision-space related to these nine categories. The interconnected relationships between environmental, anthropological, and biological entities describe a system-level view of WNv that consists of endogenous and EEVs that are difficult, if not impossible, to capture in a single model. A wide range of spatial tools such as GIS, RS, and SaTScan were also identified, as statistical analysis techniques including Multiple Criteria Decision Analysis (MCDA), principal component analysis (PCA), discriminant analysis (DA), Local Moran's I, and applied regression models.

While the Ozdenerol et al. (2013) study specifically focused on spatial-temporal studies, several findings were relevant to my research. First, the dynamic synthesis of environmental, biological, and ecological antecedents is key to the spread of the virus. Public health officials and decision makers are "increasingly challenged to assess the prevalence and to determine common risk factors, as well as to track trends over time" (p. 5420). Aligned with the geographic area of this study, Ozdenerol et al. (2013) found that climate and temperature patterns contribute to Culex Quinquefasciatus (the common southern house mosquito) presence and the spread of the virus. Finally, the researchers stated that additional research over longer timescales is required to determine the efficacy of the EEVs used in the studies reviewed.

An ongoing challenge to the development and sustainment of mixed models lies in the access to, completeness, and timing of EID data. Current EVS techniques include the clinical testing of dead birds, mosquito pools, sentinel chickens, equine, and humans. According to Hobson-Peters (2012), while these procedures provide managers with a high degree of accuracy related to the presence of WNv, there is latency associated with collection and testing that opposes dynamic, predictive modeling for the use of early warning, decision support, and disease prevention. For example, the CDC (2013) stated that the “determination that a sentinel chicken has seroconverted occurs typically 3-4 weeks after the transmission event has occurred and reporting of a positive chicken may not precede the first local case of human disease caused by WNv” (p. 22). WNv models using this type of endogenous data to predict outbreaks would affect timely predictions.

Presenting itself asymptomatically, WNv can be masked by a low-grade fever or body aches attributable to different medical causes. Because of this, the CDC and local health organizations have developed reporting metrics that provide only positive incidents of WNv that have been diagnosed through serological sampling. Rochlin et al. (2011) found that incidents of WNv are likely greater and so the statistics associated with this EID phenomenon are conservative.

This conservative approach can be problematic for researchers. Using human outbreaks alone is problematic in that data over the study period yielded only 72 WNv incidents between 1999 to 2016 in SC. With the understanding that each incident represents a potential risk for human outbreaks of WNv, I elect to use all positive incidences of WNv (human, equine, mosquito pools, and sentinel animals) in a county to provide a broader collection of WNv presence. This number highlights clinical aspects of how the virus presents itself in humans and why sample sizes for these data sets are historically low.

According to Reiner et al. (2013), few studies have focused on the benefits of a regionally tailored decision support model that examines the predictability of WNv presence using a combination of web-based, exogenous antecedents. Within the United States, most of these studies were conducted in high WNv incidence areas within the Great Plains states, Texas, and California and were oriented to disease prediction using historical EEVs. Few studies have solely addressed the rich contextual environment associated with WNv resource decisions. The southeastern United States provided the appropriate level of ecological, environmental, and anthropological complexity for the analysis of the contextual decision-space associated with EID resource management. The ecosystems of the southern states and their longitudinal data provided a scientifically useful test environment for this study.

Summary and Conclusions

In this chapter, I first examined DT using Simon’s (1960) decision-making process as a grounded theoretical framework. Snowden and Boone's (2007) CYNEFIN construct was then interwoven into the Simon's (1960) framework to show how intelligence and design worked iteratively to identify and develop ordered, unordered, and disorders contexts. Uncertainty was also introduced as a contributing factor to course of action development and action in general. I then examined the literature associated with the predictive modeling of WNv. I looked at the presence and prediction of WNv and the key variables and concepts.

Past studies associated with the predictive modeling of EIDs were examined to determine how and what variables were used in past research. The accurate and timely prediction of an EID event was consistently highlighted as a critical component of public and environmental health management. Decisions informed by event likelihood result in critical resource allocation and preventive measures. Historically, the likelihood of EID outbreaks has been determined by biologically oriented statistical tools using historical EPS and EVS factors captured in field surveillance activities or in post-treatment scenarios. However, the CDC highlighted shortfalls in this approach:

> Despite these documented associations with a variety of biotic and abiotic factors, and recognition that certain regions experience more frequent outbreaks and higher levels of human disease risk, no models have been developed to provide long-term predictions of how and where these factors will combine to produce outbreaks. (CDC, 2013, p. 7)

> The literature review revealed that despite the availability of medical data, environmental surveillance tool sets, and heightened public awareness, reliable methods for predicting WNv outbreaks remain elusive (Manore et al., 2014). According to Manore et al. (2014) and Chevalier et al. (2014), statistical predictions of local WNv outbreaks with the reliability and timeliness required for EHMs to make pro-active public health decisions is increasingly problematic. Existing research also identified shortfalls in current EID predictive models due to a lack of robust empirical data, longitudinal analysis, and research exploring the exclusive use of EEVs (see, for example, Liu et al., 2009; Manore et al., 2014; Rochlin et al., 2011).

The literature highlighted system-level WNv variables that were classified as spatial, temporal, environmental, ecological, biological, and demographic. The review also highlighted studies that used a combination of EPS and EVS explanatory variables, but no studies have exclusively used EEVs to determine the likelihood of the presence of WNv in SC localities. The shortfalls identified in the literature were compounded by the fact that no current mechanism, model, or algorithm exists for the accurate and timely prediction of WNv outbreaks without robust EPS and EVS data (Chevalier et al., 2014; Manore et al., 2014). Specifically, there was a lack of understanding of the predictive power of EEVs when used by themselves or in combination when robust EPS and EVS data are unavailable.

Suthar et al. (2013) highlighted the fact that a lack of vaccine and approved therapy exists for WNv. This has caused EHMs around the world to seek decision support tools that can predict human outbreaks of WNv within the timeframes necessary to take preventive measures. According to Pirofski and Casadevall (2012), current tools remain problematic due to the dynamic interdependencies of the pathogenic, ecological, and anthropological components. The ability to gather real-time data and develop actionable intelligence on these dynamic interdependencies is like the complex environments faced by decision makers in many different disciplines. From a practical perspective, EHMs require more holistic, accurate, and timely decision tools for predicting the outbreak of EIDs like WNv. From a theoretical perspective, the integrated synthesis and application of DT, the CYNEFIN construct, and systems-thinking may offer an alternative approach to the WNv decision-making process.

Collectively, each of these topics contributed to the analysis to address the research questions. As the decision maker considers a decision-space, they construct a mental model that is more inclusive and holistic than traditional mechanistic and reductionist approaches. An expansion of the EHM decision-making process which more broadly incorporates exogenous data provided a broader frame of discernment in EID decision-spaces, particularly when EPS and EVS is lacking. Theoretically, this type of approach drives the decision maker to a system-level orientation; an approach better suited to the systems biology of EIDs (Gatherer, 2010).

In Chapter 3, I describe the research methods associated with this study. The findings in this chapter are carried into the study methodology, as are the EEVs revealed in past studies. The next chapter will specifically address the research question, hypothesis, data collection, sampling procedures, and data analysis.

Chapter 3: Research Method

The purpose of this ex post facto quantitative research was to examine the use of EEV data in predicting outbreaks of WNv in SC when robust EPS and EVS data are unavailable. In this chapter, I operationalize the study by connecting the purpose and hypotheses to specific research design and methodology. To fully develop the overall study framework, I address the sampling approach, data collection and EDA, regression model development and fit, threats to validity, and ethical procedures. The chapter includes a review of the potential alternative study designs leading to the selection of a quantitative approach to test the research hypotheses.

Research Design and Rationale

Using a correlation design enabled through BLR and GZLM regression, I examined the importance of exogenous EEVs in providing decision-making context in complex environments associated with EIDs. With WNv as the focus EID, I sought to develop five models using 0-, 30-, 60-, 90-day lags and a 90-day moving average time interval for predicting the presence of the virus in a locality through the sole use of EEVs. Candidate EEVs were identified through a systems-level review of the WNv decision-space, which revealed historical and new EEVs. All EEV data were checked against the underlying assumptions required of a linear model, and the specified models were then developed with presence of WNv as the DV. The five models were compared using goodness-of-fit statistics (e.g., various forms of R^2), mean absolute error (MAE) and the root mean square error (RMSE).

To examine the use of EEVs in the complex EID decision-space, I developed a number of predictive models in the context of the SC ecosystem. I did this for two reasons. First, SC is representative of the broader ecosystem of the southeast United States, having coastal, lowlands, marsh, lake, and mountainous terrain. This makes the applicability of results from this research generalizable to other temperate regions. Second, in 2012, SC led all states in the southeast region with a 290% increase in incidents of WNv, marking it as a research region of interest (CDC, 2018; SC DHEC, 2013).

Quantitative analysis was used to examine the statistical relationships between the DV and the EEVs. Due to the historical nature of the data, I chose an overall nonexperimental, ex post facto research design. This type of design is particularly well suited to areas of research that study naturally occurring cause and effect relationships such as the effectiveness of ongoing health programs (Vassar & Holzmann, 2013). A retrospective look allowed me to isolate an historic effect and examine the potential causes of that outcome.

Hypotheses and Research Questions

The following research questions examined the statistical utility of EEVs in predicting outbreaks of WNv in SC. In this research, a WNv incident is the positive identification of the virus in a locality in either a human, mosquito, bird, equine, or sentinel animal.

RQ1: In the absence of robust EPS and EVS data, which EEVs accurately predict incidents of WNv in SC in a current month?

$H_0$1: When used alone or in combination, EEVs do not accurately predict incidents of WNv in SC in the same month.

H_a1: At least one EEV accurately predicts incidents of WNv in SC in the same month.

RQ2: In the absence of robust EPS and EVS data, which EEVs accurately predict incidents of WNv in SC in the future?

$H_0$2: When used alone or in combination, EEVs do not predict incidents of WNv in SC 30 days later.

H_a2: At least one EEV predicts incidents of WNv in SC 30 days later.

$H_0$3: When used alone or in combination, EEVs do not predict incidents of WNv in SC 60 days later.

H_a3: At least one EEV predicts incidents of WNv in SC 60 days later.

$H_0$4: When used alone or in combination, EEVs do not predict incidents of WNv in SC 90 days later.

H_a4: At least one EEV predicts incidents of WNv in SC 90 days later.

Study Variables

To characterize the impact of EEVs on EID decision making, I collected longitudinal data (1999 to 2016) on incidents of WNv in SC and then developed and validated predictive models using BLR and GZLM regression. The EEVs are described here for clarification. For more detailed information on the study variables, see Appendix A.

Exogenous Explanatory Variables

To test the research hypotheses, nine independent, contextually derived EEVs were examined based on a systems-level review of the WNv decision-space. All nine EEVs were examined through simple correlation and stepwise regression analysis. Of the nine EEVs, *EV1ATM*, *EV2ARN*, *EV6AWS*, *EV8USE,* and *EV9POP* were identified in past studies and are considered historical predictors of WNv.

The EEVs are defined as follows:

Average 30-day temperature in degrees Fahrenheit (^{0}F*; EV1ATM*): Average temperature data for each county seat by year and month from 1999 to 2016.

Average 30-day rainfall (*inches; EV2ARN*): Average rainfall data for each county seat by year and month from 1999 to 2016.

Average 30-day dew point in degrees Fahrenheit (^{0}F*; EV3ADP*): Average dew point data for each county seat by year and month from 1999 to 2016.

Average 30-day snow depth (*inches; EV4ASD*)*:* Average snow depth data for each county seat by year and month from 1999 to 2016.

Average 30-day barometric pressure (*millibars; EV5ABP*)*:* Average barometric data for each county seat by year and month from 1999 to 2016.

Average 30-day wind speed (*MPH; EV6AWS*)*:* Average wind speed data for each county seat by year and month from 1999 to 2016.

Topology (*elevation in feet; EV7ELV*)*:* Elevation data were extracted from the South Carolina Aeronautics Commission (2011) website, which lists elevations for all county airports. The airport locations were matched with the county seats to provide a standard measure of county elevation.

Land use per county (*EV8USE*)*:* County land use was recorded as a numerical variable—either agricultural (0) or industrial (1)—based on population density and housing density as reported by the 2010 United States Census Bureau (USCB) for each county.

Urbanization (*population per square mile; EV9POP*)*:* County populations were collected from USCB data and converted to a ratio using county square miles data. The population density ratio provided a measure of county urbanization.

Dew point depression in degrees Fahrenheit (^{0}F; *EV10ADD*)*:* Average dew point depression data for each county seat by year and month from 1999 to 2016.

Dependent Variables

The DVs were developed as $DV_{PRESENCE}$ and DV_{COUNT}. $DV_{PRESENCE}$ indicated if WNv of any quantity occurred in a locality while DV_{COUNT} was the count of WNv incidents for a specific locality. The DVs for RQ1 were computed using no time lag. The DVs for RQ2 were computed using 0-, 30-, 60-, 90-day, and 90-day moving average interval data (time lags). This resulted in four sets of hypotheses and five predictive models. I tested the significance of each predictive model using an *F* test and its associated *p* value and the significance of EEVs using a *t* test and its associated *p* value. I evaluated final models using adjusted R^2, MAE, and RMSE (or, standard error of the estimate). The DVs were defined as follows:

$DV_{PRESENCE}$ and DV_{COUNT} (*DV1*)*:* The presence and count of SC WNv incidents in a current month and a county.

$DV_{PRESENCE}$ and DV_{COUNT} (*DV2*)*:* The presence and count of SC WNv incidents 30 days after (later than) the month.

$DV_{PRESENCE}$ and DV_{COUNT} (*DV3*)*:* The presence and count of SC WNv incidents 60 days later.

$DV_{PRESENCE}$ and DV_{COUNT} (*DV4*)*:* The presence and count of SC WNv incidents 90 days later.

$DV_{PRESENCE}$ and DV_{COUNT} (*DV5*)*:* The presence and count of SC WNv incidents over a 90-day moving average.

Methodology

I used an incremental process to examine the research questions and hypotheses. This methodology included data collection and preparation, EDA of EEVs to address BLR and GZLM regression assumptions, and data analysis (model development, comparison, and validation). In this section, I detail the study population, sampling procedures, procedures for data collection to include systems-level and archival data.

Population

The target population was all reported monthly WNv incidents within each county of SC from 1999 until 2016, *N* = 9,936 (18 years x 12 months x 46 counties). In this study, I used three terms to further describe the data that comprise the study population:

1. WNv case: A WNv case is equivalent to 1 month of EEV data as well as reported WNv event data for one county. There were 9,936 WNv cases between 1999 and 2016. A WNv case may or may not have a reported event in a case.
2. WNv event: A case that has one or more reported WNv incidents. Of the 9,936 cases, there were 360 WNv events reported between 1999 and 2016.
3. WNv incident: A single reported WNv incident. Among the 360 events reported between 1999 and 2016, there were 902 reported WNv incidents.

Representing the total of statewide WNv incidents over the 18-year period, the data were compiled from publicly available sources and as all personal data are masked, all individuals associated with the human cases of WNv were protected. Reporting of these cases was done monthly by the CDC and includes all reported positive WNv events in humans, equines, birds, surveillance animals, and mosquito pools. Publicly available CDC and SC DHEC data provided the 216 months of data for both the DV and EEVs across all 46 SC counties.

Sampling and Sampling Procedures

I included the entire population in this research. Therefore, no sample or randomized sampling technique was required. I expanded WNv events to include all reported incidents of WNv in a county; the total number of positive incidents of WNv across the 46 SC counties from 1999 to 2016 was 902.

Procedures for Recruitment, Participation, and Data Collection (Primary Data)

The EEV data associated with WNv outbreaks were extracted from real-time, publicly accessible data bases to include the United States Geological Survey National Land Cover Database 2006 for topographical information, intercensal population estimates created by the Federal State Cooperative Program for Population Estimates obtained from the U.S. Census Bureau, the U.S. Geological Survey disease maps (1999 to 2016), and the Old Farmer's Almanac. These data were initially compiled via MS Excel worksheets to create a source of EEV data by year, month, day, and county. Once the collected data were compiled in MS Excel, the data preparation process began.

To prepare the data for use in SPSS, daily averages for each of the nine EEVs were summed and averaged for each month between 1999 and 2016. These EEV monthly averages were then recorded with the monthly reported incidents of WNv. The result was an MS Excel worksheet row (or case) depicting the monthly average of each EEV and the reported incidents of WNv for that month and for each and every month from 1999 to 2016. Following this process, the data were checked for errors and omissions. Outcome data were created as the data for each month is recorded and as appropriate into the *DV1-5* columns.

Frankfort-Nachmias and Nachmias (2008) described the editing process as occurring during and after the coding process to check for errors and omissions in the complied data. Once editing was complete and reviewed for completeness, the data were cleansed to correct errors and to detect inconsistent coding. Data cleansing involved a second independent party review of the data. When data editing and cleansing were complete, the validated study data were entered into SPSS.

Within SPSS, each record constituted of a row entry by record number, county, date, date code, EEV (9) and outcome data (*DV1-5*). Administrative data consisted of a record number (1-9936), county code (01-46), and the serial MS Excel DATE Function for year, month, and day. System-level data consisted of a binary (0, 1) entry for land use and numerical entries for urbanization, topology, average monthly temperature, average monthly rainfall, average monthly dew point, average monthly snow depth, average monthly wind speed, average monthly barometric pressure, and the four DVs. Administrative, system-level, and outcome data were available for each month, by county, from 1999 until 2016. While the administrative data are self-explanatory, an additional explanation of the system-level data are required.

System-Level Contextual Data Collection and Preparation

The study's system-level data represented the contextual EEVs of interest based on the literature review. In this research, I constrained system-level EEVs to data readily available to a decision maker. This required two criteria. First, the data must have been accessible via the internet or social media. Second, the data did not require transformation. For example, I calculated simple averages for data associated with rainfall and temperature, but I did not transform the data into gradients such as "growing degree days" as this would transform the collected data to another measurement.

Archival Data

The historical data used in this research were archived by several different private and government organizations (see Table 2). WNv incident data come from two primary sources, the CDC and SC DHEC. An assumption of my research was that all data were publicly accessible and explicit permissions for use were not required. For reliability, the archival data were obtained from government sources first and then private sources. If required, private sources of data are identified within the study.

Instrumentation and Operationalization of Constructs

I used SPSS Version 25 to conduct my regression analysis. The SPSS tool is widely used in quantitative research requiring predictive modeling. The SPPS software is provided by Walden University to its students and thus permission had already been obtained by the institution.

Table 2

Web-Based Publicly Accessible Data Resource

Research data	Site name	URL
Federal Information Processing Series (FIPS)	United States Census Bureau; Geography	http://www2.census.gov/geo/docs/reference/codes/files/st45_sc_cou.txt
County name	United States Census Bureau	http://www2.census.gov/geo/docs/reference/codes/files/st45_sc_cou.txt
County seat	SC Association of Counties	http://www.sccounties.org/municipalities
County seat elevation	SC Aeronautics	http://www.scaeronautics.com/AirportList.asp
County square miles		http://www.census.gov/data/tables/time-series/dec/cph-series/cph-t/cpht-1.html
County population (2000-2010)	United States Census Bureau; Population Estimates	https://www.2census.gov/datasets/time-series/demo/popest/intercensal-2000-2010-counties.html
County population (2010-2019)	United States Census Bureau; Population Estimates	https://www.2census.gov/datasets/time-series/demo/popest/2010-counties-total.html
Land Use	United States Census Bureau	https://www.census.gov/quickfacts/fact/table/SC/LND110210
Temperature Dew Point Rainfall Snow Depth Wind Speed Barometric Pressure	Old Farmer's Almanac	https://www.almanac.com/weather/history/SC
Temperature Dew Point Rainfall Snow Depth Wind speed Barometric Pressure	Weather Underground	https://www.wunderground.com/history

Data Analysis Plan

The data analysis process for this research consisted of EDA, model development using BLR and GZLM regression, model comparison, and validation. Data analysis was planned using the SPSS Analyze/Regression/ Binary Logistic Regression and Analyze/Generalized Linear Model functions. During EDA, I checked for assumptions associated with BLR and GZLM regression and examined relationships between the EEVs and the DV. The data assumptions are covered in detail in Chapter 4.

Multiple Regression Analysis

The analysis used in this study consisted of variations of multiple regression analysis (specifically, BLR and GZLM regression). Provided here are the basic components of multiple regression analysis, adapted and used in this study.

The linear regression model is the following (see Equation 1):

$$Y = \beta_0 + \beta_1 X_1 \ldots + \beta_k X_k + \varepsilon \tag{1}$$

where

Y = the DV

β_0 = the Y intercept for the population

β_k = the slope for the population (the coefficient for the EV, X_k)

X_k = the kth EV

ε = random error in Y.

Hypotheses

Null hypothesis: The hypothesis for the significance of the overall multiple regression model, regarding the influence of the *Xs* on *Y*, is there is no linear relationship between the DV and the EVs, depicted mathematically as follows:

$$H_0: \beta_1 = \beta_2 = \ldots = \beta_k = 0$$

Alternative hypothesis: There is a linear relationship between the DV and at least one EV, depicted as follows:

$$H_a: \text{at least one } \beta_k \neq 0.$$

The hypothesis is tested against the overall model to see if there is a significant relationship between the DV and the entire set of EVs using the *F* test (and its associated *p* value). The *F* test assesses whether the set of EVs predicts the DV.

R^2, the coefficient of determination, indicates the extent to which the set of EVs contributes to the variance in the DV (the portion of variation in the DV that can be attributed to variation of the model). A *t* test determines the significance of each EV, independently, when the overall model is significant (i.e., the *F* test of the aggregate regression model is significant).

Model-Building

In this study, I used both automated and manual stepwise model-building approaches in combination with subject matter expertise to develop, evaluate, and compare various predictive models (different set of EVs). The stepwise regression approach to model-building is used to evaluate various regression models when considering the influence of individual EVs, including the two-factor interactions (2FIs) between EVs, and their contribution to the strength of the overall regression model. The stepwise process selectively adds or eliminates EVs to produce a model that is the best predictive model using adjusted R^2, which accounts for the number of EVs in the model.

SPSS Automated Stepwise Method

In the BLR regression process, I used all six of the automatic stepwise methods to inform my evaluation of time lagged DVs and early model development using EVs only. These methods included the SPSS Forward Stepwise Conditional (FS COND), Likelihood Ratio (FS LR), Wald (FS WALD), Backward Elimination Conditional (BE COND), Likelihood Ratio (BE LR), and Wald (BE WALD). All automated stepwise methods run automatically through an iterative sequence of models making choices for adding or eliminating predictors until it reaches the best model according to defined criterion; thus, for each method and DV, there is only one run, and the final model in each run is, by definition, the best model that SPSS chose at the end of the process.

The forward stepwise method enters EVs that are not in the equation using the smallest p value of F. Subsequently, EVs are removed if F becomes larger. The method concludes when there are no more variables for inclusion or removal. The backward elimination stepwise method begins with all EVs. EVs are then removed based on the significance of their correlation with the DV. The method concludes when there are no more variables that meet the removal criteria.

Manual Stepwise Method

I primarily used the SPSS Enter method in both the BLR and GZLM regression process. In this method, all variables are entered in a single block, with each run producing a single regression model based on the predictors chosen. When each run was complete, I evaluated the model goodness-of-fit statistics as well as other evidence to determine the best model from among many runs. Between runs, I would eliminate predictors based on their *p* value in relation to the specified α.

When I reached a model where all predictors had met the *p* value criterion, I then reviewed all models for that stage and selected the best predictive model using a combination of subject matter expertise and best goodness-of-fit statistic. This entire process is what SPSS automates in the other methods, but they can be flawed depending on the order in which predictors are added or eliminated; the manual process is a bit less impacted from that error because it adds analysis judgment and the strategic assessment of various combinations of EVs.

Final Predictive Model

The final predictive model includes EVs and significant 2FI terms. In the predictive model, $\widehat{Y}$ increases or decreases by the coefficients (b_j) for a unit increase for each EV. The final predictive regression model is the following (see Equation 2):

$$\hat{Y} = b_0 + b_1X_1 + b_2X_2 + b_kX_k \tag{2}$$

where

$\hat{Y}$ = the predicted value of the DV

b_0 = the Y intercept for the population

b_j = the slope or coefficient for the EV X_j including 2FI terms

X_j = the jth EV.

Model Evaluation and Validation

There were several tasks in evaluating and validating the models for estimating the presence of WNv:

Task 1: Test assumptions for the data set such as (for MLR) homogeneity of variance, linearity, independence (auto correlation and multicollinearity), and normality of residuals.

Task 2: Compare regression models for each DV_j, to decide which is the best predictor of incidents of WNv; essentially, determining which is the best DV since they are different measures (time lags) for the same response.

Task 3: Select the best model for any given DV (DV_j); essentially, what is the best set of X_is that predicts DV_j? This required a stepwise regression model-building process to select the best predictive model for DV_j. The hypothesis test (significance) for any regression model was the F test and its associated p value. Adjusted R^2 = coefficient of determination (or a suitable version) was used to evaluate the goodness of fit of any model. Adjusted R^2 reports the proportion of variation in Y that is explained by the regression model (the entire set of Xs). It is a tool for selecting the best set of Xs for any one of the DVs.

Task 4: Validate the chosen model to ensure that the predictive model of WNv is reasonable, plausible, and usable.

Task 5: Re-test assumptions for the final model.

Threats to Validity

Generalizability and control are two components of research design related to validity and were important to this study (Frankfort-Nachmias & Nachmias, 2008). Generalizability addresses the validity of an inference or finding to a broader population. Control ensures that an inference or finding is as free as possible from extrinsic and intrinsic factors (2008). To mitigate these threats to validity, the researcher must design and conduct the experiment in such a way that any statistical inferences gained from the results are as free from external and internal control issues as possible.

External Validity

Reliable data allow a researcher to draw inferences and conclusions from a sample that can then be generalized to a broader population. Known as external validity, this design component can be strengthened by having representative, random samples within the experiment. This requires the research population to reflect the general characteristics of a study group (Frankfort-Nachmias & Nachmias, 2008).

Because of the historical nature of ex post facto research, the research data in this study was prearranged rather than randomly selected and there was no manipulation of the EEVs for the purposes of then measuring the DV (as in an experiment). This ultimately affected the ability to generalize results of this study to a broader community and required mitigation measures.

For this research, I collated data from different publicly available sources to create the data sets. In retrospective studies, challenges can occur with the validity of the collected data. In this research, specific indicator data had been collected by different government agencies over three decades. The means of collection and the measurement tools used were historical and I could not control but relied on the accuracy of an external collection and analysis methodology. The objective of my study was to use publicly available data to develop a system-level context to a complex problem and so the externally derived data were assumed to be accurate and, even if not, represented the data that would be available to a decision maker to use in a predictive model.

To validate my findings, I mitigated external validity through two actions. First, census data derived from a contextual process ensured that alignment of the overall study variables with the larger system-level population of variables was used. These variables were also empirical, being associated with WNv across many counties within South Carolina. This required an understanding of how the data were collected and recorded. The preponderance of the data came from historical sources and required a heightened level of content analysis during data collection. Second, I used government data that have been subject to standardized collection metric and tools, subsequently reducing recall bias and ensuring the reliability and accuracy of the data.

Internal Validity

Internal validity is concerned with the reliability of data, the control of study factors, and with the validity of the resultant findings associated with factor/criterion analysis. In this research, I sought to determine the likelihood of WNv presence and incident counts based on several contextual factors. The selection of one factor over another can lead to selection bias and confounding factors. I mitigated these risks by qualifying the EEVs, first through review of past studies and secondly through cross tabulation and bivariate regression analysis to determine correlation strength with the DVs. The backward step-wise process was used in the regression and reduced the possibility of confounding factors.

Finally, as a retrospective study, all data were historical. While this presented external validity challenges, the use of archival data controlled internal problems with selection bias, mortality, and self-reporting data.

Construct Validity

Construct validity addresses the ability of an operationalized study methodology to answer the research question. A part of construct validity, convergent validity ensures the study methodology executed will yield results similar to past research and/or tests. In tandem, discriminant validity ensures that results are suitably discriminant to separate outcomes. This study benefited from a retrospective approach as it used historical data, validated EEVs, and a statistical tool well suited to predictive modeling. The use of historical data and EEVs provided good fidelity to the overall study.

Ethical Procedures

This research was conducted under the protocols required by the Walden University Institutional Review Board (IRB). The IRB approval number for this study is 08-20-19-0102708. There were no ethical concerns related to data collection as all data will be extracted from publicly available sources. All human incidences of WNv were masked by the CDC in the reporting data and so there are no vulnerabilities to individuals.

During the data collection and analysis phase of the research, all data were stored on a separate hard drive and will be protected by computer antivirus software. At the completion of the research, the data were saved to hard drive and stored in a secure space for five years to facilitate review. This research was not conducted within my workplace environment and so there will be no conflicts of interest.

Summary

Due to the nature of the data, a longitudinal ex post facto design was planned to examine the research question and hypotheses. Archival data are collected for the years 1999 to 2016 (N=9,936) to support model development and testing. MLR was planned to perform the regression with a level of significance of $p < \alpha = .20$. Hypothesis testing was conducted using appropriate goodness-of-fit tests and associated p values. To compare the effectiveness of one regression model compared to another for any one of the five DVs, I evaluated the fitness of a model for each DV using adjusted R^2 to indicate the amount of variation in the DV attributed to the model. The regression models were developed with 2002 to 2016 data. I then compared the best predictive model for each DV using goodness-of-fit metrics, to determine the best overall predictive model of WNv.

To minimize threats to external and internal validity, several mitigation steps were taken. Census data collected for this research was based on a contextually derived process that ensures alignment of the overall study variables with the larger system-level population of variables that are historically recognized as being associated with WNv. This consisted of the entire population of reported WNv events, enhancing both external and construct validity. Second, I used government data that have been subject to standardized collection metric and tools, subsequently reducing recall bias and ensuring the reliability and accuracy of the data.

Internal validity is concerned with covariance and confounding factors that will affect any casual inference drawn from research results. I mitigated these risks by qualifying the EEVs, first through review of past studies and secondly through cross tabulation and bivariate regression analysis to determine correlation strength with the DVs. MLR was chosen as the planned statistical tool which also assumes linearity of data.

Chapter 4: Results

The purpose of this ex post facto quantitative research was to examine the use of EEV data in predicting outbreaks of WNv in SC when robust EPS and EVS data are unavailable. To address the research gap of accurate and timely predictive modeling of WNv, I developed two RQs and supporting hypotheses to examine the statistical utility of EEVs in predicting outbreaks of WNv in SC. The two RQs were developed to address both the presence and count of WNv:

RQ1: In the absence of robust EPS and EVS data, which EEVs accurately predict incidents of WNv in SC in a current month?

RQ2: In the absence of robust EPS and EVS data, which EEVs accurately predict incidents of WNv in SC in the future?

To address these RQs, I posited nine EEVs: (a) average 30-day temperature (*EV1ATM*), (b) average 30-day rainfall (*EV2ARN*), (c) average 30-day dew point (*EV3ADP)*, (d) average 30-day snow depth (*EV4ASD*), (e) average 30-day barometric pressure (*EV5ABP*), (f) average 30-day wind speed (*EV6AWS*), (g) elevation (*EV7ELV*), (h) land use (*EV8USE*), and (i) urbanization (*EV9POP*; see Table 1). The nine EEVs were proposed based on a systems-level review of the WNv decision-space and with the requirement that they be available from publicly accessible data sets. Two DVs were used to represent WNv outcomes in this study: $DV_{PRESENCE}$ and DV_{COUNT}, numerical variables suited to the selected research statistical tools of BLR and GZLM regression.

In this chapter, I provide the results of my data collection and analysis. Specifically, I address any changes to the previously defined data collection plan, provide descriptive analysis, describe the BLR and GZLM regression model-building process, and address hypotheses testing and the resultant findings.

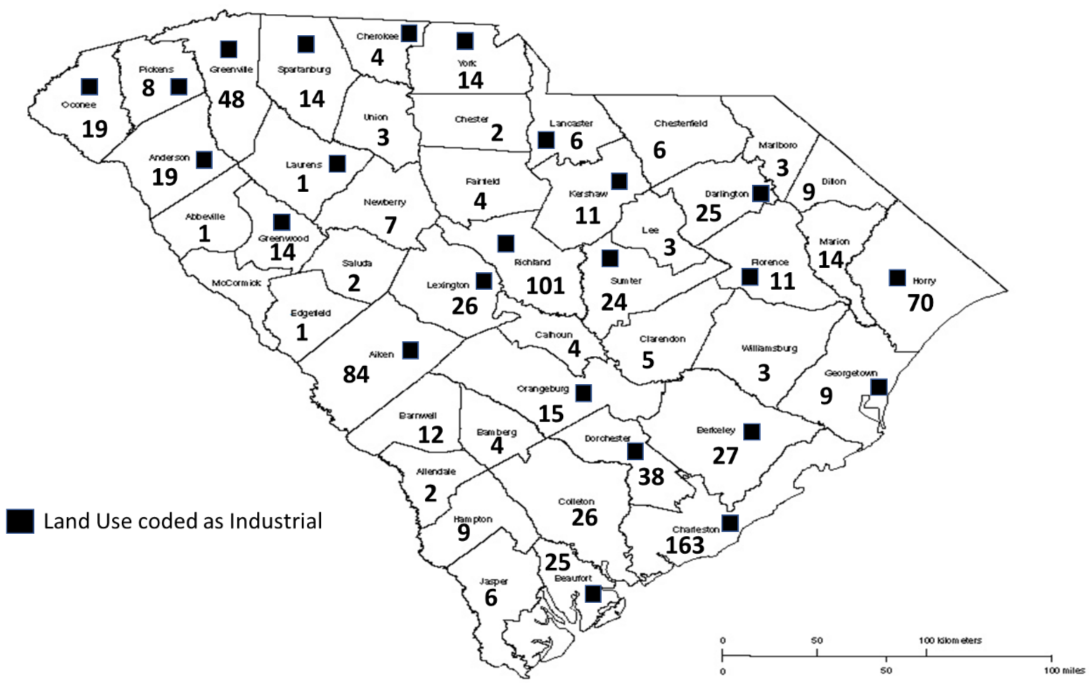
Data Collection

All WNv incident data were approved through an internal SD DHEC IRB process (SC DHEC IRB #19-011; see Figure 4). Following approval of both the Walden University and SC DHEC IRBs, data were collected from August 2019 through April 2020. EEV data used within the study were collected within the same timeframe through the internet from several primary sources that are readily accessible to the public, a primary assumption of this study.

During the EEV data collection process, more than one internet source was required to collect information on the proposed weather-related EEVs as data were sometimes missing for a particular city or region. While the Old Farmer's Almanac was used as the primary source for publicly accessible weather data, there were instances where Weather Underground (see Table 2) data were required as a secondary source. There were also instances where weather reports for a specific time period and region were not recorded. In these cases, the data from the next closest weather station were used to supplement the data.

Figure 4

Numbers of Reported WNv Incidents for All SC Counties, 2002 to 2016



Note. SC recorded WNv 360 events totaling 902 incidents. Land use determined by population per square mile.

The WNv incident and EEV data were collected for a 17-year period (1999 to 2016) resulting in N = 9,936 cases. Each of the 9,936 cases represented a month of reported WNv incidents (either zero or positive counts) and the associated EEV data (e.g., average temperature, average rainfall, elevation) for a single county. Figure 4 shows a global summary of WNv incidents by county in SC. Figure 5 shows the frequency of

WNv incidents by year over the 15-year period. This descriptive statistic guided subsequent decisions on data analysis in this study.

Data were collected for the nine EEVs, and the SC DHEC provided truth data for reported WNv events and incidents in SC from 2002 to 2016. During that time period, the state reported 360 WNv events consisting of 902 WNv positive incidents. These incidents are reported by year (Figure 5) and month (Figure 6). Data for the years of 2002 (81), 2003 (344), 2005 (80), and 2012 (130) accounted for 70% of reported WNv incidents. When the data were examined by monthly counts, July through October accounted for 90% of reported incidents and constituted the main grouping of WNv incidents over the study time period (Figure 6).

Figure 5

Histogram Showing WNv Incidents by Year, 2002 to 2016

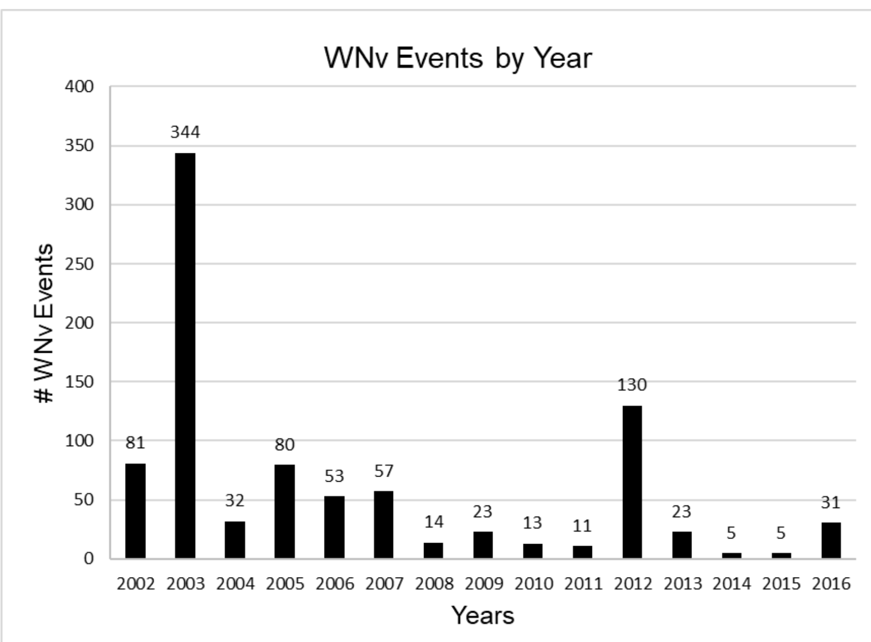
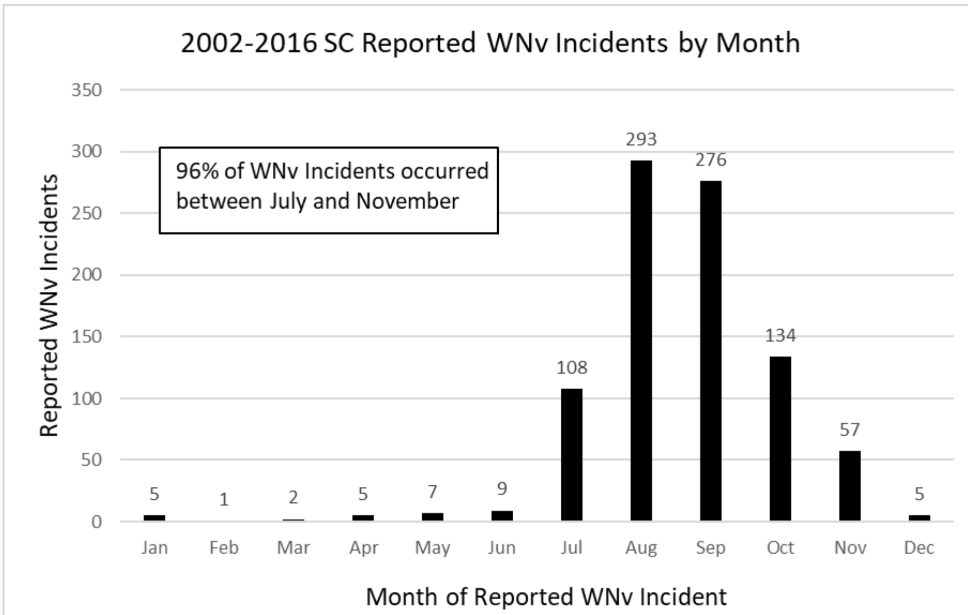


Figure 6

Histogram Showing SC WNv Incidents by Month, 2002 to 2016



Changes to the Planned Data Analysis

When all WNv data were recorded for the 9,936 cases, I performed EDA to check the assumptions associated with the planned analysis. As a result of the EDA, I found that the assumption of linearity could not be met due to the distribution of the DV. The data showed that 96% of the WNv cases resulted in a "0" count, resulting in a positively skewed (18.23), platykurtic (460.50) distribution (see Figure 7 and 8). Having violated the linearity assumption, I could not proceed with my initial plan of MLR without changes to the planned analysis.

Kurz (2017) remarked on this distribution problem in an article on health care utilization cost data. Kurz highlighted that health care data can be problematic because “the non-negative response variable is often zero because of non-users, while the positive realizations are usually right-skewed” (p. 1). He recommended a modeling approach to overcome this problem by combining binary and nonlinear probability distributions (e.g., Poisson distributions) using the family of Tweedie distributions (p. 2). The Tweedie distribution falls within the exponential family of distributions.

Ozaltin and Iyit (2016) stated that “generalized linear models (GZLMs) include regression models based on the exponential family of distributions” (p. 1). Slavkovik (2020) highlighted several advantages of GZLMs in Pennsylvania State University’s online statistics course, STAT 504, an introduction to GZLMs. First, there is no requirement to transform the DV to achieve a normal distribution, GZLMs do not assume a linear relationship between the DV and the EEVs, and the homogeneity of variance does not need to be satisfied. Also, with GZLMs, errors need to be independent but not normally distributed, they rely on large sample approximations, and goodness-of-fit measures rely on sufficiently sized samples. Based on my WNv data set and this information, I changed my planned data analysis to incorporate Kurz’s (2017) thoughts on combining binary and nonlinear probability distribution in the modeling process.

Using Kurz (2017) as a guide, I restructured my analysis to use BLR to predict the presence of WNv (RQ1) and GZLM regression to predict WNv incident counts (RQ2). I modified the approach originally proposed and documented in Chapter 1 in three ways. First, I refined the data structures for each type of regression by first removing the cases (1,656) associated with the years 1999 to 2001 as SC did not collect WNv data in those years. This left a master data set of 8,280 cases with 360 events consisting of 902 reported incidents of WNv. Secondly, I created a new set of DVs to reflect the separation of events (presence) from the number of incidents (count) and to support prediction of WNv in the future. Finally, I tested an expanded set of hypotheses to account for the modified data sets and DVs. As BLR and GZLM regression were not discussed in Chapter 3, I include a description of each in the following subsections.

BLR Overview

BLR is particularly useful in health care research when determining the likelihood that a patient has a particular disease. BLR predicts a probability of a binary outcome, using a dichotomous categorical DV, as a function of a set of continuous and categorical EVs. I used BLR to predict the presence of WNv in a county ($DV_{PRESENCE}$). Within the data set, $DV_{PRESENCE}$ was coded dichotomously: No WNv present = 0, WNv present = 1.

Field (2013) stated that “logistic regression is based on this principle: it expresses the multiple linear regression equation in logarithmic terms (called the *logit*) and thus overcomes the problem of violating the assumption of linearity” (p. 762). In logistic regression, the probability of Y is predicted given the known values of the EEVs and factors. The logistic regression equation is expressed by Equation 3:

$$\text{Logit} = L_i = B_0 + B_1X_1 + \ldots + B_KX_K \tag{3}$$

In this study, the logit was the natural log of the odds of the dichotomous response (an outbreak of WNv), see Equation 4:

$$L_i = ln\left[\frac{\hat{p}}{1-\hat{p}}\right] \tag{4}$$

where $\hat{p}$ (*p-hat*) was the predicted probability of an outbreak.

The predicted probability of an outbreak was calculated using Equation 5:

$$p_i = \frac{e^{L_i}}{1+e^{L_i}} \quad (5)$$

To interpret and evaluate the models produced in logistic regression, and for overall fit and model comparison, SPSS provides a number of statistical measures: *-2LL* (log likelihood) chi square statistics, Cox and Snell's R^2, and Nagelkerke's pseudo R^2, the Hosmer and Lemeshow test, the Wald statistic, coefficient values and associated significance, EXP(b) odds ratio, and the 95% confidence level. SPSS also allows the user to capture regression residuals for diagnostic purposes.

Based on Field's (2013) recommendation, I elected to use the pseudo $R^2_{\text{Nagelkerke}}$ rather than $R^2_{\text{Cox and Snell}}$ as my primary goodness-of-fit metric. Throughout my analysis, I followed and recorded the $R^2_{\text{Cox and Snell}}$ and the $R^2_{\text{Hosmer and Lemeshow}}$ statistic tests for completeness and consistency with the SPSS model outputs.

GZLM Overview

Javaras and Vos (2020) and Ozaltin and Iyiy (2018) stated that GZLM regression should be considered in situations where the DV values are greater than zero, the data are heavily skewed, and variables are not normally distributed. Due to the distribution of the nonnegative and positive count WNv data, I determined that statistical tools that address forms of nonlinear data such as Poisson or the Tweedie distributions could be used.

Within SPSS, GZLM regression appeared to be the appropriate analysis tool as it would allow for analysis of distributions other than normal and where the relationships between DVs and EEVs would not need to take on a simple linear form.

Javaras and Vos (2002) described GZLM regression as consisting of three components performing specific functions within the model: random component, link function, and systematic component. These are described as follows:

- The random component addresses the probability distribution of the DVs Y_1, Y_2, . . . ,Y_n and is given by (see Equation 6)

  $$E(\gamma_1) = \mu_i \tag{6}$$

- The systematic component produces a linear predictor $\underset{\sim}{\eta}$ of the covariates X_1, X_2, . . . , X_p given by (see Equation 7)

  $$\underset{\sim}{\eta} = \sum_{i-1}^{p} x_j \beta_j \tag{7}$$

- The link component describes how the systematic and random components are related and is given by (see Equation 8)

  $$\underset{\sim}{\eta} = \mathrm{g}(\mu_1), \text{ where } \mu_1 = E(\gamma_1) \tag{8}$$

Within SPSS, I structured the random component of the GZLM regression as a custom model, with a Tweedie distribution mean-variance power parameter (MVP) initially set at 1.5, and the link function set at Log $f(x) = \ln(x)$. While the link function remained the same for all GZLM regression runs, I changed the Tweedie MVP when an interim or final model was produced. SPSS allows the researcher to set the MVP between 1.1 and 1.9 to best fit the data distribution. According to Ozaltin and Iyit (2018), the

Tweedie distribution is appropriate for variables that take on nonnegative values and can mass at a single value such as zero. The log link is good for any distribution (IBM Knowledge Center, n.d.).

To enable the GZLM analysis, I created a data set that included only cases for which there was at least one recorded positive WNv incident. The GZLM data set was used to examine DV_{COUNT}, predicting the presence of WNv by count for each county using the 0- (*DV01*), 30- (*DV02*), 60- (*DV03*), 90-day (*DV04*) lags, and 90-day moving average (*DV05*) time interval.

The Analysis Process

The analysis process was accomplished in three stages (Table 3). Modeling for $DV_{PRESENCE}$ was accomplished with BLR in Stage A. Modeling for DV_{COUNT} was accomplished with GZLM regression in Stages B and C. Each stage followed an iterative process that addressed EDA, tests of assumptions, current and time lagged DV selection, and predictive modeling.

Stages A2 and B2 determined which time lagged DVs would be used for the respective BLR and GZLM regression modeling using SPSS Automated Stepwise Methods (SASMs) and the SPSS Manual Stepwise Method (SFEM). In Stages A2 and A4, SASMs were used to compare and then identify the best time lagged DV using EEVs only. Once I determined that the different time-lagged DVs were capable of developing significant models, based on their $R^2_{\text{Nagelkerke}}$ and *LL Ratio* χ^2 scores, the remainder of the $DV_{PRESENCE}$ analysis used the best time-lagged DV with all EVs (Stages A3 and A5).

In Stage B2, I compared the time lagged DVs (*DV2*, *DV3*, *DV4*, *DV5*) to select the best DV_{COUNT} variable using EEVs as the only predictors. Once I determined that the different time-lagged DVs were capable of developing significant models, the remainder of the DV_{COUNT} analysis used the best time-lagged DV with all EVs. Although the model-building process used the same α values as Stage A, GZLM regression required the random and link components of the tool to be set. To compare the DV_{COUNT} outcomes in the final GZLM regression results, I ran two separate GZLM regression analyses in Stages B and C. I then compared the results of each stages' best model.

While Stages B and C both addressed the GZLM modeling efforts using *DV5*, Stage C was an excursion with GZLM regression that started with the Stage A5 final terms. The rationale for Stage C was, if the BLR model terms could predict a WNv event, could the same EEVs then be used to determine, with an acceptable degree of accuracy, the number of WNv incidents (DV_{COUNT}) for that event. In theory, such a process would allow a simple, straightforward approach for EHMs to determine local $DV_{PRESENCE}$ and DV_{COUNT} outcomes using the same set of EEVs, but with different coefficients and different DVs.

This rationale was challenging because the BLR and GZLM regression analyses are and were vastly different. Each used a different best DV and had different data sets. It would be unlikely for two separate regression analyses to develop a final model with the same EEVs, when the data set and the DV are different.

Table 3

The Study Analysis Process by Stages

Stage	Description	Models	α
A1	$DV_{PRESENCE}$ Descriptive Statistics, EDA, and Assumptions		
A2	$DV_{PRESENCE}$ Analysis for Current Month, EEVs Only (SASM)	1-6	.20
A3	$DV_{PRESENCE}$ Analysis with *DV1* (SFEM)	7-38	.20
A4	$DV_{PRESENCE}$ Analysis for Best Time-lagged DV Selection, EEVs Only (SASM)	39-56	.20
A5	$DV_{PRESENCE}$ Analysis with *DV2*, All EVs (SFEM)	57-90	.20
A6	$DV_{PRESENCE}$ Analysis with *DV2*, All EVs (SFEM)	91-100	.05
B1	DV_{COUNT} Descriptive Statistics, EDA, and Assumptions		
B2	DV_{COUNT} Analysis with *DV1* (SFEM)	101-103	.20
B3	DV_{COUNT} Analysis with *DV1* (SFEM)	104-135	.05
B4	DV_{COUNT} Analysis for Best Time-lagged DV Selection (SFEM)	136-152	.20
B5	DV_{COUNT} Analysis with *DV5* (SFEM)	153-182	.20
B6	DV_{COUNT} Analysis with *DV5*, All EVs, α = .05 (SFEM)	183-185	.05
C1	DV_{COUNT} Analysis using $DV_{PRESENCE}$ Final Model Terms (SFEM)	182, 185	.05
C2	Comparison of DV_{COUNT} Stage B and C Models		

Note. SPSS Automated Stepwise Methods = SASM, SPSS Force Entry Method = SFEM

Moreover, beginning the model-building process with an arguably arbitrary set of EEVs is inconsistent with rigorous automated or manual stepwise model-building processes for which the outcome (final predictive model) is composed of a set of EEVs that are highly dependent on their order of elimination. Therefore, I performed both GZLM analyses with the intent to compare their final models in terms of their predictability with an understanding of the challenges.

Although I used different statistical tools for modeling $DV_{PRESENCE}$ and DV_{COUNT}, the general process was the same. The BLR Stages A2 to A5 interim models were developed using α = .20, $R^2_{\mathrm{Nagelkerke}}$ for goodness-of-fit, individual EV significance (p value compared to α = .20), and SME judgement. When models had the same or comparable $R^2_{\mathrm{Nagelkerke}}$ scores, I would also consider the model χ^2 statistic (to compare the significance of the models to the null hypothesis) as well as additional measures in the SPSS output. Using these criteria, each stage began with all nine EEVs, moving iteratively through build stages (Table 3) adding and eliminating EVs, using a combination of SASM and SFEM until an interim model was developed using α = .20 (i.e., to be included, an EEV's p < .20). The model was then examined at α = .05 (Stage A6) and compared to the interim models to determine the best and final stage model.

The choice of α = .20 initially was based on the philosophy that when trying to find the best predictive model (set of EEVs), the overriding criterion is the overall model goodness of fit, not the significance of individual EEVs. This process is more likely to produce a better goodness of fit, even if some of the terms were not individually statistically significant. This is especially appropriate since the data were a census and not a sample of the population. This philosophy also acknowledges that the true significance of any one EEV is, in fact, highly dependent on the presence of other EEVs. In other words, there are many subtle, multi-variate interactions at work in the real world, rendering a judgment on the significance of any one EEV somewhat meaningless when trying to find the best predictive model. The danger in such an approach is to over-specify the model with too many EEVs whose influence was in fact somewhat random. The check on this problem was to make many, manual model changes and to evaluate each of them for goodness of fit and for the consistent presence of EEVs throughout the analysis.

The GZLM interim models were also developed using $\alpha = .20$. I assessed Deviance (*D*), Akaike Information Criterion (*AIC*), and *LL Ratio* χ^2 scores for goodness-of-fit along with individual EEV significance (i.e., *p* value compared to $\alpha = .20$), and used SME judgement. The GZLM modeling required MVP and link settings, which were initially set at 1.5 and Log respectively. Using these criteria, each stage began with the EEVs, moving iteratively using the SFEM through build stages (Table 3) adding predictors (2FIs and *Months*) until an interim model at $\alpha = .20$ was achieved. The GZLM interim model underwent an additional step that allowed for the fine tuning of the Tweedie distribution by changing the MVP to enhance the *D* value. The model was then examined at $\alpha = .05$ and compared to the interim model to determine the best and final stage model. Once a model (BLR or GZLM) was determined to be the final (and best) model, residuals were captured, and the models expressed in a regression equation. Finally, model accuracy statistics were computed (MAE and RMSE).

It is important to note that SPSS was the only statistical tool used during the study. This meant I was reliant on either SASM (backward, forward, etc.) or the SFEM (a series of Enter method regression runs) to assess different model compositions based on pre-specified criteria (goodness-of-fit; and term significance, relative to α); or a combination of these. This made the final model composition highly dependent on the starting EEVs.

Analysis Results

This section presents the results for the $DV_{PRESENCE}$ and DV_{COUNT} analysis by stage. Each stage of analysis is addressed separately and is indexed in Table 3 by the model numbers associated with that stage of the model building process. Each stage in $DV_{PRESENCE}$ and DV_{COUNT} sections are also supported by tables and figures that describe how interim models were developed. More descriptive versions of those tables which provide more detail on the modeling process are included at Appendices B (Stages A2 to A6, Tables B1 to B12), C (Stages B2 to B6, Tables C1 to C8), and D (Stage C1, Table D1). Selected Stages A to C SPSS parameter output estimates are also provided in Appendix E (Tables E1 to E5). The study variables for all stages of modeling are shown at Table 4.

$DV_{PRESENCE}$ **Results**

The next section begins with the Stage A1 $DV_{PRESENCE}$ descriptive statistics, EDA, and assumptions. Stage A1 sets the stage for the deeper level of regression analysis required in Stages A2 to A4. Following the BLR assumptions, the $DV_{PRESENCE}$ analyses are provided.

Stage A1 $DV_{PRESENCE}$ Descriptive Statistics, EDA, and Assumptions

The Stage A variables consisted of $DV_{PRESENCE}$, nine EEVs, and one EV (Table 4). Measures of central tendency and associated variances, frequencies, and percentages are also presented as they provide valuable insights into the subsequent regression analysis and model-building process (Table 5). Figures 7 through 16 provide histograms illustrating the distribution of values for the study variables.

Stage A1 Descriptive Statistics and EDA. During the Stage A1 EDA, I determined an unacceptable collinearity between two variables representing temperature (*EV1ATM*) and dew point (*EV3ADP*). To overcome this challenge, I eliminated *EV3ADP* as an EEV and created an additional EEV (*EV10ADD)* from the baseline data collected for *EV1ATM* and *EV3ADP*.

The variable *EV10ADD* was calculated as *EV1ATM* minus *EV3ADP* providing a scale value tied to temperature. This variable is also known in the literature as *dew point depression* or *dew point deficit* and is defined as the "difference in degrees between the air temperature and the dew point" (American Meteorological Society, 2012). It is a measure of humidity—a dew point deficit of zero would equate to a moisture-saturated environment (humid conditions) whereas a large dew point deficit indicates dry conditions.

Table 4

Updated Study Variables

Variable type	Description	Label	Scale	Measure
Dependent	Presence of WNv	$DV_{PRESENCE}$	Categorical	0 (No WNv present) 1 (WNv present)
Dependent	WNv Incident Count	DV_{COUNT}	Interval	Count of WNv incidents
Independent	Average Temperature	*EV1ATM*	Ratio	Average 30-day County temperature in degrees Fahrenheit (ºF)
Independent	Average Rainfall	*EV2ARN*	Ratio	Average 30-day County rainfall in (inches)
Independent	Average Snow Depth	*EV4ASD*	Ratio	Average 30-day snow depth in inches
Independent	Average Barometric Pressure	*EV5ABP*	Ratio	Average 30-day County barometric pressure in inches of mercury (HG)
Independent	Average Wind Speed	*EV6AWS*	Ratio	Average 30-day County wind speed in miles per hour (MPH)
Independent	Topology	*EV7ELV*	Ratio	County seat elevation in (feet)
Independent	Land Use	*EV8USE*	Categorical	0 (Agricultural) 1 (Industrial/Commercial)
Independent	Urbanization	*EV9POP*	Ratio	Population density by County in (square miles)
Independent	Average Dew Point Deficit	*EV10ADD*	Ratio	Average 30-day County dew point difference in degrees Fahrenheit (ºF)
Independent	Months	*EVMonth*	Categorical	0 (Month not associated with case data) 1 (Month is associated with case data)

The Stage A1 descriptive statistics revealed a very low mean of .002 inches for *EV4ASD* (Table 5). An exploratory analysis of the *EV4ASD* variable showed that 108 of the 8,280 cases (1%) had recorded snow depth values greater than zero. Further evidence showed that only 19 of 46 counties recorded snow depth data more than once over the study period. These 19 counties were evenly distributed across the state. I elected to retain *EV4ASD* based on the collinearity and VIF scores but monitored it during Stage A.

Table 5

$DV_{PRESENCE}$ Descriptive Statistics for Post-EDA EEVs, BLR Dataset

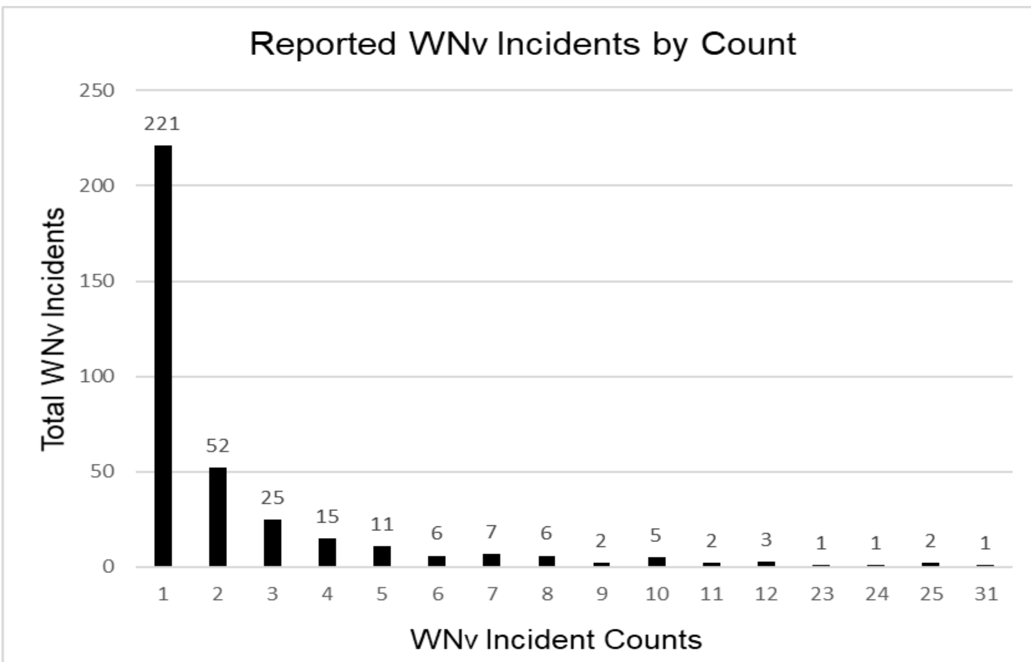
EEV	*N*	*Minimum*	*Mean*	*Maximum*	*SD*
EV1ATM	8280	23.57	63.39	97.98	13.30
EV2ARN	8280	0.00	0.14	10.52	0.30
EV4ASD	8280	0.00	0.00	1.00	0.03
EV5ABP	8280	977.02	1017.31	1025.25	4.38
EV6AWS	8280	1.51	5.20	11.57	1.33
EV7ELV	8280	10.00	353.91	1093.00	292.87
EV8USE	8280	0.00	0.50	1.00	0.50
EV9POP	8280	25.53	119.18	574.72	119.18
EV10ADD	8280	0.00	3.89	34.09	3.89
EVMonth	8280	0.00	0.50	1.00	0.50

Note. The BLR dataset included 8,280 cases.

During the descriptive statistics analysis, I noted that reported WNv incidents were heavily skewed (61%) to one reported incident per event (Figure 7). This frequency was important as I compared the GZLM WNv count results to the census data. To account for any autocorrelation or seasonal pattern in the time series data, I added coded variables for the months January to December for each case in the data sets for BLR and GZLM regression. For example, if the WNv case was associated with the month of January, the variable *January* would be coded as a 1 and the rest of the months would be coded 0.

Figure 7

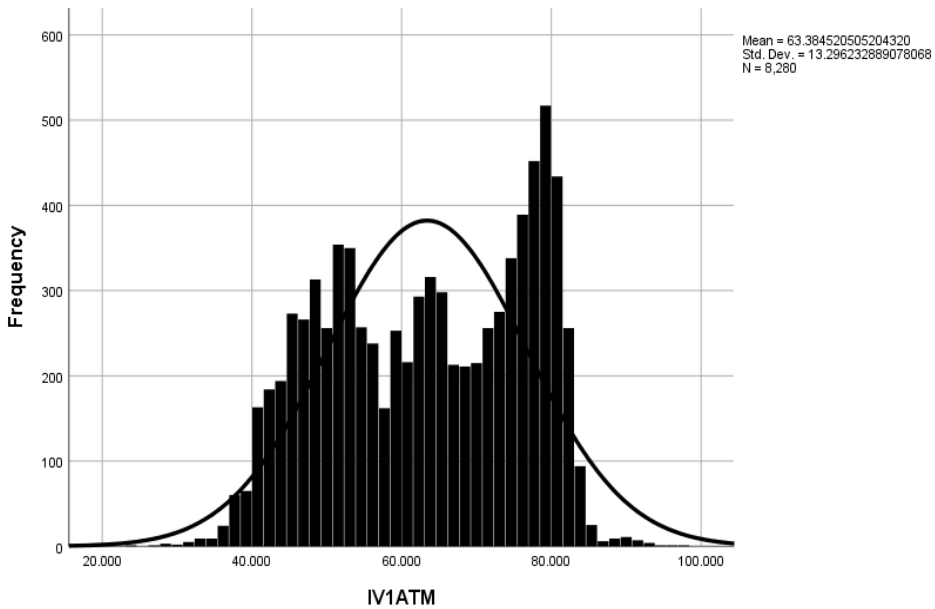
Reported WNv Incidents by Frequency and Incident Count



Average Temperature EEV (EV1ATM). The mean monthly temperature was 63.39 (SD = 13.30). The minimum temperature recorded was 23.57 and the maximum was 97.97, indicating a range of 74.40 (Figure 8).

Figure 8

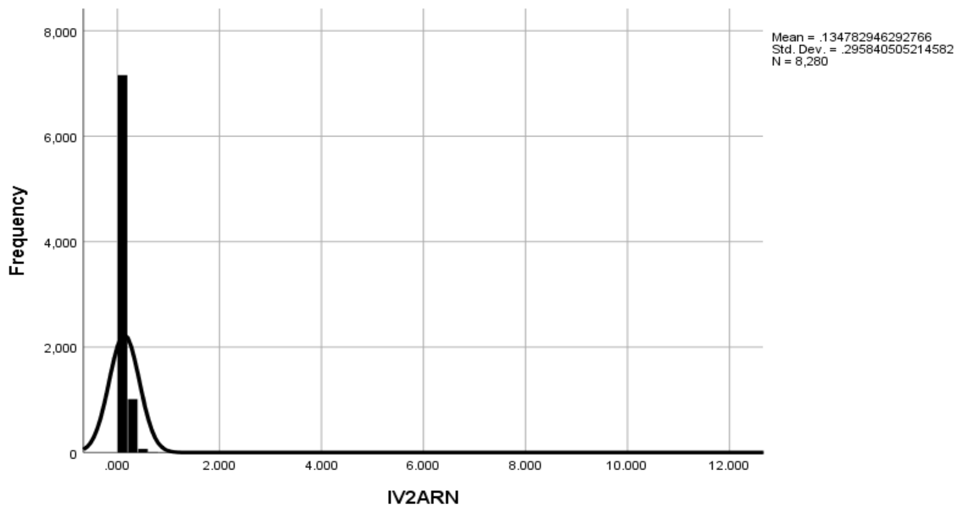
Histogram Showing the Distribution of EV1ATM



***Average Rainfall* (*EV2ARN*).** The mean rainfall across all years of the study was 0.14 (*SD* = .296). The minimum rainfall recorded was 0.00 and the maximum was 10.52, indicating a range of 10.52 (Figure 9).

Figure 9

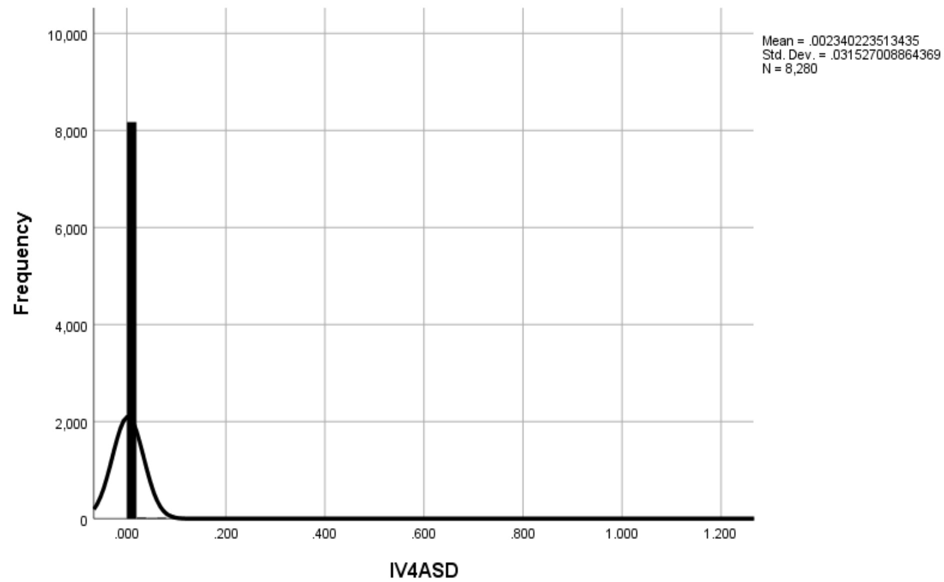
Histogram Showing the Distribution of EV2ARN



***Average Snow Depth* (*EV4ASD*).** The mean snow depth across all years of the study was 0.002 (*SD* = .032). The minimum snow depth recorded was 0.00 and the maximum was 1.00, indicating a range of 1.00 (Figure 10).

Figure 10

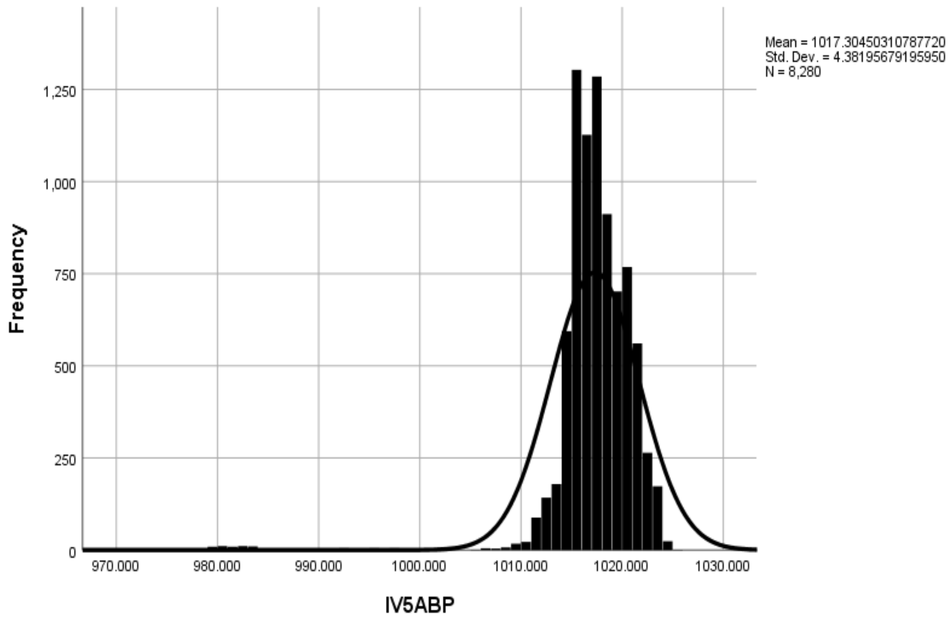
Histogram Showing the Distribution of EV4ASD



Average Barometric Pressure **(*EV5ABP*).** The mean barometric pressure in millibars across all years of the study was 1017.30 (*SD* = 4.38). The minimum barometric pressure recorded was 977.02 and the maximum was 1025.25, indicating a range of 48.23 (Figure 11).

Figure 11

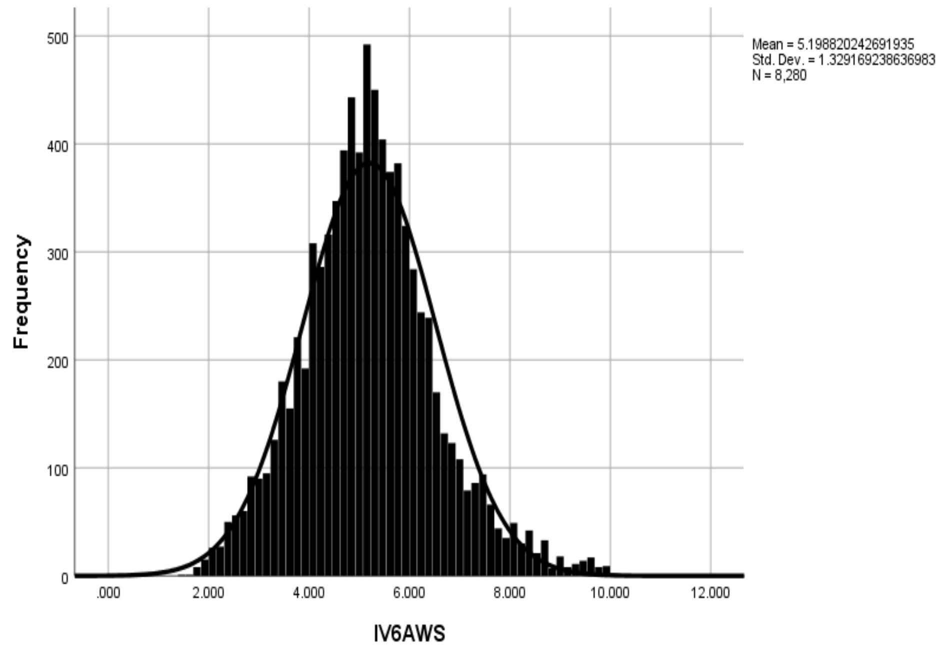
Histogram Showing the Distribution of EV5ABP



***Average Wind Speed* (*EV6AWS*).** The mean wind speed in knots across all years of the study was 5.20 (*SD* = 1.33). The minimum wind speed recorded was 1.51 and the maximum was 10.06, indicating a range of 8.55 (Figure 12).

Figure 12

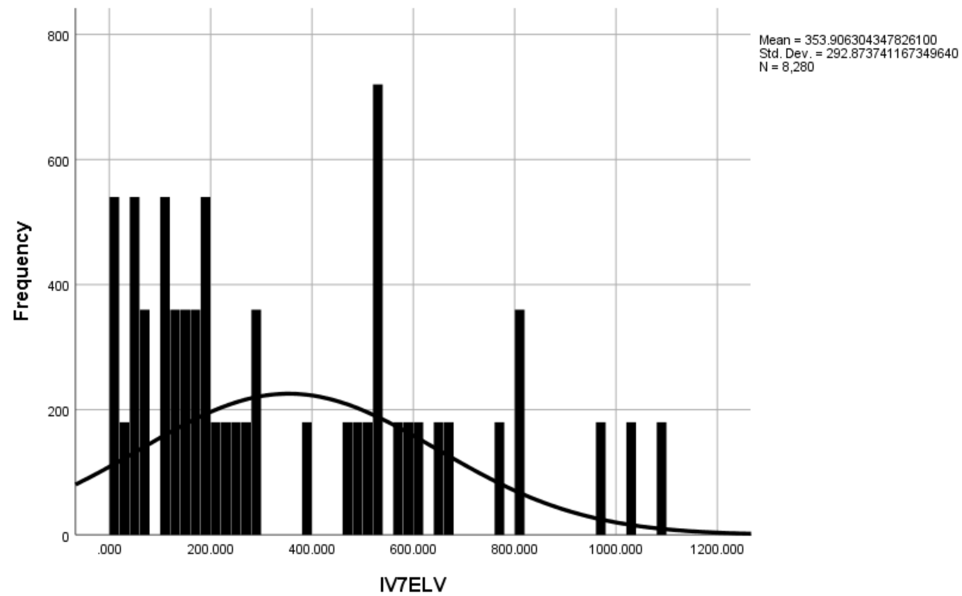
Histogram Showing the Distribution of EV6AWS



***Elevation Variable* (*EV7ELV*).** The mean elevation in feet across all 46 county seats was 353.91 (*SD* = 292.87). The minimum county elevation recorded was 10 and the maximum was 1093, indicating a range of 1083. This range reflects the rise of elevation from the southeastern coastal low country to the northwestern hills of SC (Figure 13).

Figure 13

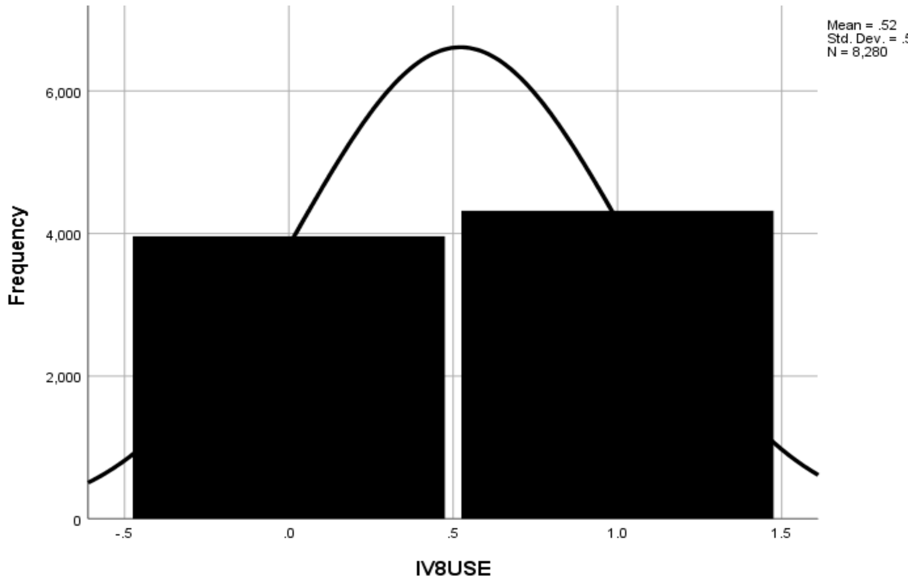
Histogram Showing the Distribution of EV7ELV



***Land Use* (*EV8USE*).** Land use was a categorical variable coded as 0 = Agricultural and 1 = Industrial Use in SPSS. This variable was calculated through the function of population per square mile. The descriptive statistics showed that based on the function above, agricultural land use accounted for 3,960 (47.8%) of the cases and industrial land use accounted for 4,320 (52.2%) cases (Figure 14).

Figure 14

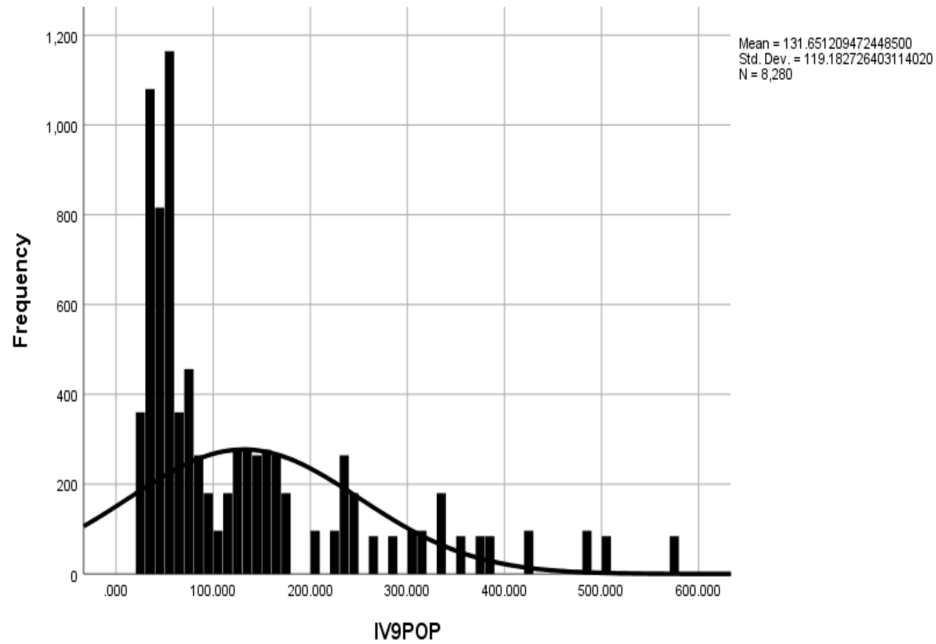
Histogram Showing the Distribution of EV8USE



Population **(*EV9POP*).** The mean population (x1000) across all 46 county seats included in the study timeframe was 131.65 (*SD* = 119.18). The minimum county population recorded was 25.53 and the maximum was 574.72, indicating a range of 549.19 (Figure 15).

Figure 15

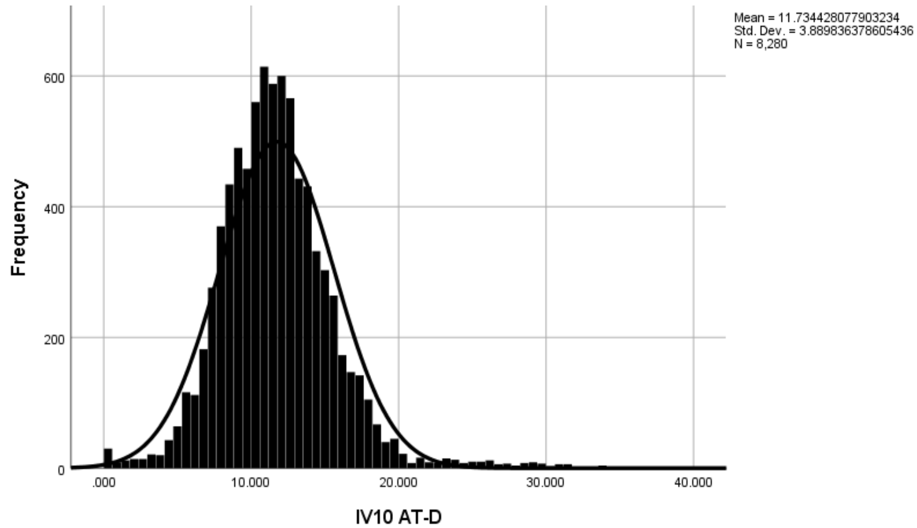
Histogram Showing the Distribution of EV9POP



***Dew Point Deficit* (*EV10ADD*).** The mean dewpoint deficit was 11.734 (*SD* = 3.89). The minimum dewpoint deficit recorded was 0.00 and the maximum was 34.09, indicating a range of 34.09 (Figure 16).

Figure 16

Histogram Showing the Distribution of EV10ADD



Stage A1 BLR Assumptions. BLR has five assumptions that required examination to ensure the correct statistical tool was being used:

1. One DV that is dichotomous (i.e., a nominal variable with two outcomes).
2. One or more EEVs that are numerical.
3. A linear relationship between the continuous EEVs and the logit transformation of the DV
4. Little or no collinearity among the EEVs.
5. Independence of observations: the categories of the dichotomous DV and all continuous EEVs should be mutually exclusive and exhaustive.

Assumptions one and two were met without requiring further analysis. The DV, $DV_{PRESENCE}$, was dichotomous and all EEVs were numerical, either continuous or discrete; categorical variables were converted to numerical dummy variables. Assumption five was also met in that the DV was dichotomous and each of the EEVs were exclusive.

Linearity of the continuous variables with respect to the logit of the DV was assessed via the Box-Tidwell (1962) test. This is a test that ensures that for every one unit increase in a continuous EEV, that the logit of the DV increases by a constant amount (Laerd Statistics, 2017). This test requires the creation of natural log transformations of the continuous EEVs which are then included as interaction terms in the subsequent test. In the Box-Tidwell test, if the natural log interaction term results in a statistically significant value ($p < \alpha = .05$), then the continuous variable violates the assumption of linearity. The Box-Tidwell results for this study showed that all variables met the assumption of linearity test at $p > \alpha = .05$ and did not require further analysis.

The assumption of no collinearity was met using tolerance values and Variable Inflation Factor (VIF) outputs from SPSS. Field (2013) stated that “tolerance values less than .01 and VIF values greater than 10 indicate a problem” (p. 795). The test results revealed no evidence of multicollinearity (Table 6).

Table 6

BLR Collinearity Assumption Test

Variable	Tolerance	VIF
EV1ATM	.833	1.184
EV2ARN	.966	1.035
EV4ASD	.985	1.015
EV5ABP	.870	1.149
EV6AWS	.799	1.252
EV7ELV	.760	1.315
EV8USE	.515	1.941
EV9POP	.522	1.915
EV10ADD	.872	1.147

Note. *EV3ADP* removed from regression based on EDA.

Stage A2 $DV_{PRESENCE}$ Analysis for Current Month, EEVs Only

I began the $DV_{PRESENCE}$ analysis by examining whether the combination of the nine study EEVs (*EV1ATM, EV2ARN, EV4ASD, EV5ABP, EV6AWS, EV7ELV, EV8USE, EV9POP, EV10ADD*) could predict the presence of WNv in the same month. *DV1* (0-day) was the only outcome variable supporting this $DV_{PRESENCE}$ analysis. I used two SPSS stepwise methods to analyze *DV1*, SASM first and then SFEM. This approach allowed me to observe *DV1* model composition using all six SASM techniques while avoiding the dangers of incorrect model specification due to the criteria used by the individual selection methods or the arbitrariness of the order of selection or elimination (Field, 2013).

All six *DV1* SASM models (Table 7) produced the same result, $R^2_{\text{Nagelkerke}}$ = .171, *LL Ratio* $\chi^2(8) = 436.763$, $p < .001$ The following EEVs were retained in each model, *EV1ATM*, *EV2ARN*, *EV5ABP*, *EV6AWS*, *EV7ELV*, *EV8USE*, *EV9POP*, *EV10ADD*. In each of the SASM models, *EV4ASD* was not significant at $p < \alpha = .20$. This corroborated the EDA findings and informed the Stage A SFEM analyses. This analysis was followed by an examination of *DV1* using the SFEM method.

Table 7

Stage A2 $DV_{PRESENCE}$ Analysis With DV1 Using EVs Only

$DV_{PRESENCE}$	Stepwise method	$R^2_{\text{Nagelkerke}}$	*LL Ratio* χ^2
DV1 Model 1	FS (COND)	.171	$\chi^2(8) = 436.763$, $p < .001$
DV1 Model 2	FS (LR)	.171	$\chi^2(8) = 436.763$, $p < .001$
DV1 Model 3	FS (WALD)	.171	$\chi^2(8) = 436.763$, $p < .001$
DV1 Model 4	BE (COND)	.171	$\chi^2(8) = 436.763$, $p < .001$
DV1 Model 5	BE (LR)	.171	$\chi^2(8) = 436.763$, $p < .001$
DV1 Model 6	BE (WALD)	.171	$\chi^2(8) = 436.763$, $p < .001$

Prior to the *DV1* SFEM analysis, I ran diagnostics on the *DV1* residuals. The studentized residuals (SRE) for *DV1* were within limits, with 2.5% of the cases falling outside ±1.96 and .006% of the cases falling outside ±2.58. Cook's Distance was also checked and was 0.003, within the < 1.0 limit.

Stage A3 DV_{PRESENCE} Analysis With DV1 and EEVs Only

Two models were produced in the first step of the *DV1* SFEM model building process with the results mirroring Stage A2, $R^2_{\text{Nagelkerke}} = .171$, *LL Ratio* $\chi^2(8) = 436.763$, $p < .001$ (Table 8). The interim model predictors at $p < \alpha = .20$ were *EV1ATM, EV2ARN, EV5ABP, EV6AWS, EV7ELV, EV8USE, EV9POP, EV10ADD*. These predictors and their 2FIs were carried into the 2FI step of *DV1* model building.

Table 8

Stage A3 SFEM With DV1 and EEVs Only

$DV_{PRESENCE}$	$R^2_{\text{Nagelkerke}}$	*LL Ratio* χ^2
DV1 Model 7	.171	$\chi^2(9) = 436.929$, $p < .001$
DV1 Model 8	.171	$\chi^2(8) = 436.763$, $p < .001$

Stage A3 DV_{PRESENCE} With DV1, EEVs, and 2FIs

The addition of the relevant 2FIs added 28 predictors to this step of the *DV1* model (Table 9). These additional EVs increased the model goodness-of-fit.

Table 9

Stage A3 2FIs

	EV1ATM	*EV2ARN*	*EV5ABP*	*EV6AWS*	*EV7ELV*	*EV8USE*	*EV9POP*	*EV10ADD*
EV1ATM	**	*X*	*X*	*X*	*X*	*X*	*X*	*X*
EV2ARN	**	**	*X*	*X*	*X*	*X*	*X*	*X*
EV5ABP	**	**	**	*X*	*X*	*X*	*X*	*X*
EV6AWS	**	**	**	**	*X*	*X*	*X*	*X*
EV7ELV	**	**	**	**	**	*X*	*X*	*X*
EV8USE	**	**	**	**	**	**	*X*	*X*
EV9POP	**	**	**	**	**	**	**	*X*

Note. *X* = 2FI interaction analysis performed, ** = redundant

Eighteen models were produced using the eight EEVs and 2FIs with Model 20 providing the most parsimonious and best interim results at $R^2_{\text{Nagelkerke}}$ = .202, *LL Ratio* $\chi^2(23)$ = 517.625, $p < .001$ (Table 10). Model 20 retained the following predictors at $p < \alpha = .20$: *EV1ATM, EV2ARN, EV5ABP, EV6AWS, EV7ELV, EV8USE, EV9POP, EV10ADD, EV1·EV2, EV1·EV5, EV1·EV6, EV1·EV7, EV1·EV9, EV2·EV5, EV2·EV7, EV2·EV8, EV5·EV7, EV5·EV8, EV6·EV7, EV6·EV10, EV7·EV8, EV7·EV10, EV8·EV9.* These predictors were carried into the final step of *DV1* model building where the *Months* EVs were added.

Table 10

Stage A3 $DV_{PRESENCE}$ With DV1, EEVs, and 2FIs

$DV_{PRESENCE}$	$R^2_{\text{Nagelkerke}}$	*LL Ratio* χ^2
DV1 Model 9	.202	$\chi^2(34) = 519.047, p < .001$
DV1 Model 10	.202	$\chi^2(33) = 519.047, p < .001$
DV1 Model 11	.202	$\chi^2(32) = 519.045, p < .001$
DV1 Model 12	.202	$\chi^2(31) = 519.013, p < .001$
DV1 Model 13	.202	$\chi^2(30) = 518.943, p < .001$
DV1 Model 14	.202	$\chi^2(29) = 518.877, p < .001$
DV1 Model 15	.202	$\chi^2(28) = 518.787, p < .001$
DV1 Model 16	.202	$\chi^2(27) = 518.662, p < .001$
DV1 Model 17	.202	$\chi^2(26) = 518.419, p < .001$
DV1 Model 18	.202	$\chi^2(25) = 518.144, p < .001$
DV1 Model 19	.202	$\chi^2(24) = 517.905, p < .001$
DV1 Model 20	.202	$\chi^2(23) = 517.625, p < .001$
DV1 Model 21	.201	$\chi^2(22) = 517.012, p < .001$
DV1 Model 22	.201	$\chi^2(21) = 516.266, p < .001$
DV1 Model 23	.201	$\chi^2(20) = 515.617, p < .001$
DV1 Model 24	.201	$\chi^2(19) = 515.617, p < .001$
DV1 Model 25	.200	$\chi^2(18) = 514.286, p < .001$
DV1 Model 26	.200	$\chi^2(17) = 512.694, p < .001$

Stage A3 $DV_{PRESENCE}$ Analysis With DV1, All EVs

Thirteen models (Models 26-38) were produced in the final step of the *DV1* modeling (Table 11). The best overall *DV1* model was Model 36 at $R^2_{\text{Nagelkerke}} = .285$, *LL Ratio* $\chi^2(25) = 739.250$, $p < .001$. Model 36 consisted of the following final EVs at $p < \alpha$

= .20, *EV1ATM, EV2ARN, EV5ABP, EV8USE, EV9POP, EV10ADD, EV1·EV2, EV1·EV5, EV1·EV7, EV2·EV5, EV2·EV7, EV2·EV8, EV5·EV7, EV5·EV8, EV6·EV10, EV7·EV10, EV8·EV9, April, May, June, July, August, September, October, November.*

Table 11

Stage A3 $DV_{PRESENCE}$ Analysis With DV1 and All EEVs

$DV_{PRESENCE}$	$R^2_{\text{Nagelkerke}}$	*LL Ratio* χ^2
DV1 Model 26	.285	$\chi^2(34) = 743.225, p < .001$
DV1 Model 27	.285	$\chi^2(34) = 743.225, p < .001$
DV1 Model 28	.285	$\chi^2(33) = 743.221, p < .001$
DV1 Model 29	.285	$\chi^2(32) = 743.194, p < .001$
DV1 Model 30	.285	$\chi^2(31) = 743.103, p < .001$
DV1 Model 31	.285	$\chi^2(30) = 742.924, p < .001$
DV1 Model 32	.285	$\chi^2(29) = 742.863, p < .001$
DV1 Model 33	.285	$\chi^2(28) = 742.323, p < .001$
DV1 Model 34	.285	$\chi^2(27) = 741.539, p < .001$
DV1 Model 35	.285	$\chi^2(26) = 740.692, p < .001$
DV1 Model 36	.285	$\chi^2(25) = 739.250, p < .001$
DV1 Model 37	.284	$\chi^2(24) = 738.117, p < .001$
DV1 Model 38	.283	$\chi^2(23) = 737.226, p < .001$

Stage A3 DV1 Final Model

DV1 model development produced 38 models. Model 36 at $R^2_{\text{Nagelkerke}} = .285$, *LL Ratio* $\chi^2(25) = 739.250, p < .001$, retained six of the eight SASM EVs while *EV4ASD,* *EV6AWS*, and *EV7ELV* were removed at $p < \alpha = .20$. The final *DV1* model was expressed as follows (see Equation 9):

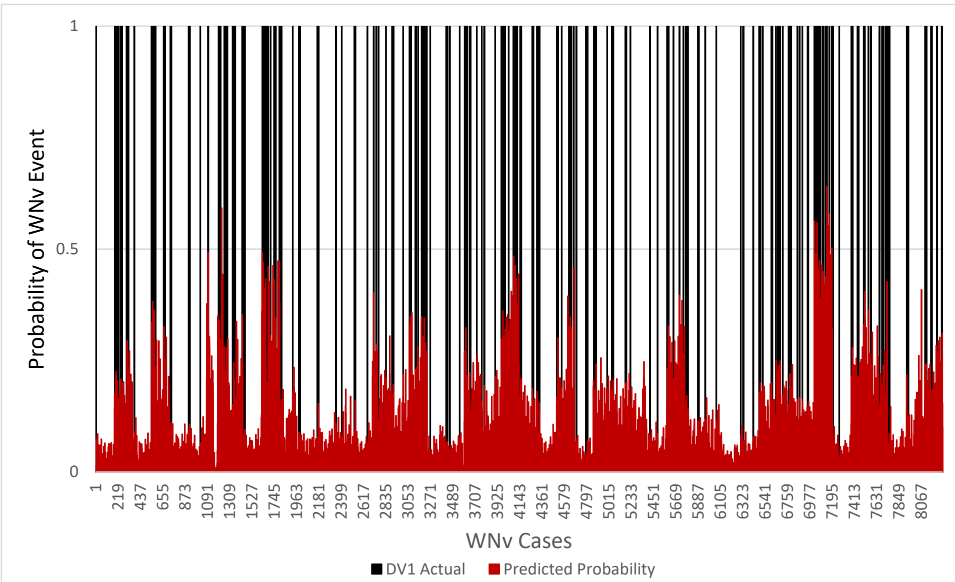
$$\hat{Y} = 214.881 + (-4.371 \cdot EV1ATM) + (234.861 \cdot EV2ARN) + (-0.145 \cdot EV5ABP) + (-0.67 \cdot EV8USE) + (0.015 \cdot EV9POP) + (-0.151 \cdot EV10ADD) + (0.192 \cdot EV1 \cdot EV2) + (0.004 \cdot EV1 \cdot EV5) + (0.000067 \cdot EV1 \cdot EV7) + (-0.247 \cdot EV2 \cdot EV5) + (0.003 \cdot EV2 \cdot EV7) + (-1.168 \cdot EV2 \cdot EV8) + (-0.000008 \cdot EV5 \cdot EV7) + (-0.064 \cdot EV5 \cdot EV8) + (-0.016 \cdot EV6 \cdot EV10) + (0.000215 \cdot EV7 \cdot EV10) + (-0.012 \cdot EV8 \cdot EV9) + (1.386 \cdot April)\ (1.597 \cdot May) + (1.652 \cdot June) + (3.361 \cdot July) + (4.158 \cdot August) + (4.360 \cdot September) + (3.401 \cdot October) + (2.577 \cdot November) \qquad (9)$$

The *DV1* final model produced a MAE = .072 and RMSE = .847 which compared favorably to later modeling with *DV2*. When Model 36 predicted probability results were plotted against the actual WNv events, model accuracy was good, but the model's precision was poor (Figure 17). As a result, only 4% of WNv events were correctly identified at the SPSS default classification cutoff value (CCV) of .500. At a CCV of .250, 23% of WNV events within the predicted group range, highlighting the future role of the EHM in any local configuration settings.

Based on these results, I rejected the RQ1 $DV_{PRESENCE}$ null hypothesis (H_01) at $p < \alpha = .05$. I concluded there was evidence that at least one coefficient in the final regression model was not equal to zero and that the final model was a statistically significant predictor of *DV1* (the likelihood of a WNv event in the current month). The operationally significant predictors associated with the final model can account for 28% of the variance of *DV1*, providing reasonable accuracy of WNv incident occurrence in the same month.

Figure 17

DV1 Actual WNv Events Versus Final Model Probability



Stage A4 $DV_{PRESENCE}$ Analysis for Best Time-lagged DV Selection, EVs Only

In the Stage A4 $DV_{PRESENCE}$ analysis, I initially examined the time-lagged DVs, 30- (*DV2*), 60- (*DV3*), and 90-day (*DV4*) to identify the best DV for Stage A5. The initial analysis employed the six SASM methods: FS (COND), FS (LR), FS (WALD), BE (COND), BE (LR), and BE (WALD). This approach allowed me to observe and compare the results of all time-lagged DVs using SASM techniques prior to Stage A5.

This analysis was conducted with the nine post-EDA study EEVs at $\alpha = .20$. Goodness-of-fit and *LL Ratio* χ^2 scores were used to select the best model. Eighteen models (Appendix B, Tables B5-B7) were produced in this step with *DV2* outperforming the other outcome variables (Table 12).

Model 44 produced the best Stage A4 model at $R^2_{\text{Nagelkerke}}$ = .257, *LL Ratio* $\chi^2(6)$ = 661.645, $p < .001$, with all predictors significant ($p < \alpha = .20$). *EV4ASD* (p = .990) and *EV7ELV* (p = .290) were removed from *DV2* Model 44. Model 50 produced the best *DV3* model at $R^2_{\text{Nagelkerke}}$ = .247 with all predictors significant ($p < \alpha = .20$). *EV4ASD* (p = .991), *EV7ELV* (p = .829), and *EV2ARN* (p = .281) were removed from Model 50. Model 54 produced the best *DV4* model at $R^2_{\text{Nagelkerke}}$ score = .163. *EV4ASD* (p = .990) was removed from Model 54. *EV2ARN* (p = .314), *EV6AWS* (p = .221), and *EV5ABP* (p = .223) remained in the model but were not significant ($p < \alpha = .20$).

Table 12

Stage A4 Comparison of the $DV_{PRESENCE}$ Time-Lagged DVs

$DV_{PRESENCE}$	$R^2_{\text{Nagelkerke}}$	*LL Ratio* χ^2
DV2 Model 44	.257	$\chi^2(6) = 661.645$, $p < .001$
DV3 Model 50	.245	$\chi^2(5) = 634.453$, $p < .001$
DV4 Model 54	.162	$\chi^2(5) = 412.961$, $p < .001$

In a comparison of the time-lagged DVs (Models 44, 50, and 54), *DV2* produced the best goodness-of-fit scores with six EEVs only. Although, *DV3* and *DV4* were not selected as the best time-lagged DVs, the $R^2_{\text{Nagelkerke}}$ and *LL Ratio* χ^2 scores showed that the RQ2 $DV_{PRESENCE}$ null hypotheses for 60- and 90-day time-lags could be rejected based on the use of EEVs alone at $p < \alpha = .20$. Based on the results of the Stage A modeling, I was confident that the addition of the remaining 2FIs and *Months* EVs would continue to enhance the goodness-of-fit scores for *DV3* and *DV4* as they did for *DV2*.

The time lagged DV analysis also provided additional context for the rest of the Stage A *DV2* analysis. For example, the variable *EV4ASD* was not significant ($p < \alpha =$ .20) in any of the models, corroborating the EDA findings. Additionally, *EV1ATM*, *EV8USE*, *EV9POP*, and *EV10ADD* were the only EEVs present in all of the Stage A4 $DV_{PRESENCE}$ analyses.

Based on these results, I selected *DV2* as the best time lagged DV for answering RQ2. I then examined the *DV2* residuals. The studentized residuals (SRE) were within limits, with 2.5% of the cases falling outside ±1.96 and .006% of the cases falling outside ±2.58. Cook's Distance.

Stage A5 $DV_{PRESENCE}$ Analysis With DV2, All EVs

In the Stage A5 analysis, I used the SFEM and began with *DV2* and the nine EEVs. Predictors were removed from the model after each run based on the $R^2_{\text{Nagelkerke}}$ score, significance at $p < \alpha = .20$, and subject matter expertise. This allowed for a more inclusive model-building process throughout Stage A5.

Model 59 produced the best model in Stage A5 based on a goodness-of-fit value of $R^2_{\text{Nagelkerke}} = .257$ and considering Stage A4 results (Table 13). While the remaining seven EEVs were significant, the removal of *EV4ASD* ($p = .990$) and *EV7ELV* ($p = .290$) had little effect on the Model 59 χ^2 statistic at $\chi^2(7) = 665.685$, $p < .001$. This meant that without these predictors, the model still significantly outperformed the constant-only model.

Table 13

Stage A5 $DV_{PRESENCE}$ Analysis With DV2, EEVs Only

Model	$R^2_{\text{Nagelkerke}}$	*LL Ratio* χ^2
57	.257	$\chi^2(9) = 667.271, p < .001$
58	.257	$\chi^2(8) = 666.817, p < .001$
59	.257	$\chi^2(7) = 665.685, p < .000$

In Stage A1, the *EV4ASD* descriptive statistics revealed a very low mean of .002 inches for the entirety of 8,280 WNv cases. In Models 57, *EV4ASD* showed no strength of association between the predictor and *DV*2. At this point, I decided to remove the *EV4ASD* from the Stage A5 modeling.

In Model 58, *EV7ELV* did not reach the required significance level and produced the lowest strength of association with *DV2* of all the remaining model predictors. Combined with the results of Stage A4, I elected to also eliminate *EV7ELV* from the remaining Stage A5 analysis.

Stage A5 $DV_{PRESENCE}$ Analysis With DV2, EEVs, 2FIs

The Stage A5 analysis began with the seven remaining EEVs (*EV1ATM*, *EV2ARN*, *EV5ABP*, *EV6AWS, EV8USE*, *EV9POP*, *EV10ADD*) and 21 2FIs (Table 14). The EEV and 2FI analysis significance criterion remained the same at $p < \alpha = .20$.

Table 14

Stage A5 DV2 Two-Factor Interactions

	EV1ATM	*EV2ARN*	*EV5ABP*	*EV6AWS*	*EV8USE*	*EV9POP*	*EV10ADD*
EV1ATM	··	*X*	*X*	*X*	*X*	*X*	*X*
EV2ARN	· ·	··	*X*	*X*	*X*	*X*	*X*
EV5ABP	· ·	· ·	··	*X*	*X*	*X*	*X*
EV6AWS	· ·	· ·	· ·	··	*X*	*X*	*X*
EV8USE	· ·	· ·	· ·	· ·	··	*X*	*X*
EV9POP	· ·	· ·	· ·	· ·	· ·	··	*X*

Note. *X* = 2FI interaction analysis performed, = not applicable, ··= redundant

With the inclusion of the 2FIs variables, the Stage A5 $R^2_{\text{Nagelkerke}}$ and *LL Ratio* χ^2 values increased significantly. The Stage A5 modeling process produced 15 models (Table 15). I continued the modeling process until all EVs were at $p < \alpha = .20$, regardless of the $R^2_{\text{Nagelkerke}}$ score. I then evaluated all models in the stage to determine the best model.

Fourteen EVs were removed using the $p < \alpha = .20$ criterion. Four of these variables were original EEVs (*EV2ARN*, *EV5ABP*, *EV8USE*, *EV10ADD*). The predictors *EV2ARN*, *EV5ABP*, and *EV8USE* were also removed in several Stage A4 models using the SASM method. In Stage A5, the removal of these predictors had little effect on the Stage A5 overall model *LL Ratio* χ^2 statistic. However, *EV10ADD* was removed late in the Stage A4 modeling (Model 69) at $p = .203$. This EV was retained in the Stage A modeling, as I selected Model 67 as the best Stage A5 interim model with a goodness-of-fit value at $R^2_{\text{Nagelkerke}}$ = .275, *LL Ratio* $\chi^2(20) = 713.352$, $p < .001$.

Table 15

Stage A5 DV$_{PRESENCE}$ Analysis With DV2, EEVs, 2FIs

Model	$R^2_{\text{Nagelkerke}}$	*LL Ratio* χ^2
60	.275	$\chi^2(27) = 713.876$, $p < .001$
61	.275	$\chi^2(26) = 713.875$, $p < .001$
62	.275	$\chi^2(25) = 713.873$, $p < .001$
63	.275	$\chi^2(24) = 713.862$, $p < .001$
64	.275	$\chi^2(23) = 713.839$, $p < .001$
65	.275	$\chi^2(22) = 713.714$, $p < .001$
66	.275	$\chi^2(21) = 713.574$, $p < .001$
67	.275	$\chi^2(20) = 713.352$, $p < .001$
68	.274	$\chi^2(19) = 713.309$, $p < .001$
69	.274	$\chi^2(18) = 712.878$, $p < .001$
70	.274	$\chi^2(17) = 712.379$, $p < .001$
71	.274	$\chi^2(16) = 711.688$, $p < .001$
72	.273	$\chi^2(15) = 710.292$, $p < .001$
73	.273	$\chi^2(14) = 710.292$, $p < .001$
74	.273	$\chi^2(13) = 708.378$, $p < .001$

Stage A5 DV$_{PRESENCE}$ Analysis With DV2, All EVs

The analysis began with the Model 67 with the addition of the *Month* EVs (Table 13), *EV1ATM*, *EV6AWS*, *EV9POP*, *EV10ADD*, *EV1·EV2*, *EV1·EV5*, *EV1·EV6*, *EV1·EV8*, *EV1·EV10*, *EV2·EV6*, *EV2·EV10*, *EV5·EV6*, *EV5·EV8*, *EV5·EV9*, *EV5·EV10*, *EV6·EV8*,

EV6·EV9, *EV6·EV10*, *EV8·EV9*, *EV9·EV10*, *January*, *February*, *March*, *April*, *May*, *June*, *July*, *August*, *September*, *October*, *November*, *December*.

In this part of the Stage A5 analysis, 16 models were developed. During the model-building process, *EV10ADD*, five 2FIs, and eight *Month* EVs were removed at $p > \alpha = .20$. For completeness, I continued the modeling process until all predictors were at $p < \alpha = .20$, regardless of the $R^2_{\text{Nagelkerke}}$ score. I then evaluated all models in the stage to determine the best model.

Model 84 was the best Stage A5 model with a $R^2_{\text{Nagelkerke}} = .293$ (Table 16). The model included the following predictors: *EV1ATM*, *EV6AWS*, *EV9POP*, *EV1·EV5*, *EV1·EV6*, *EV1·EV10*, *EV2·EV6*, *EV2·EV10*, *EV5·EV6*, *EV5·EV8*, *EV5·EV9*, *EV5·EV10*, *EV6·EV8*, *EV6·EV9*, *EV6·EV10*, *EV8·EV9*, *EV9·EV10*, *May*, *June*, *August*, *September*, *October*, *November*. Although *May* only had 8 reported incidents, it was statistically significant based on the census data used in this study. The Stage B2 interim model also excluded the month of *July*, as was the case here. In the EDA, the month of *July* accounted for 108 WNv incidents. However, in both sets of analysis, the *July* EV was removed very early for lack of significance ($p < \alpha = .20$) after the Month EVs were added to the model. This occurred with both the $DV_{PRESENCE}$ and DV_{COUNT} datasets and with different statistical tools. I decided to follow the statistics produced by SPSS and remove the *July* EV in both cases.

Based on the strength and consistency of predictors, and as Model 84 was developed with the SFEM and subject matter expertise, I selected Model 84 as the *DV2* best interim model at $R^2_{\text{Nagelkerke}} = .293$, *LL Ratio* $\chi^2(23) = 762.344$, $p < .001$ at $\alpha = .20$.

Table 16

Stage A5 DV_{PRESENCE} With DV2, All EVs

Model	$R^2_{\text{Nagelkerke}}$	*LL Ratio* χ^2
75	.293	$\chi^2(31) = 763.626$, $p < .001$
76	.293	$\chi^2(31) = 763.626$, $p < .001$
77	.293	$\chi^2(30) = 763.626$, $p < .001$
78	.293	$\chi^2(29) = 763.622$, $p < .001$
79	.293	$\chi^2(28) = 763.608$, $p < .001$
80	.293	$\chi^2(27) = 763.454$, $p < .001$
81	.293	$\chi^2(26) = 763.271$, $p < .001$
82	.293	$\chi^2(25) = 763.095$, $p < .001$
83	.293	$\chi^2(24) = 762.793$, $p < .001$
84	.293	$\chi^2(23) = 762.344$, $p < .001$
85	.292	$\chi^2(22) = 761.838$, $p < .001$
86	.292	$\chi^2(21) = 761.838$, $p < .001$
87	.292	$\chi^2(20) = 760.093$, $p < .001$
88	.291	$\chi^2(19) = 760.093$, $p < .001$
89	.291	$\chi^2(18) = 758.244$, $p < .001$
90	.291	$\chi^2(17) = 758.244$, $p < .001$

Stage A6 $DV_{PRESENCE}$ Analysis With DV2, All EVs, α = .05

Stage A6 analysis (Table 17) of the EEVs, 2FIs, and Months variables began using the Model 84 predictors excluding the month of *November* (p = .491), *EV1ATM*, *EV6AWS*, *EV9POP*, *EV1·EV5*, *EV1·EV6*, *EV1·EV10*, *EV2·EV6*, *EV2·EV10*, *EV5·EV6*, *EV5·EV8*, *EV5·EV9*, *EV5·EV10*, *EV6·EV8*, *EV6·EV9*, *EV6·EV10*, *EV8·EV9*, *EV9·EV10*, *May*, *June*, *August*, *September*, *October*. For this stage of modeling, I used $p < \alpha = .05$ to raise the threshold for predictor inclusion (for a simpler model, with less chance of over-specification) and to provide a comparison of the previous stage models using $R^2_{\text{Nagelkerke}}$ scores as well.

Table 17

Stage A6 $DV_{PRESENCE}$ Analysis With DV2, EEVs, 2FIs, and Months, α = .05

Model	$R^2_{\text{Nagelkerke}}$	*LL Ratio* χ^2
91	.292	$\chi^2(22) = 761.838$, $p < .001$
92	.292	$\chi^2(21) = 761.270$, $p < .001$
93	.292	$\chi^2(20) = 760.093$, $p < .001$
94	.291	$\chi^2(19) = 759.069$, $p < .001$
95	.291	$\chi^2(19) = 758.244$, $p < .001$
96	.291	$\chi^2(17) = 756.960$, $p < .001$
97	.290	$\chi^2(16) = 755.020$, $p < .001$
98	.289	$\chi^2(15) = 753.069$, $p < .001$
99	.287	$\chi^2(15) = 746.698$, $p < .001$
100	.285	$\chi^2(13) = 750.953$, $p < .001$

In the Stage A6 analysis, 10 models were developed at $\alpha = .05$. During the model-building process, the months of *November* (Model 91), *September* (Model 93), *August* (Model 44), and seven 2FIs were removed at $p > .05$. For completeness, I continued the modeling process until all predictors were at $p < \alpha = .05$, regardless of the $R^2_{\text{Nagelkerke}}$ score. I then evaluated all models in the stage to determine the best model.

Model 93 was the best Stage A7 model with a $R^2_{\text{Nagelkerke}} = .292$. The model included the following predictors, *EV1ATM*, *EV6AWS*, *EV9POP*, *EV1·EV5*, *EV1·EV6*, *EV2·EV6*, *EV2·EV10*, *EV5·EV6*, *EV5·EV8*, *EV5·EV9*, *EV6·EV8*, *EV6·EV9*, *EV6·EV10*, *EV8·EV9*, *EV9·EV10*, *May*, *June*, *August*, *September*, *October*. While producing the best Stage A6 $R^2_{\text{Nagelkerke}}$ score, it also included key *Month* EVs based on the data provided in the descriptive statistics.

$DV_{PRESENCE}$ Final Model Selection

To select a final $DV_{PRESENCE}$ model, I compared the best interim models from Stages A4 through A6 using $R^2_{\text{Nagelkerke}}$ as my primary goodness-of-fit metric, *LL Ratio* χ^2 as a supporting metric, and subject matter expertise to consider issues such as predictor inclusion and model parsimony. The Stage A5-6 results reflected the iterative model building process (i.e., EEVs only, then 2FIs EVs, and then *Month* EVs) using the SFEM at $p < \alpha = .20$. Stage A6 was a continuation of Model 84 using $p < \alpha = .05$.

In the subsequent analysis of the Stage A5 and A6 models, I found Model 84 to be the most compelling from a statistical and subject matter expert perspective (Table 18). First, Stage A5 (Model 59) and A5 (Model 67) produced $R^2_{\text{Nagelkerke}}$ values consistent with the EVs included in those stages. Model 67 contained a more complete set of EVs of interest than earlier stages and this was reflected in a higher $R^2_{\text{Nagelkerke}}$ value. Model 93, while more parsimonious than Model 84, had a lower $R^2_{\text{Nagelkerke}}$ value.

Model 84 was achieved using all EVs at $p < \alpha = .20$ While Model 56 retained more of the EEVs, *EV7ELV*, *EV8USE*, and *EV10ADD* using the SASM approach, they had also been removed at various stages of the modeling process. *EV7ELV* had been removed from all the Stage A4 *DV2* and *DV3* models (Models 39-50) and was removed at $p = .990$ in Stage A4 Model 58. *EV8USE* had been removed in Stage A5 Model 65 and *EV10ADD* also been removed in Model 72 at $p = .203$. While all three EEVs were retained in the SASM method, I determined their exclusion in the SFEM process to be guided by an appropriate combination of variable removal criterion and subject matter expertise.

Based on these results, I selected Model 84 as final model for the Stage A $DV_{PRESENCE}$ analysis because it produced the best $R^2_{\text{Nagelkerke}}$ value and included the months of *May*, *June*, *August*, *September*, *October*, and *November*. Although *May* only had 8 reported incidents, it was statistically significant based on the census data used in this study.

Table 18

Comparison of Stages A5-A6 Best Models

Stage	Model	$R^2_{\text{Nagelkerke}}$	*LL Ratio* χ^2
A5	59	.257	$\chi^2(7) = 665.685$, $p < .001$
A5	67	.275	$\chi^2(20) = 713.352$, $p < .001$
A5	84	.293	$\chi^2(23) = 762.344$, $p < .001$
A6	93	.292	$\chi^2(20) =760.093$, $p < .001$

Following this analysis, I performed a diagnostic test on the final model residuals checking Cook's Distance (COO·1), leverage (LEV·1), absolute values of the standardized residual (ZRE·1), and deviance (DEV·1). The diagnostic test revealed no unusually high values of Cook's distance (mean = 0.003). All 8,280 case values of the leverage residual (LEV1) were less than 1 (mean = .0029). The ZRE·1 (.02) and DEV·1 (.006) absolute values were also within the 5% limit at ± 1.96 and 1% limit for ±2.58.

Based on the diagnostic test results, I confirmed Model 84 as the final Stage A $DV_{PRESENCE}$ model. The final model resulted in the following forecast accuracies: MAE = 0.072, RMSE = 0.192. The final logistic regression model was expressed as follows (see Equation 10):

$$\hat{Y} = -13.277 + (-8.380 \cdot EV1ATM) + (78.978 \cdot EV6AWS) + (0.397 \cdot EV9POP) + (0.008 \cdot EV1 \cdot EV5) + (-0.005 \cdot EV1 \cdot EV6) + (0.001304 \cdot EV1 \cdot EV10) + (-0.332 \cdot EV2 \cdot EV6) + (0.183 \cdot EV2 \cdot EV10) + (-0.077 \cdot EV5 \cdot EV6) + (0.004 \cdot EV5 \cdot EV8) + (-0.0004 \cdot EV5 \cdot EV9) + (-0.0002 \cdot EV5 \cdot EV10) + (-$$

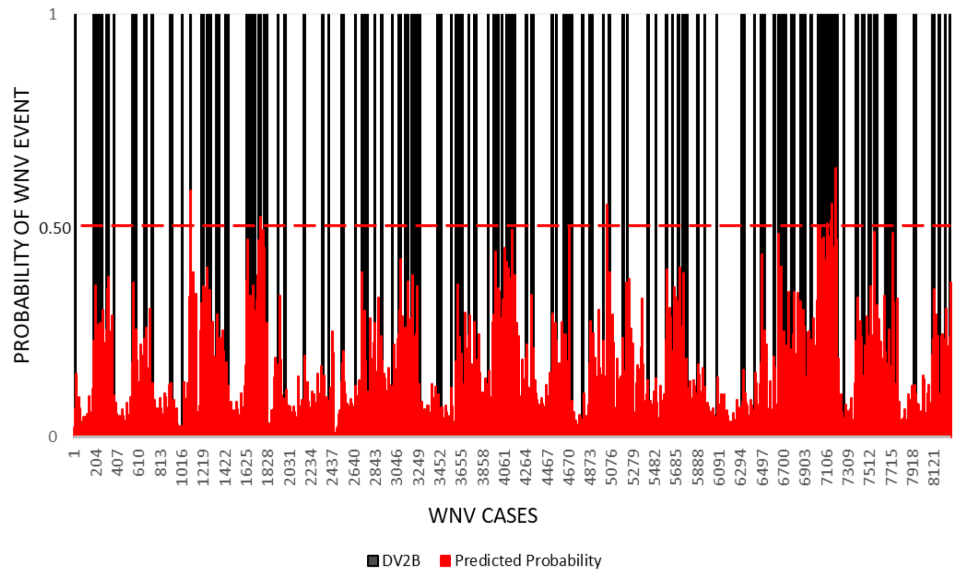
$0.420 \cdot EV6 \cdot EV8) + (0.000640 \cdot EV6 \cdot EV9) + (-0.033 \cdot EV6 \cdot EV10) + (-0.019 \cdot EV8 \cdot EV9) + (0.0002 \cdot EV9 \cdot EV10) + (-1.629 \cdot May) + (-0.449 \cdot Jun) + (0.200 \cdot August) + (0.163 \cdot September) + (0.659 \cdot October) + (-0.355 \cdot November)$ (10)

When Model 84 predicted probability results were plotted against the actual *DV2* WNv events, model accuracy was good, but the model's precision was poor (Figure 18). Reported WNv events are shown in black, representing any incident of WNv that occurred in a case. Predicted probability of WNv events are shown in red by case.

The results show that while the predicted probabilities show good correlation with actual events, the ability of the predictors to meet the default model classification cut-off value (CCV) of .500 was lacking. Environmental health decision makers have the flexibility to set a prediction threshold based on historical data or on risk level. In this study, lowering the CCV to .250 would have resulted in 13% of the WNv incidents being correctly identified.

Figure 18

DV2 Actual Events Versus BLR Final Model Probability



Note. WNv Present (Actual) = Black, WNv Predicted Probability= Red. The red dashed line depicts the SPSS default .500 classification cutoff determination.

Based on these results, I rejected the RQ2 $DV_{PRESENCE}$ null hypothesis at $p < \alpha = .05$. I concluded there was evidence that at least one coefficient in the final regression model was not equal to zero, and that the final model was a statistically significant predictor of *DV2* (which was the best DV for predicting the likelihood of a future WNv event). The operationally significant predictors associated *DV2* model can account for 29% of the variance of the WNv outcome variable, providing reasonable accuracy of WNv incident occurrence in the same month.

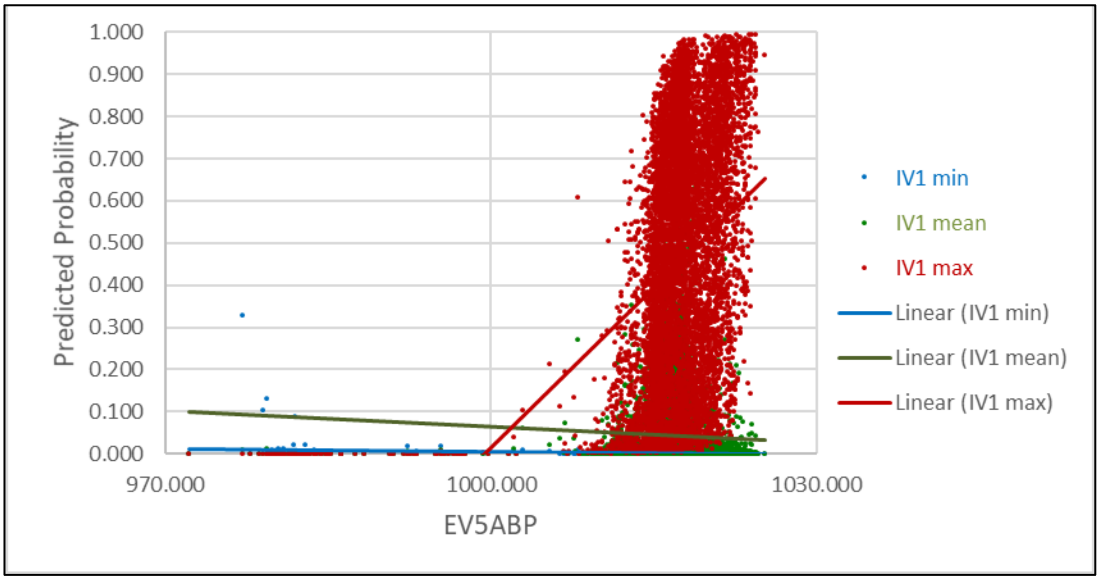
DV_{PRESENCE} Final Model 2FI Analysis

Upon completion of the Stage A6, I conducted an analysis of the final model 14 2FIs: *EV1·EV5, EV1·EV6, EV1·EV10, EV2·EV6, EV2·EV10, EV5·EV6, EV5·EV8, EV5·EV9*, *EV5·EV10*, *EV6·EV8*, *EV6·EV9*, *EV6·EV10*, *EV8·EV9*, *EV9·EV10*. The three EEVs that were not in the final model individually but exerted moderating effects on the other predictors were *EV5ABP*, *EV8USE*, and *EV10ADD*. I include four of the most interesting interactions in Figures 19 to 26.

The relationship between *EV5ABP* and *DV2* was influenced by minimum, mean, and maximum average temperatures. Figure 19 shows that when average temperature was high and the average barometric pressure increased above 1000.0, the probability of an occurrence of WNv rose sharply. As the temperature decreased from mean to minimum values, the probability of occurrence was at or near zero for all levels of barometric pressure.

Figure 19

EV1ATM Influence on the Relationship Between EV5ABP and DV2

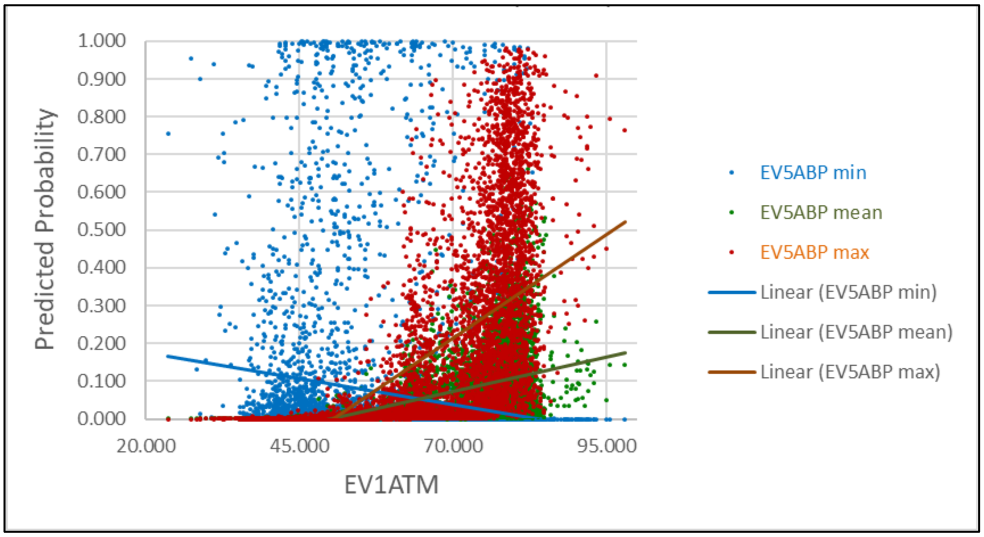


Note. *EV1ATM* values, minimum (23.57°F), mean (63.39°F), and maximum (97.97°F).

The relationship between *EV1ATM* and *DV2* was influenced by minimum, mean, and maximum average barometric pressures (Figure 20). When average barometric pressure was at its maximum value, the probability of WNv occurrence increased sharply as average temperature rose past 45^0F. When average barometric pressure was at its minimum value, the probability of WNv occurrence decreased with rising temperatures. When average barometric pressure was at its mean, the probability of WNv occurrence also increased as temperature rose above 45^0F but at a lower rate than when average barometric pressure was at its maximum value. While average barometric pressure itself did not contribute to the model goodness-of-fit, it did exert a moderating effect on temperature as a predictor of probability of occurrence.

Figure 20

The Influence of EV5ABP on the EV1ATM and DV2 Relationship



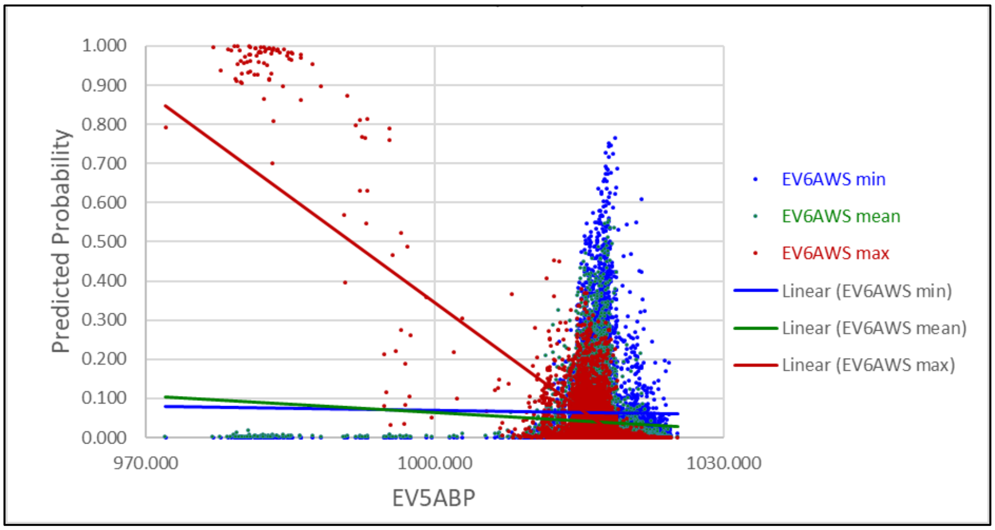
Note. *EV5ABP* values, minimum (972.12), mean (1017.30), and maximum (1025.25).

The relationship between *EV5ABP* and *DV2* was influenced by minimum, mean, and maximum average wind speeds (Figure 21). When average barometric pressure was at its maximum value and wind speed average wind speed was rising, the probability of WNv occurrence decreased sharply. When average barometric pressure was at its mean value and wind speed was rising, WNv occurrence decreased moderately. When average barometric pressure was at its minimum value and wind speed rose, the probability of the occurrence of WNv increased sharply above five knots average wind speed.

The interaction showed that when pressure was high, probability of outbreak decreases with increases in wind speed, and when pressure is low, probability of outbreak increases with increases in wind speed. While pressure was not a significant predictor by itself, it did exert a moderating effect on wind speed as a predictor of probability of WNv occurrence.

Figure 21

The Influence of EV6AWS on the EV5ABP and DV2 Relationship



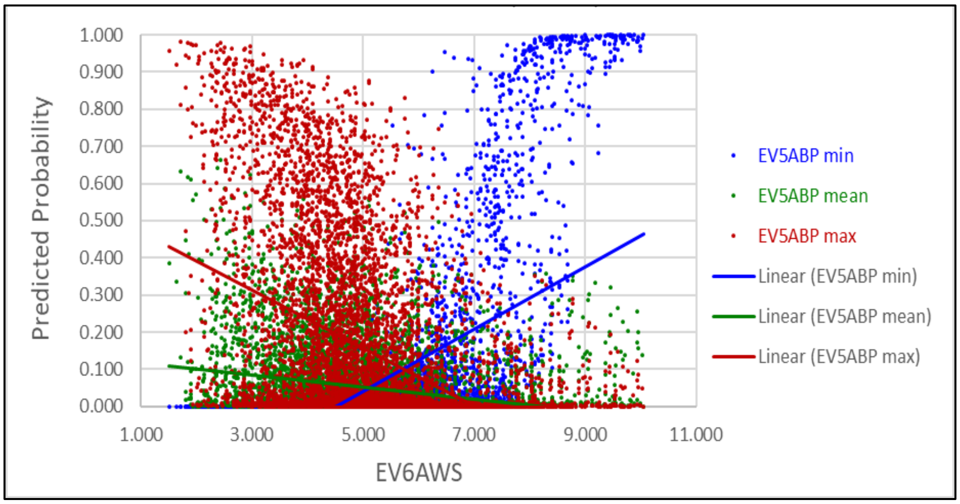
Note. *EV5ABP* values, minimum (972.12), mean (1017.30), and maximum (1025.25).

The relationship between *EV6AWS* and *DV2* was influenced by minimum, mean, and maximum average barometric pressure (Figure 22). When average wind speed was at its maximum value and the average barometric pressure rose, the probability of WNv occurrence decreased sharply. When average wind speed was at its mean or minimum values, there were low probabilities of WNv occurrence, which decreased slightly with increases in average barometric pressure. This interaction showed that when wind speed was high, the probability of an outbreak decreased sharply with increases in pressure.

When the winds are average or low, barometric pressure was not influential on the probability of WNv occurrence.

Figure 22

The Influence of EV5ABP on the EV6AWS and DV2 Relationship



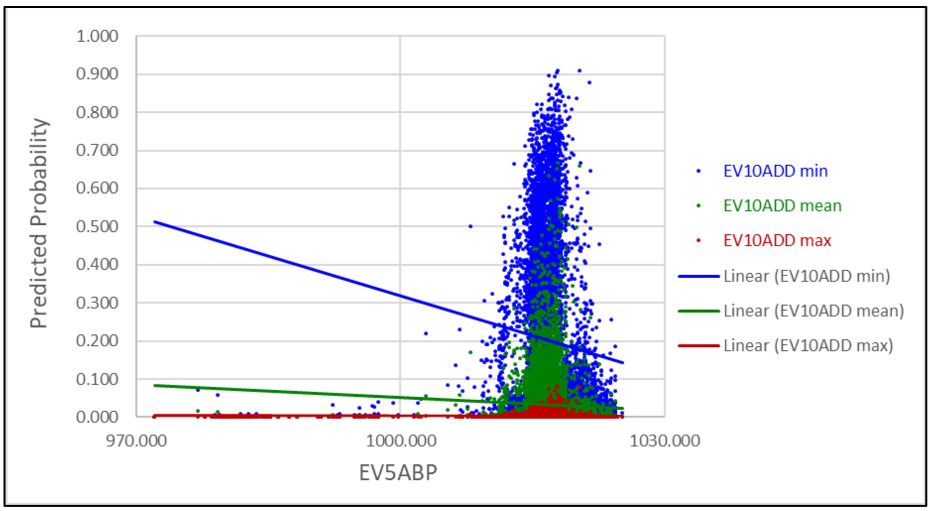
Note. *EV6AWS* values, minimum (1.51), mean (5.20), and maximum (10.06).

The relationship between *EV5ABP* and *DV2* was influenced by minimum, mean, and maximum average dewpoint deficits (Figure 23). When the dew point deficit was at its maximum value, the probability of WNv occurrence was low and remained low irrespective of changes in average barometric pressure. When the dew point deficit was at its minimum, the probability of WNv occurrence decreased with an increase in average barometric pressure. When the dew point deficit was at its mean value, the probability of WNv decreased at a lower rate when average barometric pressure increased.

When the dew point deficit was low (indicating humid conditions), probability of an event decreased as average barometric pressure increased. The lower the dew point deficit the more influential barometric pressure was on the probability of WNv occurrence. While neither dew point deficit nor pressure alone were significant predictors, their interaction was a significant predictor of the probability of WNv occurrence.

Figure 23

EV10ADD Influence on the Relationship Between EV5ABP and DV2 Relationship



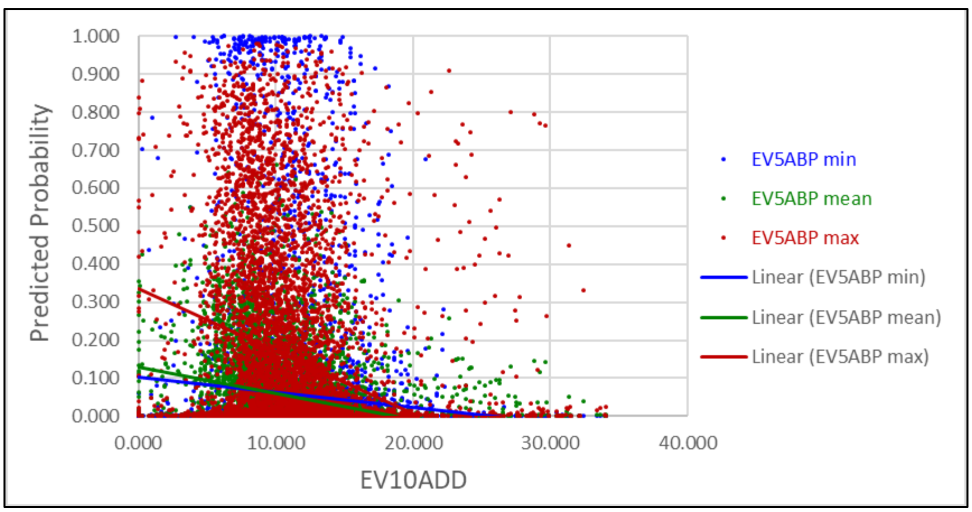
Note. *EV10ADD* values, minimum (0.00), mean (11.73), and maximum (34.09).

The relationship between *EV10ADD* and *DV2* was influenced by minimum, mean, and maximum average barometric pressures (Figure 24). When average barometric pressure was at its maximum value, the probability of WNv occurrence decreased sharply as the average dew point deficit increased. When average barometric pressure was held to either its mean or minimum values, the probability of WNv occurrence was low and decreased slightly with an increase in average dew point deficit.

Predicted probability decreased with higher dew point deficits (drier conditions) but that decrease was more pronounced the higher the barometric pressure. While barometric pressure and dew point deficit were not by themselves significant predictors, the interaction between the two predictors was a significant influence on the probability of WNv occurrence.

Figure 24

EV5ABP Influence on the Relationship Between EV10ADD and DV2 Predicted Probability



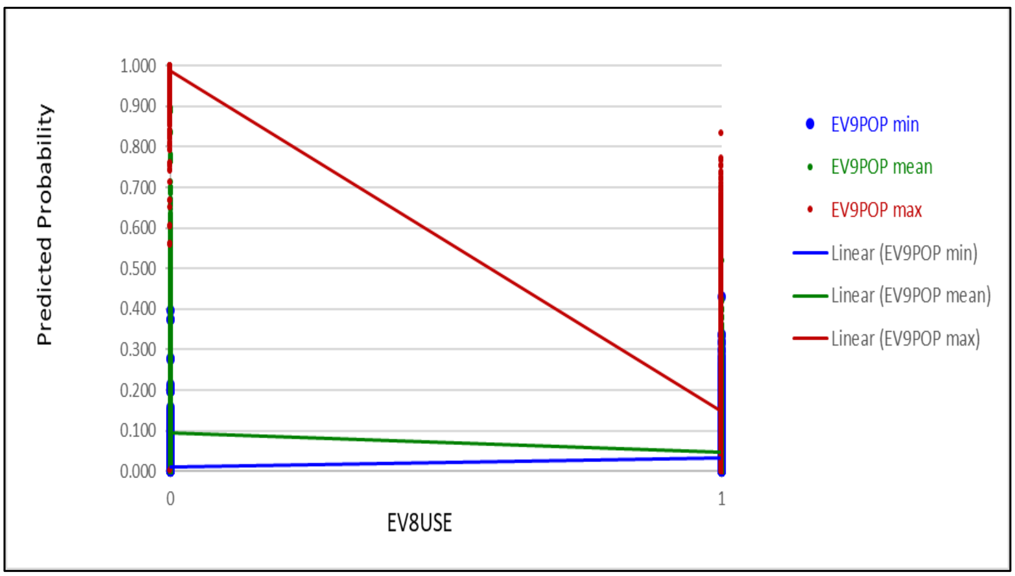
Note. *EV5ABP* values, minimum (972.12), mean (1017.30), and maximum (1025.25).

The relationship between *EV8USE* and *DV2* was influenced by minimum, mean, and maximum average populations (Figure 25). When population density was at its maximum value and land use was agricultural, the probability of WNv occurrence was at its highest. When land use is industrial and the population density was at its minimum or mean values, the probability of WNv occurrence remained low.

In this interaction, when population density was high, the probability was significantly higher in agricultural land use counties than in industrial counties. When population density was moderate or low, land use was not influential on probability.

Figure 25

EV9POP Influence on the Relationship Between EV8USE and DV2 Predicted Probability

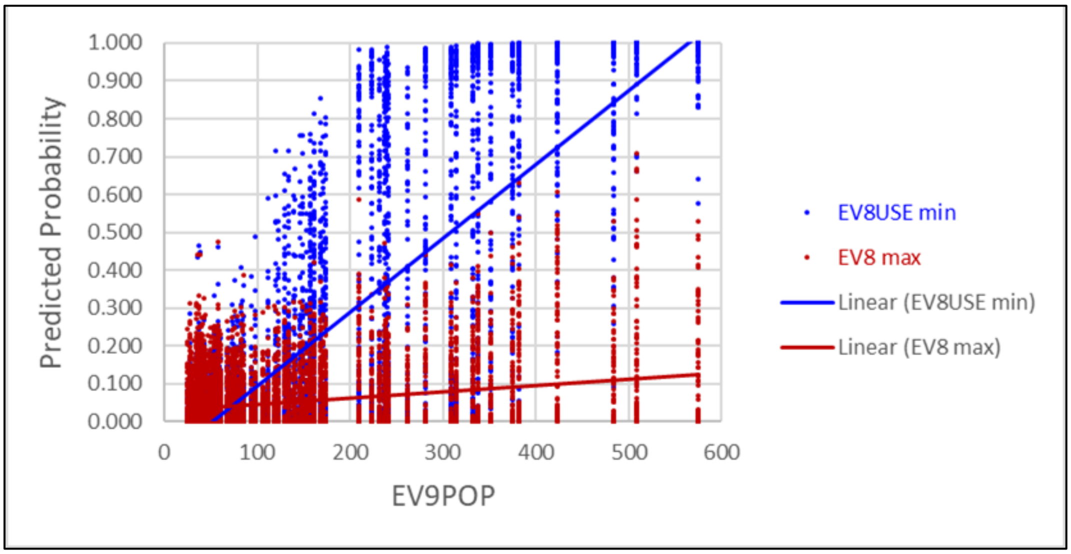


Note. *EV9POP* values, minimum (25.53), mean (131.68), and maximum (574.72).

The relationship between *EV9POP* and *DV2* was influenced by minimum, mean, and maximum average land use (Figure 26). When county land use was classified as agricultural (minimum value of 0), the probability of a WNv occurrence increased with an increase in population density. When a county land use was predominantly industrial, population density was not significantly influential to the probability of a WNv occurrence. While land use itself was not a significant predictor, the interaction with population density did produce an influential effect on probability of WNv occurrence. That probability was the highest in more densely populated agricultural areas.

Figure 26

EV8USE Influence on the Relationship Between EV9POP and DV2 Predicted Probability



Note. *EV8USE* values, minimum (0.00) and maximum (1.0).

DV$_{PRESENCE}$ Additional Findings

The *DV$_{PRESENCE}$* analysis produced a *DV2* final model that included three operationally significant predictors: average temperature, average wind speed, and average population. There were 10 2FIs in final model. The final model produced additional findings relevant to its operationally significant EVs.

Average Temperature (*$DV_{PRESENCE}$*). Manore et al. (2014) and Ozdenerol et al. (2013) identified the importance of temperature in the life cycle of WNv. In my study, the census mean average temperature was an operationally significant predictor. The final model average temperature coefficient (β = -8.380), showed a negative influence on the presence of WNv; for every one-unit increase in average temperature, the odds of the presence of WNv decreased by -.0002 (see Appendix E for selected $DV_{PRESENCE}$ SPSS parameter estimates).

Average Wind Speed (*$DV_{PRESENCE}$*). Average wind speed was an historical EEV gleaned from the literature, and it was an operationally significant predictor for $DV_{PRESENCE}$. In the final model SPSS parameter estimates, the average wind speed coefficient (β = 78.978) showed a significant positive correlation between average wind speed and $DV_{PRESENCE}$.

The census mean was 5.20 knots (SD = 1.33), with a minimum (1.51) and maximum (10.06) average wind speed. In the years reporting 50 or more WNV events (for all counties and all months: 2002, 2003, 2005, 2006, 2007, 2012), the yearly average wind speed values were all within one *SD* of the census mean. For the 360 WNv events, average wind speeds fell between a minimum and maximum of 1.64 to 9.95 (mean = 4.51). In 27 percent of the high-count cases, average wind speed was at or above the mean.

Population (*$DV_{PRESENCE}$*). Ozdenerol et al. (2013) and Manore et al. (2014) found that population and population densities were operationally significant predictors of a WNv occurrence. In my study, the population census means (x1,000) was 131.65 (*SD* = 119.18), with minimum and maximum values of 25.53 and 574.72 respectively. In years reporting 50 or more WNv incidents (2002, 2003, 2005, 2006, 2007, 2012), the yearly minimum population values were all consistent with the census means except for 2012. The year 2012, was the first year of increased WNv incidents since the 2010 census and the mean population for that year exceeded the census mean by 9.44. As the population in SC increased during the 2003 to 2012 timeframe, this increase was foreseeable.

In the $DV_{PRESENCE}$ final model, population was an operationally significant predictor at β = .397, *EXP(B)* = 1.485. The odds ratio showed that for every one-unit (or 1,000 people) increase in population, the odds of the presence of WNv increased by 1.5. Throughout this study, population proved to be an operationally significant predictor of WNv in SC, confirming its worth as a predictor variable for $DV_{PRESENCE}$.

Average Barometric Temperature and Average Temperature Interaction (*$DV_{PRESENCE}$*). In the interaction between average temperature and average barometric pressure, barometric pressure moderated the effect of temperature on the DV. When the average barometric pressure was at its maximum value (1025.25), the probability of WNv occurrence increased sharply as the average temperature rose past 45°F. When average barometric pressure was at its minimum value (972.12) and with rising temperatures, the probability of WNv occurrence decreased. When average barometric pressure was at its mean (1017.30) and the temperature rose past 45^0F, the probability of WNv occurrence also increased but at a lower rate than when the average barometric pressure was at its maximum value. The barometric pressure interaction had a moderate to significant effect on the DV for temperature ranges > 45^0F.

Average Barometric Pressure and Average Wind Speed Interaction ($DV_{PRESENCE}$). When average barometric pressure was at its maximum value (1025.25) and wind speed rose above 1 MPH, the probability of WNv occurrence decreased sharply. When average barometric pressure was at its mean value (1017.30) and wind speed rose, WNv occurrence decreased moderately. When average barometric pressure was at its minimum value (972.12) and wind speed rose, WNv occurrence increased sharply between 5-9 MPH average wind speed. This interaction shows that when wind speed was high and barometric pressure at its minimum, the probability of an WNv outbreak increased. When the winds were average or low, barometric pressure was not as influential on WNv event occurrence.

Average Barometric Pressure and Dew Point Deficit Interaction ($DV_{PRESENCE}$). When average barometric pressure was at its maximum value (1025.25), the probability of a WNv occurrence decreased sharply as the average dew point deficit increased. When average barometric pressure was held to either its mean or minimum values, the probability of a WNv occurrence was low and decreased slightly with an increase in average dew point deficit. In this interaction, the probability of a WNv occurrence was highest when average barometric pressure was high and the values for average dew point deficit were low. When the barometric pressure was moderate or low, dew point deficit was not a significant influence on probability.

Month EV ($DV_{PRESENCE}$). In the final model, the months of *May*, *June*, *August*, *September, October, and November* were significant at $p < \alpha = .05$. The coefficients and associated odds ratios for each of the months increased from May to October in the final model estimates; *May* (β = -1.63, Exp(B) = .196), *June* (β = -0.449, Exp(B) = 0.638), *August* (β = 0.199, Exp(B) = 1.221), *September* (β = 0.163, Exp(B) = 1.177), *October* (β = .659, Exp(B) = 1.933) and *November* (β = -0.354, Exp(B) = 0.702). The increasing odds ratios for the months *June* through *October* tracked with the EDA that found 96% of WNv occurring in these months. The months of *May*, *June*, and *November* were the only months correlated negatively to the DV. May had only seven WNv events over the study timeframe from 2002-2016.

DV_{COUNT} Results

The section begins with the DV_{COUNT} descriptive statistics, EDA, and assumptions. Stage B1 sets the stage for the deeper level of regression analysis required in Stages B2 to B4. Following the GZLM assumptions, the DV_{COUNT} analyses are provided.

Models for DV_{COUNT} were examined with a SFEM process using goodness-of-fit statistics for EVs. The primary goodness of fit statistics was *D* (Deviance). Buro (2020) stated that “the Deviance of a model is based on the difference between the log-likelihood of the model of interest, L_M, and the log-likelihood of the most complex model that perfectly fits the data (one parameter per "measurement", saturated model), L_S" (p. 17).

$$Deviance = -2(L_M - L_S)$$

Deviance is a χ^2 statistic that can be used to test if the saturated model gives a better model fit than a proposed model. A D value of 1.0 would reflect a perfect fit of the data. Values of < 1.0 or > 1.0 would reflect under and overdispersion of the data. For completeness, I also monitored and documented the Pearson Chi Square (χ^2_P) and the omnibus test Log Likelihood Ratio Chi Square (*LL2 Ratio* χ^2) statistics.

In Stage B, models for each DV were evaluated using a Tweedie MPV of 1.5 with the Log link function selected. When all of a model's EEVs were significant ($p < \alpha = .20$), I compared the D, *LL2 Ratio* χ^2 scores, and conducted a subject matter expert review of the model composition.

Stage B1 DV_{COUNT} Descriptive Statistics, EDA, and Assumptions

To build to a final DV_{COUNT} model for the current month and time-lagged DVs, I used the same data collection, preparation, data analysis, regression model-building, and validation process described earlier for the $DV_{PRESENCE}$ analysis. I created a DV_{COUNT} data set from the study master data set to include only cases for which there was a count of at least one recorded positive WNv incident. This approach varied from Kurz (2017) in that I excluded all zero count cases to focus my analysis on the specific WNv positive count data. This resulted in a DV_{COUNT} data set of 970 cases and 360 events that included 902 incidents.

***DV_{COUNT}* Dataset Descriptive Statistics and EDA**. While the descriptive statistics (Table 19) and distributions for this data set (Figures 27 to 36) were a necessary component of the Stage B analysis, they also provided a comparison of the *$DV_{PRESENCE}$* (BLR) and *DV_{COUNT}* (GZLM) datasets. Within the descriptive statistics, the *ΔMean* values represented the differences in mean values between the two data sets which included 8,280 and 970 cases respectively.

Table 19

Descriptive Statistics for the DV_{COUNT} (GZLM) Dataset

EEV	*Mean*	*Minimum*	*Maximum*	*SD*	*ΔMean*
EV1ATM	72.84	36.99	92.01	8.76	9.48
EV2ARN	0.14	0.00	0.48	0.09	0.00
EV4ASD	0.00	0.00	0.28	0.01	0.00
EV5ABP	1016.61	979.57	1023.80	2.73	6.49
EV6AWS	5.04	1.64	11.57	1.61	-.06
EV7ELV	349.19	10.00	1093.00	317.84	4.72
EV8USE	0.50	0.00	1.00	0.50	0.00
EV9POP	197.64	27.47	574.72	142.41	78.46
EV10ADD	10.32	0.00	33.43	3.89	6.43
EVMonth	0.50	0.00	1.0	0.50	0.00

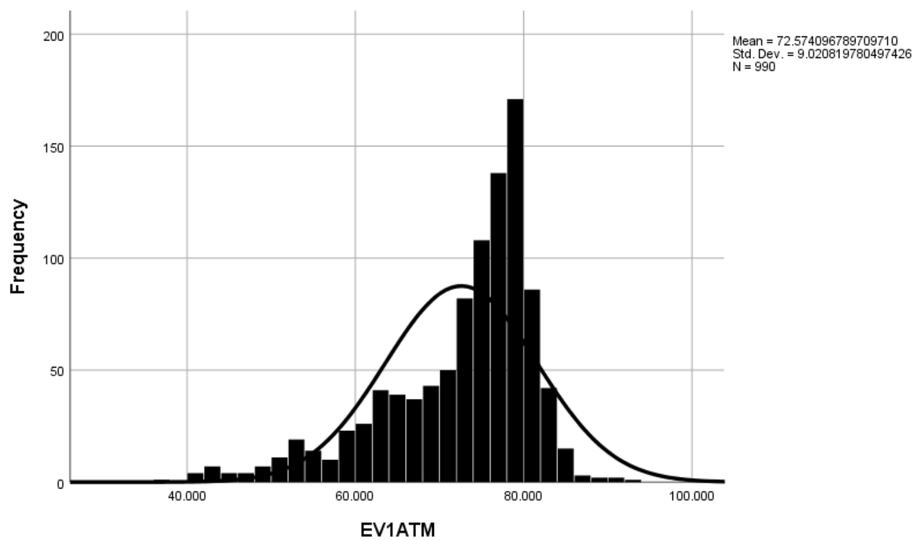
The statistics showed that the average temperature used in the *DV_{COUNT}* analysis was higher for cases with recorded WNv incidents than in the population of all cases. There was also a considerable difference in the average population means. In the

DV_{COUNT} analysis, the sample average population was higher. These findings informed the Stage B analysis.

Average Temperature (EV1ATM). The mean monthly temperature was 72.57 (*SD = 9.02*). The minimum temperature recorded was 36.99 and the maximum was 92.01, indicating a range of 55.02 (Figure 27).

Figure 27

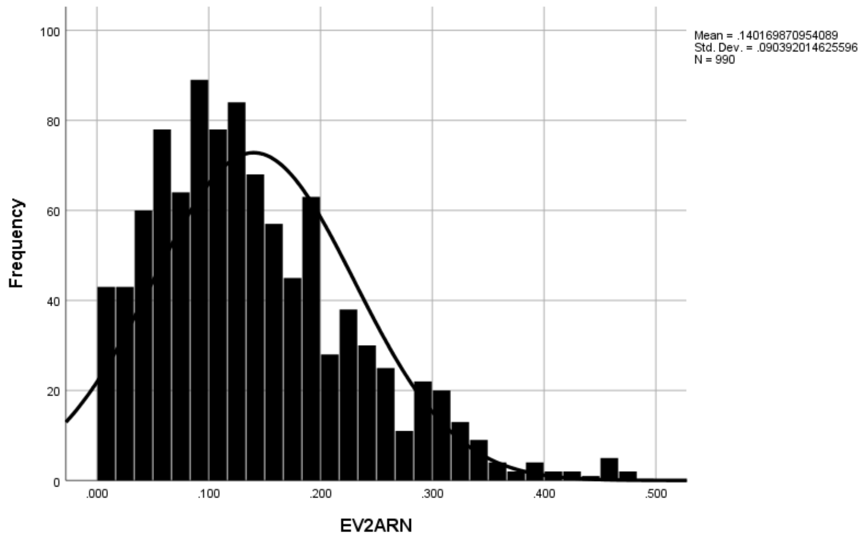
GZLM Dataset EV1ATM Histogram



Average Rainfall (EV2ARN). The mean monthly rainfall was .140 (*SD = .09*). The minimum rainfall recorded was 0.00 and the maximum was 0.48, indicating a range of 0.48 (Figure 28).

Figure 28

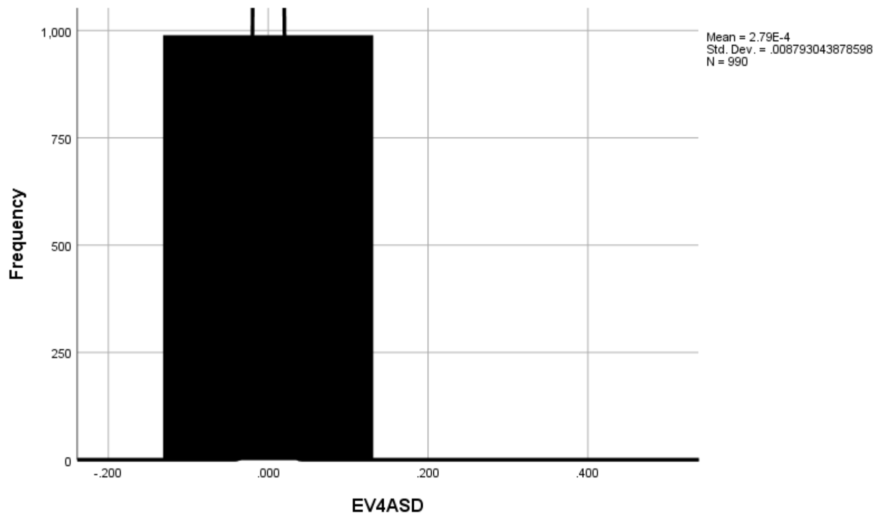
GZLM Dataset EV2ARN Histogram



Average Snow Depth (EV4ASD). The mean snow depth was zero (*SD = .009*). The minimum snow depth recorded was zero and the maximum was 0.28, indicating a range of 0.28 (Figure 29).

Figure 29

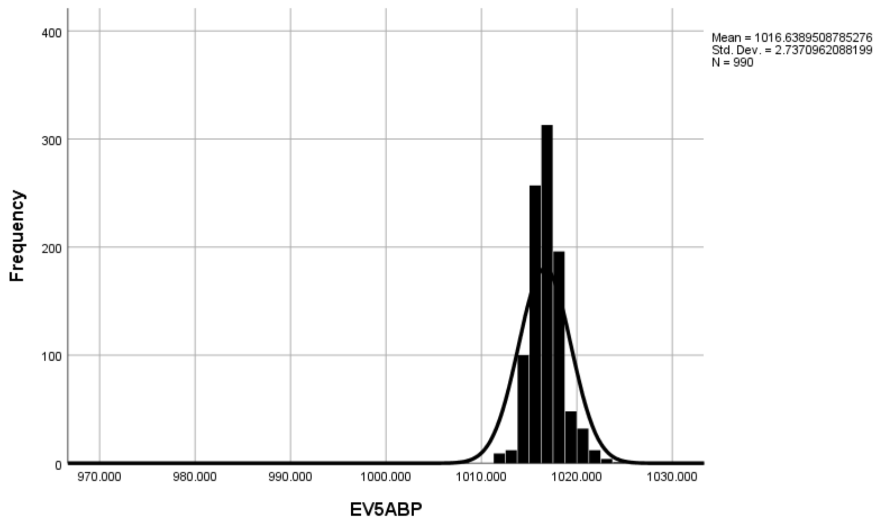
GZLM Dataset EV4ASD Histogram



Average Barometric Pressure (EV5ABP). The mean barometric pressure was 1016.64 (SD = .2.74). The minimum barometric pressure recorded was 979.57 and the maximum was 1023.78, indicating a range of 44.21 (Figure 30).

Figure 30

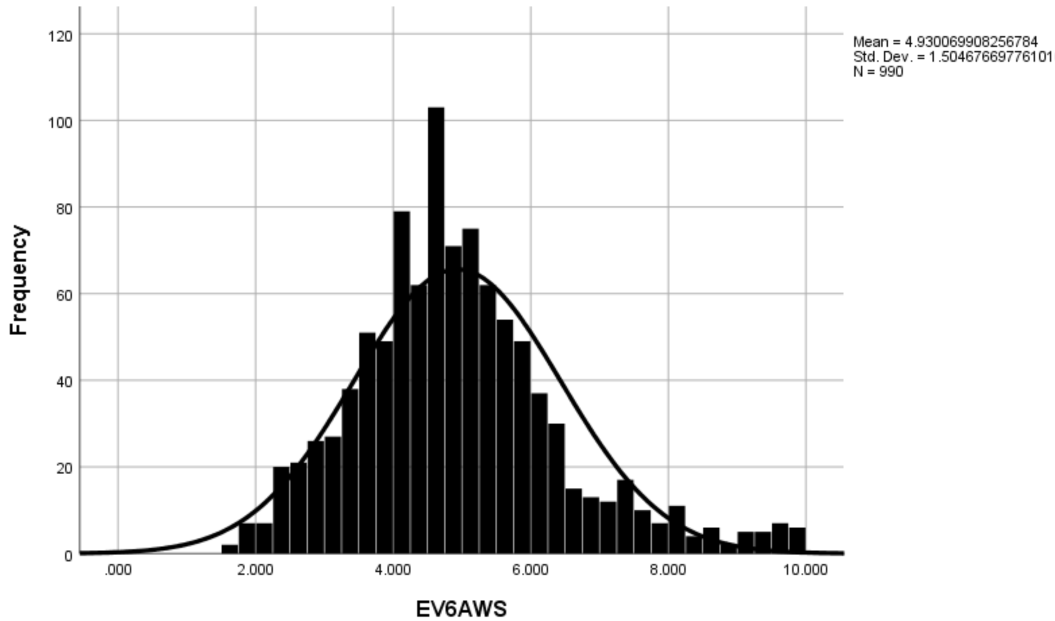
GZLM Dataset EV5ABP Histogram



Average Wind Speed (EV6AWS). The mean wind speed was 4.93 (SD = 1.50). The minimum wind speed recorded was 1.64 and the maximum was 9.96, indicating a range of 8.32 (Figure 31).

Figure 31

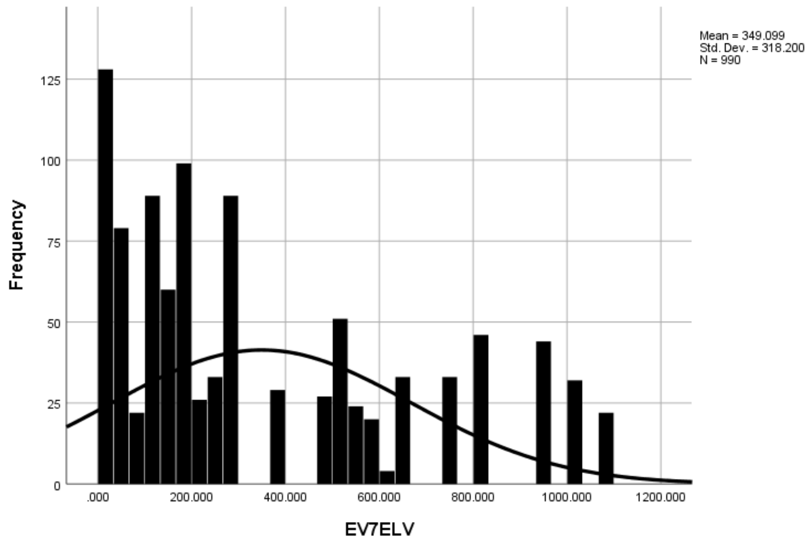
GZLM Dataset EV6AWS Histogram



Average Elevation (EV7ELV). The mean elevation was 349.10 (SD = 318.20). The minimum elevation recorded was 10 and the maximum was 1093, indicating a range of 1083 (Figure 32).

Figure 32

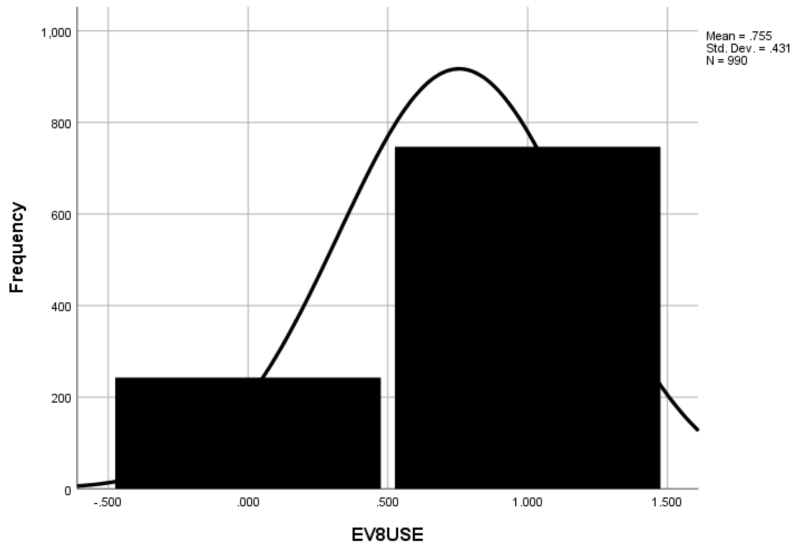
GZLM Dataset EV7ELV Histogram



Land Use (EV8USE). Land use was classified as Agricultural (29%) and Industrial (71%) based on population per square mile (Figure 33).

Figure 33

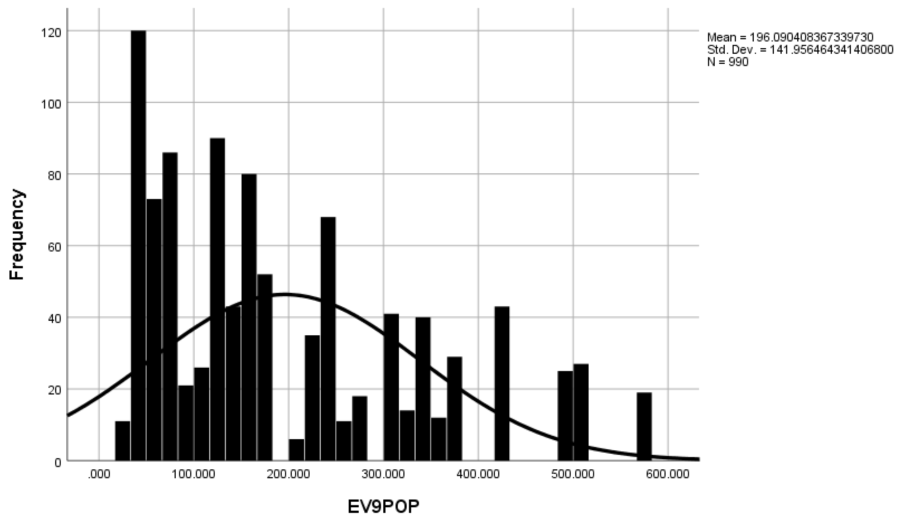
GZLM Dataset EV8USE Histogram



Average Population (*EV9POP*). The mean population was 196.24 (SD = 141.96). The minimum population recorded was 27.48 and the maximum was 574.72, indicating a range of 547.24 (Figure 34).

Figure 34

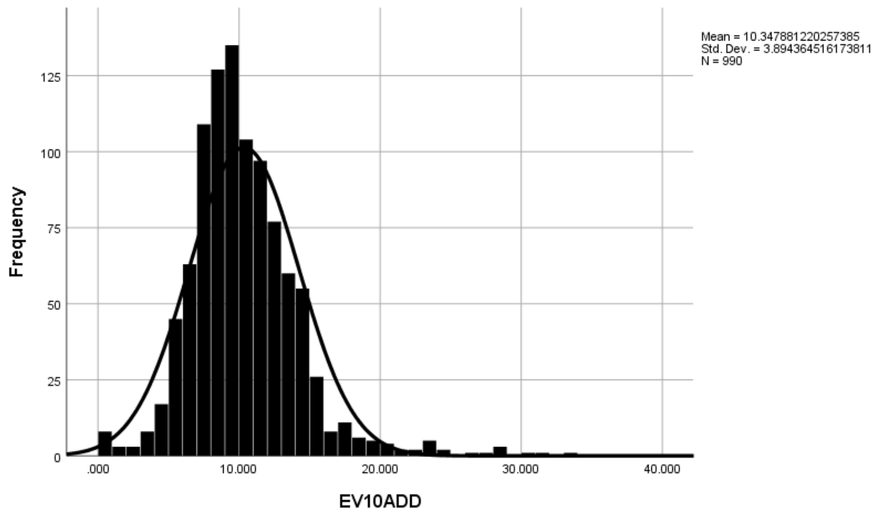
GZLM Dataset EV9POP Histogram



Average Dew Point Deficit (EV10ADD). The mean dew point deficit was 10.35 (*SD* = 3.90). The minimum dew point deficit recorded was 0.00 and the maximum was 33.43, indicating a range of 33.43 (Figure 35).

Figure 35

GZLM Dataset EV10ADD Histogram



***DV_{COUNT}* GZLM Assumption Testing**. Although less restrictive than MLR, GZLM regression relies upon assumptions that required examination to ensure the correct statistical tool was being used (Javaras & Vos, 2002):

1. The data $Y_1, Y_2, \ldots, Y_n$ are independent (i.e., cases are independent).
2. The DV does not need to be normally distributed, but it typically assumes a distribution from an exponential family (e.g., binomial, Poisson, multinomial, etc.).
3. There is a linear relationship between the transformed response (using the link function) and the explanatory variables.
4. The homogeneity of variance does not need to be satisfied.

5. Residuals need to be independent but do not need to be normally distributed.

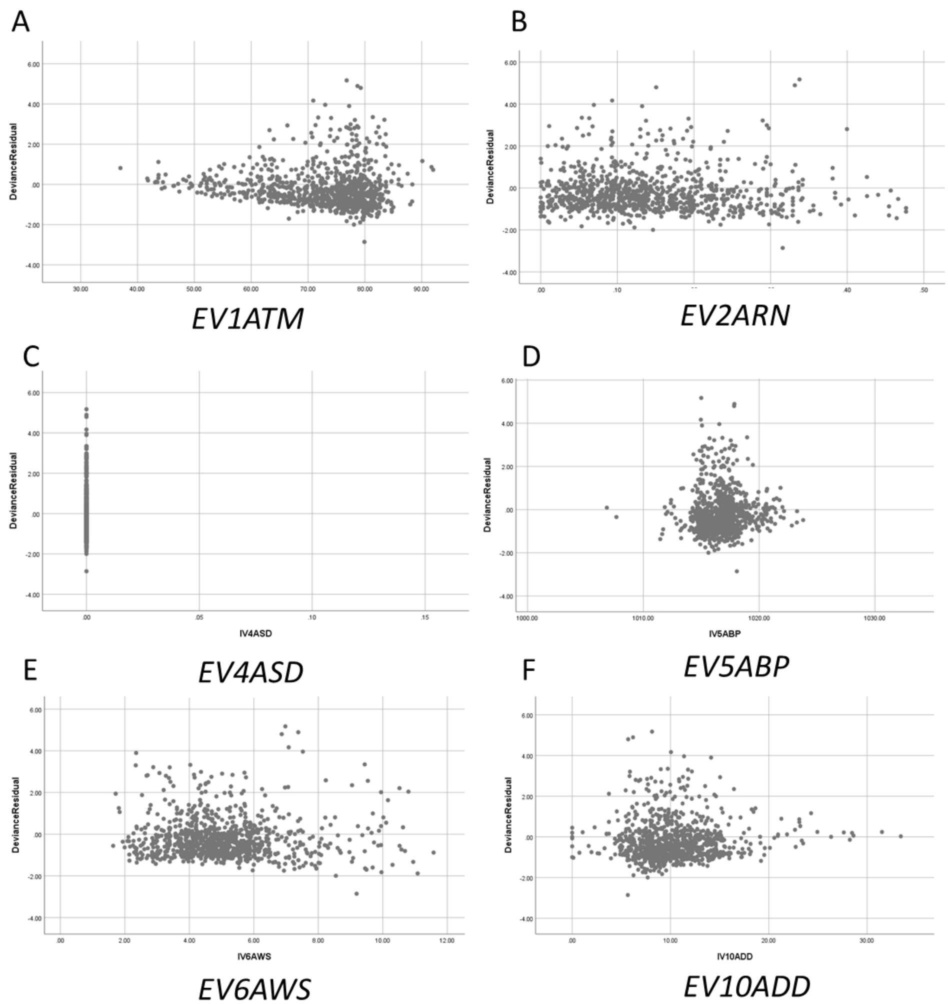
6. The regression relies on large sample approximations.

Assumptions 1 and 5 were met based on the independence of the observations used to collect the overall WNv data. To test this assumption, I plotted the raw residuals against all of the continuous EEVs and noted no patterns in the plots (Figure 36).

Assumption 2 was met during the EDA where I discovered that the data were heavily skewed, favoring a Poisson distribution. The assumption of linearity (Assumption 3) is modified with GZLM. In describing GZLM regression assumptions, Javaras and Vos (2002) stated that, “the essence of linear models is that the response variable is continuous and normally distributed: here we relax these assumptions and consider cases where the response variable is non-normal and in particular has a discrete distribution” (p. 1).

Figure 36

GZLM Test for Independence of Observations



Note. Panel A-F scatterplots related to continuous EEVs used in GZLM analysis.

To meet this assumption, I transformed the DV data and then assessed the relationship between the transformed data and the EEVs. The results are provided in scatterplots in Figures 37 to 43. Assumption 6 was met as the GZLM dataset included all reported incidents of WNv from the census population.

Figure 37

Scatterplot for Test of Linearity Between EV6 and the DV

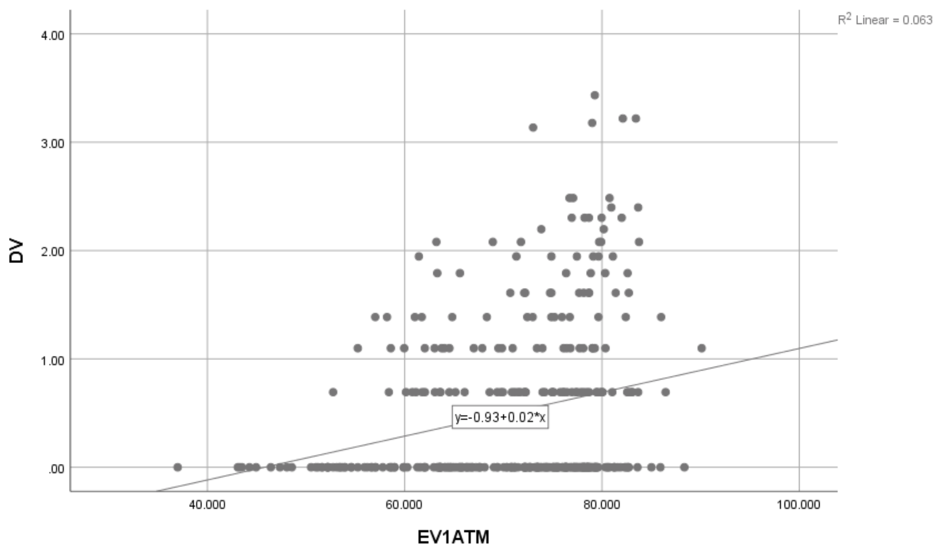


Figure 38

Scatterplot for the Test of linearity Between EV2ARN and the DV

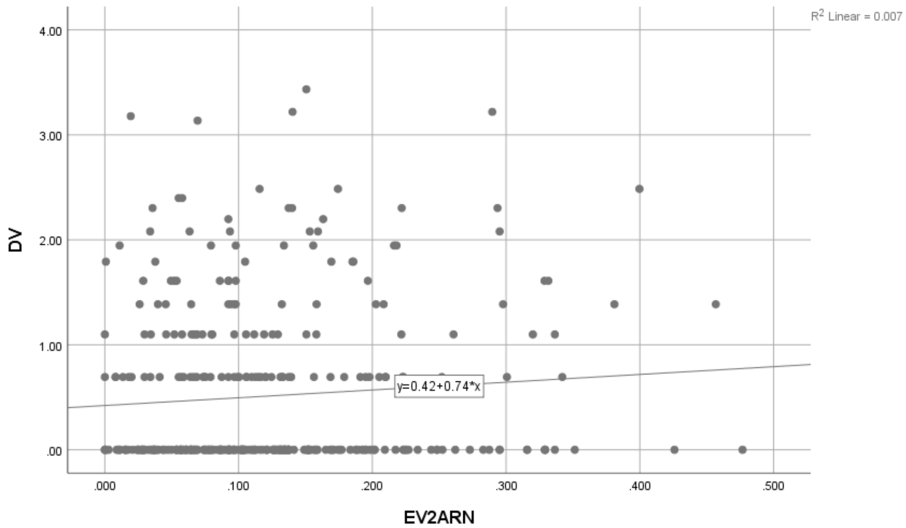


Figure 39

Scatterplot for the Test of Linearity Between EV5ABP and the DV

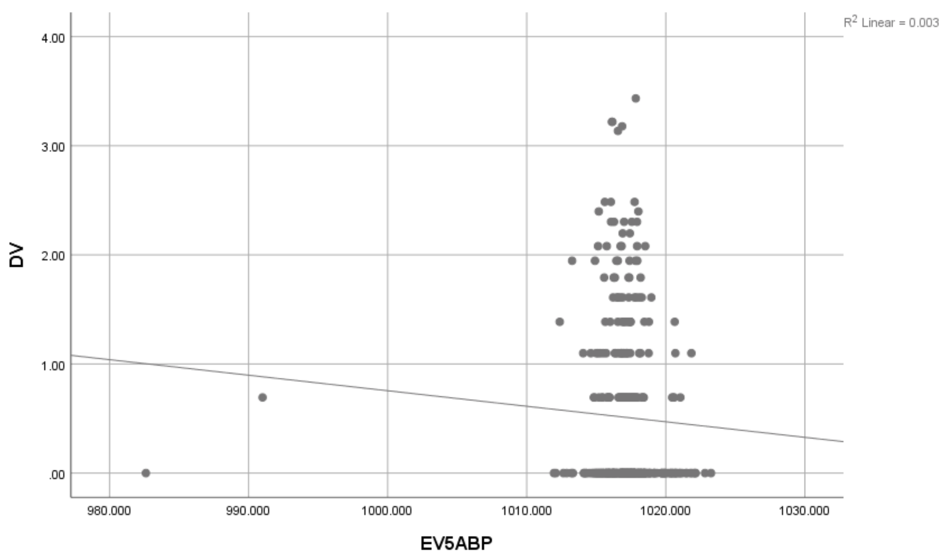


Figure 40

Scatterplot for the Test of Linearity Between EV6AWS and the DV

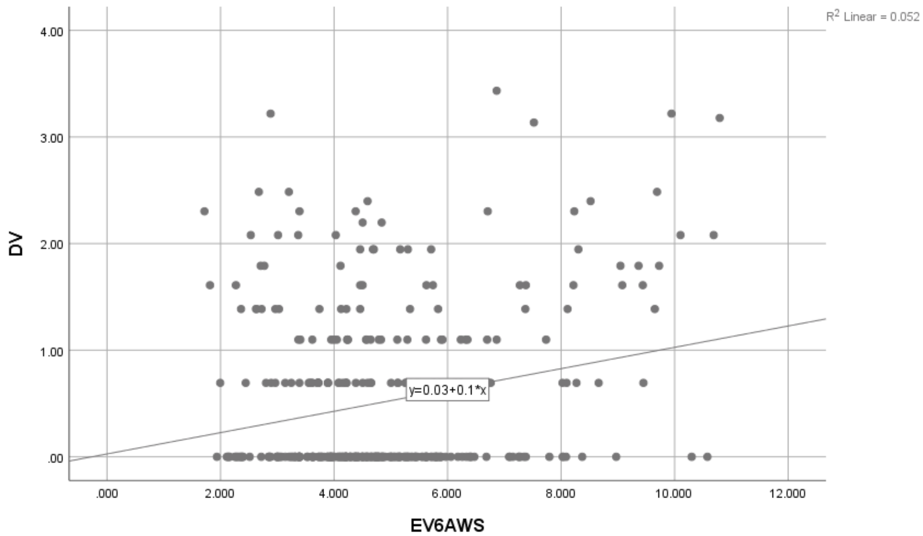


Figure 41

Scatterplot for the Test of Linearity Between EV7ELV and the DV

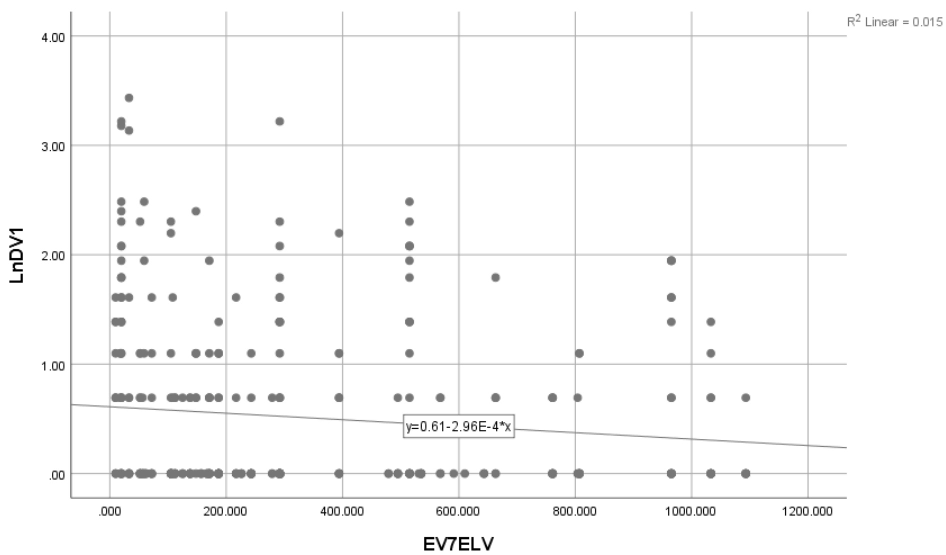


Figure 42

Scatterplot for the Test of Linearity Between EV9POP and the DV

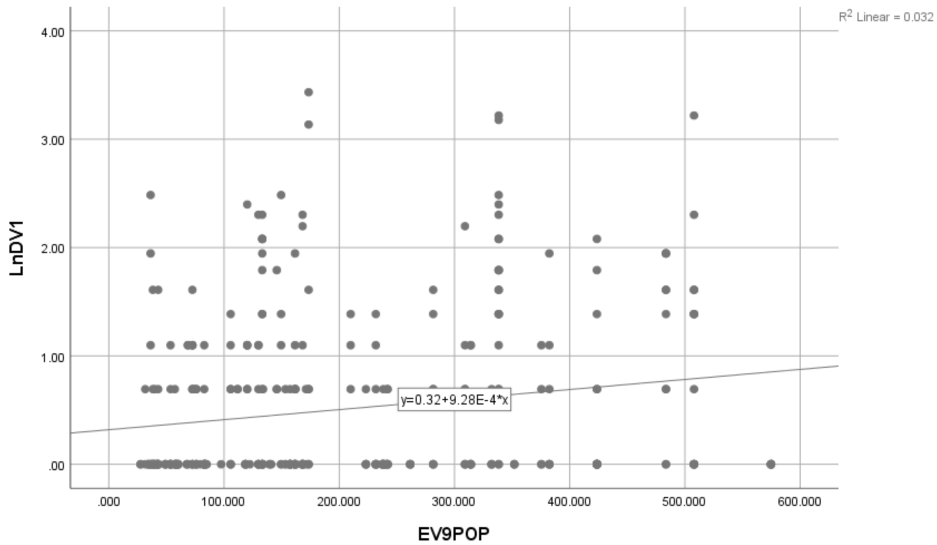
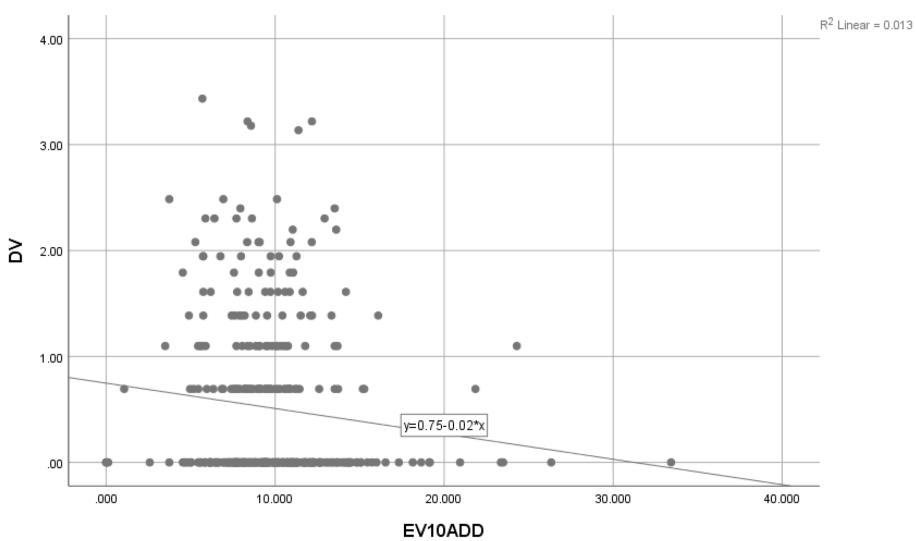


Figure 43

Scatterplot for the Test of Linearity Between EV10ADD and the DV



Stage B2 DV$_{COUNT}$ Current Month DV Selection

As *DV1* was the only current month (0-day) outcome variable, an initial comparative analysis of DVs was not required. The Stage B2 analysis began with the following nine EEVs: *EV1ATM*, *EV2ARN*, *EV4ASD*, *EV5ABP*, *EV6AWS*, *EV7ELV*, *EV8USE*, *EV9POP*, and *EV10ADD*.

Stage B2 DV$_{COUNT}$ for Current Month With DV1 and EVs

Three models were developed in the initial step of the model building process (Table 20). Model 102 was the best model of the three with a *D* = 3.271. Although, this *D* was the lowest of the three it reflected overdispersion of the data. *EV6AWS* (p = .756) and *EV4ASD* (p = .289) were removed under the $p < \alpha = .20$ criterion.

Table 20

Stage B2 DV$_{COUNT}$ Analysis With DV1 and EVs

DV$_{COUNT}$	Tweedie MPV	*D*	χ^2_P	*AIC*	*LL2 Ratio* χ^2	*df*	*.sig*
DV1 Model 101	1.5	3.274	4519.157	2608.507	71.917	9	.000
DV1 Model 102	1.5	3.271	4522.391	2606.603	71.820	8	.000
DV1 Model 103	1.5	3.272	4533.620	2606.779	70.645	7	.000

Stage B2 DV$_{COUNT}$ for Current Month with DV1, EEVs, and 2FIs

This step of model development required the addition of the 21 2FI predictors (Table 21).

Table 21

Stage B DV1 2FIs

	EV1ATM	*EV2ARN*	*EV5ABP*	*EV7ELV*	*EV8USE*	*EV9POP*	*EV10ADD*
EV1ATM	**	*X*	*X*	*X*	*X*	*X*	*X*
EV2ARN	**	**	*X*	*X*	*X*	*X*	*X*
EV5ABP	**	**	**	*X*	*X*	*X*	*X*
EV7ELV	**	**	**	**	*X*	*X*	*X*
EV8USE	**	**	**	**	**	*X*	*X*
EV9POP	**	**	**	**	**	**	*X*

The Stage B2 analysis with *DV1*, *EEVs*, and *2FIs* produced 20 models (Table 22). Model 117 was the best interim model at $D = 3.146$, *LL2 Ratio* $\chi^2(15) = 108.206$, $p < .001$. The model was composed of the following EVs: *EV1ATM*, *EV7ELV*, *EV8USE*, *EV1·EV2*, *EV1·EV5*, *EV1·EV7*, *EV1·EV8*, *EV1·EV9*, *EV2·EV5*, *EV2·EV8*, *EV2·EV9*, *EV5·EV7*, *EV5·EV10*, *EV7·EV10*, *EV8·EV10*, *EV9·EV10*. *EV10ADD* ($p = .911$), *EV2ARN* ($p = .493$), *EV9POP* ($p = .697$), *EV5ABP* ($p = .433$), *EV7ELV* ($p = .209$).

Table 22

Stage B2 DV_{COUNT} Analysis With DV1, EVs, and 2FIs

DV_{COUNT}	Tweedie MPV	*D*	χ^2_P	*AIC*	*LL2* Ratio χ^2	*df*	*.sig*
DV1 Model 104	1.5	3.171	4153.884	2605.794	112.629	28	.000
DV1 Model 105	1.5	3.168	4157.235	2603.806	112.617	27	.000
DV1 Model 106	1.5	3.164	4171.766	2601.839	112.584	26	.000
DV1 Model 107	1.5	3.161	4174.960	2599.881	112.543	25	.000
DV1 Model 108	1.5	3.158	4175.527	2597.934	112.490	24	.000
DV1 Model 109	1.5	3.155	4194.011	2596.090	112.334	23	.000
DV1 Model 110	1.5	3.153	4195.498	2594.384	112.039	22	.000
DV1 Model 111	1.5	3.151	4247.277	2592.761	111.662	21	.000
DV1 Model 112	1.5	3.150	4284.068	2591.261	111.163	20	.000
DV1 Model 113	1.5	3.149	4288.419	2589.736	110.688	19	.000
DV1 Model 114	1.5	3.148	4268.160	2588.434	109.990	18	.000
DV1 Model 115	1.5	3.149	4283.300	2587.449	108.974	17	.000
DV1 Model 116	1.5	3.146	4274.052	2585.601	108.823	16	.000
DV1 Model 117	1.5	3.146	4265.900	2584.217	108.206	15	.000
DV1 Model 118	1.5	3.148	4287.538	2583.711	106.712	14	.000
DV1 Model 119	1.5	3.151	4283.710	2583.312	105.112	13	.000
DV1 Model 120	1.45	2.805	4187.574	2535.911	128.793	15	.000
DV1 Model 121	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	.000
DV1 Model 122	1.15	2.442	4087.575	2579.901	216.068	15	.000
DV1 Model 123	1.1	2.421	4083.084	2638.255	253.262	15	.000

Stage B2 DV_{COUNT} for Current Month With DV1, All EVs

This final step in the *DV1* produced six models (Table 23). The months of *December* (*p* = .951), *July* (*p* = .951), *and January* (*p* = .527) were removed at $p < \alpha = .20$ along with two 2FIs. Model 129 was the best model at *D* = 1.898, *LL2 Ratio* $\chi^2(22) = 582.309$, $p < .001$. The model was composed of the following EEVs: *EV1ATM*, *EV7ELV*, *EV8USE*, *EV1·EV2, EV1·EV7, EV1·EV8, EV1·EV9, EV2·EV5, EV2·EV9, EV5·EV7, EV5·EV10, EV7·EV10, EV8·EV10, EV9·EV10, February*, *March*, *April*, *May*, *June*, *August*, *September*, *October*, *November*.

Table 23

Stage B2 DV_{COUNT} Analysis With DV1, All EVs

DV_{COUNT}	Tweedie MPV	*D*	χ^2_P	*AIC*	*LL2 Ratio* χ^2	*df*	*.sig*
DV1 Model 124	1.1	1.902	3060.265	2328.891	584.626	26	.000
DV1 Model 125	1.1	1.902	3060.265	2328.891	584.626	26	.000
DV1 Model 126	1.1	1.900	3060.049	2326.895	584.623	25	.000
DV1 Model 127	1.1	1.899	3057.706	2325.309	584.209	24	.000
DV1 Model 128	1.1	1.899	3081.439	2324.941	582.576	23	.000
DV1 Model 129	1.1	1.898	3081.895	2323.209	582.309	22	.000

This Stage B2 interim model also excluded the month of *July*, as was the case in the $DV_{PRESENCE}$ modeling. In the EDA, the month of *July* accounted for 108 WNv incidents. However, in both sets of analysis, the *July* EV was removed very early for significance when the Month EVs were added to the model. This occurred with both the

DV$_{PRESENCE}$ and *DV*$_{COUNT}$ datasets and with different statistical tools. I decided to follow the statistics produced by SPSS and remove the *July* EV in both cases.

Stage B3 DV$_{COUNT}$ *for Current Month Final Model, α = .05*

The analysis of *DV1* at $\alpha = .05$ produced six models (Table 24). The month EVs of *March* ($p = .142$) and *February* ($p = .146$) were removed at $p < \alpha = .05$ along with two 2FIs. Model 132 was the best model at $D = 1.898$, *LL2 Ratio* $\chi^2(20) = 579.499$, $p < .001$. This model retained the months of *March and February*. The final model was composed of the following EEVs: *EV1ATM, EV7ELV, EV8USE, EV1·EV2, EV1·EV7, EV1·EV8, EV2·EV5*, *EV2·EV9*, *EV5·EV10, EV7·EV10, EV8·EV10*, *February*, *March*, *April*, *May*, *June*, *August*, *September*, *October*, *November*.

Table 24

Stage B3 DV$_{COUNT}$ *Analysis With DV1, All EVs, α = .05*

DV$_{COUNT}$	Tweedie MPV	*D*	χ^2_P	*AIC*	*LL2 Ratio* χ^2	*df*	*.sig*
DV1 Model 130	1.1	1.898	3081.895	2323.209	582.309	22	.000
DV1 Model 131	1.1	1.899	3062.322	2323.304	580.213	21	.000
DV1 Model 132	1.1	1.898	3062.937	2022.018	579.499	20	.000
DV1 Model 133	1.1	1.901	3067.056	2322.826	576.692	19	.000
DV1 Model 134	1.1	1.903	3061.548	2323.877	573.648	18	.000
DV1 Model 135	1.1	1.903	3077.320	2323.045	572.472	17	.000

DV_{COUNT} for Current Month Final Model

After reviewing all the Stages B2 to B3 models, I selected Model 129 as the final model at D = 1.898, *LL2 Ratio* χ^2 = 582.309 (22), p < .001. Following this analysis, I performed a diagnostic test on the residuals checking Cook's Distance (*COO1*), and leverage (*LEV1*). There were no unusually high values of Cook's distance (mean = 0.0026). All 970 case values of the leverage residual (*LEV1*) were less than 1 (mean = .023). Based on the results of the diagnostics test, I used this version of the regression analysis to select my final Stage A model. The final model resulted in the following forecast accuracies: MAE = 1.074, RMSE = 2.490. The final logistic regression model was expressed as follows:

The final model consisted of the following EEVs: *EV1ATM*, *EV7ELV*, *EV8USE*, *EV1·EV2*, *EV1·EV7*, *EV1·EV8*, *EV2·EV5*, *EV2·EV8*, *EV2·EV9*, *EV5·EV10*, *EV7·EV10*, *EV8·EV10*, *February*, *March*, *April*, *May*, *June*, *August*, *September*, *October*, *November*. The final model was expressed as (see Equation 11):

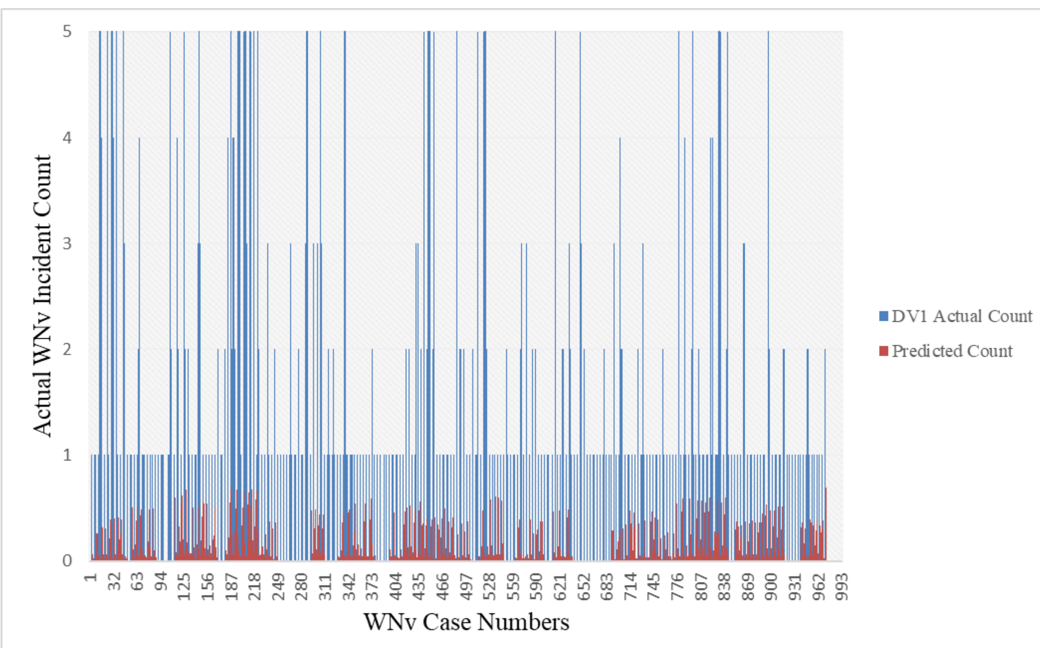
$$\hat{Y} = -4.936 + (-0.029 \cdot EV1ATM) + (0.003 \cdot EV7ELV) + (2.929 \cdot EV8USE) + (0.332 \cdot EV1 \cdot EV2) + (-0.00007 \cdot EV1 \cdot EV7) + (0.030 \cdot EV1 \cdot EV8) + (0.00003 \cdot EV1 \cdot EV9) + (-0.028 \cdot EV2 \cdot EV5) + (0.009 \cdot EV2 \cdot EV9) + (-0.00012 \cdot EV5 \cdot EV10) + (0.00008 \cdot EV7 \cdot EV10) + (-0.123 \cdot EV8 \cdot EV10) + (-0.0002 \cdot EV9 \cdot EV10) + (-0.019 \cdot February) + (1.853 \cdot March) + (1.328 \cdot April) + (1.940 \cdot May) + (2.202 \cdot June) + (-0.930 \cdot August) + (-1.157 \cdot September) + (-0.952 \cdot October) + (-0.905 \cdot November) \quad (11)$$

Based on these results, I rejected the RQ1 DV_{COUNT} null hypothesis ($H_0 1$) at $p < \alpha = .05$. I concluded there was evidence that at least one coefficient in the final regression model was not equal to zero, and that the final model was a statistically significant predictor of *DV1* for incident counts in the current month.

An examination of actual *DV1* WNv incident counts versus the predicted counts from the final model equation reveals a very low predicted count mean of .499 (Figure 44). Actual WNv incident counts are shown in black by case. The final model predicted counts of WNv incidents are shown in red by case.

Figure 44

DV1 Actual WNv Counts Versus Predicted WNv Counts, All Counties



Stage B4 DV_{COUNT} Time-Lagged DV Selection

Stage B4 began with the following EEVs: *EV1ATM*, *EV2ARN*, *EV5ABP*, *EV6AWS*, *EV7ELV*, *EV8USE*, *EV9POP*, and *EV10ADD*. This stage was conducted to determine which of the time-lagged DVs (*DV2* through *DV5*) produced the best model goodness-of-fit and *LL Ratio* χ^2 scores. These were the same time-lagged DVs I examined in Stage A, with the addition of *DV5* which was computed as a 90-day moving average.

Stage B4 produced 17 models for comparison (Models 136-152). Comparing the interim models for each DV (Table 25), Model 152 (*DV5*) provided the best result at D = 1.023 and with an *AIC* = 1731.274. The Model 88 build process removed *EV4ASD* (p = .928) and *EV5ABP* (p = .782).

Final models for all the DVs excluded *EV4ASD* for a lack of significance. I ran a separate GZLM analysis of *EV4ASD* with *DV5* and found that *EV4ASD* was not significant at p = .928. *EV5ABP* was also excluded in Model 86 at p = .782, as it was earlier in Model 80 by *DV3* at p = .962. Based on the EDA findings and the Stage B2-3 results, I elected to remove *EV4ASD and EV5ABP* in the Stage B4 modeling. Based on these results, I selected *DV5* for Stage B4. Model 152 consisted of the following EEVs: *EV1ATM*, *EV2ARN*, *EV6AWS*, *EV7ELV*, *EV8USE*, *EV9POP*, and *EV10ADD*.

Table 25

Stage B4 DV_{COUNT} Time-Lagged Model Results Using EEVs Only

Model	DV	Tweedie MPV	*D*	χ^2_P	*AIC*	*LL2 Ratio* χ^2	*df*	*.sig*
141	*DV2*	1.5	2.825	3824.156	2426.375	190.984	7	.000
145	*DV3*	1.5	2.952	4226.178	2527.626	145.55	4	.000
147	*DV4*	1.5	3.104	4559.273	2569.364	112.038	8	.000
152	*DV5*	1.5	1.023	1731.274	2471.097	424.903	8	.000

Based on the Stage B4 results, I rejected the RQ2 DV_{COUNT} null hypotheses for each of the time lagged DVs (H_02 H_03, H_04) at $p < \alpha = .05$. I concluded there was evidence that at least one coefficient in the final regression model was not equal to zero and that the final model was a statistically significant predictor of DV_{COUNT}. I continued with *DV5* to examine the best DV_{COUNT} model for future WNv count prediction.

Stage B5 DV_{COUNT} Analysis With DV5, EEVs, and 2FIs

Stage B5 began with seven significant EEVs for *DV5*: *EV1ATM*, *EV2ARN*, *EV6AWS*, *EV7ELV*, *EV8USE*, *EV9POP*, and *EV10ADD*. Along with these seven EEVs, 21 2FIs were added (Table 26).

Table 26

Stage B5 DV_{COUNT} DV5 2FIs

	EV1ATM	*EV2ARN*	*EV6AWS*	*EV7ELV*	*EV8USE*	*EV9POP*	*EV10ADD*
EV1ATM	·	X	X	X	X	X	X
EV2ARN			X	X	X	X	X
EV6AWS				X	X	X	X
EV7ELV					X	X	X
EV8USE						X	X
EV9POP							X

Modeling in Stage B4 was done using SFEM (Table 27). At each step, EEVs (*EV6AWS*, *IV7ELV*, *IV9POP*) were eliminated based on an $p < \alpha = .20$, as well as the model goodness of fit (*D*). When all predictors were at $p < \alpha = .20$, I ran a series of model runs to adjust the Tweedie MVP. Model 169 produced the best model at Tweedie MVP = 1.44 with a *D* = .999 and *AIC* = 2487.825. This model consisted of the following EEVs: *EV1ATM*, *EV2ARN*, *EV8USE*, *EV10ADD*, *EV1·EV2*, *EV1·EV6*, *EV1·EV8*, *EV1·EV10*, *EV2·EV7*, *EV2·EV9*, *EV6·EV7*, *EV6·EV10*, *EV7·EV8*, *EV7·EV9*, *EV8·EV10*, *EV9·EV10*. The interim Model 169 predictors were moved into Stage B5 with all EVs.

Table 27

Stage B5 DV$_{COUNT}$ Analysis With DV5, EEVs, and 2FIs

Model	Tweedie MPV	*D*	χ^2_P	*AIC*	*LL2* *Ratio* χ^2	*df*	*.sig*
153	1.5	0.981	1555.298	2433.181	504.819	28	.000
154	1.5	0.980	1554.768	2431.187	504.813	27	.000
155	1.5	0.979	1554.717	2429.206	504.794	26	.000
156	1.5	0.978	1555.581	2427.324	504.676	25	.000
157	1.5	0.977	1561.365	2425.522	504.478	24	.000
158	1.5	0.976	1563.805	2423.921	504.079	23	.000
159	1.5	0.975	1563.087	2422.156	503.843	22	.000
160	1.5	0.975	1563.464	2420.635	503.365	21	.000
161	1.5	0.974	1559.727	2419.144	502.856	20	.000
162	1.5	0.974	1565.207	2417.839	502.161	19	.000
163	1.5	0.973	1570.098	2416.673	501.327	18	.000
164	1.5	0.973	1575.309	2416.107	499.893	17	.000
165	1.5	0.973	1571.462	2414.268	499.732	16	.000
166	1.45	0.995	1598.914	2475.274	513.636	16	.000
167	1.40	1.019	1627.926	2539.175	531.239	16	.000
168	1.43	1.006	1613.221	2506.868	521.876	16	.000
169	1.44	0.999	1604.589	2487.825	516.812	16	.000

Stage B5 DV$_{COUNT}$ Analysis with DV5, All EVs; $\alpha = .20$

This step of the analysis began with the Model 169 predictors and added the *Months* EEVs: *EV1ATM*, *EV2ARN*, *EV8USE*, *EV10ADD*, *EV1ATM·EV2ARN*,

EV1ATM·EV6AWS, *EV1ATM·EV8USE*, *EV1ATM·EV10ADD*, *EV2ARN·EV7ELV*, *EV2ARN·EV8USE*, *EV2ARN·EV9POP*, *EV6AWS·EV7ELV*, *EV6AWS·EV10ADD*, *EV7ELV·EV9POP*, *EV8USE·EV10ADD*, *EV9POP·EV10ADD*, *January*, *February*, *March*, *April*, *May*, *June*, *July*, *August*, *September*, *October*, *November*, *December*. The Tweedie MVP was set at 1.44 and significance level remained at $p < \alpha = .20$.

Table 28

Stage B5 DV_{COUNT} Results for DV5, All EVs

Model	Tweedie MPV	*D*	χ^2_P	*AIC*	*LL2 Ratio* χ^2	*df*	*.sig*
170	1.44	.984	1482.918	2475.447	551.190	27	.000
171	1.44	.984	1482.918	2475.447	551.190	27	.000
172	1.44	.983	1482.900	2474.448	551.190	26	.000
173	1.44	.982	1484.133	2471.659	550.979	25	.000
174	1.44	.981	1485.264	2469.775	550.862	24	.000
175	1.44	.980	1485.510	2467.936	550.701	23	.000
176	1.44	.979	1487.497	2466.097	550.540	22	.000
177	1.44	.979	1492.014	2464.508	550.129	21	.000
178	1.44	.978	1491.176	2462.855	549.782	20	.000
179	1.44	.978	1492.774	2462.272	548.365	19	.000
180	1.44	.969	1481.610	2437.651	541.793	19	.000
181	1.40	.997	1515.908	2512.804	563.609	19	.000
182	1.40	.997	1523.859	2512.378	562.035	18	.000

Stage B5 produced Models 170 to 182 (Table 28). The following EVs were not significant at $p < \alpha = .20$ and were removed from the model in the following order:

January, *December*, *November*, *March*, *February*, *October*, *September*, *August*, *July*, *EV2·EV7*.

The removal of *Months* EVs was of particular interest. From a subject matter expertise perspective, the months were removed in the correct order leaving the months of *April*, *May*, and *June*. These three months would logically establish the environmental antecedent conditions for the development of WNv. Model 182 was selected as the best model with a D = .997 and AIC = 2512.378, and consisted of the following terms: *EV1ATM*, *EV2ARN*, *EV8USE*, *EV10ADD*, *EV1·EV2*, *EV1·EV6*, *EV1·EV8*, *EV1·EV10*, *EV2·EV9*, *EV6·EV7*, *EV6·EV10*, *EV7·EV8*, *EV7·EV9*, *EV8·EV10*, *EV9·EV10*, *Apr*, *May*, *June*.

Stage B6 DV_{COUNT} Analysis With All EVs, α = .05

In Stage B6, the significance level changed to α = .05 and the Tweedie MVP started at 1.40 (Table 29). It began with the Model 182 predictors except for the removal of *EV1·EV8* (p = .061) which was not significant at $p < \alpha$ =.05. The purpose of this stage and the change in significance level was to assess models with a more stringent requirement for EV significance, for comparison to previous models, and to ensure that the final model was not over-specified with EVs contributing very little to goodness-of-fit.

In this stage of modeling, only the 2FI, *EV1ATM·EV8USE*, was not significant at $p < \alpha$ =.05. and was removed from the model. The removal of this 2FI had a small but positive effect on the D score as the model progressed with Tweedie MVP adjustments at

1.39 and 1.395. Model 185 was the best model with the following scores: *D* = 1.000 and *AIC* = 2520.159. Model 185 consisted of the following EEVs: *EV1ATM*, *EV2ARN*, *EV8USE*, *EV10ADD*, *EV1ATM·EV2ARN*, *EV1ATM·EV6AWS*, *EV1ATM·EV10ADD*, *EV2ARN·EV7ELV*, *EV2ARN·EV8USE*, *EV2ARN·EV9POP*, *EV6AWS·EV7ELV*, *EV6AWS·EV10ADD*, *EV7ELV·EV9POP*, *EV8USE·EV10ADD*, *EV9POP·EV10ADD*, *April*, *May*, *June*.

Table 29

Stage B6 DV_{COUNT} Analysis With DV5, EVs, 2FIs, and Months, α = .05

Model	Tweedie MPV	*D*	χ^2_P	*AIC*	*LL2 Ratio* χ^2	*df*	*.sig*
183	1.400	.998	1533.391	2513.754	558.659	17	.000
184	1.390	1.003	1539.083	2526.590	563.058	17	.000
185	1.395	1.000	1536.228	2520.159	560.828	17	.000

Comparison of DV_{COUNT} Stage B5 and B6 Models

The Stage B6 analysis looked at the best interim models in Stages B5 and B6. Models 182 and 185 included all but one of the same predictors, (*EV1·EV8*), which was removed in Stage B4 Model 182. Based on the goodness-of-fit statistics, I selected the Stage B6 Model 185 as the final DV_{COUNT} model with *D* = 1.000 and *LL Ratio* χ^2 560.828 (17), $p < .001$ (Table 30). This model consisted of the following terms: *EV1ATM*, *EV2ARN*, *EV8USE*, *EV10ADD*, *EV1ATM·EV2ARN*, *EV1ATM·EV6AWS*, *EV1ATM·EV10ADD*, *EV2ARN·EV7ELV*, *EV2ARN·EV8USE*, *EV2ARN·EV9POP*,

EV6AWS·EV7ELV, *EV6AWS·EV10ADD*, *EV7ELV·EV9POP*, *EV8USE·EV10ADD*, *EV9POP·EV10ADD*, *April*, *May*, *June*. Model 185 produced the following measures: MAE = 0.895, RMSE = 2.199.

Based on these results, I rejected the RQ2 DV_{COUNT} null hypothesis at $p < \alpha = .05$. I concluded there is evidence that at least one coefficient in the final regression model is not equal to zero, and that the final model is a statistically significant predictor of *DV5* (the best DV for predicting the count of WNv incidents in the future).

Table 30

Comparison Of Stage B5 and B6 Models

Model	Tweedie MPV	*D*	χ^2_P	*AIC*	*LL2 Ratio* χ^2	*df*	*.sig*
182	1.400	.997	1523.859	2512.378	562.035	18	.000
185	1.395	1.000	1536.228	2520.159	560.828	17	.000

Stage C1 DV_{COUNT} Analysis Using $DV_{PRESENCE}$ Final Model EVs

In the Stage C excursion, I started with the $DV_{PRESENCE}$ final model predictors (Model 84): *EV1ATM*, *EV6AWS*, *EV9POP*, *EV1·EV5, EV1·EV6, EV1·EV10, EV2·EV6, EV2·EV10, EV5·EV6, EV5·EV8, EV5·EV9, EV5·EV10, EV6·EV8, EV6·EV9*, *EV6·EV10*, *EV8·EV9*, *EV9·EV10*, *May*, *June*, *August*, *September*, *October*, *November*. As in Stages A and B, I used a significance level of $p < \alpha = .20$ and compared the *D* values. Thirteen models were produced (Models 186-201) in the model building process. When all predictors were significant, I then adjusted the Tweedie MVP. The best model for Stage C was Model 190, D = 1.000 and *LL Ratio* $\chi^2(19)$ =468.833, p < .001 (Table 31).

Table 31

Stage C1 DV_{COUNT} With $DV_{PRESENCE}$ Final Model EVs

Model	*Tweedie MPV*	*D*	χ^2_P	*AIC*	*LL2 Ratio* χ^2	*df*	*.sig*
186	1.5	1.004	1559.274	2458.608	469.392	23	.000
187	1.5	1.003	1559.212	2456.608	469.391	22	.000
188	1.5	1.002	1561.148	2454.658	469.341	21	.000
189	1.5	1.000	1563.837	2452.930	469.070	20	.000
190	1.5	1.000	1562.524	2451.161	468.833	19	.000
191	1.5	.999	1564.192	2449.426	468.574	18	.000
192	1.5	.998	1564.231	2447.783	468.216	17	.000
193	1.5	.998	1563.929	2445.957	468.043	16	.000
194	1.5	.997	1574.935	2444.426	467.574	15	.000
195	1.5	.996	1575.527	2442.851	467.148	14	.000
196	1.5	.996	1580.569	2442.352	466.151	13	.000
197	1.5	.996	1584.252	2440.697	465.503	12	.000
198	1.5	.995	1583.150	2439.738	464.262	11	.000
199	1.4	1.041	1640.38	2565.422	494.991	11	.000
200	1.395	1.044	1643.344	2572.033	496.955	11	.000
201	1.45	1.017	1610.892	2501.028	477.882	11	.000

Stage C2 Comparison Stage B and C DV_{COUNT} Final Models

To complete *the DV_{COUNT}* analysis, I examined the Stage B and Stage C1 final models (Table 32). While the Stage C1 model (same predictors as the final BLR model) result was better than I had expected, considering the differences in data sets and statistical tools, I retained the Stage B Model 185 as the final model for several reasons. Model 185 was vetted through the entire GZLM analysis process, retaining a more inclusive set of EEVs and predictors. The *D* was virtually the same as Model 185, although the Stage C model *had a much lower LL Ratio χ^2* score. The Model 185 significance when compared to the null model was better than the Stage C model.

Table 32

Stage C2 Comparison of Stage B and C Final Models

Model	Tweedie MPV	*D*	χ^2_P	*AIC*	*LL2 Ratio χ^2*	*df*	*.sig*
185	1.395	1.000	1536.228	2520.159	560.828	17	.000
190	1.450	1.000	1562.524	2451.161	468.833	19	.000

I also compared the two models using MAE and RMSE. Model 185 produced the following error metrics: MAE = 0.895, RMSE = 2.199. The difference between the MAE and RMSE scores (1.304) is reflective of the variance of error in the sample (Eumetrain, 2017). Model 190 produced the following: MAE = 0.875, RMSE = 2.227. The difference between MAE and RMSE scores was 1.352. Both error metrics are considered to be better when the score is lower. In this instance, Model 185 outperformed Model 190.

Finally, from a subject matter perspective, I found the Stage B model to be more robust with regards the predictors of interest. The final DV_{COUNT} model remained Model 185 which had the following predictors: *EV1ATM*, *EV2ARN*, *EV8USE*, *EV10ADD*, *EV1ATM·EV2ARN*, *EV1ATM·EV6AWS*, *EV1ATM·EV8USE*, *EV1ATM·EV10ADD*, *EV2ARN·EV7ELV*, *EV2ARN·EV8USE*, *EV2ARN·EV9POP*, *EV6AWS·EV7ELV*, *EV6AWS·EV10ADD*, *EV7ELV·EV9POP*, *EV8USE·EV10ADD*, *EV9POP·EV10ADD*, *April*, *May*, *June*.

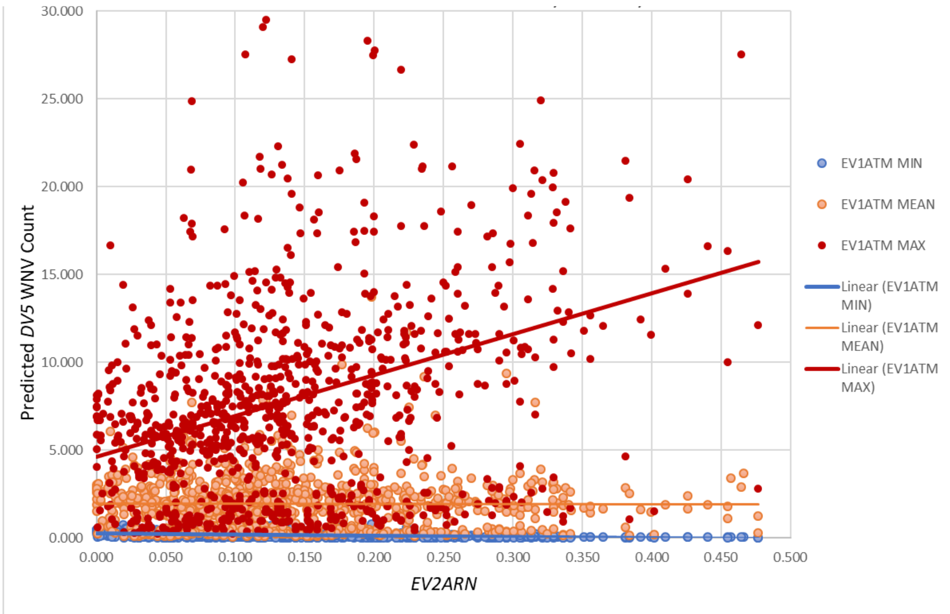
DV_{COUNT} Final Model 2FI Analysis

Upon selection of the final DV_{COUNT} Model 185, I conducted an analysis of the final model 2FIs of interest. Four EEVs were significant ($p < .05$) and contributed to the model goodness-of-fit ($D = 1.000$). Three EEVs (*EV6AWS*, *EV7ELV*, and *EV9POP*) were not in the final model but exerted moderating effects on the other EVs because they were part of 2FIs in the final model. In this section I address five of the 10 2FIs as they are directly related to the remaining four EEVs (Figures 45-52).

The relationship between *EV2ARN* and *DV5* was influenced by minimum, mean, and maximum average temperature (Figure 45). When the average temperature was at its minimum, the predicted number of WNv incidents remained low as average rainfall increased. When the average temperature was at its mean, the predicted number of WNv incidents rose but remained steady as the average rainfall increased. When the average temperature was at its maximum, the predicted number of WNv incidents rose sharply as rainfall increased.

Figure 45

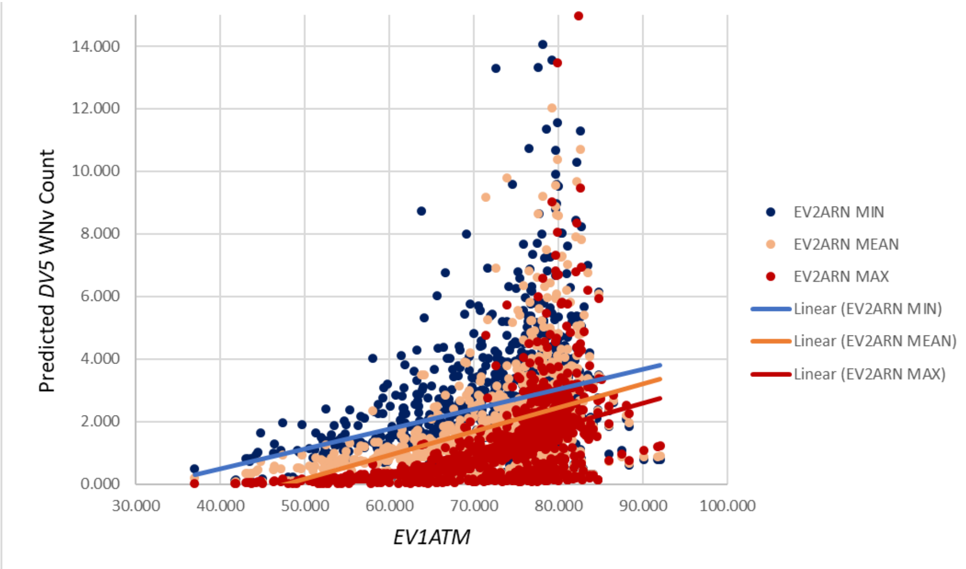
Predicted DV5 WNv Count as a Function of EV1ATM for EV2ARN



The relationship between *EV1ATM* and *DV5* was influenced by minimum, mean, and maximum average rainfall (Figure 46). When *EV2ARN* was at its minimum, mean, and maximum values, the predicted number of WNv incidents increased sharply as the average temperature exceeded 40^0F.

Figure 46

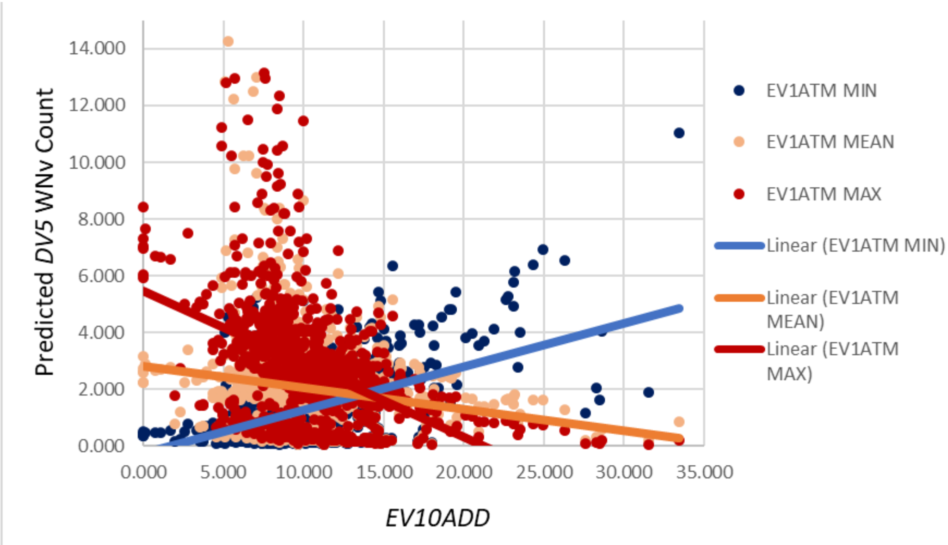
Predicted DV5 WNv Count as a Function of EV2ARN for EV1ATM



The relationship between *EV10ADD* and *DV5* was influenced by minimum, mean, and maximum average temperature (Figure 47). When the average temperature was its minimum, predicted WNv incidents rose sharply with an increase in dew point deficit. When the average temperature was at its mean, predicted WNv incidents decreased as the dew point deficit increased. When the average temperature was at its maximum value, predicted WNv incidents decreased sharply as the dew point deficit increased.

Figure 47

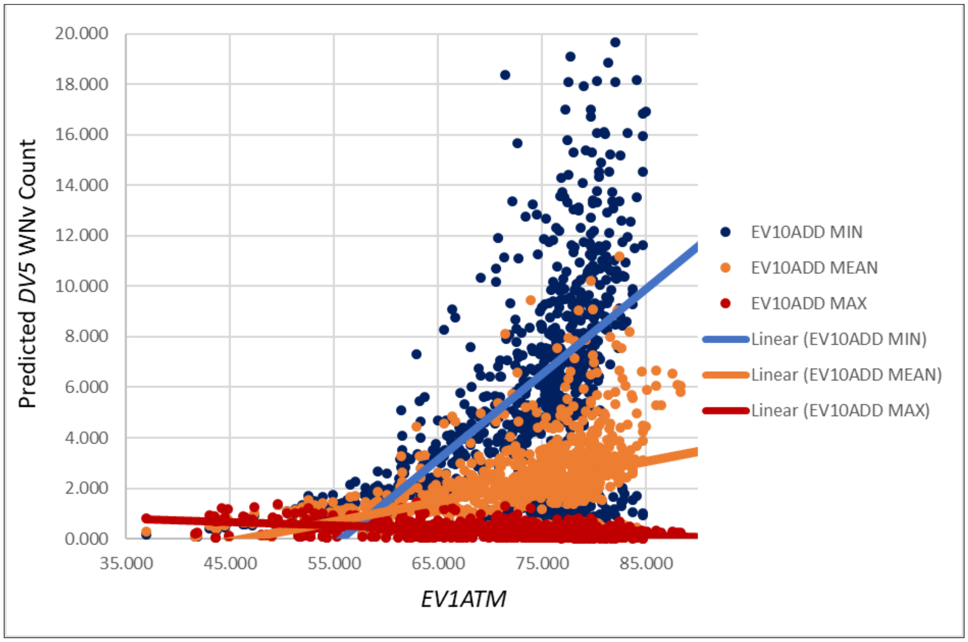
Predicted DV5 WNv Count as a Function of EV10ADD for EV1ATM



The relationship between *EV1ATM* and *DV5* was influenced by minimum, mean, and maximum average dew point deficit (Figure 48). When the average dew point deficit was at its minimum value, predicted WNv incidents rose sharply as the temperature exceeded 55^{0}F. When the average dew point deficit was at its mean, the count of predicted WNv incidents increased as the average temperature increased. When the average dew point deficit was at its maximum value, the count of predicted WNv incidents remained low (< 2.0) as temperature increased. This reflects the importance of moisture and relative humidity (lower dew point deficits) in the WNv environment.

Figure 48

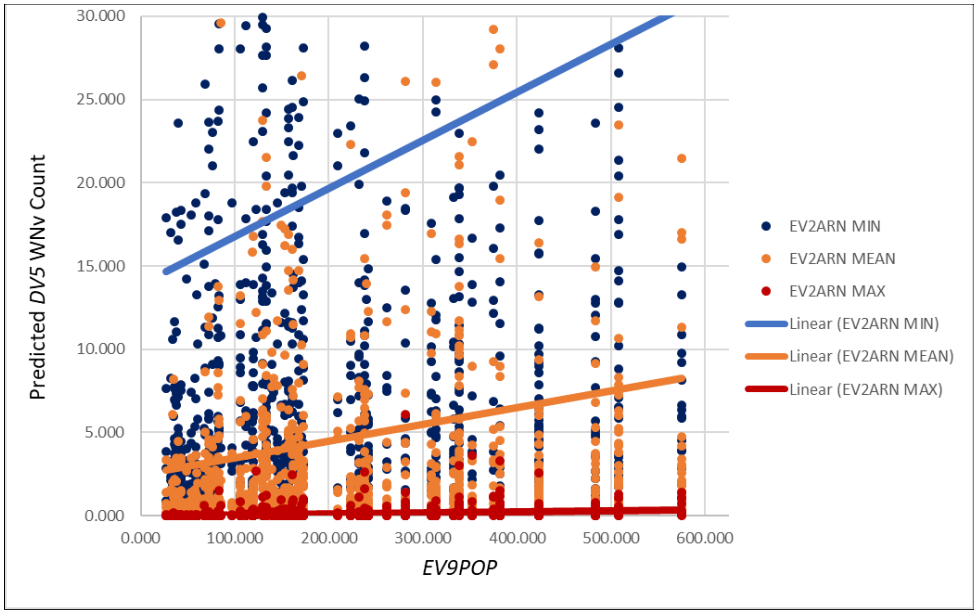
Predicted DV5 Count as a Function of EV1ATM for EV10ADD



The relationship between *EV9POP* and *DV5* was influenced by minimum, mean, and maximum average rainfall (Figure 49). When average rainfall was at its minimum value, the count of predicted WNv incidents increased sharply as the average population increased. When average rainfall was at its mean, the count of predicted WNv incidents increased slightly as the average population increased. When average rainfall was at its maximum value, the count of predicted WNv incidents remained low as the average population increased. In summary, light to moderate average rainfall appeared to provide a better environment for WNv, particularly in more largely populated areas.

Figure 49

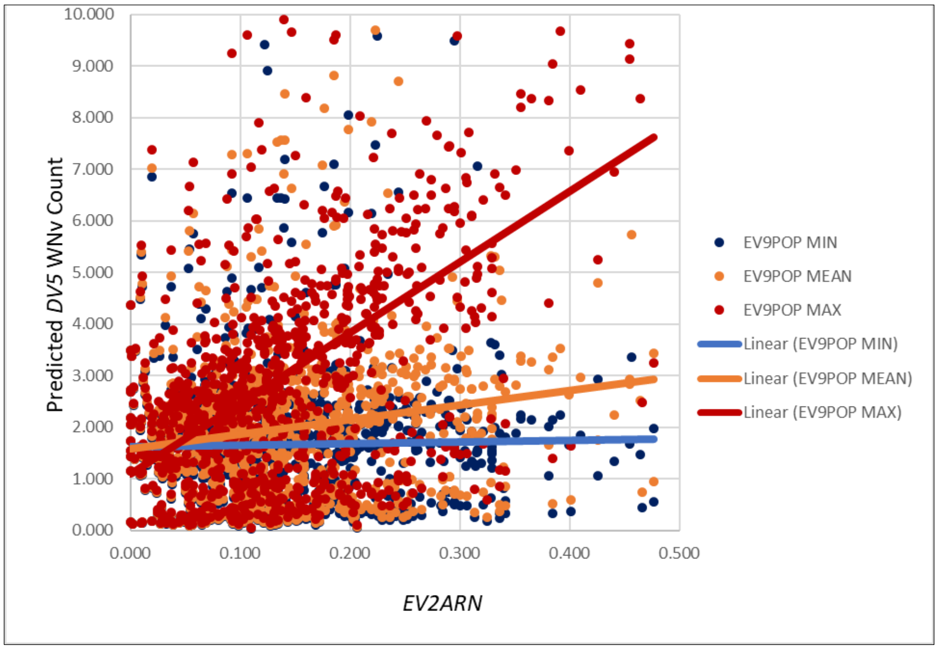
Predicted DV5 WNv Count as a Function of EV9POP for EV2ARN



The relationship between *EV2ARN* and *DV5* was influenced by minimum, mean, and maximum average population (Figure 50). When the average population was at its minimum value, the count of predicted WNv incidents remained stable as the average rainfall increased. When the average population was at its mean, the count of predicted WNv incidents rose as the average rainfall increased. When the average population was at its maximum value, the count of predicted WNv incidents increased sharply as the average rainfall increased. In summary, the relationship between average rainfall and predicted WNv incidents was positively affected by larger average populations.

Figure 50

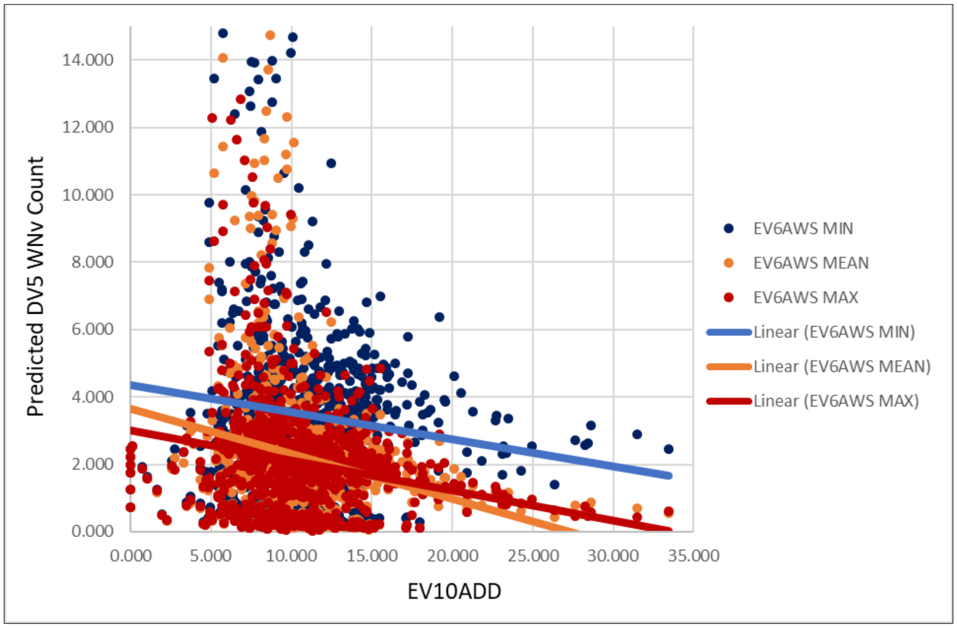
Predicted DV5 WNv Count as a Function of EV2ARN for EV9POP



The relationship between *EV10ADD* and *DV5* was influenced by minimum, mean, and maximum wind speed (Figure 51). When the average wind speed was at its minimum, mean, and maximum values, the count of predicted WNv incidents decreased as the average dew point deficit increases. In summary, average wind speed negatively moderated the relationship between predicted WNv count and average dew point deficit.

Figure 51

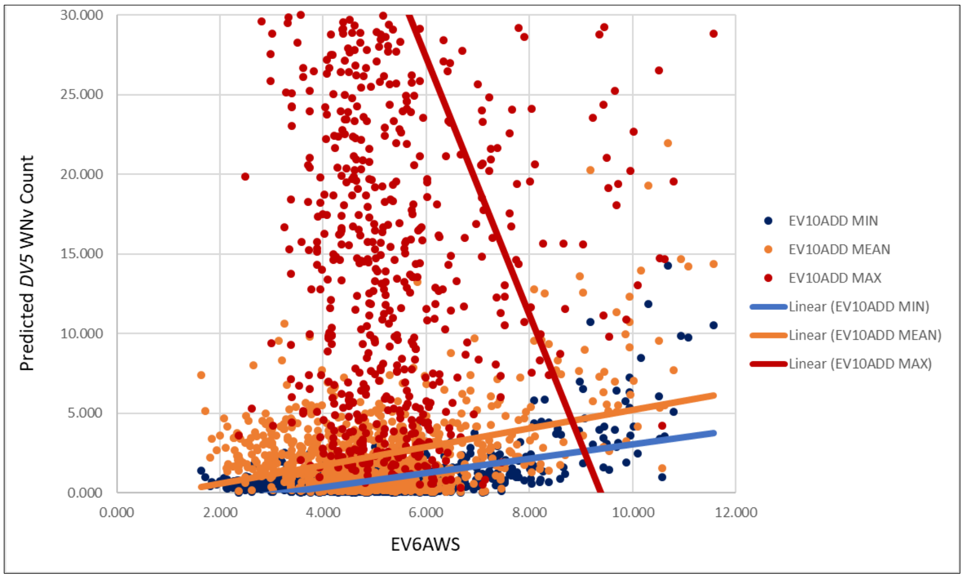
Predicted DV5 WNv Count as a Function of EV10ADD for EV6AWS



The relationship between *EV6AWS* and *DV5* was influenced by minimum, mean, and maximum average dew point deficit (Figure 52). When the average dew point deficit was at its minimum and mean value, the count of predicted WNv incidents rose slightly with higher average wind speed. When the average dew point deficit was at its maximum, the count of predicted WNv incidents decreased dramatically with higher average wind speeds > 6 mph. In summary, the relationship between predicted WNv incidents and average wind speed was moderated by average dew point deficit in two ways, 1) higher humidity positively affected the relationship between relationship between the outcome variable and average wind speed, 2) lower humidity had a negative effect on the relationship.

Figure 52

Predicted DV5 WNv Count as a Function of EV6AWS for EV10ADD



DV$_{COUNT}$ Final Model

Model 185 (*D* = 1.000, *AIC* =2520.159) was selected as the final model. I revisited the underlying assumptions for the GZLM regression and found all assumptions associated with the EEVs that had been previously checked to hold true. The final model consisted of these terms: *EV1ATM*, *EV2ARN*, *EV8USE*, *EV10ADD*, *EV1ATM·EV2ARN*, *EV1ATM·EV6AWS*, *EV1ATM·EV8USE*, *EV1ATM·EV10ADD*, *EV2ARN·EV7ELV*, *EV2ARN·EV8USE*, *EV2ARN·EV9POP*, *EV6AWS·EV7ELV*, *EV6AWS·EV10ADD*, *EV7ELV·EV9POP*, *EV8USE·EV10ADD*, *EV9POP·EV10ADD*, *April*, *May*, *June*.

The final model equation predicting the DV_{COUNT} was as follows (see Equation 12):

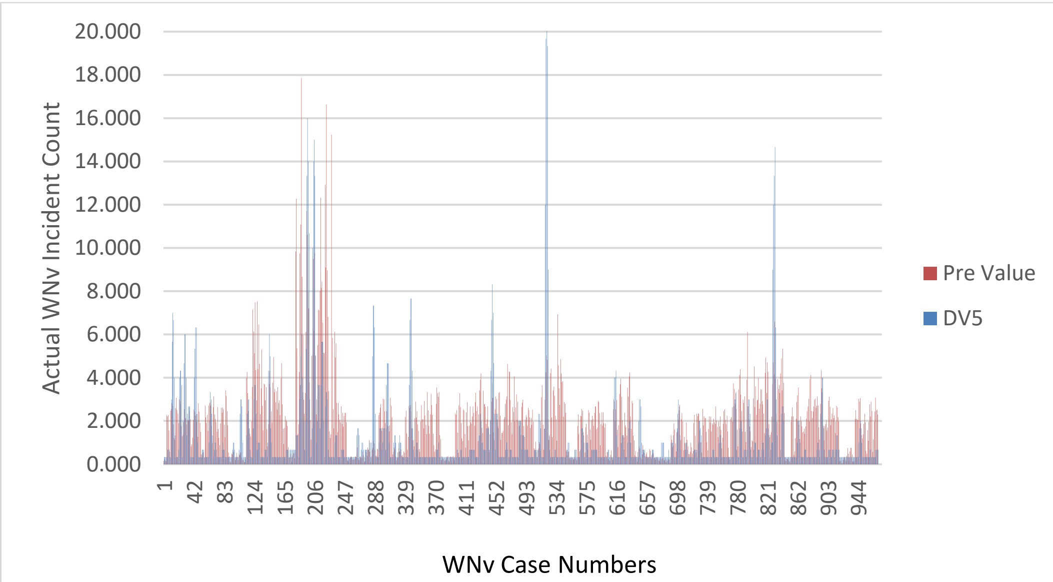
$$DV_{COUNT} = -4.368 + (0.051 \cdot EV1ATM) + (-12.111 \cdot EV2ARN) + (0.629 \cdot EV8USE) + (0.184 \cdot EV10ADD)\ (0.142 \cdot EV1ATM \cdot EV2ARN) + (-0.003 \cdot EV1 \cdot EV6) + (-0.0037 \cdot EV1 \cdot EV10) + (0.007 \cdot EV2 \cdot EV9) + (-0.0002 \cdot EV6 \cdot EV7) + (-0.013 \cdot EV6 \cdot EV10) + (0.000 \cdot EV7 \cdot EV8) + (-0.000002 \cdot EV7 \cdot EV9) + (-0.070 \cdot EV8 \cdot EV10) + (-0.003 \cdot EV9 \cdot EV10) + (-0.489 \cdot April) + (-0.404 \cdot May) + (-0.282 \cdot June) \quad (12)$$

DV_{COUNT} Results

Based on the final model, I examined all cases with positive WNv counts. This is depicted in Figure 53 where *DV5* actual WNv counts (blue) are plotted against the predicted WNv counts (red). Like the findings in *the* $DV_{PRESENCE}$ modeling, the use of EEVs alone resulted in predictions of the DV_{COUNT} across the 902 incidents when and where they occurred, but the ability to accurately predict specific counts > 4 was not achieved with the final model predictors. High counts of WNv incidents in the 2002, 2003, 2005, 2006, and 2012 were not predicted accurately by the model. While these high-count cases were unusual across the entirety of the study timeframe, they likely reflect the limitations of the current model accuracy using the chosen EEVs alone.

Figure 53

DV5 Actual WNv Counts Versus Predicted WNv Counts, All Counties



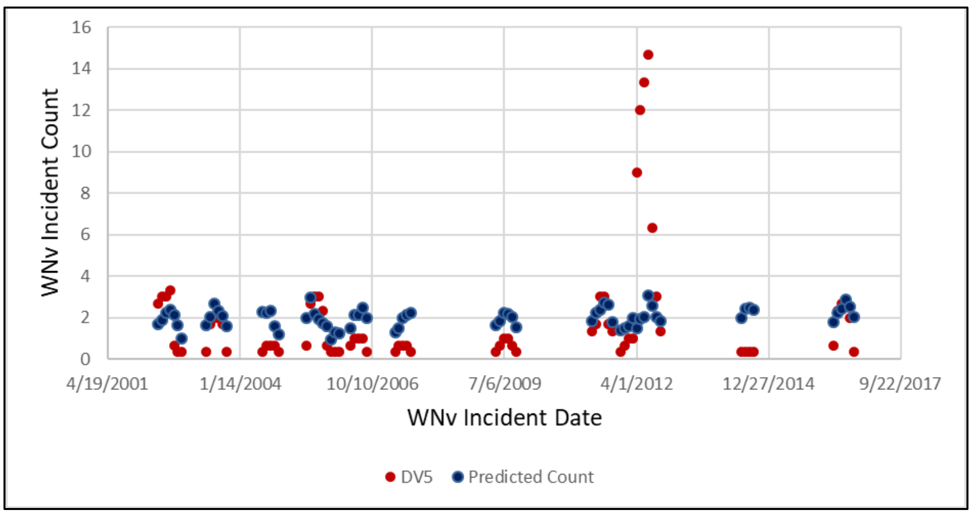
Note. *DV5* WNv actual counts (blue) versus predicted counts (red).

To visualize the results of the DV_{COUNT} data, I constructed 2003, 2012, and 2016 scatterplots for the counties of Aiken, Charleston, Dorchester, Greenville, Horry, and Richland (Appendix F). These counties reported the highest numbers of WNv incidents in SC over the years 2002-2016. Figure 54 is an example of that analysis for Richland County. In the years 2002-2016, 70 incidents were reported in the county. The Model 184 maximum prediction count was 3.06, falling below the actual count numbers. This was representative of the other counties which predicted maximum counts from 1.79 to 4.05. Although the DV_{COUNT} regression final model accurately predicts incident counts of WNv

< 4.0 over the total of WNv incidents, these were consistently lower than actual *DV5* incident counts. This was also consistent with the low *DV1* incident counts.

Figure 54

DV5 Actual WNv Incident Count Versus Predicted WNv Incident Count for Richland County 2002-2016



Note. *DV5* Actual vs Predicted Incident Count Richland County 2003, 2012, 2016. Incidents were reported in 2003, 2012, and 2016. Stage B Model 121, Tweedie 1.395.

DV_{COUNT} Additional Findings

The DV_{COUNT} final model included four operationally significant predictors: average temperature, average rainfall, land use, and average dew point deficit. In the GZLM final model, *EV7ELV* and *EV9POP* did not contribute to the model goodness-of-fit but did exert moderating effects on model EEVs as a predictor of probability of WNv occurrence. In this section, I summarize the more interesting interactions from the DV_{COUNT} final model.

Average Temperature (DV_{COUNT}). The GZLM regression final model average temperature coefficient ($\beta = 0.051$, $p < .05$) using the 90-Day moving average *DV5*, showed a slightly positive correlation on the presence of WNv with a .0002 odds ratio. The average temperature descriptive data showed the GZLM regression parameters captured the influence of temperature on the DV across a smaller temperature range (36.98–92.01) versus that of the BLR model (23.57–97.97). The difference in means 72.84 (GZLM regression) and 63.40 (BLR) coupled with different regression types and different time lags would have contributed to the different BLR and GZLM regression model outcomes.

Average Rainfall (DV_{COUNT}). Average rainfall was an historical EEV gleaned from literature, and it was an operationally significant predictor for DV_{COUNT} only. In the GZLM final estimates, the average rainfall ($\beta = -12.11$, $p < .05$) was derived from 970 cases. In the GZLM modeling, the mean for average rainfall was 0.14 inches ($SD = 0.09$) with minimum and maximum values of 0.00 and 0.48 respectively. In the years reporting 50 or more WNv incidents (2002, 2003, 2005, 2006, 2007, 2012), the mean was 0.14 inches ($SD = .08$) with minimum and maximum values of 0.00 and 0.46 inches respectively.

Although, average rainfall was not a significant predictor in the $DV_{PRESENCE}$ modeling, it did interact as a moderator of the relationships between the DV and two EVs in the final model. In the relationship between *DV5* and average wind speed, average rainfall was negatively correlated to the DV (β = -0.332, *EXP(B)* = 0.717). In the relationship between *DV5* and average dew point deficit, average rainfall was positively correlated to the DV (β = 0.182, *EXP(B)* = 1.200).

Land Use (DV_{COUNT}). Land use was coded as agricultural use (24%) and industrial use (76%). In the years reporting 50 or more WNv incidents (2002, 2003, 2005, 2006, 2007, 2012), 83% percent of all WNv incidents occurred. Within that percentage, industrial land use accounted for 71% of the WNv incidents while agricultural land use accounted for 12%. The GZLM final model produced a land use β = 0.63, *EXP (B)* = 1.86, showing a positive correlation between land use and DV_{COUNT}.

Land use was based on population per square mile in this study and as 71% of all WNv incidents occurred in land coded as industrial use, I found that areas of dense human activity promoted the occurrence of WNv incidents.

Dew Point Deficit (DV_{COUNT}). Dew point deficit was a new variable introduced for this study. It was defined as the "difference in degrees between the air temperature and the dew point" (American Meteorological Society, 2012). Dew point deficit values reflect levels of moisture saturation in the air. When the dew point deficit is 5 units of ºF different or lower, the air moisture content is nearly saturated. Higher dew point deficit values reflect dryer conditions.

The census mean for average dew point deficit was 11.73, with the minimum and maximum values at 0.00 and 34.09 respectively. In the years reporting 50 or more WNv incidents (2002, 2005, 2006, 2007, 2012), the dew point deficit mean (10.06) was lower than the census mean, as was the maximum value (26.33). This would mean that in the high-count years for WNv, the air moisture content would have been higher in high count years than the overall census means. This environmental predictor could support the genesis and maturation of WNv in those high producing years.

Months (*DV_{COUNT}*). In the GZLM final model, the months of April (β = .489, Exp(B) = 1.630), May (β = .404, Exp(B) = 1.497), and June (β = .282, Exp(B) = 1.326), were all positively correlated with the DV. The β values and the odds ratio decreased steadily from April to June. While May and June were operationally significant predictors in the BLR and GZLM analysis, the months of August, September, October and November were not included in the GZLM analysis. After reviewing the model building process, the months of August, September, and October were the last predictors to be removed from the model ($p < \alpha = .20$). These results are likely grounded in the differences in the data sets and time lagged DV.

Average Temperature and Average Rainfall Interaction (*DV_{COUNT}*). When the average temperature was at its minimum, the predicted number of WNv incidents remained low as average rainfall increased. When the average temperature was at its mean, the predicted number of WNv incidents rose but remained steady as the average rainfall increased. When the average temperature was at its maximum, the predicted number of WNv incidents rose sharply as rainfall increased.

When average rainfall was at its minimum, mean, or maximum values, the predicted number of WNv incidents increased sharply as the average temperature exceeded 40^{0}F.

Average Temperature and Average Dew Point Dewpoint Interaction (*DV_{COUNT}*). When the average temperature was its minimum, predicted WNv incidents rose sharply with an increase in dew point deficit (Figure 44). When the average temperature was at its mean, predicted WNv incidents decreased as the dew point deficit increased. When the average temperature was at its maximum value, predicted WNv incidents decreased sharply as the dew point deficit increased.

When the average dew point deficit was at its minimum value, predicted WNv incidents rose sharply as the temperature exceeded 55^{0}F (Figure 45). When the average dew point deficit was at its mean, the count of predicted WNv incidents increased as the average temperature increased. When the average dew point deficit was at its maximum value, the count of predicted WNv incidents remained low (< 2.0) as temperature increased. This reflects the importance of moisture and relative humidity (lower dew point deficits) in the WNv environment.

Urbanization and Average Rainfall Interaction *(DV_{COUNT}).* When average rainfall was at its minimum value, the count of predicted WNv incidents increased sharply as average population increased (Figure 46). When average rainfall was at its mean, the count of predicted WNv incidents increased slightly as the average population increased. When average rainfall was at its maximum value, the count of predicted WNv incidents remained low as the average population increased. In summary, light to moderate average rainfall appeared to provide a better environment for WNv, particularly in more largely populated areas.

Average Wind Speed and Average Dew Point Deficit Interaction *(DV_{COUNT})*. When the average wind speed was at its minimum, mean, and maximum values, the count of predicted WNv incidents decreased as the average dew point deficit increases. In summary, average wind speed negatively moderated the relationship between predicted WNv count and average dew point deficit.

When the average dew point deficit was at its minimum and mean value, the count of predicted WNv incidents rose slightly with higher average wind speeds (Figure 49). When the average dew point deficit was at its maximum, the count of predicted WNv incidents decreased dramatically with higher average wind speeds > 6 mph. In summary, the relationship between predicted WNv incidents and average wind speed was moderated by average dew point deficit in two ways, 1) higher humidity positively affected the relationship between relationship between the outcome variable and average wind speed, 2) lower humidity had a negative effect on the relationship.

Land Use and Dew Point Deficit (*DV_{COUNT}*). In the land use and average dew point deficit interaction, when average dew point deficit was at its maximum value (33.48), the predicted number of WNv occurrences increased slightly when land use was industrial. When average dew point deficit was at its minimum value (0.00), the predicted number of WNv occurrences increased moderately when land use was classified industrial. In summary, average dew point deficit had a moderating effect on the relationship between the DV and land use. The predicted number of WNv occurrences increased significantly when land use was classified as agricultural, and the dew point deficit was at its minimum value.

Summary

The purpose of this ex post facto quantitative research was to examine the use of EEV data in predicting outbreaks of West Nile Virus in South Carolina when robust EPS and EVS data are unavailable. The research questions were formulated to understand the effects of the EEVs on the DVs, examining the statistical utility of EEVs in predicting outbreaks of WNv in SC. In this research, a WNv incident was the positive identification of the virus in a locality in either a human, mosquito, bird, equine, or sentinel animal.

To address the research gap of accurate and timely predictive modeling of WNv, I posited 10 EEVs: (a) average 30-day temperature, (b) average 30-day rainfall, (c) average 30-day dew point, (d)average 30-day snow depth, (e) average 30-day barometric pressure, (f) average 30-day wind speed, (g) elevation, (h) land use, (i) urbanization and (j) average dew point deficit. During EDA, average dew point was removed, and average dew point deficit was added. Nine EEVs were subsequently used in the $DV_{PRESENCE}$ and DV_{COUNT} regression analyses based on a systems-level review of the WNv decision-space in the literature and the requirement for all predictors to be readily available from publicly accessible data sets.

Historical EEV data were collected from publicly accessible sites while WNv incident data were collected from SC DHEC. The complete data set consisted of 9,936 cases covering 1999 to 2016. Each case included monthly EEV and WNv incident data. Following review of the data, cases for the years 1999 to 2001 were excluded as no WNv incident reporting occurred in SC during those years SC. The final data set consisted of 8,280 cases (46 counties x 15 years x 12 months). From that population, there were 902 reported incidents of WNv associated with 360 events across all counties in the years 2002-2016.

The distribution of the data required a change to the planned analysis, moving the study away from MLR toward a combination of BLR (for the presence of WNv) and GZLM (for the count of WNv incidents). For the $DV_{PRESENCE}$ analysis, the predicted outcome was oriented to the likelihood of the presence of WNv in an environment. The DV_{COUNT} regression outcome was oriented to the count of WNv incidents in 46 SC counties. The $DV_{PRESENCE}$ analysis was performed on a modified data set of 970 cases which were extracted from the data set for the DV_{COUNT} analysis, focusing only on those cases for which there was a WNv event.

RQ1 was, in the absence of robust EPS and EVS data, which EEVs are predictors of incidents of WNv in SC in a current month? Based on the regression models developed for $DV_{PRESENCE}$ and DV_{COUNT}, I rejected the RQ1 null hypothesis and concluded there was evidence that at least one coefficient in the final regression models was not equal to zero and that each of the final models was a statistically significant predictor of $DV_{PRESENCE}$ and DV_{COUNT}.

RQ2 was, in the absence of robust EPS and EVS data, which EEVs accurately predict incidents of WNv in SC in the future? I rejected the RQ2 null hypotheses and concluded there was evidence that at least one coefficient in the final regression model was not equal to zero and that each of the final models was a statistically significant predictor of $DV_{PRESENCE}$ and DV_{COUNT}.

Among the candidate predictors, *EV1ATM* was the only operationally significant EEV for both $DV_{PRESENCE}$ and DV_{COUNT}. The following EEVs were not operationally significant predictors of WNv: average snow depth, average barometric pressure, and elevation.

In both the BLR and GZLM regression analyses, I found that EEVs alone can predict $DV_{PRESENCE}$. In the study findings, I considered predictors that were included in mathematical models that proved to be good predictors of the response, as *operationally significant predictors* of the response. Both regression methods highlighted the significance of certain EEVs in predicting the $DV_{PRESENCE}$ and DV_{COUNT}. The variables *EV1ATM*, *May*, and *June* were operationally significant in both types of regression and throughout the model-building process.

Within the GZLM analysis, I took an excursion which started with the BLR final model predictors and ended with a comparison of GZLM model outcomes. While the results of a predictive model of WNv incident counts using the same predictors as the final BLR model for predicting WNv events was better than I had expected, considering the differences in data sets and statistical tools, I considered the final model developed through the entire GZLM analysis process to be superior. It retained a more inclusive set of predictors, and the measurement accuracy was better than the model consisting of only the predictors from the BLR final model.

As a result of the study analysis, I found that specific combinations of operationally significant EEVs can predict WNv presence in the environment and provide count data with good precision but with lower accuracy. Both regression methods highlighted the significance of average temperature in predicting $DV_{PRESENCE}$ and DV_{COUNT}.

I interpret these findings in Chapter 5, expanding on the study limitations, the generalizability of the final models, as well as recommendations for future use of EEVs in predicting the presence of WNv. I also include recommendations for future research and the implications of the study for practice and social change.

Chapter 5: Discussion, Conclusions, and Recommendations

The purpose of this ex post facto quantitative research was to examine the use of EEV data in predicting outbreaks of WNv in SC when robust EPS and EVS data are unavailable. The study analysis revealed three key findings. First, in the absence of EVS using sentinel animals and human EPS data, EEV data captured from publicly available sources can provide useful predictions of the presence and count of WNv in SC prior to an outbreak. Based on some EHM-determined threshold, this level of awareness could provide an early indication of WNv in local environments to take steps to mitigate the severity of the outbreak.

Second, although the BLR and GZLM tools were used to examine different data sets and measures of WNv outbreaks in SC, some predictors were operationally significant in both types of regression. The data collection process and subsequent regression analysis reflected past findings with regards to the importance of locally/regionally derived data. The predictors average temperature and the months of May and June were included in predictive models throughout the respective model-building process.

Third, in SC, I found the complexity of the WNv lifecycle to be influenced by complex interdependencies. This finding corroborated the widely accepted viewpoint that the significance of individual predictors is heavily dependent on the other predictors in any regression model.

This chapter includes a summary of key findings and interpretation of findings. I describe limitations of the study and offer recommendations for future research. I conclude with implications for positive social change and recommendations for practice.

Interpretation of Findings

The key findings are addressed in three sections below. The interpretations and narrative of the findings flow from a broad general finding to more specific lessons learned. Together, they address the challenges of accurately predicting the presence of WNv in the state of SC.

A Novel Process for Predicting WNv

Predicting outbreaks in SC required a longitudinal approach to data collection. As the state first started to record WNv incidents in 2002, my data collection efforts spanned the 2002 to 2016 timeframe. I aligned count data received from SC DHEC in cases by county, year, and month. Each case contained variables and predictors with data associated to the month in question. This resulted in 8,280 cases to be used in the regression analysis. These cases were also organized by 0-, 30-, 60- and 90-day time-lags to account for any serial correlation.

The nine EEVs chosen for the study were either historical or new to previous WNv research. Average temperature, rainfall, snow depth, elevation, land use, and population EEVs were either directly related to past studies or were variations of the same (see Cotar et al., 2016; Ozdenerol et al., 2013; Rochlin et al., 2011; Young et al., 2013). Average wind speed, barometric pressure, and dew point deficit were new to the research topic. In both the BLR and GZLM, these EEVs, their 2FIs, and the Month EVs acted as the set of contextual variables for predicting the presence and count of WNv in SC.

The study design resulted in a two-step approach to analysis using BLR and GZLM. This was required because the empirical data set distribution was not as expected and based on the nature of the EID was most probably incomplete (meaning that it was highly probable that there were WNv incidents over time not reported or recorded). As a result, the data set was highly skewed, with most response values equal to zero. For the purposes of this study, the data could not support classic MLR or satisfy its assumptions. The nature of the data forced an alteration of the original plan into two directions with two modified data sets using BLR to predict the probability of an outbreak of any dimension and GZLM using the Tweedie distribution to predict the magnitude of an outbreak.

The new plan was to use BLR analysis ($DV_{PRESENCE}$) and modeling technique to give decision makers an indication, based on a predictive, mathematical model composed of a set of EEVs, that sometime in the future an outbreak was likely. Then, given a likely outbreak, the GZLM (DV_{COUNT}) analysis and modeling technique would provide an indication of the magnitude of an outbreak (in terms of a predicted number of incidents).

Because the empirical data provided by the state were flawed, the precision with which to predict an outbreak or its magnitude could not be any better than the accuracy of the data upon which the models were developed. However, the BLR predictive model was capable of accounting for 29% of the variance of the outcome variable. This provided a reasonably useful prediction of the probability of WNv presence 30 days in advance of an outbreak. Practically, if the predicted probability of an outbreak, based on values for the EEVs, exceeds some EHM applied threshold, decision makers would have the evidence to act on that probability of an outbreak and take steps to mitigate the severity.

Likewise, the GZLM regression model could provide a useful prediction of the number of incidents in any month, given the values for the EEVs. Realizing the GZLM regression model relied on empirical data for which incidents were most likely underreported, the magnitude of the underreporting could be quantified so that predictions of the number of incidents based on values for the EEVs might be adjusted to reflect a more accurate prediction of the expected number.

In addition, given the lack of accuracy in the GZLM regression model predictions, the GZLM prediction could be characterized by categories of severity—for example, by quartile: low intensity, moderate, high, and severe (without trying to pin down exact predicted numbers based on flawed empirical data); but providing a good indicator nevertheless would be useful to EHMs. It would be possible to quantify the error in each of the severity quartiles, perhaps demonstrating less error in the low numbers as the predictions are likely closer to the empirical data.

It is important to note that the analysis showed the best predictive models included a time lag, meaning models consisting of a set of EEVs were most effective at predicting an outbreak or the number of incidents 30 days in the future. This seems logical because the environmental conditions, described by the EEVs, are likely to spawn and spread the virus, not immediately, but over time.

Predicting the Presence and Counts of WNv in SC

In SC, I found that in the years 2002 to 2016, the WNv life cycle was most active between the months of July and November and was most prevalent in counties with 133,000 or more inhabitants. To support the mosquito-borne virus, the average temperature needed to range between 40.0^{0}F and 95.1^{0}F (4.4^{0}C to 35.0^{0}C). WNv incidents were greater when average rainfall remained at or close to 0.14 inches per month. Average barometric pressures above 1,000.00 millibars increased WNv counts when temperature was above 45.0^{0}F (7.2^{0}C).

I also found that positive WNv events in SC occurred across a range of wind speeds from 1.64 to 9.95 mph. A quarter or more of the incidents were associated with wind speeds greater than 5 mph and by the influence of barometric pressure above 1,000.00 millibars at increased wind speeds. Wind speeds between 4.39 and 5.29 mph affected WNv incident counts like past research findings (see Cheong et al., 2013). Cheong et al. (2013) found that rising windspeeds increased WNv mosquito dispersion while decreasing the number of blood meals. Land classified as industrial use was also positively correlated to WNv outcomes. Finally, in the years in which the virus was most active in the state, the average dew point deficit pointed to more humid conditions. All of these conditions acted as antecedents to the presence of WNv.

In combination, the study EEVs defined a contextual decision-space for EHMs when considering the complexities of predicting the presence of WNv. Many of the findings directly correlate with previous studies. Cotar et al. (2016), Ozdenerol et al. (2013), and Rochlin et al. (2011) found that temperature was an operationally significant predictor of WNv. In SC, average temperature was a significant predictor of WNv in both the BLR and GZLM analysis. In years reporting 50 or more WNv incidents (2002, 2003, 2005, 2006, 2007, 2012), the WNV ecosystem in SC thrived within an average temperature of 40.0^{0}F to 95.1^{0}F.

This finding is like a 2020 study examining temperature-dependence on mosquito-borne diseases (see Shocket et al., 2020). The researchers found mosquitoes, such as the SC native Culex Quinquefasciatus, were biologically temperature-dependent within a range of 50.0^{0}F to 78.0^{0}F. When temperatures were within this range, vector-borne diseases were supported by the environment. Consequently, in future modeling of WNv in SC, EHMs could refine their preventive activities to average temperatures within the interval 40.0^{0}F to 95.1^{0}F. This range of temperature would also affect the predictors' relationship with the outcome variable depending on the temperature means of the months in question. This would explain why temperature within the BLR final model was negatively correlated with the DV, while the GZLM model had a positive correlation.

The influence of average temperature on WNv was moderated by average barometric pressure. When barometric pressure is high, WNv incidents are highly correlated with temperature, but at low temperatures, the relationship between WNv and temperature is less pronounced. When average barometric pressure was decreasing and temperatures rose, the probability of WNv occurrence decreased. This barometric pressure interaction had a moderate to significant effect on the DV for temperature ranges > 45.0^0F. This interaction was new to WNv research.

The mean average rainfall in the most productive WNv years was consistent with the census mean and may suggest that rainfall is more important to the genesis and maturation of mosquito larvae development than it is to the virus transmission. In fact, once the virus is mature and resident in the female mosquito, increased amounts of rainfall may hinder the spread of WNv vector. Similar findings were highlighted by Paull et al. (2016) where they posited increased rainfall could either increase or decrease mosquito abundance.

An interval of average wind speeds between 4.38 and 5.29 existed where WNv cases were recorded as either positive or negative. These results appear to be supported by a 2013 study where both decreases and increases in wind speed increased the relative risk of dengue fever cases (see Cheong et al., 2013). In this 2013 study, wind speed was examined over 0, 30, 60, and 90-day time lags. Depending on the time lag, the relative risk of dengue fever rose and fell with decreases and increases in wind speed. The researchers found that lower wind speeds supported oviposition (expulsion of the egg into the environment) and contact with humans raising the risk of dengue fever (Cheong et al., 2013). However, higher wind speeds also supported mosquito vector dispersion (potentially to more populated areas) and oviposition (2013). In both instances, the relative risk of dengue fever rose within certain wind speed ranges.

There was also an interaction between barometric pressure and wind speed. When wind speed was high and barometric pressure was at its minimum, the probability of an WNv outbreak was highest. When the winds were average or low, barometric pressure was not as influential on WNv event occurrence.

Land use and population were connected in this study. Exploratory data analysis revealed that six SC counties accounted for 56% (504/902) of all reported WNv incidents. Each of these counties met the coding classification of industrial land use, a reflection of population per square mile in this study. Thus, in addition to the weather related EEVs used in this study, areas with greater populations (>133,000 per county) were more susceptible to the WNv incidents. This was true also for agricultural areas with denser populations.

In the interaction between average barometric pressure and dew point deficit, the probability of a WNv occurrence was highest when average barometric pressure was high and the values for average dew point deficit were low. When the barometric pressure was moderate or low, dew point deficit was not a significant influence on probability. This finding was new to WNv research.

The increasing odds ratios for the months June through October tracked with the EDA that found 96% of WNv occurring in these months. The month of May and November correlated negatively to the DV. May had only seven WNv events over the study timeframe from 2002 to 2016 while November represented the end of the WNv cycle for the years of the study.

The GZLM approach to determine the magnitude of that presence through the prediction of WNv incident counts was useful but mixed with the predictors used in this study. Although the final model was precise with regards to the occurrence of WNv, it could not accurately predict counts greater than four. This was likely due to incomplete empirical data, which lacked precision, partly as a result of a lack of accurate reporting and accounting.

Petersen et al. (2013) found that within the United States, biological footprints can vary within the diverse ecological conditions in individual states. Reiner et al. (2013) found inherent problems with current mathematical models of mosquito-borne pathogens like WNv. As seen in my study, the ability to adequately define biological and ecological factors remained difficult. While nine EEVs were used as contextual variables in examining the abiotic conditions necessary for WNv, the BLR final model $R^2_{\text{Nagelkerke}}$ statistic of 29% reflected the complexities found in previous studies and highlights the limitation of defining the correct EEVs in a complex environment such as WNv.

WNv Interdependencies and the Complex Nature of Predictor Selection

The complex nature of the WNv introduces potential pathogenic, ecological, and anthropological predictors that need to be captured within a regional or local context. In other words, the influence of any individual EEV is moderated by many local/regional interactions between and among multiple EEVs. This is particularly true when the researcher tries to capture the context surrounding a complex problem such as WNv. Some of these interactions are quantified, but there are many more subtle and complex real-world interactions that influence (moderate) the effect any one variable has on the response. That was the case in this study.

Limitations of the Study

My findings showed that predictive modeling of WNv in the state of SC requires accurate and persistent reporting of WNv incidents (of all categories) and environmental data to provide the robust truth data to train a contextually driven predictive model. I used WNv incident truth data captured by SC DHEC, but the low counts of reported WNv events likely reveals a combination of reporting constraints such as asymptomatic presentation by the virus, as well as a lack of consistent and persistent surveillance activities that are equally relevant to other states and geographic areas. This is reflective of the findings of previous studies such as Rochlin et al. (2011).

Although the climate and topography of the region is like other temperate areas, the ability to generalize the study was limited by the use of regionally focused publicly accessible data (Liu et al., 2009; Ozdenerol et al., 2013). The scope of my study was limited to the state of SC, and the data collected were primarily historical, raising issues of external validity about the numerous means of collection as well as accuracy. Hence, any attempt to infer results beyond the scope of this study should be done with caution.

The analysis of the best time lagged *DV* for both BLR and GZLM regression was limited. The BLR DV was selected only by comparing the $R^2_{\text{Nagelkerke}}$ and *LL Ratio* χ^2

scores for models using the post-EDA EEVs; no other predictors were included in the initial analysis. The GZLM regression DV was selected by comparing *D* and *LL Ratio* χ^2 scores adjusted for different Tweedie distributions.

Generalizability and Trustworthiness

The analytical process used in this longitudinal study could potentially generalize to another state or geographic area that enjoys a temperate climate. The use of a decision-theory construct which expands decision data to include contextual elements is also generalizable to topics outside of emerging infectious diseases. The EEV data were collected each month over a 15-year time period in all 46 SC counties. All data were accessed through public means using standard, acceptable measures. I found that EEVs such as average temperature, average rainfall, average wind speed, land use, population and dew point deficit, can be significant predictors of an otherwise complex biological phenomenon.

The most vexing problem with WNv is that many of the cases are asymptomatic. Using a very inclusive incident definition (human, equine, mosquito pools, birds, other mammals), over a 15-year period, resulted in only 902 reported events. This highlights the importance of WNv incident reporting and analysis for purposes of prediction. It also raises the question of trustworthiness due to accuracy in the reported data.

The CDC's and other studies have highlighted the importance of local environmental conditions to the biology of the WNv. This means that EEVs used in the southeast United States may not be the same or have the same relevance for the virus in other geographic regions. I also found that the reporting of environmental factors is not without errors and missing data. In many instances, I found environmental data to be missing for a particular county during a specific time period. This required the use of secondary sources to confirm the missing data.

Validity and Reliability

As planned, I used publicly accessible sources for EEV capture and WNv incident truth data from the SC DHEC. While the metrics for collecting the data are standardized, the collection means are not. In some county locations, data were not collected for certain time periods, and I was required to use a secondary source. However, I found that contextually rich environmental data are readily available to the public through several publicly accessible government and private websites.

Recommendations

Results of this study suggest that a model composed of contextually derived EEVs can provide operationally significant predictors of the presence of WNv in SC. The final models, generated through BLR and GZLM regression, could predict when WNv was likely to be present but the ability to predict the count of WNv incidents was limited to counts of three and below. Future researchers should consider the following:

- Expand EEVs to include additional predictor variables that are contextually related to WNv. For example, data associated with storm drains would be interesting to explore in an EEV-only study.

- Given the lack of accuracy in the GZLM regression model predictions, another approach to interpreting the GZLM predictions could have been to categorize the results by levels of severity – for example by quartiles: low intensity, moderate, high, and severe.
- Explore the benefits of a network centric Service Oriented Architecture (SOA) that would automatically collect EEVs based on predictive modeling requirements.
- Conduct a more robust analysis of time lagged DV selection
- Explore the use of artificial intelligence and machine learning to predict WNv incidents.

Implications

Positive Social Change Implications

The social change resulting from this study provides EHMs with another approach to EID predictive modeling, particularly when EVS and EPS data are latent or not available. Using a tailored, contextual approach to decision making would allow EHMs to supplement current modelling capabilities using publicly available data in a near real-time fashion. The ability to understand when WNv outbreaks may occur would allow preventive actions to occur in a more pro-active and resource friendly manner.

The use of contextually derived EEV data may also generalize to other EID predictive modeling efforts such as the Zika virus. Although, the epidemiological cycle associated with WNv incubation, transport, and transmission reflects stochastic ecological and environmental conditions that differ from region to region, the framework developed in this study could be adapted to include those EEVs that were relevant to a specific locality.

Methodological, Theoretical, and Empirical Implication

In this research, I examined the relationship between DT, decision-making context (CYNEFIN), and systems theory to understand the impact of exogenous data on complex decision making. The tenets of these theories were combined to provide a theoretical framework that challenges traditional linear-causal approaches to decision making, to expand the manager's perception of the decision-space, and to provide greater fidelity to the design and choice phases of the decision-making process.

This theoretical approach was accomplished using an emphasis on data intelligence and specifically the use of systems-level exogenous data to provide context and reduce uncertainty in the decision-making process. The expansion of contextually related variables for the predictive modeling of WNv can provide EHMs and public health officials with another method to implement preventive measures in a timely fashion.

Implications for Practice

Managers of all professions are required to make decisions daily. These decisions are made within decision-spaces that range from simple to chaotic (Snowden & Boone, 2007). The ability for a manager to make decisions when dealing with complicated and complex decision-spaces is dependent on their ontological understanding, breadth of intelligence, and analytic support.

This research used a contextually based theoretical foundation that leveraged the dynamic presence of publicly accessible data in forming intelligence collection strategies for decision making for WNv. The theoretical foundation approached the decision-space in a way that allows EHM practitioners to make decisions when empirical data are not available. In this study, I found that contextually derived EEV data could account for some variance in a WNv outcome. The use of DT, the CYNEFIN construct, and system-thinking can provide EHMs with different approach to predicting WNv in a locality.

Conclusions

The outcomes of this analysis were revealing and important, based on a rigorous application of sound mathematical techniques. The original objective was achieved. It is possible to predict an outbreak of WNv based on a set of EEVs. Moreover, it is possible to predict with some confidence both the likelihood of an outbreak (of any magnitude) and the severity of an outbreak, based on external, environmental conditions. This may also be an analytical process that can be used in similar situations, to predict other real-world phenomena based on external, environmental factors.

This analysis also corroborated the widely accepted viewpoint that the significance of individual predictors is heavily dependent on the other predictors in any regression model. In other words, the influence of any individual EEV is moderated by many complex interactions between and among multiple EEVs. Some of these interactions are quantified, but there are many more subtle and complex real-world interactions that influence (moderate) the effect that any one variable has on the response.

Historically, the timely and accurate prediction of WNv in a locality requires robust EPS and EVS programs. These programs produce surveillance data that populate predictive models allowing EHMs to make informed decisions on preventive measures (Manore et al., 2014). With EIDs like WNv and more recently Zika, simple, reliable predictive tools are required to ensure public health measures can be taken before an outbreak occurs. To address this scholarly gap, I examined the accuracy and timeliness of contextually based exogenous explanatory data in predicting outbreaks of WNv in South Carolina. In doing so, I also examined the importance of context and system-level thinking in decision making.

I had three key findings in this study. First, in the absence of EVS and EPS data, EEVs captured from publicly available sources can provide indications of the presence of WNv in SC 30 days prior to an incident. The final BLR model which consisted of three predictor variables (*EV1ATM*, *EV6AWS*, *EV9POP*) and the monthly factors (*May*, *June*, *August*, *September*, *October*, *November*) explained 29% of the variation in $DV_{PRESENCE}$. This level of awareness could provide EHMs with an early indication of WNv in the local environment. It is important to note that the analysis showed the best predictive models included a time lag; meaning, models consisting of a set of EEVs were most effective at predicting an outbreak or the number of incidents 30 days into the future. This seems logical since the environmental conditions, described by the EEVs, are likely to spawn and spread the virus, not immediately, but sometime in the future.

The GZLM approach to determine the magnitude of that presence through the prediction of WNv incident counts was mixed. Although the best DV_{COUNT} model was precise with regards to the occurrence of WNv, it could not accurately predict counts greater than 4.0 with any consistency. This was likely due to incomplete empirical data, which could have only provided the precision and accuracy resident in the study data.

Second, while the BLR and GZLM tools were used to examine different RQs and used different data sets, it should be noted that the variable average temperature remained an operationally significant predictor ($p < \alpha = .05$) in both regression types and throughout the respective model-building process. This was consistent with past study findings associated with using temperature as a WNv predictor.

Third, is that the study findings also corroborated the widely accepted viewpoint that the significance of individual predictors is heavily dependent on the other predictors in any regression model. In other words, the influence of any individual EEV is moderated by many complex interactions between and among multiple EEVs. This is particularly true when the researcher tries to capture the context surrounding a complex problem such as WNv. Some of these interactions are quantified, but there are many more subtle and complex real-world interactions that influence (moderate) the effect that any one variable has on the response. Therefore, in the study findings, I considered predictors that are included in a mathematical model (i.e., a set of predictors) and prove to be good predictors of the response, as operationally significant predictors of the response.

Management tools that predict trends and services need to adapt to the complexity of today's information environment and to the systems-level data it produces. A systems-level, context driven approach to the complex decision-space of WNv offers an answer to these data challenges. When required, this practical approach can allow EHMs to place a decision within a broader systems-level context, using exogenous data to enrich and define a less ordered decision-space. This is particularly relevant for decision makers and managers who work within the complex field of EIDs.

The findings of the study may prove useful to environmental health decision makers in understanding the role of EEVs in the dynamic temporal and spatial interdependencies of the pathogenic, ecological, and anthropological components of the virus (Pirofski & Casadevall, 2012). These interdependencies currently present a complex decision-space for EHM decision makers.

The findings may also prove valuable for decision makers across several professions. The complex decision-space presented by WNv, and the contextual framework used to address the predictability of the virus in this study, is directly transferable to the complicated, complex, and chaotic decision spaces identified by Snowden and Boone (2007). The use of contextual elements to frame and clarify a complex decision-space in new ways contributes to DT.

References

Ahmadnejad, F., Otarod, V., Fathnia, A., Ahmadabadi, A., Fallah, M. H., Zavareh, A., Miandehi, N., Durand, B., & Sabatier, P. (2016). Impact of climate and environmental factors on West Nile Virus circulation in Iran. *Journal of Arthropod-Borne Diseases*, *10*(3), 315-327. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4906738/

Abiotic. (n.d.). In *Merriam-Webster's online dictionary*. Retrieved September 24, 2021, from http://www.merriam-webster.com/dictionary/abiotic

Allison, P. (2013, March 5). *Why I don't trust the Hosmer-Lemeshow test for logistic regression*. Statistical Horizons. http://statisticalhorizons.com/hosmer-lemeshow

American Meteorological Society. (2012, January 26). *Dewpoint depression*. https://glossary.ametsoc.org/wiki/Dewpoint_depression

Arbovirus. (n.d.). In *Merriam-Webster's online dictionary*. Retrieved September 24, 2021, from http://www.merriam-webster.com/dictionary/arbovirus

Arthropod. (n.d.) In *Merriam-Webster online dictionary*. Retrieved September 24, 2021, from http://www.merriam-webster.com/dictionary/arthropod

Artikis, A., Etzion, O., Feldman, Z., & Fournier, F. (2012). Event processing under uncertainty. *Proceedings of the 6th Annual ACM International Conference on Distributed Event-Based Systems, USA*, 32-43. https://doi.org/10.1145/2335484.2335488

Austin, S. K., & Dowd, K. A. (2014). B cell response and mechanisms of antibody protection to West Nile Virus. *Viruses*, *6*(3), 1015-1035. https://doi.org/10.3390/v6031015

Azvine, B., Cui, Z., Nauck, D. D., & Majeed, B. (2006). Real time business intelligence for the adaptive enterprise. *Proceedings of the 8th IEEE International Conference on E-Commerce Technology and the 3rd IEEE International Conference on Enterprise Computing*, *E-Commerce, and E-Services*, *USA*, 29. https://doi.org/10.1109/CEC-EEE.2006.73

Babu, M., & Sastry, S. (2014). Big data and predictive analytics in ERP systems for automating decision making process. *Proceedings of the 2014 5th IEEE International Conference on Software Engineering and Service Science (ICSECC), China*, 259-262. https://doi.org/ 10.1109/ICSESS.2014.6933558

Bar-Yam, Y. (1997). *Dynamics of complex systems*. Westview Press.

Biotic. (n.d.). In *Merriam-Webster's online dictionary*. Retrieved September 24, 2021, from http://www.merriam-webster.com/dictionary/biotic

Boisot, M. H. (1999). *Knowledge assets*: *Securing competitive advantage in the information economy*. Oxford University Press.

Bower, V. (2001). Vector-borne diseases and global warming: Are both on the upward swing? *EMBO Reports, 21*(9), 755-756. https://doi.org/10.1093/embo-reports/kve193

Box, G.E P., & Tidwell, P. W. (1962). Transformation of the independent variables. *Technometrics*, *4*(4), 531-550. https://doi.org/10.1080/00401706.1962.10490038

Buchanan, L., & O'Connell, A. (2006, January). A brief history of decision making. *Harvard Business Review*. https://hbr.org/2006/01/a-brief-history-of-decision-making

Bunge, M. (1973). *Method, model, and matter*. D. Reiding Publishing.

Buro, K. (2020). *Stat 371*. https://academic.macewan.ca/burok/Stat371/notes/glm.pdf

Campbell, G. L., Marfin, A. A., Lanciotti, R. S., & Gubler, D. J. (2002, Sep). West Nile Virus. *The Lancet*, *2*(9), 519-529. https://doi.org/10.1016/S1473-3099(02)00368-7

Casti, J. L. (2017). Complexity. *In Encyclopedia Britannica*. https://www.britannica.com/science/complexity-scientific-theory

Centers for Disease Control and Prevention. (2013). *West Nile Virus in the United States*: *Guidelines for surveillance, prevention, and control*, *4th*, 1-69. U.S. Department of Health and Human Services, Public Health Service. https://www.cdc.gov/westnile/resourcepages/pubs.html

Centers for Disease Control and Prevention. (2018). *West Nile Virus disease cases reported by CDC by state, 1999-2016*. U.S. Department of Health and Human Services, Public Health Service. https://www.cdc.gov/westnile/statsmaps/cumMapsData.html#one

Cheong, Y. L., Burkart, K., Leitão, P. J., & Lakes, T. (2013). Assessing weather effects on dengue disease in Malaysia. *International Journal of Environmental Research and Public Health*, *10*(12), 6319–6334. https://doi.org/10.3390/ijerph10126319

Chevalier, V., Tran, A., & Durand, B. (2014). Predictive modeling of West Nile Virus transmission risk in the Mediterranean Basin: How far from landing? *International Journal of Environmental Research and Public Health*, *11*(1), 67-90. https://doi.org/ 10.3390/ijerph1101000067

Collett, D. (2003) *Modeling binary data* (2nd Ed.). CRC Press.

Cotar, A.I., Falcuta, E., Prioteasa, L.F., Dinu, S., Ceianu, C. S., & Paz, S. (2016). Transmission dynamics of the West Nile virus in mosquito vector populations under the influence of weather factors in the Danube Delta, Romania. *EcoHealth*, *13*(4), 796-807. https://doi.org/10.1007/s10393-016-1176-y

Decision-support tools. (n.d.). *Medical Dictionary for the Health Professions and Nursing*. Retrieved September 23, 2021, from https://medical-dictionary.thefreedictionary.com/decision-support+tools

DeLong, J. (2010, September 6). Recession takes a bite out of mosquito programs. *USA Today*. http://usatoday30.usatoday.com/news/nation/environment/2010-06-08-mosquitoes·N.htm

Devinney, T., & Siegel, D. (2012). Perspectives on the art and science of management scholarship. *Academy of Management Perspectives*, *26*(1), 6-11. https://www.jstor.org/stable/23208281

Dietrich, C. (2010). Decision making: Factors that influence decision making, heuristics used, and decision outcomes. *Inquiries Journal/Student Pulse, 2*(02). http://www.inquiriesjournal.com/a?=180

Dillon, S. M. (1998, August 30–1 September). *Descriptive decision-making*: *comparing theory with practice*. Paper presented at the 33 Annual Conference of the New Zealand Operational Research Society, Auckland, New Zealand. http://www.orsnz.org.nz/conf33/papers/p61.pdf

Divekar, A. A., Bangal, S., & Sumangala, D. (2012). The study of prescriptive and descriptive models of decision making. *International Journal of Advanced Research in Artificial Intelligence, 1*(1), 71-74. https://doi.org/10.14569/IJARAI.2012.010112

Elith, J., Phillips, S. J., Hastie, T., Dudlik, M., Chee, Y. E., & Yates, C. J. (2011). A statistical explanation of MaxEnt for ecologists. *Diversity and Distributions, 17,* 43-57. https://doi.org/10.1111/j.1472-4642.2010.00725.x

Emerging Infectious Disease. (n.d.). Medical definition of emerging infectious disease. Retrieved October 15, 2015, from https://www.medicinenet.com/script/main/art.asp?articlekey=22801

Endogenous [Def 2a]. (n.d.). In *Merriam-Webster online dictionary*. Retrieved February 10, 2019, from http://www.merriam-webster.com/dictionary/endogenous

Eumetrain. (2017, February 28). *Mean Absolute Error (MAE) and Root Mean Square Error RMSE*.

http://www.eumetrain.org/data/4/451/english/msg/ver_cont_var/uos3/uos3_ko1.htm

Exogenous [Def 2b]. (n.d.). In *Merriam-Webster online dictionary*. Retrieved February 10. 2019, from http://www.merriam-webster.com/dictionary/exogenous

Field, A. (2013). *Discovering statistics using IBM SPSS Statistics*. SAGE Publications.

Frankfort-Nachmias, C., & Nachmias, D. (2008). *Research Methods in the Social Sciences* (7 ed.). Worth Publishers.

French, S., & Rios Insua, D. (2000). *Statistical Decision Theory*. Oxford University Press Inc.

Gatherer, D. (2010). So what do we really mean when we say that systems biology is holistic? *BMC Systems Biology*, *4*(22), 1-12. https://doi.org/10.1186/1752-0509-4-22

Geerdink, B. (2013). A reference architecture for big data solutions introducing a model to perform predictive analytics using big data technology. *8th International Conference for Internet Technology and Secured Transactions, (ICITST-2013)*, 71-76. https://doi.org/10.1109/ICITST.2013.6750165

Gigerenzer, G., & Selten, R. (2001). Rethinking rationality. In G. Gigerenzer, & R. Selten (Eds.), *Bounded Rationality*: *The Adaptive Toolbox* (pp. 1-12). MIT Press.

Gorzen-Mitka, I., & Okreglicka, M. (2014). Improving decision-making in complexity environments. *Procedia Economics and Finance*, *16*, 402-409. https://doi.org/10.1016/S2212-5671(14)00819-3

Groen, F. J., & Mosleh, A. (2005). Foundations of probabilistic inference with uncertain evidence. *International Journal of Approximate Reasoning*, *39*(1), 49-83. https://doi.org/10.1016/j.ijar.2004.09.000

Gubler, D. J. (2007). The continuing spread of West Nile Virus in the western hemisphere. *Clinical Infectious Diseases*, *45*(8), 1039-1046. https://doi.org/10.1086/521911

Hadler, J. L., Patel, D., Nasci, R. S., Petersen, L. R., Hughes, J. M., Bradley, K., Etkind, P., Kan, L., & Engel, J. (2015). Assessment of arbovirus surveillance 13 years after introduction of West Nile Virus, United States. *Emerging Infectious Diseases*, *21*(7), 1159-1166. https://doi.org/10.3201/eid2107.140858

Hewett, R. (2004). Decision making using incomplete data. *Proceedings of the 2004 IEEE International Conference on Systems, Man, and Cybernetics, The Hague*, 182-187. https://doi.org/10.1109/ICSMC.2004.1398294

Hobson-Peters, J. M. (2012). Approaches for the development of rapid serological assays for surveillance and diagnosis of infections caused by zoonotic flaviviruses of the Japanese encephalitis virus serocomplex. *Journal of Biomedicine and Biotechnology*, *2012*, 1-15. https://doi.org/10.1155/2012/379738

Howden, L., & Meyer, J. (2011) *Age and sex composition*: *2010* (C2010BR-03). https://www.census.gov/library/publications/2011/dec/c2010br-03.html

IBM Knowledge Center. (2020, January 21). *Generalized Linear Models; SPSS Statistics 24.0.0*. https://www.ibm.com/support/knowledgecenter/zh/SSLVMB·24.0.0/spss/

Javaras, K. & Vos, W. (2002). *Introduction to Generalized Linear Models*. Modelling Lecture 5; Introduction to Statistical Modelling. Department of Statistics University of Oxford. https://www.stats.ox.ac.uk/pub/bdr/IAUL/ModellingLecture5.pdf

Juarrero, A. (2010). *Complex Dynamical Systems Theory*. https://www.cognitive-edge.com/uploads/articles/100608 complex·Dynamical·Systems·Theory.pdf

Kempermann, G. (2017). CYNEFIN as reference framework to facilitate insight and decision-making in complex contexts of biomedical research. *Frontiers in Neuroscience*, *11*(634), 1-8. https://doi.org/10.3389/fnins.2017.00634

Koehler, G. (2014). Defining and exploring a complex system's relational spaces. *Emergence*: *Complexity and Organization*, *16*(1), 100-130.

Koopmans, M. (2014). Change, self-organization and the search for causality in educational research and practice. *Complicity*: *An International Journal of Complexity and Education*, *11*(1), 20-39. https://files.eric.ed.gov/fulltext/EJ1074485.pdf

Kurz, C. F. (2017). Tweedie Distributions for fitting semicontinuous health care utilization cost data. *BMC Medical Research Methodology*, *17*(171)1-8. https://doi.org/10.1186/s12874-017-0445-y

Kurtz, C. F., & Snowden, D. J. (2003). The new dynamics of strategy: Sense-making in a complex and complicated world. *IBM Systems Journal*, *42*(3). https://doi.org/10.1147/sj.423.0462

Kwan, J. L., Kluh, S., & Reisen, W. K. (2012). Antecedent Avian Immunity Limits Tangential Transmission of West Nile Virus to Humans. *PLOS One*, *7*(3). https://doi.org/10.1371/journal.pone.0034127

Kwan, J. L., Park, B. K., Carpenter, T. E., Ngo, V., Civen, R., & Reisen, W. K. (2012). Comparison of enzootic risk measures for predicting West Nile disease, Los Angeles, California, 2004-2010. *Emerging Infectious Disease*, *18*(8), 1298-1306. https://doi.org/10.3201/eid1808.111558

Laerd Statistics (2017). Binomial logistic regression using SPSS Statistics. Statistical tutorials and software guides. https://statistics.laerd.com/

Lichtenstein, B. B., Uhl-Bien, M., Marion, R., Seers, A., & Orton, J. D. (2006). Complexity leadership theory; An interactive perspective on leading in complex adaptive systems. *Emergence*, *8*(4), 2-12. http://digitalcommons.unl.edu/managementfacpub/8/?utm·source=digitalcommons.unl.edu%2Fmanagementfacpub%2F8&utm·medium=PDF&utm·campaign=PDFCoverPages

Liu, A., Lee, V., Galusha, D., Slade, M. D., Diuk-Wasser, M., Andreadis, T., Scotch, M., & Rabinowitz, P. M. (2009). Risk factors for human infection with West Nile Virus in Connecticut: a multi-year analysis. *International Journal of Health Geographics*, *8*(68), 1-11. https://doi.org/10.1186/1476-072X-8-67

Mancayo, A. (2014). *Lecture Topics and Descriptions*. Retrieved July 31, 2014, from American Society for Microbiology: http://www.asm.org/index.php/72-membership/asm-branch-lectureships-prog/91953-dr-abelardo-moncayo-15

Manore, C. A., Davis, J., Christofferson, R. C., Wesson, D., Hyman, J. M., & Mores, C. N. (2014). Towards an early warning system for forecasting West Nile Virus incidence. *PLOS*, *6*, 1-13. https://doi.10.1371/currents.outbreaks.f0b3978230599a56830ce30cb9ce0500

Meadows, D. H. (2008). *Thinking in Systems; A primer*. (D. Wright, Ed.) White River Junction, Vermont, USA: Chelsea Green Publishing.

Menard, S. (2010). *Logistic regression. From introductory to advanced concepts and applications*. Sage Publications.

Montgomery, D. C. (2019). *Design and analysis of experiments*. New York: Wiley.

Morcol, G. (2001). What is complexity science? Postmodernist or Postpositivist? *Emergence*, *3*(1), 104-119. https://doi.org/10.1207/S15327000EM0301

Moreno-Jimenez, J. M., & Vargas, L. G. (2018). Cognitive multiple-criteria decision making and the legacy of the analytic hierarchy process. *Estudios De Economia Aplicada*, *36*(1), 67-80. http://www.revista-eea.net/volumen.php?Id=98&vol=36&ref=1

Morgan, G. (2007). *Images of Organization*. SAGE Publications.

Motta, S., & Pappalardo, F. (2013). Mathematical modeling of biological systems. *Briefings in Bioinformatics*, *14*(4), 411-422. https://doi.org/10.1093/bib/bbs061

National Association of County & City Health Officials (NACCO) Epidemiology and Infectious Disease Prevention and Control Workgroups. (2014). *Statement of Policy*: *Arbovirus Surveillance, Prevention, and Control*. https://www.naccho.org/advocacy/activities

National Association of County & City Health Official (NACCHO) Epidemiology and Infectious Disease Prevention and Control Workgroups. (2017). *Statement of Policy*: *Arbovirus surveillance, prevention, and control*. https://www.naccho.org/uploads/downloadable-resources/Mosquito-control-in-the-U.S.-Report.pdf

Nice, K. (2019). How fuel gauges work [Web log post]. http://auto.howstuffworks.com/fuel-gauge1.htm

North, D. W. (1968, September). A tutorial introduction to decision theory. *IEEE Transactions on Systems Science and Cybernetics*, *SSC-4*(3), 200-210. https://doi.org/10.1109/TSSC.1968.300114

Osman, A., El-Refaey, M., & Elnaggar, A. (2013). Towards real-time analytics in the cloud. *Proceedings of the 2013 IEEE Ninth World Congress on Services*, *USA*, 428-435. http://doi.ieeecomputersociety.org/10.1109/SERVICES.2013.36

Ozaltin, O., & Iyit, N. (2018). Modelling the US Diabetes Mortality Rates via Generalized Linear Model with the Tweedie Distribution. International Journal of Science and Research, *7*(2), 1326-1334. https://doi.org/10.21275/ART2018368

Ozdenerol, E., Taff, G. N., & Akkus, C. (2013). Exploring the spatio-temporal dynamics of reservoir hosts, vectors, and human hosts of West Nile Virus: A review of recent literature. *International Journal of Environmental Research and Public Health*, *10*, 5399-5432. https://doi.org/10.3390/ijerph10115399

Paull, S. H., Horton, D. E., Ashfaq, M., Rastogi, D., Kramer, L. D., Diffenbaugh, N. S., & Kilpatrick, A. M. (2016). Drought and immunity determine the intensity of West Nile virus epidemics and climate change impacts. *Proceedings of the Royal Society B*, *284:*20162078. http://dx.doi.org/10.1098/rspb.2016.2078

Pearl, J. (2009). *Causality*: *Models*, *reasoning*, *and inference*. Cambridge University Press.

Petersen, L. R., Brault, A. C., & Nasci, R. S. (2013). West Nile Virus: Review of the literature. *The Journal of American Medical Association*, *310*(3), 308-315. https://doi.org/10.1001/jama.2013.8042

Pirofski, L. A., & Casadevall, A. (2012). Q&A: What is a pathogen? A question that begs the point. *BMC Biology*, *10*(6), 1-3. https://doi.org/10.1186/1741-7007-10-6

Radde, N. E., & Hutt, M. T. (2016). The physics behind systems biology. *EPJ Nonlinear Biomedical Physics*, *4*(7), 1-19. https://doi.org/10.1140/epjnbp/s40366-016-0034-8

Raia, F. (2008). Causality in complex dynamic systems: A challenge in earth systems science education. *Journal of Geoscience Education*, *56*(1), 81-94. https://doi.org/10.5408/1089-9995-56.1.81

Reiner, R. C., Perkins, T. A., Barker, C. M., Niu, T., Chaves, L. F., Ellis, A. M., George, D.B., Le Menach, A., Pulliam, J.R., Bisanzio, D., Buckee, C., Chiyaka, C., Cummings, D.A., Garcia, A.J., Gatton, M. L., Gething, P. W., Hartley, D. M., Johnston, G., Klein, E.Y., Michael, E.,…Smith, D. L. (2013). A systematic review of mathematical models of mosquito-borne pathogen transmission: 1970-2010. *Journal of the Royal Society Interface*, *10*(81), 1-25. https://doi.org/10.1098/rsif.2012.0921

Riabacke, A. (2006). Managerial decision making under risk and uncertainty. *IAENG International Journal of Computer Science*, *32*(4), 1-7. Advanced online publication. http://scholar.google.com/scholar·url?url=http://www.diva-portal.org/smash/get/diva2:29138/FULLTEXT01.pdf&hl=en&sa=X&scisig=AAGBfm06BTXOygeTyYdJ0pzIo9aVCtdldQ&nossl=1&oi=scholar

Rissino, S., & Lambert-Torres, G. (2009). Rough set theory - Fundamental concepts, principals, data extraction, and applications. In J. &. Ponce (Ed.), *Data Mining and Knowledge Discovery in Real Life Applications* (pp. 1-58). https://doi.org/10.5772/6440

Rochlin, I., Turbow, D., Gomez, F., Ninivaggi, D. V., & Campbell, S. R. (2011). Predictive mapping of human risk for West Nile Virus (WNV) based on environmental and socioeconomic factors. *PLOS One*, *6*(8), 1-10. https://doi.org/10.1371/journal.pone.0023280

Rosenberg, R. (2015). Detecting the emergence of novel, zoonotic viruses pathogenic to humans. *Cellular and Molecular Life Sciences*, *72*(6), 1115-1125. https://doi.org/10.1007/s00018-014-1785-y

Russell, R. E., Katz, R. A., Richgels, K. L., Walsh, D. P., & Grant, E. H. (2017). A framework for modelling emerging diseases to inform management. *Emerging Infectious Diseases*, *23*(1), 1-6. https://doi.org/10.3201/eid2301.161452

Sargut, G., & McGrath, R. G. (2011). Learning to live with complexity. *Harvard Business Review*, 1-15. https://hbr.org/2011/09/learning-to-live-with-complexity

Schafer, G. (1976). *The mathematical theory of evidence*. Princeton, NJ: Princeton University Press.

Senge, P. M. (2006). *The Fifth Discipline*: *The art and practice of the learning organization*. New York, NY: Currency Doubleday.

Shocket, M.S, Verwillow, A. B., Numazu, M. G., Slamani, H., Cohen, J. M., Moustaid, F. E., Rohr, J., Johnson, L. R., & Mordecai, E. A. (2020). Transmission of West Nile and five other temperate mosquito-borne viruses peaks at temperatures between 23°C and 26°C. *eLife 2020*, *9*, 1-67. https://doi.org/10.7554/elife.58511

Simon, H. A. (1955). A behavioral model of rational choice. *The Quarterly Journal of Economics*, *69*(1), 99-118. https://doi.org/10.2307/1884852

Simon, H. A. (1956). Rational choice and the structure of the environment. *Psychological Review*, *63*(2), 129-138. https://doi.org/http://dx.doi.org/10.1037/h0042769

Simon, H. A. (1960). *The new science of management decision* (First Edition ed.). Harper & Brothers Publishers.

Simon, M. K., & Goes, J. (2013). Assumptions, limitations, delimitations, and scope of the study [Web log post]. http://www.dissertationrecipes.com/wp-content/uploads/2011/04/Assumptions-Limitations-Delimitations-and-Scope-of-the-Study.pdf

Slavkovik, A. (2020). Statistics Online: STAT ONLINE. Retrieved August 31, 2020. https://online.stat.psu.edu/statprogram/

Smit, E. H., & Derksen, J. J. (2017). Vignette research on messy and confusing problems in primary mental healthcare. *Europe's Journal of Psychology*, *13(2)*, 300-313. https://doi.org/10.5964/ejop.v13i2.1212

Snowden, D. J., & Boone, M. E. (2007, November). A leader's framework for decision making. *Harvard Business Review*. https://hbr.org/2007/11/a-leaders-framework-for-decision-making

South Carolina Aeronautics Commission. (2011). *Aviation Facilities*. http://www.scaeronautics.com/AirportList.asp

South Carolina Association of Counties. (n.d.). *Incorporated Municipalities*. http://www.sccounties.org/municipalities

South Carolina Department of Health and Environmental Control. (2013, November 1). *Medical Entomology*. https://www.scdhec.gov/health/lab/micro/medical·entomology/arbovirus·data.htm

South Carolina Department of Health and Environmental Control. (2018). *Current Budget Plans*. https://admin.sc.gov/files/J040%-%20DHEC.pdf

Soverow, J. E., Wellenius, G. A., Fisman, D. N., & Mittleman, M. A. (2009). Infectious disease in a warming world: How weather influenced West Nile Virus in the United States (2001-2005). *Environmental Health Perspectives*, *117*(7), 1049-1052. https://doi.org/ 10.1289/ehp.0800487

Straub, R. (2013, May). Why managers haven't embraced complexity. *Harvard Business Review*, 1-4. https://hbr.org/2013/05/why-managers-havent-embraced-c

Suthar, M. S., Diamond, M. S., & Gale Jr., M. (2013, February). West Nile virus infection and immunity. *Nature Reviews| Microbiology*, *11*(2), 115-128. https://doi.org/10.1038/nrmicro2950

Tversky, A., & Kahneman, D. (1981). The framing of decisions and the psychology of choice. *Science*, *211*(4481), 453-458. https://doi.org/10.1126/science.7455683

Udelll, I. (2017). The four dimensions of complexity: Using framing in resolving wicked problems (Poverty in the U.S.). *Policy and Complex Systems*, *3*(1), 84-100. https://doi.org/10.18278/jpcs.3.1.5

United States Census Bureau (2012). *South Carolina*: *2010, Population and housing unit counts* (CPH 2-42). https://www2.census.gov/prod/cen2010/cph-2-42.pdf

United States Census Bureau. (2018). *State area measurements and internal point coordinates*. http://www.census.gov/geographies/reference-files/2010/geo/state-area.html

United States Census Bureau (2019). *2010 Census urban and rural classification and urban area criteria*. https://www.census.gov/programs-surveys/geography/guidance/geo-areas/urban-rural/2010-urban-rural-2010.html

United States Census Bureau. (2020). *American National Standards Institute (ANSI) Codes for States* (INCITS 446-2008). http://www.census.gov/library/reference/code-lists/ansi.html

United States Census Bureau (2021). *Methodology for the United States population estimates*: *Vintage 2018, Nation, States, Counties, and Puerto Rico – April 1, 2010 – Jul 1, 2018*. https://www2.census.gov/programs-surveys/popest/technical-documentation/methodology/2010-2018/2018-natstcopr-meth.pdf

van Doorn, H.R. (2014). Emerging infectious diseases. *Medicine (Abingdon, England*: *UK ed.), 42*(1), 60-63. https://doi.org/10.1016/j.mpmed.2013.10.014: 10.1016/j.mpmed.2013.10.014

Vassar, M., & Holzmann, M. (2013, Oct 12). The retrospective chart review: Important methodological considerations. *Journal of Educational Evaluation for Health Professions*, *10*(12). https://doi.org/10.3352/jeehp.2013.10.12

Walczak, B., & Massart, D. L. (1999). Rough sets theory. *Chemometrics and Intelligent Laboratory Systems*, *47*(1), 1-16. https://doi.org/10.1016/S0169-7439(98)00200-7

Walton, M. (2016). Setting the context for using complexity theory in evaluation: boundaries, governance, and utilization. *Evidence & Policy*: *A Journal of*

Research, Debate, & Practice, *12*(1), 73-89. https://doi.org/10.1332/174426415X14298726247211

Weather Underground. (2016). *Historical Weather*. https://www.wunderground.com/history/

Wolkenhauer, O., & Green, S. (2013, April 22). The search for organizing principles as a cure against reductionism in systems medicine. *FEBS Journal*, *280*(23), 5938-5948. https://doi.org/10.1111/febs.12311

Wolkenhauer, O. (2014a). Pushing limits by embracing complexity. *IET Systems Biology*, *8*(6), 244-250. https://doi.org/10.1049/iet-syb.2014.0031

Wolkenhauer, O. (2014b). Why model? *Frontiers in Physiology*, *5*(21), 1-7. https://doi.org/10.3389/fphys.2014.00021

Young, S. G., Tullis, J. A., & Cothern, J. (2013). A remote sensing and GIS-assisted landscape epidemiology approach to West Nile Virus. *Applied Geography*, *45*, 241-249. https://doi.org/10.1016/j.apgeog.2013.09.022

Zalta, E. N. (Ed.) (2016). Decision Theory. (E. N. Zalta, Ed.) *The Stanford encyclopedia of philosophy* (Winter 2016 ed.). Stanford University. https://plato.stanford.edu/archives/win2016/entries/decision-theory/

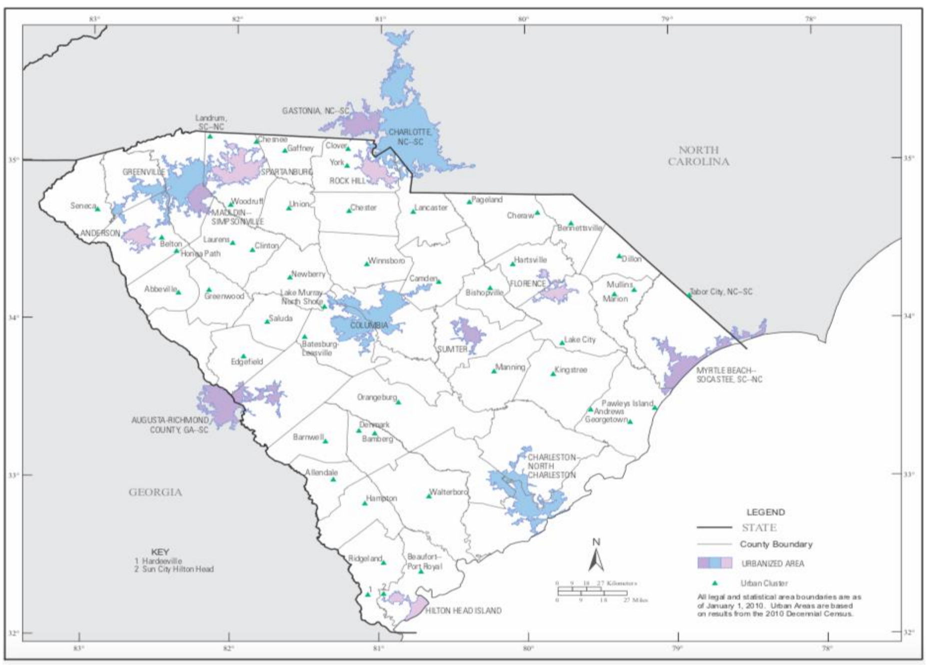
Appendix A: Research Data Overview

State Overview

Located in the southeast region of the United States, South Carolina is organized into 46 counties (Figure A1) incorporating land mass and water areas totaling 32,020 square miles (United States Census Bureau, 2018). Within the state boundaries, 30,061 square miles (94% of the state) are measured as land mass (2018). The remaining 1,960 square miles (6%) are divided into inland (1,064 square miles), coastal (110 square miles), and territorial waters (786 square miles) (2018).

Figure A1

The State of South Carolina



Note. The State of South Carolina (United States Census Bureau, 2012). This map shows the distribution of the urban areas and urban clusters within the state.

In the last U.S. census of 2010, the total population of South Carolina was recorded as 4,625,364 (United States Census Bureau, 2012). As seen in Figure A1, the majority of the population is distributed across urban areas (population > 50,000) in 23 of the 46 counties. The 2010 census attributes 3,067,809 or 66.3% of the South Carolina population to these areas. Of interest to this study, 1,875,097 (40%) of the population total was recorded as being greater than 44 years of age (Howden & Meyer, 2011). Previous studies and reports place this population segment as an at-risk group for WNv.

Research Data Overview

For this longitudinal study, I collected data to populate values for the DV and EEVs over the period 1999 to 2016. This required the creation of a data set that recorded the following information by row according to its variable symbol, math variable, type of variable (e.g., numerical, categorical, etc.), and number of variables. Data were entered into the spreadsheet using the following column headings:

1. Data Record Number (DRN). Numerical entry value (1-9936).
2. SC County (CTY). Categorical entry value (e.g., Aiken).
3. Date of Record (DOR). Numerical entry value (19990101-20161231).
4. Land Use (LAN). Categorical entry value (agricultural, industrial).
5. Converted Land Use (USE). Numerical entry value (0=agricultural, 1=industrial)
6. Topology (TOP). Numerical value (county seat elevation in feet)

7. West Nile Virus events by county
8. Average 30-Day Temperature in degrees Fahrenheit (°F) by county
9. Average 30-Day Rainfall in inches by county
10. Average 30-Day Dew Point in degrees Fahrenheit (°F) by county
11. Average 30-Day Snow Depth in (inches) by county
12. Average 30-Day Wind Speed in (mile per hour) by county
13. Average 30-Day Barometric Pressure in (millibars) by county
14. Housing Units per (square mile) by county
15. Population density per (square miles) by county

This archival data was collected in a single spreadsheet which acted as the 1999 to 2016 data repository for each of the data sets above. The next section provides specific information on the administrative and SPSS data sets.

Data Record Number (DRC) - SC County Names, FIPS codes, County Seat Locations, and Square Mileage

County names and FIPS codes are readily available from the USCB. County seat locations are available from the South Carolina Association of Counties website. The data was combined into two data field entries for each of the eleven archival data sets in the previous section.

Federal Information Processing Series (FIPS). These codes are standardized numeric or alphabetic codes issued by the American National Standards Institute Codes (ANSI) to ensure uniform identification of geographic entities and are readily available from the USCB website (United States Census Bureau, 2020). Within these standardized codes, ANSI uses the Federal Information Processing Series (FIPS). While FIPS codes are no longer the standard, they are still used within ANSI and provide historical continuity for this research providing a catalogue number for data collection.

I extracted FIPS information for each county to manage data entries over a 15-year period. The data will be recorded in the following format:

State Postal Code – SC

State FIPS Code – 45

County FIPS Code – NNN (e.g., Abbeville County – 001)

County Name – Abbeville County

County Seat. County seats were identified using the South Carolina Association of Counties website and will provide the county centroid for mapping purposes (South Carolina Association of Counties, n.d.). The 46 county seats will be cross-referenced with the Department of Commerce, Bureau of Census, Qualifying Urban Area for the 2010 and 2000 (United States Census Bureau, 2019) census to determine whether a county seat was classified as an urban or rural area (2019).

County Square Miles. This USCB data provided information on the square miles associated with each county. This was used to form ratio data for population density and housing density. These two ratios were used to populate the Land Use and Urbanization variables.

West Nile Virus Events by County

This data came directly from the South Carolina Department of Health and Environmental Control (SC DHEC) and cross checked with total numbers from the federal CDC**.** The data was provided in the form of positive West Nile Virus events by year, month, county, and type (i.e., human, equine, mosquito pool, dead bird, sentinel animal, other). These data reflected WNv events that have been recorded over the 2002-2016 timeframe. These data will be matched to the monthly entries for each county by year of data and logged by number of WNv events.

Climate Data

Historical climate data came from three sources, the websites of the Weather Underground, Old Farmer's Almanac, and the South Carolina Department of Natural Resources (DNR). An IBM company, the Weather Underground mission "is to provide climate data to every person in the world" (Weather Underground, 2016). Using a system called BestForecast™, Weather Underground collects weather data from over 180,000 proprietary weather stations and compares that data to the National Weather Service's National Digital Forecast Database (NDFD). Weather Underground also provides historical climate data by county for each year of the proposed study. Old Farmer's Almanac provides historical data by county for the United States. The South Carolina DNR will be used as a tertiary source for missing or incomplete data.

Average 30-Day Temperature in Degrees Fahrenheit (^{0}F) by County. Average 30-Day temperature data was aligned to the county seat by year and month from 1999 to 2016.

Average 30-Day Rainfall in (inches) by county. Average 30-day rainfall data was aligned to the county seat by year and month from 1999 to 2016.

Average 30-Day Dew Point in degrees Fahrenheit (^{0}F) by county. Average 30-day dew point data was aligned to the county seat by year and month from 1999 to 2016.

Average 30-Day Snow Depth in (inches) by county. Average 30-day snow depth data was aligned to the county seat by year and month from 1999 to 2016.

Average 30-Day Barometric Pressure in (millibars) by county. Average 30-day barometric data was aligned to the county seat by year and month from 1999 to 2016.

Elevation in (Feet) For Each County Seat

Elevation data was extracted from the South Carolina Aeronautics Commission website which lists elevations for all county airports (South Carolina Aeronautics Commission, 2011). The airport locations were matched with the county seats to provide a standard measure of elevation.

Housing Units Per (Square Mile) by County

These data came from the USCB 2000 and 2010 census. Housing units per square mile were computed for each census. The variation between the 2000 and 2010 figures will be computed and then extrapolated for the years 2000 to 2009 and for 2010 to 2015.

These data were used by year to compute housing per square mile metrics as entry criteria for the Land Use variable.

Population Density Per (Square Miles) by County

County populations were collected and converted to a ratio using the total county square mile data. The population density ratio provided a measure of county urbanization. compared to the 2010 U.S. Census Bureau national average of 88.4 people/square mile. County population estimates were captured for the years 1999 to 2016 from data available through the USCB. The USCB uses a program that annually estimates population and housing by nation, state, and county based on the last decennial census. For the period 2010 to 2018, the USCB used a cohort component methodology (Population Estimate = Base population + Births – Deaths + Net domestic migration) for computing national and county population estimates (United States Census Bureau, 2021).

Appendix B: Stage A Model Building Process

Stage A2 *$DV_{PRESENCE}$* Analysis for Current Month (SASM)

Table B1

Stage A2 $DV_{PRESENCE}$ Analysis with DV1 using EVs Only (SASM)

$DV_{PRESENCE}$	Stepwise Method	$R^2_{\text{Nagelkerke}}$	*LL Ratio* χ^2
DV1 Model 1	FS (COND)	.171	$\chi^2(8) = 436.763$, $p < .001$
DV1 Model 2	FS (LR)	.171	$\chi^2(8) = 436.763$, $p < .001$
DV1 Model 3	FS (WALD)	.171	$\chi^2(8) = 436.763$, $p < .001$
DV1 Model 4	BE (COND)	.171	$\chi^2(8) = 436.763$, $p < .001$
DV1 Model 5	BE (LR)	.171	$\chi^2(8) = 436.763$, $p < .001$
DV1 Model 6	BE (WALD)	.171	$\chi^2(8) = 436.763$, $p < .001$

Stage A3 *$DV_{PRESENCE}$* Analysis, *DV1*, EVs (SFEM)

Table B2

Stage A3 $DV_{PRESENCE}$ Analysis with DV1 and EVs Only

$DV_{PRESENCE}$	$R^2_{\text{Nagelkerke}}$	*LL Ratio* χ^2
DV1 Model 7	.171	$\chi^2(9) = 436.929$, $p < .001$
DV1 Model 8	.171	$\chi^2(8) = 436.763$, $p < .001$

a. Model 7 Predictors: (Constant), *EV1ATM*, *EV2ARN*, *EV4ASD*, *EV5ABP*, *EV6AWS*, *EV7ELV*, *EV8USE*, *EV9POP*, *EV10ADD*. *EV4ASD* not significant at $p = .644$.

b. Model 8 Predictors: (Constant), *EV1ATM*, *EV2ARN*, *EV5ABP*, *EV6AWS*, *EV7ELV*, *EV8USE*, *EV9POP*, *EV10ADD*. All predictors significant at $p < .20$.

Table B3

Stage A3 Results with DV1, EVs, and 2FIs

DV$_{PRESENCE}$	$R^2_{\text{Nagelkerke}}$	*LL Ratio* χ^2
DV1 Model 9	.202	$\chi^2(34) = 519.047$, $p < .001$
DV1 Model 10	.202	$\chi^2(33) = 519.047$, $p < .001$
DV1 Model 11	.202	$\chi^2(32) = 519.045$, $p < .001$
DV1 Model 12	.202	$\chi^2(31) = 519.013$, $p < .001$
DV1 Model 13	.202	$\chi^2(30) = 518.943$, $p < .001$
DV1 Model 14	.202	$\chi^2(29) = 518.877$, $p < .001$
DV1 Model 15	.202	$\chi^2(28) = 518.787$, $p < .001$
DV1 Model 16	.202	$\chi^2(27) = 518.662$, $p < .001$
DV1 Model 17	.202	$\chi^2(26) = 518.419$, $p < .001$
DV1 Model 18	.202	$\chi^2(25) = 518.144$, $p < .001$
DV1 Model 19	.202	$\chi^2(24) = 517.905$, $p < .001$
DV1 Model 20	.202	$\chi^2(23) = 517.625$, $p < .001$
DV1 Model 21	.201	$\chi^2(22) = 517.012$, $p < .001$
DV1 Model 22	.201	$\chi^2(21) = 516.266$, $p < .001$
DV1 Model 23	.201	$\chi^2(20) = 515.617$, $p < .001$
DV1 Model 24	.201	$\chi^2(19) = 515.617$, $p < .001$
DV1 Model 25	.200	$\chi^2(18) = 514.286$, $p < .001$
DV1 Model 26	.200	$\chi^2(17) = 512.694$, $p < .001$

a. Model 9 Predictors: (Constant), *EV1ATM, EV2ARN, EV5ABP, EV6AWS, EV7ELV, EV8USE, EV9POP, EV10ADD, EV1·EV2, EV1·EV5, EV1·EV6, EV1·EV7, EV1·EV8, EV1·EV9, EV1·EV10, EV2·EV5, EV2·EV6, EV2·EV7, EV2·EV8, EV2·EV9, EV2·EV10, EV5·EV6, EV5·EV7, EV5·EV8, EV5·EV9, EV5·EV10, EV6·EV7, EV6·EV8, EV6·EV9, EV6·EV10, EV7·EV8, EV7·EV9, EV7·EV10, EV8·EV9, EV8·EV10. EV2·EV6* not significant at p = .989.

b. Model 10 Predictors: (Constant), *EV1ATM, EV2ARN, EV5ABP, EV6AWS, EV7ELV, EV8USE, EV9POP, EV10ADD, EV1·EV2, EV1·EV5, EV1·EV6, EV1·EV7, EV1·EV8, EV1·EV9, EV1·EV10, EV2·EV5, EV2·EV7, EV2·EV8, EV2·EV9, EV2·EV10, EV5·EV6, EV5·EV7, EV5·EV8, EV5·EV9, EV5·EV10, EV6·EV7, EV6·EV8, EV6·EV9, EV6·EV10, EV7·EV8, EV7·EV9, EV7·EV10, EV8·EV9, EV8·EV10. EV2·EV10* not significant at p = .989.

c. Model 11 Predictors: (Constant), *EV1ATM, EV2ARN, EV5ABP, EV6AWS, EV7ELV, EV8USE, EV9POP, EV10ADD, EV1·EV2, EV1·EV5, EV1·EV6, EV1·EV7, EV1·EV8, EV1·EV9, EV1·EV10, EV2·EV5, EV2·EV7, EV2·EV8, EV2·EV9, EV5·EV6, EV5·EV7, EV5·EV8, EV5·EV9, EV5·EV10, EV6·EV7, EV6·EV8, EV6·EV9, EV6·EV10, EV7·EV8, EV7·EV9, EV7·EV10, EV8·EV9, EV8·EV10. EV7·EV9* not significant at p = .859.

d. Model 12 Predictors: (Constant), *EV1ATM, EV2ARN, EV5ABP, EV6AWS, EV7ELV, EV8USE, EV9POP, EV10ADD, EV1·EV2, EV1·EV5, EV1·EV6, EV1·EV7, EV1·EV8, EV1·EV9, EV1·EV10, EV2·EV5, EV2·EV7, EV2·EV8, EV2·EV9, EV5·EV6, EV5·EV7, EV5·EV8, EV5·EV9, EV5·EV10, EV6·EV7, EV6·EV8, EV6·EV9, EV6·EV10, EV7·EV8, EV7·EV10, EV8·EV9, EV8·EV10. EV5·EV6* not significant at p = .791.

e. Model 13 Predictors: (Constant), *EV1ATM, EV2ARN, EV5ABP, EV6AWS, EV7ELV, EV8USE, EV9POP, EV10ADD, EV1·EV2, EV1·EV5, EV1·EV6, EV1·EV7, EV1·EV8, EV1·EV9, EV1·EV10, EV2·EV5, EV2·EV7, EV2·EV8, EV2·EV9, EV5·EV7, EV5·EV8, EV5·EV9, EV5·EV10, EV6·EV7, EV6·EV8, EV6·EV9, EV6·EV10, EV7·EV8, EV7·EV10, EV8·EV9, EV8·EV10. EV1·EV10* not significant at p = .796.

f. Model 14 Predictors: (Constant), *EV1ATM, EV2ARN, EV5ABP, EV6AWS, EV7ELV, EV8USE, EV9POP, EV10ADD, EV1·EV2, EV1·EV5, EV1·EV6, EV1·EV7, EV1·EV8, EV1·EV9, EV2·EV5, EV2·EV7, EV2·EV8, EV2·EV9, EV5·EV7, EV5·EV8, EV5·EV9, EV5·EV10, EV6·EV7, EV6·EV8, EV6·EV9, EV6·EV10, EV7·EV8, EV7·EV10, EV8·EV9, EV8·EV10. EV5·EV10* not significant at p = .765.

g. Model 15 Predictors: (Constant), *EV1ATM, EV2ARN, EV5ABP, EV6AWS, EV7ELV,*

EV8USE, EV9POP, EV10ADD, EV1·EV2, EV1·EV5, EV1·EV6, EV1·EV7, EV1·EV8, EV1·EV9, EV2·EV5, EV2·EV7, EV2·EV8, EV2·EV9, EV5·EV7, EV5·EV8, EV5·EV9, EV6·EV7, EV6·EV8, EV6·EV9, EV6·EV10, EV7·EV8, EV7·EV10, EV8·EV9, EV8·EV10. EV5·EV9 not significant at p = .723.

h. Model 16 Predictors: (Constant), *EV1ATM, EV2ARN, EV5ABP, EV6AWS, EV7ELV, EV8USE, EV9POP, EV10ADD, EV1·EV2, EV1·EV5, EV1·EV6, EV1·EV7, EV1·EV8, EV1·EV9, EV2·EV5, EV2·EV7, EV2·EV8, EV2·EV9, EV5·EV7, EV5·EV8, EV5·EV10, EV6·EV7, EV6·EV8, EV6·EV9, EV6·EV10, EV7·EV8, EV7·EV10, EV8·EV9, EV8·EV10. EV2·EV9* not significant at p = .622.

i. Model 17 Predictors: (Constant), *EV1ATM, EV2ARN, EV5ABP, EV6AWS, EV7ELV, EV8USE, EV9POP, EV10ADD, EV1·EV2, EV1·EV5, EV1·EV6, EV1·EV7, EV1·EV8, EV1·EV9, EV2·EV5, EV2·EV7, EV2·EV8, EV5·EV7, EV5·EV8, EV5·EV10, EV6·EV7, EV6·EV8, EV6·EV9, EV6·EV10, EV7·EV8, EV7·EV10, EV8·EV9, EV8·EV10. EV6·EV9* not significant at p = .600.

j. Model 18 Predictors: (Constant), *EV1ATM, EV2ARN, EV5ABP, EV6AWS, EV7ELV, EV8USE, EV9POP, EV10ADD, EV1·EV2, EV1·EV5, EV1·EV6, EV1·EV7, EV1·EV8, EV1·EV9, EV2·EV5, EV2·EV7, EV2·EV8, EV5·EV7, EV5·EV8, EV5·EV10, EV6·EV7, EV6·EV8, EV6·EV10, EV7·EV8, EV7·EV10, EV8·EV9, EV8·EV10. EV6·EV8* not significant at p = .625.

k. Model 19 Predictors: (Constant), *EV1ATM, EV2ARN, EV5ABP, EV6AWS, EV7ELV, EV8USE, EV9POP, EV10ADD, EV1·EV2, EV1·EV5, EV1·EV6, EV1·EV7, EV1·EV8, EV1·EV9, EV2·EV5, EV2·EV7, EV2·EV8, EV5·EV7, EV5·EV8, EV5·EV10, EV6·EV7, EV6·EV10, EV7·EV8, EV7·EV10, EV8·EV9, EV8·EV10. EV1·EV8* not significant at p = .598.

l. Model 20 Predictors: (Constant), *EV1ATM, EV2ARN, EV5ABP, EV6AWS, EV7ELV, EV8USE, EV9POP, EV10ADD, EV1·EV2, EV1·EV5, EV1·EV6, EV1·EV7, EV1·EV9, EV2·EV5, EV2·EV7, EV2·EV8, EV5·EV7, EV5·EV8, EV6·EV7, EV6·EV10, EV7·EV8, EV7·EV10, EV8·EV9. EV5·EV8* not significant at p = .437.

m. Model 21 Predictors: (Constant), *EV1ATM*, *EV2ARN*, *EV5ABP*, *EV6AWS*, *EV7ELV*, *EV8USE*, *EV9POP*, *EV10ADD*, *EV1·EV2*, *EV1·EV5*, *EV1·EV6*, *EV1·EV7*, *EV1·EV9*, *EV2·EV5, EV2·EV7, EV2·EV8, EV5·EV7, EV5·EV10, EV6·EV7, EV6·EV10, EV7·EV8, EV7·EV10*, *EV8·EV9, EV8·EV10. EV6·EV7* not significant at $p = .387$.

n. Model 22 Predictors: (Constant), *EV1ATM*, *EV2ARN*, *EV5ABP*, *EV6AWS*, *EV7ELV*, *EV8USE*, *EV9POP*, *EV10ADD*, *EV1·EV2*, *EV1·EV5*, *EV1·EV6*, *EV1·EV7*, *EV1·EV9*, *EV2·EV5, EV2·EV7, EV2·EV8, EV5·EV7, EV5·EV10, EV6·EV10, EV7·EV8, EV7·EV10*, *EV8·EV9, EV8·EV10. EV2·EV7* not significant at $p = .420$.

o. Model 23 Predictors: (Constant), *EV1ATM*, *EV2ARN*, *EV5ABP*, *EV6AWS*, *EV7ELV*, *EV8USE*, *EV9POP*, *EV10ADD*, *EV1·EV2*, *EV1·EV5*, *EV1·EV6*, *EV1·EV7*, *EV1·EV9*, *EV2·EV5, EV2·EV8, EV5·EV7, EV5·EV10, EV6·EV10, EV7·EV8, EV7·EV10*, *EV8·EV9*, *EV8·EV10. EV5·EV7* not significant at $p = .413$.

p. Model 24 Predictors: (Constant), *EV1ATM*, *EV2ARN*, *EV5ABP*, *EV6AWS*, *EV7ELV*, *EV8USE*, *EV9POP*, *EV10ADD*, *EV1·EV2*, *EV1·EV5*, *EV1·EV6*, *EV1·EV7*, *EV1·EV9*, *EV2·EV5, EV2·EV8, EV5·EV10, EV6·EV10, EV7·EV8, EV7·EV10, EV8·EV9*, *EV8·EV10. EV2·EV8* not significant at $p = .412$.

q. Model 25 Predictors: (Constant), *EV1ATM*, *EV2ARN*, *EV5ABP*, *EV6AWS*, *EV7ELV*, *EV8USE*, *EV9POP*, *EV10ADD*, *EV1·EV2*, *EV1·EV5*, *EV1·EV6*, *EV1·EV7*, *EV1·EV9*, *EV2·EV5, EV5·EV10, EV6·EV10, EV7·EV8, EV7·EV10*, *EV8·EV9, EV8·EV10. EV7·EV8* not significant at $p = .214$.

r. Model 25 Predictors: (Constant), *EV1ATM*, *EV2ARN*, *EV5ABP*, *EV6AWS*, *EV7ELV*, *EV8USE*, *EV9POP*, *EV10ADD*, *EV1·EV2*, *EV1·EV5*, *EV1·EV6*, *EV1·EV7*, *EV1·EV9*, *EV2·EV5*, *EV5·EV10*, *EV6·EV10*, *EV7·EV10*, *EV8·EV9*, *EV8·EV10.* All predictors significant at $p < .20$.

Table B4

Stage A3 $DV_{PRESENCE}$ Analysis with DV1 and All EEVs

$DV_{PRESENCE}$	$R^2_{\text{Nagelkerke}}$	*LL Ratio* χ^2
DV1 Model 26	.285	$\chi^2(34) = 743.225, p < .001$
DV1 Model 27	.285	$\chi^2(34) = 743.225, p < .001$
DV1 Model 28	.285	$\chi^2(33) = 743.221, p < .001$
DV1 Model 29	.285	$\chi^2(32) = 743.194, p < .001$
DV1 Model 30	.285	$\chi^2(31) = 743.103, p < .001$
DV1 Model 31	.285	$\chi^2(30) = 742.924, p < .001$
DV1 Model 32	.285	$\chi^2(29) = 742.863, p < .001$
DV1 Model 33	.285	$\chi^2(28) = 742.323, p < .001$
DV1 Model 34	.285	$\chi^2(27) = 741.539, p < .001$
DV1 Model 35	.285	$\chi^2(26) = 740.692, p < .001$
DV1 Model 36	.285	$\chi^2(25) = 739.250, p < .001$
DV1 Model 37	.284	$\chi^2(24) = 738.117, p < .001$
DV1 Model 38	.283	$\chi^2(23) = 737.226, p < .001$

a. Model 26 Predictors: (Constant), *EV1ATM, EV2ARN, EV5ABP, EV6AWS, EV7ELV, EV8USE, EV9POP, EV10ADD, EV1·EV2, EV1·EV5, EV1·EV6, EV1·EV7, EV1·EV9, EV2·EV5, EV2·EV7, EV2·EV8, EV5·EV7, EV5·EV8, EV6·EV7, EV6·EV10, EV7·EV8, EV7·EV10, EV8·EV9, January, February, March, April, May, June, July, August, September, October, November, December. March* not significant at $p = .951$.

b. Model 27 Predictors: (Constant), *EV1ATM, EV2ARN, EV5ABP, EV6AWS, EV7ELV, EV8USE, EV9POP, EV10ADD, EV1·EV2, EV1·EV5, EV1·EV6, EV1·EV7, EV1·EV9, EV2·EV5, EV2·EV7, EV2·EV8, EV5·EV7, EV5·EV8, EV6·EV7, EV6·EV10, EV7·EV8, EV7·EV10, EV8·EV9, January, February, April, May, June, July, August, September, October, November, December. December* not significant at $p = .951$.

c. Model 28 Predictors: (Constant), *EV1ATM*, *EV2ARN*, *EV5ABP*, *EV6AWS*, *EV7ELV*, *EV8USE*, *EV9POP*, *EV10ADD*, *EV1·EV2*, *EV1·EV5*, *EV1·EV6*, *EV1·EV7*, *EV1·EV9*, *EV2·EV5*, *EV2·EV7*, *EV2·EV8*, *EV5·EV7*, *EV5·EV8*, *EV6·EV7*, *EV6·EV10*, *EV7·EV8*, *EV7·EV10*, *EV8·EV9*, *January*, *February*, *April*, *May*, *June*, *July*, *August*, *September*, *October*, *November*. *January* not significant at $p = .869$.

d. Model 29 Predictors: (Constant), *EV1ATM*, *EV2ARN*, *EV5ABP*, *EV6AWS*, *EV7ELV*, *EV8USE*, *EV9POP*, *EV10ADD*, *EV1·EV2*, *EV1·EV5*, *EV1·EV6*, *EV1·EV7*, *EV1·EV9*, *EV2·EV5*, *EV2·EV7*, *EV2·EV8*, *EV5·EV7*, *EV5·EV8*, *EV6·EV7*, *EV6·EV10*, *EV7·EV8*, *EV7·EV10*, *EV8·EV9*, *February*, *April*, *May*, *June*, *July*, *August*, *September*, *October*, *November*. *EV7ELV* not significant at $p = .764$.

e. Model 30 Predictors: (Constant), *EV1ATM*, *EV2ARN*, *EV5ABP*, *EV6AWS*, *EV8USE*, *EV9POP*, *EV10ADD*, *EV1·EV2*, *EV1·EV5*, *EV1·EV6*, *EV1·EV7*, *EV1·EV9*, *EV2·EV5*, *EV2·EV7*, *EV2·EV8*, *EV5·EV7*, *EV5·EV8*, *EV6·EV7*, *EV6·EV10*, *EV7·EV8*, *EV7·EV10*, *EV8·EV9*, *February*, *April*, *May*, *June*, *July*, *August*, *September*, *October*, *November*. *EV6AWS* not significant at $p = .672$.

f. Model 31 Predictors: (Constant), *EV1ATM*, *EV2ARN*, *EV5ABP*, *EV8USE*, *EV9POP*, *EV10ADD*, *EV1·EV2*, *EV1·EV5*, *EV1·EV6*, *EV1·EV7*, *EV1·EV9*, *EV2·EV5*, *EV2·EV7*, *EV2·EV8*, *EV5·EV7*, *EV5·EV8*, *EV6·EV7*, *EV6·EV10*, *EV7·EV8*, *EV7·EV10*, *EV8·EV9*, *February*, *April*, *May*, *June*, *July*, *August*, *September*, *October*, *November*. *EV1·EV6* not significant at $p = .806$.

g. Model 32 Predictors: (Constant), *EV1ATM*, *EV2ARN*, *EV5ABP*, *EV8USE*, *EV9POP*, *EV10ADD*, *EV1·EV2*, *EV1·EV5*, *EV1·EV7*, *EV1·EV9*, *EV2·EV5*, *EV2·EV7*, *EV2·EV8*, *EV5·EV7*, *EV5·EV8*, *EV6·EV7*, *EV6·EV10*, *EV7·EV8*, *EV7·EV10*, *EV8·EV9*, *February*, *April*, *May*, *June*, *July*, *August*, *September*, *October*, *November*. *EV1·EV9* not significant at $p = .464$.

h. Model 33 Predictors: (Constant), *EV1ATM*, *EV2ARN*, *EV5ABP*, *EV8USE*, *EV9POP*, *EV10ADD*, *EV1·EV2*, *EV1·EV5*, *EV1·EV7*, *EV2·EV5*, *EV2·EV7*, *EV2·EV8*, *EV5·EV7*, *EV5·EV8*, *EV6·EV7*, *EV6·EV10*, *EV7·EV8*, *EV7·EV10*, *EV8·EV9*, *February*, *April*, *May, June, July, August, September, October, November. EV6·EV7* not significant at *p* = .375.

i. Model 34 Predictors: (Constant), *EV1ATM, EV2ARN, EV5ABP, EV8USE, EV9POP, EV10ADD, EV1·EV2, EV1·EV5, EV1·EV7, EV2·EV5, EV2·EV7, EV2·EV8, EV5·EV7, EV5·EV8, EV6·EV10, EV7·EV8, EV7·EV10, EV8·EV9, February, April, May, June, July, August, September, October, November. EV7·EV8* not significant at *p* = .363.

j. Model 35 Predictors: (Constant), *EV1ATM, EV2ARN, EV5ABP, EV8USE, EV9POP, EV10ADD, EV1·EV2, EV1·EV5, EV1·EV7, EV2·EV5, EV2·EV7, EV2·EV8, EV5·EV7, EV5·EV8, EV6·EV10, EV7·EV8, EV7·EV10, EV8·EV9, February, April, May, June, July, August, September, October, November. February* not significant at *p* = .301.

k. Model 36 Predictors: (Constant), *EV1ATM, EV2ARN, EV5ABP, EV8USE, EV9POP, EV10ADD, EV1·EV2, EV1·EV5, EV1·EV7, EV2·EV5, EV2·EV7, EV2·EV8, EV5·EV7, EV5·EV8, EV6·EV10, EV7·EV10, EV8·EV9, April, May, June, July, August, September, October, November. EV2·EV8* not significant at *p* = .281.

l. Model 37 Predictors: (Constant), *EV1ATM, EV2ARN, EV5ABP, EV8USE, EV9POP, EV10ADD, EV1·EV2, EV1·EV5, EV1·EV7, EV2·EV5, EV2·EV7, EV5·EV7, EV5·EV8, EV6·EV10, EV7·EV10, EV8·EV9, April, May, June, July, August, September, October, November. EV5·EV8* not significant at *p* = .354.

m. Model 38 Predictors: (Constant), *EV1ATM, EV2ARN, EV5ABP, EV8USE, EV9POP, EV10ADD, EV1·EV2, EV1·EV5, EV1·EV7, EV2·EV5, EV2·EV7, EV5·EV7, EV6·EV10, EV7·EV10, EV8·EV9, April, May, June, July, August, September, October, November.* All predictors significant at *p* < .20.

Stage A4 *DV_{PRESENCE}* Analysis for Future Months (SASM)

Table B5

Stage A4 $DV_{PRESENCE}$ Analysis for Best Time-lagged DV Selection, DV2

$DV_{PRESENCE}$	*SASM*	$R^2_{\text{Nagelkerke}}$	*LL Ratio* χ^2
DV2 Model 39	FS (COND)	.255	$\chi^2(6) = 661.645$, $p < .001$
DV2 Model 40	FS (LR)	.255	$\chi^2(6) = 661.645$, $p < .001$
DV2 Model 41	FS (WALD)	.255	$\chi^2(6) = 661.645$, $p < .001$
DV2 Model 42	BE (COND)	.257	$\chi^2(7) = 665.685$, $p < .001$
DV2 Model 43	BE (LR)	.257	$\chi^2(7) = 665.685$, $p < .001$
DV2 Model 44	BE (WALD)	.257	$\chi^2(6) = 661.645$, $p < .001$

a. Model 39 Predictors: (Constant), *EV1ATM*, *EV5ABP*, *EV6AWS*, *EV8USE*, *EV9POP*, *EV10ADD*.
b. Model 40 Predictors: (Constant), *EV1ATM*, *EV5ABP*, *EV6AWS*, *EV8USE*, *EV9POP*, *EV10ADD*.
c. Model 41 Predictors: (Constant), *EV1ATM*, *EV5ABP*, *EV6AWS*, *EV8USE*, *EV9POP*, *EV10ADD*.
d. Model 42 Predictors: (Constant), *EV1ATM*, *EV2ARN*, *EV5ABP*, *EV6AWS*, *EV8USE*, *EV9POP*, *EV10ADD*.
e. Model 43 Predictors: (Constant), *EV1ATM*, *EV2ARN*, *EV5ABP*, *EV6AWS*, *EV8USE*, *EV9POP*, *EV10ADD*.
f. Model 44 Predictors: (Constant), *EV1ATM*, *EV5ABP*, *EV6AWS*, *EV8USE*, *EV9POP*, *EV10ADD*.

Table B6

Stage A4 $DV_{PRESENCE}$ Analysis for Best Time-lagged DV Selection, DV3

$DV_{PRESENCE}$	*SASM*	$R^2_{\text{Nagelkerke}}$	*LL Ratio* χ^2
DV3 Model 45	FS (COND)	.245	$\chi^2(5) = 634.453$, $p < .001$
DV3 Model 46	FS (LR)	.245	$\chi^2(5) = 634.453$, $p < .001$
DV3 Model 47	FS (WALD)	.245	$\chi^2(5) = 634.453$, $p < .001$
DV3 Model 48	BE (COND)	.245	$\chi^2(5) = 634.453$, $p < .001$
DV3 Model 49	BE (LR)	.245	$\chi^2(5) = 634.453$, $p < .001$
DV3 Model 50	BE (WALD)	.245	$\chi^2(5) = 634.453$, $p < .001$

a. Model 45 Predictors: (Constant), *EV1ATM*, *EV5ABP*, *EV8USE*, *EV9POP*, *EV10ADD*.
b. Model 46 Predictors: (Constant), *EV1ATM*, *EV5ABP*, *EV8USE*, *EV9POP*, *EV10ADD*.
c. Model 47 Predictors: (Constant), *EV1ATM*, *EV5ABP*, *EV8USE*, *EV9POP*, *EV10ADD*.
d. Model 48 Predictors: (Constant), *EV1ATM*, *EV5ABP*, *EV8USE*, *EV9POP*, *EV10ADD*.
e. Model 49 Predictors: (Constant), *EV1ATM*, *EV5ABP*, *EV8USE*, *EV9POP*, *EV10ADD*.
f. Model 50 Predictors: (Constant), *EV1ATM*, *EV5ABP*, *EV8USE*, *EV9POP*, *EV10ADD*.

Table B7

Stage A4 DV_{PRESENCE} Analysis for Best Time-lagged DV Selection, DV4

$DV_{PRESENCE}$	*SASM*	$R^2_{\text{Nagelkerke}}$	*LL Ratio* χ^2
DV4 Model 51	FS (COND)	.161	$\chi^2(4) = 410.104$, $p < .001$
DV4 Model 52	FS (LR)	.161	$\chi^2(4) = 410.104$, $p < .001$
DV4 Model 53	FS (WALD)	.161	$\chi^2(4) = 410.104$, $p < .001$
DV4 Model 54	BE (COND)	.162	$\chi^2(5) = 412.961$, $p < .001$
DV4 Model 55	BE (LR)	.161	$\chi^2(5) = 412.961$, $p < .001$
DV4 Model 56	BE (WALD)	.161	$\chi^2(5) = 412.961$, $p < .001$

a. Model 51 Predictors: (Constant), *EV1ATM*, *EV8USE*, *EV9POP*, *EV10ADD*.
b. Model 52 Predictors: (Constant), *EV1ATM*, *EV8USE*, *EV9POP*, *EV10ADD*.
c. Model 53 Predictors: (Constant), *EV1ATM*, *EV8USE*, *EV9POP*, *EV10ADD*.
d. Model 54 Predictors: (Constant), *EV1ATM*, *EV7ELV*, *EV8USE*, *EV9POP*, *EV10ADD*.
e. Model 55 Predictors: (Constant), *EV1ATM*, *EV7ELV*, *EV8USE*, *EV9POP*, *EV10ADD*.
f. Model 56 Predictors: (Constant), *EV1ATM*, *EV7ELV*, *EV8USE*, *EV9POP*, *EV10ADD*.

Stage A5 $DV_{PRESENCE}$ Analysis (SFEM)

Table B8

Stage A5 $DV_{PRESENCE}$ Analysis with DV2 and EVs

$DV_{PRESENCE}$	$R^2_{\text{Nagelkerke}}$	*LL Ratio* χ^2
DV2 Model 57	.257	$\chi^2(9) = 667.271$, $p < .001$
DV2 Model 58	.257	$\chi^2(8) = 666.817$, $p < .001$
DV2 Model 59	.257	$\chi^2(7) = 665.685$, $p < .001$

a. Model 57 Predictors: (Constant), *EV1ATM*, *EV2ARN*, *EV4ASD*, *EV5ABP*, *EV6AWS*, *EV7ELV*, *EV8USE*, *EV9POP*, *EV10ADD*. *EV4ASD* not significant at $p = .990$.
b. Model 58 Predictors: (Constant), *EV1ATM*, *EV2ARN*, *EV5ABP*, *EV6AWS*, *EV7ELV*, *EV8USE*, *EV9POP*, *EV10ADD*. *EV7ELV* not significant at $p = .990$.
c. Model 59 Predictors: (Constant), *EV1ATM*, *EV2ARN*, *EV5ABP*, *EV6AWS*, *EV8USE*, *EV9POP*, *EV10ADD*. All variables significant ($p < \alpha = .20$).

Table B9

Stage A5 DV$_{PRESENCE}$ *Analysis with DV2, EVs, and 2FIs*

Model	$R^2_{\text{Nagelkerke}}$	*LL Ratio* χ^2
DV2 Model 60	.275	$\chi^2(27) = 713.876$, $p < .001$
DV2 Model 61	.275	$\chi^2(26) = 713.875$, $p < .001$
DV2 Model 62	.275	$\chi^2(25) = 713.873$, $p < .001$
DV2 Model 63	.275	$\chi^2(24) = 713.862$, $p < .001$
DV2 Model 64	.275	$\chi^2(23) = 713.839$, $p < .001$
DV2 Model 65	.275	$\chi^2(22) = 713.714$, $p < .001$
DV2 Model 66	.275	$\chi^2(21) = 713.574$, $p < .001$
DV2 Model 67	.275	$\chi^2(20) = 713.352$, $p < .001$
DV2 Model 68	.274	$\chi^2(19) = 713.309$, $p < .001$
DV2 Model 69	.274	$\chi^2(18) = 712.878$, $p < .001$
DV2 Model 70	.274	$\chi^2(17) = 712.379$, $p < .001$
DV2 Model 71	.274	$\chi^2(16) = 711.688$, $p < .001$
DV2 Model 72	.273	$\chi^2(15) = 710.292$, $p < .001$
DV2 Model 73	.273	$\chi^2(14) = 710.292$, $p < .001$
DV2 Model 74	.273	$\chi^2(13) = 708.378$, $p < .001$

a. Model 60 Predictors, (Constant), *EV1ATM*, *EV2ARN*, *EV5ABP*, *EV6AWS*, *EV8USE*, *EV9POP*, *EV10ADD*, *EV1·EV2*, *EV1·EV5*, *EV1·EV6*, *EV1·EV8*, *EV1·EV9*, *EV1·EV10*, *EV2·EV5*, *EV2·EV6*, *EV2·EV8*, *EV2·EV9*, *EV2·EV10*, *EV5·EV6*, *EV5·EV8*, *EV5·EV9*, *EV5·EV10*, *EV6·EV8*, *EV6·EV9*, *EV6·EV10*, *EV8·EV9*, *EV8·EV10*, *EV9·EV10*. *EV2ARN* not significant at $p = .979$.

b. Model 61 Predictors, (Constant), *EV1ATM*, *EV5ABP*, *EV6AWS*, *EV8USE*, *EV9POP*, *EV10ADD*, *EV1·EV2*, *EV1·EV5*, *EV1·EV6*, *EV1·EV8*, *EV1·EV9*, *EV1·EV10*, *EV2·EV5*, *EV2·EV6*, *EV2·EV8*, *EV2·EV9*, *EV2·EV10*, *EV5·EV6*, *EV5·EV8*, *EV5·EV9*, *EV5·EV10*, *EV6·EV8*, *EV6·EV9*, *EV6·EV10*, *EV8·EV9*, *EV9·EV10*. *EV5ABP* not significant at $p = .919$.

c. Model 62 Predictors, (Constant), *EV1ATM*, *EV6AWS*, *EV8USE*, *EV9POP*, *EV10ADD*, *EV1·EV2*, *EV1·EV5*, *EV1·EV6*, *EV1·EV8*, *EV1·EV9*, *EV1·EV10*, *EV2·EV5*, *EV2·EV6*,

EV2·EV8, *EV2·EV9*, *EV2·EV10*, *EV5·EV6*, *EV5·EV8*, *EV5·EV9*, *EV5·EV10*, *EV6·EV8*, *EV6·EV9*, *EV6·EV10*, *EV8·EV9*, *EV9·EV10*. *EV2·EV8* not significant at p = .920.

d. Model 63 Predictors, (Constant), *EV1ATM*, *EV6AWS*, *EV8USE*, *EV9POP*, *EV10ADD*, *EV1·EV2*, *EV1·EV5*, *EV1·EV6*, *EV1·EV8*, *EV1·EV9*, *EV1·EV10*, *EV2·EV5*, *EV2·EV6*, *EV2·EV9*, *EV2·EV10*, *EV5·EV6*, *EV5·EV8*, *EV5·EV9*, *EV5·EV10*, *EV6·EV8*, *EV6·EV9*, *EV6·EV10*, *EV8·EV9*, *EV9·EV10*. *EV1·EV9* not significant at p = .879.

e. Model 64 Predictors, (Constant), *EV1ATM*, *EV6AWS*, *EV8USE*, *EV9POP*, *EV10ADD*, *EV1·EV2*, *EV1·EV5*, *EV1·EV6*, *EV1·EV8*, *EV1·EV10*, *EV2·EV5*, *EV2·EV6*, *EV2·EV9*, *EV2·EV10*, *EV5·EV6*, *EV5·EV8*, *EV5·EV9*, *EV5·EV10*, *EV6·EV8*, *EV6·EV9*, *EV6·EV10*, *EV8·EV9*, *EV9·EV10*. *EV2·EV9* not significant at p = .723.

f. Model 65 Predictors, (Constant), *EV1ATM*, *EV6AWS*, *EV8USE*, *EV9POP*, *EV10ADD*, *EV1·EV2*, *EV1·EV5*, *EV1·EV6*, *EV1·EV8*, *EV1·EV10*, *EV2·EV5*, *EV2·EV6*, *EV2·EV10*, *EV5·EV6*, *EV5·EV8*, *EV5·EV9*, *EV5·EV10*, *EV6·EV8*, *EV6·EV9*, *EV6·EV10*, *EV8·EV9*, *EV9·EV10*. *EV8USE* not significant at p = .707.

g. Model 66 Predictors, (Constant), *EV1ATM*, *EV6AWS*, *EV9POP*, *EV10ADD*, *EV1·EV2*, *EV1·EV5*, *EV1·EV6*, *EV1·EV8*, *EV1·EV10*, *EV2·EV5*, *EV2·EV6*, *EV2·EV10*, *EV5·EV6*, *EV5·EV8*, *EV5·EV9*, *EV5·EV10*, *EV6·EV8*, *EV6·EV9*, *EV6·EV10*, *EV8·EV9*, *EV9·EV10*. *EV2·EV5* not significant at p = .641.

h. Model 67 Predictors, (Constant), *EV1ATM*, *EV6AWS*, *EV9POP*, *EV10ADD*, *EV1·EV2*, *EV1·EV5*, *EV1·EV6*, *EV1·EV8*, *EV1·EV10*, *EV2·EV6*, *EV2·EV10*, *EV5·EV6*, *EV5·EV8*, *EV5·EV9*, *EV5·EV10*, *EV6·EV8*, *EV6·EV9*, *EV6·EV10*, *EV8·EV9*, *EV9·EV10*. *EV1·EV2* not significant at p = .838.

i. Model 68 Predictors, (Constant), *EV1ATM*, *EV6AWS*, *EV9POP*, *EV10ADD*, *EV1·EV5*, *EV1·EV6*, *EV1·EV8*, *EV1·EV10*, *EV2·EV6*, *EV2·EV10*, *EV5·EV6*, *EV5·EV8*, *EV5·EV9*, *EV5·EV10*, *EV6·EV8*, *EV6·EV9*, *EV6·EV10*, *EV8·EV9*, *EV9·EV10*. *EV1·EV8* not significant at p = .508.

j. Model 69 Predictors, (Constant), *EV1ATM*, *EV6AWS*, *EV9POP*, *EV10ADD*, *EV1·EV5*, *EV1·EV6*, *EV1·EV10*, *EV2·EV6*, *EV2·EV10*, *EV5·EV6*, *EV5·EV8*, *EV5·EV9*, *EV5·EV10*, *EV6·EV8*, *EV6·EV9*, *EV6·EV10*, *EV8·EV9*, *EV9·EV10*. *EV1·EV10* not significant at p = .473.

k. Model 70 Predictors, (Constant), *EV1ATM*, *EV6AWS*, *EV9POP*, *EV10ADD*, *EV1·EV5*, *EV1·EV6*, *EV2·EV6*, *EV2·EV10*, *EV5·EV6*, *EV5·EV8*, *EV5·EV9*, *EV5·EV10*, *EV6·EV8*, *EV6·EV9*, *EV6·EV10*, *EV8·EV9*, *EV9·EV10*. *EV1·EV6* not significant at p = .397.

l. Model 71 Predictors, (Constant), *EV1ATM*, *EV6AWS*, *EV9POP*, *EV10ADD*, *EV1·EV5*, *EV2·EV6*, *EV2·EV10*, *EV5·EV6*, *EV5·EV8*, *EV5·EV9*, *EV5·EV10*, *EV6·EV8*, *EV6·EV9*, *EV6·EV10*, *EV8·EV9*, *EV9·EV10*. *EV2·EV10* not significant at p = .253.

m. Model 72 Predictors, (Constant), *EV1ATM*, *EV6AWS*, *EV9POP*, *EV10ADD*, *EV1·EV5*, *EV2·EV6*, *EV5·EV6*, *EV5·EV8*, *EV5·EV9*, *EV5·EV10*, *EV6·EV8*, *EV6·EV9*, *EV6·EV10*, *EV8·EV9*, *EV9·EV10*. *IV10ADD* not significant at $p = .203$.
n. Model 73 Predictors, (Constant), *EV1ATM*, *EV6AWS*, *EV9POP*, *EV1·EV5*, *EV2·EV6*, *EV5·EV6*, *EV5·EV8*, *EV5·EV9*, *EV5·EV10*, *EV6·EV8*, *EV6·EV9*, *EV6·EV10*, *EV8·EV9*, *EV9·EV10*. *IV5·IV10* not significant at $p = .464$.
o. Model 74 Predictors, (Constant), *EV1ATM*, *EV6AWS*, *EV9POP*, *EV1·EV5*, *EV2·EV6*, *EV5·EV6*, *EV5·EV8*, *EV5·EV9*, *EV6·EV8*, *EV6·EV9*, *EV6·EV10*, *EV8·EV9*, *EV9·EV10*. All predictors significant at $p < .20$.

Table B10

Stage A5 $DV_{PRESENCE}$ Analysis With DV2, All EVs, $\alpha = .20$

Model	$R^2_{\text{Nagelkerke}}$	*LL Ratio* χ^2
DV2 Model 75	.293	$\chi^2(31) = 763.626$, $p < .001$
DV2 Model 76	.293	$\chi^2(31) = 763.626$, $p < .001$
DV2 Model 77	.293	$\chi^2(30) = 763.626$, $p < .001$
DV2 Model 78	.293	$\chi^2(29) = 763.622$, $p < .001$
DV2 Model 79	.293	$\chi^2(28) = 763.608$, $p < .001$
DV2 Model 80	.293	$\chi^2(27) = 763.454$, $p < .001$
DV2 Model 81	.293	$\chi^2(26) = 763.271$, $p < .001$
DV2 Model 82	.293	$\chi^2(25) = 763.095$, $p < .001$
DV2 Model 83	.293	$\chi^2(24) = 762.793$, $p < .001$
DV2 Model 84	.293	$\chi^2(23) = 762.344$, $p < .001$
DV2 Model 85	.292	$\chi^2(22) = 761.838$, $p < .001$
DV2 Model 86	.292	$\chi^2(21) = 761.838$, $p < .001$
DV2 Model 87	.292	$\chi^2(20) = 760.093$, $p < .001$
DV2 Model 88	.291	$\chi^2(19) = 760.093$, $p < .001$
DV2 Model 89	.291	$\chi^2(18) = 758.244$, $p < .001$
DV2 Model 90	.291	$\chi^2(17) = 758.244$, $p < .001$

a. Model 75 Predictors, (Constant), *EV1ATM*, *EV6AWS*, *EV9POP*, *EV10ADD*, *EV1·EV5*, *EV1·EV6*, *EV1·EV8*, *EV1·EV10*, *EV2·EV6*, *EV2·EV10*, *EV5·EV6*, *EV5·EV8*, *EV5·EV9*, *EV5·EV10*, *EV6·EV8*, *EV6·EV9*, *EV6·EV10*, *EV8·EV9*, *EV9·EV10*, *January*, *February*, *March*, *April*, *May*, *June*, *July*, *August*, *September*, *October*, *November*, *December*. *July* not significant at p = .985.
b. Model 76 Predictors, (Constant), *EV1ATM*, *EV6AWS*, *EV9POP*, *EV10ADD*, *EV1·EV5*, *EV1·EV6*, *EV1·EV8*, *EV1·EV10*, *EV2·EV6*, *EV2·EV10*, *EV5·EV6*, *EV5·EV8*, *EV5·EV9*, *EV5·EV10*, *EV6·EV8*, *EV6·EV9*, *EV6·EV10*, *EV8·EV9*, *EV9·EV10*, *January*, *February*, *March*, *April*, *May*, *June*, *August*, *September*, *October*, *November*, *December*. *December* not significant at p = .985.
c. Model 77 Predictors, (Constant), *EV1ATM*, *EV6AWS*, *EV9POP*, *EV10ADD*, *EV1·EV5*, *EV1·EV6*, *EV1·EV8*, *EV1·EV10*, *EV2·EV6*, *EV2·EV10*, *EV5·EV6*, *EV5·EV8*, *EV5·EV9*, *EV5·EV10*, *EV6·EV8*, *EV6·EV9*, *EV6·EV10*, *EV8·EV9*, *EV9·EV10*, *January*, *February*, *March*, *April*, *May*, *June*, *August*, *September*, *October*, *November*. *March* not significant at p = .953.
d. Model 78 Predictors, (Constant), *EV1ATM*, *EV6AWS*, *EV9POP*, *EV10ADD*, *EV1·EV5*, *EV1·EV6*, *EV1·EV8*, *EV1·EV10*, *EV2·EV6*, *EV2·EV10*, *EV5·EV6*, *EV5·EV8*, *EV5·EV9*, *EV5·EV10*, *EV6·EV8*, *EV6·EV9*, *EV6·EV10*, *EV8·EV9*, *EV9·EV10*, *January*, *February*, *April*, *May*, *June*, *August*, *September*, *October*, *November*. *February* not significant at p = .905.
e. Model 79 Predictors, (Constant), *EV1ATM*, *EV6AWS*, *EV9POP*, *EV10ADD*, *EV1·EV5*, *EV1·EV6*, *EV1·EV8*, *EV1·EV10*, *EV2·EV6*, *EV2·EV10*, *EV5·EV6*, *EV5·EV8*, *EV5·EV9*, *EV5·EV10*, *EV6·EV8*, *EV6·EV9*, *EV6·EV10*, *EV8·EV9*, *EV9·EV10*, *January*, *April*, *May*, *June*, *August*, *September*, *October*, *November*. *January* not significant at p = .710.
f. Model 80 Predictors, (Constant), *EV1ATM*, *EV6AWS*, *EV9POP*, *EV10ADD*, *EV1·EV5*, *EV1·EV6*, *EV1·EV8*, *EV1·EV10*, *EV2·EV6*, *EV2·EV10*, *EV5·EV6*, *EV5·EV8*, *EV5·EV9*, *EV5·EV10*, *EV6·EV8*, *EV6·EV9*, *EV6·EV10*, *EV8·EV9*, *EV9·EV10*, *April*, *May*, *June*, *August*, *September*, *October*, *November*. *EV1·EV2* not significant at p = .679.
g. Model 81 Predictors, (Constant), *EV1ATM*, *EV6AWS*, *EV9POP*, *EV10ADD*, *EV1·EV5*, *EV1·EV6*, *EV1·EV8*, *EV1·EV10*, *EV2·EV6*, *EV2·EV10*, *EV5·EV6*, *EV5·EV8*, *EV5·EV9*, *EV5·EV10*, *EV6·EV8*, *EV6·EV9*, *EV6·EV10*, *EV8·EV9*, *EV9·EV10*, *April*, *May*, *June*, *August*, *September*, *October*, *November*. *EV1·EV8* not significant at p = .673.

h. Model 82 Predictors, (Constant), *EV1ATM*, *EV6AWS*, *EV9POP*, *EV10ADD*, *EV1·EV5*, *EV1·EV6*, *EV1·EV10*, *EV2·EV6*, *EV2·EV10*, *EV5·EV6*, *EV5·EV8*, *EV5·EV9*, *EV5·EV10*, *EV6·EV8*, *EV6·EV9*, *EV6·EV10*, *EV8·EV9*, *EV9·EV10*, *April*, *May*, *June*, *August*, *September*, *October*, *November*. *EV10ADD* not significant at p = .572.
i. Model 83 Predictors, (Constant), *EV1ATM*, *EV6AWS*, *EV9POP*, *EV1·EV5*, *EV1·EV6*, *EV1·EV10*, *EV2·EV6*, *EV2·EV10*, *EV5·EV6*, *EV5·EV8*, *EV5·EV9*, *EV5·EV10*, *EV6·EV8*, *EV6·EV9*, *EV6·EV10*, *EV8·EV9*, *EV9·EV10*, *April*, *May*, *June*, *August*, *September*, *October*, *November*. *April* not significant at p = .514.
j. Model 84 Predictors, (Constant), *EV1ATM*, *EV6AWS*, *EV9POP*, *EV1·EV5*, *EV1·EV6*, *EV1·EV10*, *EV2·EV6*, *EV2·EV10*, *EV5·EV6*, *EV5·EV8*, *EV5·EV9*, *EV5·EV10*, *EV6·EV8*, *EV6·EV9*, *EV6·EV10*, *EV8·EV9*, *EV9·EV10*, *May*, *June*, *August*, *September*, *October*, *November*. *November* not significant at p = .491.
k. Model 85 Predictors, (Constant), *EV1ATM*, *EV6AWS*, *EV9POP*, *EV1·EV5*, *EV1·EV6*, *EV1·EV10*, *EV2·EV6*, *EV2·EV10*, *EV5·EV6*, *EV5·EV8*, *EV5·EV9*, *EV5·EV10*, *EV6·EV8*, *EV6·EV9*, *EV6·EV10*, *EV8·EV9*, *EV9·EV10*, *May*, *June*, *August*, *September*, *October*. *EV1·EV10* not significant at p = .443.
l. Model 86 Predictors, (Constant), *EV1ATM*, *EV6AWS*, *EV9POP*, *EV1·EV5*, *EV1·EV6*, *EV2·EV6*, *EV2·EV10*, *EV5·EV6*, *EV5·EV8*, *EV5·EV9*, *EV5·EV10*, *EV6·EV8*, *EV6·EV9*, *EV6·EV10*, *EV8·EV9*, *EV9·EV10*, *May*, *June*, *August*, *September*, *October*. *EV5·EV10* not significant at p = .281.
m. Model 87 Predictors, (Constant), *EV1ATM*, *EV6AWS*, *EV9POP*, *EV1·EV5*, *EV1·EV6*, *EV2·EV6*, *EV2·EV10*, *EV5·EV6*, *EV5·EV8*, *EV5·EV9*, *EV6·EV8*, *EV6·EV9*, *EV6·EV10*, *EV8·EV9*, *EV9·EV10*, *May*, *June*, *August*, *September*, *October*. *September* not significant at p = .309.
n. Model 88 Predictors, (Constant), *EV1ATM*, *EV6AWS*, *EV9POP*, *EV1·EV5*, *EV1·EV6*, *EV2·EV6*, *EV2·EV10*, *EV5·EV6*, *EV5·EV8*, *EV5·EV9*, *EV6·EV8*, *EV6·EV9*, *EV6·EV10*, *EV8·EV9*, *EV9·EV10*, *May*, *June*, *August*, *October*. *August* not significant at p = .363.
o. Model 89 Predictors, (Constant), *EV1ATM*, *EV6AWS*, *EV9POP*, *EV1·EV5*, *EV1·EV6*, *EV2·EV6*, *EV2·EV10*, *EV5·EV6*, *EV5·EV8*, *EV5·EV9*, *EV6·EV8*, *EV6·EV9*, *EV6·EV10*, *EV8·EV9*, *EV9·EV10*, *May*, *June*, *October*. *EV1·EV6* not significant at p = .253.
p. Model 90 Predictors, (Constant), *EV1ATM*, *EV6AWS*, *EV9POP*, *EV1·EV5*, *EV2·EV6*, *EV2·EV10*, *EV5·EV6*, *EV5·EV8*, *EV5·EV9*, *EV6·EV8*, *EV6·EV9*, *EV6·EV10*, *EV8·EV9*, *EV9·EV10*, *May*, *June*, *October*. *All predictors significant at* $p < \alpha = .20$

Stage A6 *DV*$_{PRESENCE}$ Analysis with *DV2*, All EVs, α = .05

Table B11

Stage A6 DV$_{PRESENCE}$ Analysis with DV, All EVs, α = .05

Model	$R^2_{\text{Nagelkerke}}$	*LL Ratio* χ^2
DV2 Model 91	.292	$\chi^2(22) = 761.838$, $p < .001$
DV2 Model 92	.292	$\chi^2(21) = 761.270$, $p < .001$
DV2 Model 93	.292	$\chi^2(20) = 760.093$, $p < .001$
DV2 Model 94	.291	$\chi^2(19) = 759.069$, $p < .001$
DV2 Model 95	.291	$\chi^2(19) = 758.244$, $p < .001$
DV2 Model 96	.291	$\chi^2(17) = 756.960$, $p < .001$
DV2 Model 97	.290	$\chi^2(16) = 755.020$, $p < .001$
DV2 Model 98	.289	$\chi^2(15) = 753.069$, $p < .001$
DV2 Model 99	.287	$\chi^2(15) = 746.698$, $p < .001$
DV2 Model 100	.285	$\chi^2(13) = 750.953$, $p < .001$

a. Model 91 Predictors, (Constant), *EV1ATM*, *EV6AWS*, *EV9POP*, *EV1·EV5*, *EV1·EV6*, *EV1·EV10*, *EV2·EV6*, *EV2·EV10*, *EV5·EV6*, *EV5·EV8*, *EV5·EV9*, *EV5·EV10*, *EV6·EV8*, *EV6·EV9*, *EV6·EV10*, *EV8·EV9*, *EV9·EV10*, *May*, *June*, *August*, *September*, *October*. EV1·EV10 not significant at p = .443.

b. Model 92 Predictors, (Constant), *EV1ATM*, *EV6AWS*, *EV9POP*, *EV1·EV5*, *EV1·EV6*, *EV2·EV6*, *EV2·EV10*, *EV5·EV6*, *EV5·EV8*, *EV5·EV9*, *EV5·EV10*, *EV6·EV8*, *EV6·EV9*, *EV6·EV10*, *EV8·EV9*, *EV9·EV10*, *May*, *June*, *August*, *September*, *October*. *EV5·EV10* not significant at p = .281.

c. Model 93 Predictors, (Constant), *EV1ATM*, *EV6AWS*, *EV9POP*, *EV1·EV5*, *EV1·EV6*, *EV2·EV6*, *EV2·EV10*, *EV5·EV6*, *EV5·EV8*, *EV5·EV9*, *EV6·EV8*, *EV6·EV9*, *EV6·EV10*, *EV8·EV9*, *EV9·EV10*, *May*, *June*, *August*, *September*, *October*. *September* not significant at p = .309.

d. Model 94 Predictors, (Constant), *EV1ATM*, *EV6AWS*, *EV9POP*, *EV1·EV5*, *EV1·EV6*, *EV2·EV6*, *EV2·EV10*, *EV5·EV6*, *EV5·EV8*, *EV5·EV9*, *EV6·EV8*, *EV6·EV9*, *EV6·EV10*, *EV8·EV9*, *EV9·EV10*, *May*, *June*, *August*, *October*. *August* not significant at p = .363.

e. Model 95 Predictors, (Constant), *EV1ATM*, *EV6AWS*, *EV9POP*, *EV1·EV5*, *EV1·EV6*, *EV2·EV6*, *EV2·EV10*, *EV5·EV6*, *EV5·EV8*, *EV5·EV9*, *EV6·EV8*, *EV6·EV9*, *EV6·EV10*, *EV8·EV9*, *EV9·EV10*, *May*, *June*, *October*. *EV1·EV6* not significant at $p = .253$.
f. Model 96 Predictors, (Constant), *EV1ATM*, *EV6AWS*, *EV9POP*, *EV1·EV5*, *EV2·EV6*, *EV2·EV10*, *EV5·EV6*, *EV5·EV8*, *EV5·EV9*, *EV6·EV8*, *EV6·EV9*, *EV6·EV10*, *EV8·EV9*, *EV9·EV10*, *May*, *June*, *October*. *EV2·EV10* not significant at $p = .195$.
g. Model 97 Predictors, (Constant), *EV1ATM*, *EV6AWS*, *EV9POP*, *EV1·EV5*, *EV2·EV6*, *EV5·EV6*, *EV5·EV8*, *EV5·EV9*, *EV6·EV8*, *EV6·EV9*, *EV6·EV10*, *EV8·EV9*, *EV9·EV10*, *May*, *June*, *October*. *EV9·EV10* not significant at $p = .195$.
h. Model 98 Predictors, (Constant), *EV1ATM*, *EV6AWS*, *EV9POP*, *EV1·EV5*, *EV2·EV6*, *EV5·EV6*, *EV5·EV8*, *EV5·EV9*, *EV6·EV8*, *EV6·EV9*, *EV6·EV10*, *EV8·EV9*, *May*, *June*, *October*. *EV2·EV6* not significant at $p = .114$.
i. Model 99 Predictors, (Constant), *EV1ATM*, *EV6AWS*, *EV9POP*, *EV1·EV5*, *EV5·EV6*, *EV5·EV8*, *EV5·EV9*, *EV6·EV8*, *EV6·EV9*, *EV6·EV10*, *EV8·EV9*, *May*, *June*, *October*. *EV6·EV9* not significant at $p = .062$.
j. Model 100 Predictors, (Constant), *EV1ATM*, *EV6AWS*, *EV9POP*, *EV1·EV5*, *EV5·EV6*, *EV5·EV8*, *EV5·EV9*, *EV6·EV8*, *EV6·EV9*, *EV6·EV10*, *EV8·EV9*, *May*, *June*, *October*. All predictors significant at $p < \alpha = .05$.

$DV_{PRESENCE}$ Final Model Selection

Table B12

Comparison of Stages A5-A6 Best Models

Stage	Model	$R^2_{\text{Nagelkerke}}$	χ^2
A5	59	.257	$\chi^2(7) = 665.685$, $p < .001$
A5	67	.275	$\chi^2(20) = 713.352$, $p < .001$
A5	84	.293	$\chi^2(23) = 762.344$, $p < .001$
A6	93	.292	$\chi^2(20) = 760.093$, $p < .001$

Appendix C: Stage B2-B5 Model Building Process

Stage B2 *DV_{COUNT}* Analysis with *DV1*, All EVs, *α* = .20

Table C1

Stage B2 DV_{COUNT} Analysis with DV1 and EVs

DV_{COUNT}	Tweedie MPV	*D*	χ^2_P	*AIC*	*LL2 Ratio χ^2*	*df*	*.sig*
DV1 Model 101	1.5	3.274	4519.157	2608.507	71.917	9	.000
DV1 Model 102	1.5	3.271	4522.391	2606.603	71.820	8	.000
DV1 Model 103	1.5	3.272	4533.620	2606.779	70.645	7	.000

a. Model 101 Predictors: Constant, *EV1ATM*, *EV2ARN*, *EV4ASD*, *EV5ABP*, *EV6AWS*, *EV7ELV*, *EV8USE*, *EV9POP*, *EV10ADD*. *EV6AWS* not significant at $p = .756$.

b. Model 102 Predictors: Constant, *EV1ATM*, *EV2ARN*, *EV4ASD*, *EV5ABP*, *EV7ELV*, *EV8USE*, *EV9POP*, *EV10ADD*. *EV4ASD* not significant at $p = .289$.

c. Model 103 Predictors: Constant, *EV1ATM*, *EV2ARN*, *EV5ABP*, *EV7ELV*, *EV8USE*, *EV9POP*, *EV10ADD*. All predictors significant at $p < \alpha = .20$.

Table C2

Stage B2 DV_{COUNT} Analysis with DV1, EVs, and 2FIs

DV_{COUNT}	Tweedie MPV	*D*	χ^2_P	*AIC*	*LL2* Ratio χ^2	*df*	*.sig*
DV1 Model 104	1.5	3.171	4153.884	2605.794	112.629	28	.000
DV1 Model 105	1.5	3.168	4157.235	2603.806	112.617	27	.000
DV1 Model 106	1.5	3.164	4171.766	2601.839	112.584	26	.000
DV1 Model 107	1.5	3.161	4174.960	2599.881	112.543	25	.000
DV1 Model 108	1.5	3.158	4175.527	2597.934	112.490	24	.000
DV1 Model 109	1.5	3.155	4194.011	2596.090	112.334	23	.000
DV1 Model 110	1.5	3.153	4195.498	2594.384	112.039	22	.000
DV1 Model 111	1.5	3.151	4247.277	2592.761	111.662	21	.000
DV1 Model 112	1.5	3.150	4284.068	2591.261	111.163	20	.000
DV1 Model 113	1.5	3.149	4288.419	2589.736	110.688	19	.000
DV1 Model 114	1.5	3.148	4268.160	2588.434	109.990	18	.000
DV1 Model 115	1.5	3.149	4283.300	2587.449	108.974	17	.000
DV1 Model 116	1.5	3.146	4274.052	2585.601	108.823	16	.000
DV1 Model 117	1.5	3.146	4265.900	2584.217	108.206	15	.000
DV1 Model 118	1.5	3.148	4287.538	2583.711	106.712	14	.000
DV1 Model 119	1.5	3.151	4283.710	2583.312	105.112	13	.000
DV1 Model 120	1.45	2.805	4187.574	2535.911	128.793	15	.000
DV1 Model 121	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	**.000**
DV1 Model 122	1.15	2.442	4087.575	2579.901	216.068	15	.000
DV1 Model 123	1.1	2.421	4083.084	2638.255	253.262	15	.000

a. Model 104 Predictors: Constant, *EV1ATM, EV2ARN, EV5ABP, EV7ELV, EV8USE, EV9POP, EV10ADD, EV1·EV2, EV1·EV5, EV1·EV7, EV1·EV8, EV1·EV9, EV1·EV10, EV2·EV5, EV2·EV7, EV2·EV8, EV2·EV9, EV2·EV10, EV5·EV7, EV5·EV8, EV5·EV9,*

EV5·EV10, EV7·EV8, EV7·EV9, EV7·EV10, EV8·EV9, EV8·EV10, EV9·EV10. EV10ADD not significant at p = .911.

b. Model 105 Predictors: Constant, *EV1ATM*, *EV2ARN*, *EV5ABP, EV7ELV, EV8USE, EV9POP, EV1·EV2, EV1·EV5, EV1·EV7, EV1·EV8, EV1·EV9, EV1·EV10, EV2·EV5, EV2·EV7, EV2·EV8, EV2·EV9, EV2·EV10, EV5·EV7, EV5·EV8, EV5·EV9, EV5·EV10, EV7·EV8, EV7·EV9, EV7·EV10, EV8·EV9, EV8·EV10, EV9·EV10. EV2·EV10* not significant at p = .856.

c. Model 106 Predictors: Constant, *EV1ATM*, *EV2ARN*, *EV5ABP*, *EV7ELV*, *EV8USE, EV9POP, EV1·EV2, EV1·EV5, EV1·EV7, EV1·EV8, EV1·EV9, EV1·EV10, EV2·EV5, EV2·EV7, EV2·EV8, EV2·EV9, EV5·EV7, EV5·EV8, EV5·EV9, EV5·EV10, EV7·EV8, EV7·EV9, EV7·EV10, EV8·EV9, EV8·EV10, EV9·EV10. EV5·EV8* not significant at p = .839.

d. Model 107 Predictors: Constant, *EV1ATM*, *EV2ARN*, *EV5ABP, EV7ELV, EV8USE, EV9POP, EV1·EV2, EV1·EV5, EV1·EV7, EV1·EV8, EV1·EV9, EV1·EV10, EV2·EV5, EV2·EV7, EV2·EV8, EV2·EV9, EV5·EV7, EV5·EV9, EV5·EV10, EV7·EV8, EV7·EV9, EV7·EV10, EV8·EV9, EV8·EV10, EV9·EV10. EV8·EV9* not significant at p = .818.

e. Model 108 Predictors: Constant, *EV1ATM*, *EV2ARN*, *EV5ABP, EV7ELV, EV8USE, EV9POP, EV1·EV2, EV1·EV5, EV1·EV7, EV1·EV8, EV1·EV9, EV1·EV10, EV2·EV5, EV2·EV7, EV2·EV8, EV2·EV9, EV5·EV7, EV5·EV9, EV5·EV10, EV7·EV8, EV7·EV9, EV7·EV10, EV8·EV10, EV9·EV10. EV2·EV8* not significant at p = .693.

f. Model 109 Predictors: Constant, *EV1ATM*, *EV2ARN*, *EV5ABP*, *EV7ELV*, *EV8USE, EV9POP, EV1·EV2, EV1·EV5, EV1·EV7, EV1·EV8, EV1·EV9, EV1·EV10, EV2·EV5, EV2·EV7, EV2·EV9, EV5·EV7, EV5·EV9, EV5·EV10, EV7·EV8, EV7·EV9, EV7·EV10, EV8·EV10, EV9·EV10. EV1·EV10* not significant at p = .588.

g. Model 110 Predictors: Constant, *EV1ATM*, *EV2ARN*, *EV5ABP, EV7ELV, EV8USE, EV9POP, EV1·EV2, EV1·EV5, EV1·EV7, EV1·EV8, EV1·EV9, EV2·EV5, EV2·EV7, EV2·EV9, EV5·EV7, EV5·EV9, EV5·EV10, EV7·EV8, EV7·EV9, EV7·EV10, EV8·EV10, EV9·EV10. EV2·EV7* not significant at p = .540.

h. Model 111 Predictors: Constant, *EV1ATM*, *EV2ARN*, *EV5ABP*, *EV7ELV*, *EV8USE*, *EV9POP*, *EV1·EV2*, *EV1·EV5*, *EV1·EV7*, *EV1·EV8*, *EV1·EV9*, *EV2·EV5*, *EV2·EV9*, *EV5·EV7*, *EV5·EV9*, *EV5·EV10*, *EV7·EV8*, *EV7·EV9*, *EV7·EV10*, *EV8·EV10*, *EV9·EV10*. *EV2ARN* not significant at p = .493.

i. Model 112 Predictors: Constant, *EV1ATM*, *EV5ABP*, *EV7ELV*, *EV8USE*, *EV9POP*, *EV1·EV2*, *EV1·EV5*, *EV1·EV7*, *EV1·EV8*, *EV1·EV9*, *EV2·EV5*, *EV2·EV9*, *EV5·EV7*, *EV5·EV9*, *EV5·EV10*, *EV7·EV8*, *EV7·EV9*, *EV7·EV10*, *EV8·EV10*, *EV9·EV10*. *EV7·EV8* not significant at p = .441.

j. Model 113 Predictors: Constant, *EV1ATM*, *EV5ABP*, *EV7ELV*, *EV8USE*, *EV9POP*, *EV1·EV2*, *EV1·EV5*, *EV1·EV7*, *EV1·EV8*, *EV1·EV9*, *EV2·EV5*, *EV2·EV9*, *EV5·EV7*, *EV5·EV9*, *EV5·EV10*, *EV7·EV9*, *EV7·EV10*, *EV8·EV10*, *EV9·EV10*. *EV7·EV9* not significant at p = .403.

k. Model 114 Predictors: Constant, *EV1ATM*, *EV5ABP*, *EV7ELV*, *EV8USE*, *EV9POP*, *EV1·EV2*, *EV1·EV5*, *EV1·EV7*, *EV1·EV8*, *EV1·EV9*, *EV2·EV5*, *EV2·EV9*, *EV5·EV7*, *EV5·EV9*, *EV5·EV10*, *EV7·EV10*, *EV8·EV10*, *EV9·EV10*. *EV5·EV9* not significant at p = .312.

l. Model 115 Predictors: Constant, *EV1ATM*, *EV5ABP*, *EV7ELV*, *EV8USE*, *EV9POP*, *EV1·EV2*, *EV1·EV5*, *EV1·EV7*, *EV1·EV8*, *EV1·EV9*, *EV2·EV5*, *EV2·EV9*, *EV5·EV7*, *EV5·EV10*, *EV7·EV10*, *EV8·EV10*, *EV9·EV10*. *EV9POP* not significant at p = .697.

m. Model 116 Predictors: Constant, *EV1ATM*, *EV5ABP*, *EV7ELV*, *EV8USE*, *EV1·EV2*, *EV1·EV5*, *EV1·EV7*, *EV1·EV8*, *EV1·EV9*, *EV2·EV5*, *EV2·EV9*, *EV5·EV7*, *EV5·EV10*, *EV7·EV10*, *EV8·EV10*, *EV9·EV10*. *EV5ABP* not significant at p = .433.

n. Model 117 Predictors: Constant, *EV1ATM*, *EV7ELV*, *EV8USE*, *EV1·EV2*, *EV1·EV5*, *EV1·EV7*, *EV1·EV8*, *EV1·EV9*, *EV2·EV5*, *EV2·EV9*, *EV5·EV7*, *EV5·EV10*, *EV7·EV10*, *EV8·EV10*, *EV9·EV10*. *EV1·EV7* not significant at p = .220.

o. Model 118 Predictors: Constant, *EV1ATM*, *EV7ELV*, *EV8USE*, *EV1·EV2*, *EV1·EV5*, *EV1·EV8*, *EV1·EV9*, *EV2·EV5*, *EV2·EV9*, *EV5·EV7*, *EV5·EV10*, *EV7·EV10*, *EV8·EV10*, *EV9·EV10*. *EV7ELV* not significant at p = .209.

p. Model 119 Predictors: Constant, *EV1ATM*, *EV8USE*, *EV1·EV2*, *EV1·EV5*, *EV1·EV8*, *EV1·EV9*, *EV2·EV5*, *EV2·EV9*, *EV5·EV7*, *EV5·EV10*, *EV7·EV10*, *EV8·EV10*, *EV9·EV10.* All predictors significant at $p < \alpha = .20$.

q. Model 120 Predictors: Constant, *EV1ATM*, *EV7ELV*, *EV8USE*, *EV1·EV2*, *EV1·EV5*, *EV1·EV7*, *EV1·EV8*, *EV1·EV9*, *EV2·EV5*, *EV2·EV9*, *EV5·EV7*, *EV5·EV10*, *EV7·EV10*, *EV8·EV10*, *EV9·EV10.* All predictors significant at $p < \alpha = .20$.

r. Model 121 Predictors: Constant, *EV1ATM*, *EV7ELV*, *EV8USE*, *EV1·EV2*, *EV1·EV5*, *EV1·EV7*, *EV1·EV8*, *EV1·EV9*, *EV2·EV5*, *EV2·EV9*, *EV5·EV7*, *EV5·EV10*, *EV7·EV10*, *EV8·EV10*, *EV9·EV10.* All predictors significant at $p < \alpha = .20$.

s. Model 122 Predictors: Constant, *EV1ATM*, *EV7ELV*, *EV8USE*, *EV1·EV2*, *EV1·EV5*, *EV1·EV7*, *EV1·EV8*, *EV1·EV9*, *EV2·EV5*, *EV2·EV9*, *EV5·EV7*, *EV5·EV10*, *EV7·EV10*, *EV8·EV10*, *EV9·EV10.* All predictors significant at $p < \alpha = .20$.

t. Model 123 Predictors: Constant, *EV1ATM*, *EV7ELV*, *EV8USE*, *EV1·EV2*, *EV1·EV5*, *EV1·EV7*, *EV1·EV8*, *EV1·EV9*, *EV2·EV5*, *EV2·EV9*, *EV5·EV7*, *EV5·EV10*, *EV7·EV10*, *EV8·EV10*, *EV9·EV10.* All predictors significant at $p < \alpha = .20$.

Table C3

Stage B2 DV_{COUNT} Analysis with DV1, All EVs, α = .20

DV_{COUNT}	Tweedie MPV	*D*	χ^2_P	*AIC*	*LL2 Ratio* χ^2	*df*	*.sig*
DV1 Model 124	1.1	1.902	3060.265	2328.891	584.626	26	.000
DV1 Model 125	1.1	1.902	3060.265	2328.891	584.626	26	.000
DV1 Model 126	1.1	1.900	3060.049	2326.895	584.623	25	.000
DV1 Model 127	1.1	1.899	3057.706	2325.309	584.209	24	.000
DV1 Model 128	1.1	1.899	3081.439	2324.941	582.576	23	.000
DV1 Model 129	1.1	1.898	3081.895	2323.209	582.309	22	.000

a. Model 124 Predictors: Constant, *EV1ATM*, *EV7ELV*, *EV8USE*, *EV1·EV2*, *EV1·EV5*, *EV1·EV7*, *EV1·EV8*, *EV1·EV9*, *EV2·EV5*, *EV2·EV9*, *EV5·EV7*, *EV5·EV10*, *EV7·EV10*, *EV8·EV10*, *EV9·EV10*, *January*, *February*, *March*, *April*, *May*, *June*, *July*, *August*, *September*, *October*, *November*, *December*. *December* not significant at *p* = .951.

b. Model 125 Predictors: Constant, *EV1ATM*, *EV7ELV*, *EV8USE*, *EV1·EV2*, *EV1·EV5*, *EV1·EV7*, *EV1·EV8*, *EV1·EV9*, *EV2·EV5*, *EV2·EV9*, *EV5·EV7*, *EV5·EV10*, *EV7·EV10*, *EV8·EV10*, *EV9·EV10*, *January*, *February*, *March*, *April*, *May*, *June*, *July*, *August*, *September*, *October*, *November*. *July* not significant at *p* = .951.

c. Model 126 Predictors: Constant, *EV1ATM*, *EV7ELV*, *EV8USE*, *EV1·EV2*, *EV1·EV5*, *EV1·EV7*, *EV1·EV8*, *EV1·EV9*, *EV2·EV5*, *EV2·EV9*, *EV5·EV7*, *EV5·EV10*, *EV7·EV10*, *EV8·EV10*, *EV9·EV10*, *January*, *February*, *March*, *April*, *May*, *June*, *August*, *September*, *October*, *November*. *January* not significant at *p* = .527.

d. Model 127 Predictors: Constant, *EV1ATM*, *EV7ELV*, *EV8USE*, *EV1·EV2*, *EV1·EV5*, *EV1·EV7*, *EV1·EV8*, *EV1·EV9*, *EV2·EV5*, *EV2·EV9*, *EV5·EV7*, *EV5·EV10*, *EV7·EV10*, *EV8·EV10*, *EV9·EV10*, *February*, *March*, *April*, *May*, *June*, *August*, *September*, *October*, *November*. *EV1·EV5* not significant at *p* = .204.

e. Model 128 Predictors: Constant, *EV1ATM*, *EV7ELV*, *EV8USE*, *EV1·EV2*, *EV1·EV7*, *EV1·EV8*, *EV1·EV9*, *EV2·EV5*, *EV2·EV9*, *EV5·EV7*, *EV5·EV10*, *EV7·EV10*, *EV8·EV10*, *EV9·EV10*, *February*, *March*, *April*, *May*, *June*, *August*, *September*, *October*, *November*. *EV5·EV7* not significant at $p = .590$.

f. Model 129 Predictors: Constant, *EV1ATM*, *EV7ELV*, *EV8USE*, *EV1·EV2*, *EV1·EV7*, *EV1·EV8*, *EV1·EV9*, *EV2·EV5*, *EV2·EV9*, *EV5·EV10*, *EV7·EV10*, *EV8·EV10*, *EV9·EV10*, *February*, *March*, *April*, *May*, *June*, *August*, *September*, *October*, *November.* All EEVs significant at $p < \alpha = .20$.

Stage B3 *DV*$_{COUNT}$ Analysis with *DV1*, All EVs, α = .05

Table C4

Stage B3 DV$_{COUNT}$ Analysis with DV1, All EVs, α = .05

DV$_{COUNT}$	Tweedie MPV	*D*	χ^2_P	*AIC*	*LL2 Ratio* χ^2	*df*	*.sig*
DV1 Model 130	1.1	1.898	3081.895	2323.209	582.309	22	.000
DV1 Model 131	1.1	1.899	3062.322	2323.304	580.213	21	.000
DV1 Model 132	1.1	1.898	3062.937	2022.018	579.499	20	.000
DV1 Model 133	1.1	1.901	3067.056	2322.826	576.692	19	.000
DV1 Model 134	1.1	1.903	3061.548	2323.877	573.648	18	.000
DV1 Model 135	1.1	1.903	3077.320	2323.045	572.472	17	.000

a. Model 130 Predictors: Constant, *EV1ATM*, *EV7ELV*, *EV8USE*, *EV1·EV2*, *EV1·EV7*, *EV1·EV8*, *EV1·EV9*, *EV2·EV5*, *EV2·EV8*, *EV2·EV9*, *EV5·EV10*, *EV7·EV10*, *EV8·EV10*, *EV9·EV10*, *February*, *March*, *April*, *May*, *June*, *August*, *September*, *October*, *November*. *EV9·EV10* not significant at *p* = .150.

b. Model 131 Predictors: Constant, *EV1ATM*, *EV7ELV*, *EV8USE*, *EV1·EV2*, *EV1·EV7*, *EV1·EV8*, *EV1·EV9*, *EV2·EV5*, *EV2·EV8*, *EV2·EV9*, *EV5·EV10*, *EV7·EV10*, *EV8·EV10*, *February*, *March*, *April*, *May*, *June*, *August*, *September*, *October*, *November*. *EV1·EV9* not significant at *p* = .397.

c. Model 132 Predictors: Constant, *EV1ATM*, *EV7ELV*, *EV8USE*, *EV1·EV2*, *EV1·EV7*, *EV1·EV8*, *EV2·EV5*, *EV2·EV8*, *EV2·EV9*, *EV5·EV10*, *EV7·EV10*, *EV8·EV10*, *February*, *March*, *April*, *May*, *June*, *August*, *September*, *October*, *November*. *March* not significant at *p* = .142.

d. Model 133 Predictors: Constant, *EV1ATM*, *EV7ELV*, *EV8USE*, *EV1·EV2*, *EV1·EV7*, *EV1·EV8*, *EV2·EV5*, *EV2·EV8*, *EV2·EV9*, *EV5·EV10*, *EV7·EV10*, *EV8·EV10*, *February*, *April*, *May*, *June*, *August*, *September*, *October*, *November*. *February* not significant at *p* = .146.

e. Model 134 Predictors: Constant, *EV1ATM, EV7ELV, EV8USE, EV1·EV2, EV1·EV7, EV1·EV8, EV2·EV5, EV2·EV8, EV2·EV9, EV5·EV10, EV7·EV10, EV8·EV10, April*, *May*, *June*, *August*, *September*, *October*, *November*. *EV1ATM* not significant at p = .271.

f. Model 135 Predictors: Constant, *EV7ELV, EV8USE, EV1·EV2, EV1·EV7, EV1·EV8, EV2·EV5, EV2·EV8, EV2·EV9, EV5·EV10, EV7·EV10, EV8·EV10, April*, *May*, *June*, *August*, *September*, *October*, *November*. All predictors significant at p = .146.

Stage B4 DV_{COUNT} Analysis for Best Time-lagged DV Selection

Table C5

Stage B4 DV_{COUNT} Time-Lagged Model Results Using EEVs Only

Model	DV	Tweedie MPV	*D*	χ^2_P	*AIC*	*LL2 Ratio* χ^2	*df*	*.sig*
141	*DV2*	1.5	2.825	3824.156	2426.375	190.984	7	.000
145	*DV3*	1.5	2.952	4226.178	2527.626	145.55	4	.000
147	*DV4*	1.5	3.104	4559.273	2569.364	112.038	8	.000
152	*DV5*	1.5	1.023	1731.274	2471.097	424.903	8	.000

Stage B5 *DV*$_{COUNT}$ Analysis with *DV5*, All EVs, α = .20

Table C6

Stage B5 DV$_{COUNT}$ Analysis with DV5, EEVs, and 2FIs

Model	Tweedie MPV	*D*	χ^2_P	*AIC*	*LL2* *Ratio* χ^2	*df*	*.sig*
153	1.5	0.981	1555.298	2433.181	504.819	28	.000
154	1.5	0.980	1554.768	2431.187	504.813	27	.000
155	1.5	0.979	1554.717	2429.206	504.794	26	.000
156	1.5	0.978	1555.581	2427.324	504.676	25	.000
157	1.5	0.977	1561.365	2425.522	504.478	24	.000
158	1.5	0.976	1563.805	2423.921	504.079	23	.000
159	1.5	0.975	1563.087	2422.156	503.843	22	.000
160	1.5	0.975	1563.464	2420.635	503.365	21	.000
161	1.5	0.974	1559.727	2419.144	502.856	20	.000
162	1.5	0.974	1565.207	2417.839	502.161	19	.000
163	1.5	0.973	1570.098	2416.673	501.327	18	.000
164	1.5	0.973	1575.309	2416.107	499.893	17	.000
165	1.5	0.973	1571.462	2414.268	499.732	16	.000
166	1.45	0.995	1598.914	2475.274	513.636	16	.000
167	1.40	1.019	1627.926	2539.175	531.239	16	.000
168	1.43	1.006	1613.221	2506.868	521.876	16	.000
169	1.44	0.999	1604.589	2487.825	516.812	16	.000

a. Model 153 Predictors: Constant, *EV1ATM*, *EV2ARN*, *EV6AWS*, *EV7ELV*, *EV8USE*, *EV9POP*, *EV10ADD*, *EV1ATM·EV2ARN*, *EV1ATM·EV6AWS*, *EV1ATM·EV7ELV*, *EV1ATM·EV8USE*, *EV1ATM·EV9POP*, *EV1ATM·EV10ADD*, *EV2ARN·EV6AWS*, *EV2ARN·EV7ELV*, *EV2ARN·EV8USE*, *EV2ARN·EV9POP*, *EV2ARN·EV10ADD*, *EV6AWS·EV7ELV*, *EV6AWS·EV8USE*, *EV6AWS·EV9POP*, *EV6AWS·EV10ADD*,

EV7ELV·EV8USE, *EV7ELV·EV9POP*, *EV7ELV·EV10ADD*, *EV8USE·EV9POP*, *EV8USE·EV10ADD*, *EV9POP·EV10ADD*. *EV6AWS* not significant at p = .941.

b. Model 154 Predictors: Constant, *EV1ATM*, *EV2ARN*, *EV7ELV*, *EV8USE*, *EV9POP*, *EV10ADD*, *EV1ATM·EV2ARN*, *EV1ATM·EV6AWS*, *EV1ATM·EV7ELV*, *EV1ATM·EV8USE*, *EV1ATM·EV9POP*, *EV1ATM·EV10ADD*, *EV2ARN·EV6AWS*, *EV2ARN·EV7ELV*, *EV2ARN·EV8USE*, *EV2ARN·EV9POP*, *EV2ARN·EV10ADD*, *EV6AWS·EV7ELV*, *EV6AWS·EV8USE*, *EV6AWS·EV9POP*, *EV6AWS·EV10ADD*, *EV7ELV·EV8USE*, *EV7ELV·EV9POP*, *EV7ELV·EV10ADD*, *EV8USE·EV9POP*, *EV8USE·EV10ADD*, *EV9POP·EV10ADD*. *EV2ARN·EV6AWS* not significant at p = .889.

c. Model 155 Predictors: *EV1ATM*, *EV2ARN*, *EV7ELV*, *EV8USE*, *EV9POP*, *EV10ADD*, *EV1ATM·EV2ARN*, *EV1ATM·EV6AWS*, *EV1ATM·EV7ELV*, *EV1ATM·EV8USE*, *EV1ATM·EV9POP*, *EV1ATM·EV10ADD*, *EV2ARN·EV7ELV*, *EV2ARN·EV8USE*, *EV2ARN·EV9POP*, *EV2ARN·EV10ADD*, *EV6AWS·EV7ELV*, *EV6AWS·EV8USE*, *EV6AWS·EV9POP*, *EV6AWS·EV10ADD*, *EV7ELV·EV8USE*, *EV7ELV·EV9POP*, *EV7ELV·EV10ADD*, *EV8USE·EV9POP*, *EV8USE·EV10ADD*, *EV9POP·EV10ADD*. *EV2ARN·EV10ADD* not significant at p = .732.

d. Model 156 Predictors: Constant, *EV1ATM*, *EV2ARN*, *EV7ELV*, *EV8USE*, *EV9POP*, *EV10ADD*, *EV1ATM·EV2ARN*, *EV1ATM·EV6AWS*, *EV1ATM·EV7ELV*, *EV1ATM·EV8USE*, *EV1ATM·EV9POP*, *EV1ATM·EV10ADD*, *EV2ARN·EV7ELV*, *EV2ARN·EV8USE*, *EV2ARN·EV9POP*, *EV6AWS·EV7ELV*, *EV6AWS·EV8USE*, *EV6AWS·EV9POP*, *EV6AWS·EV10ADD*, *EV7ELV·EV8USE*, *EV7ELV·EV9POP*, *EV7ELV·EV10ADD*, *EV8USE·EV9POP*, *EV8USE·EV10ADD*, *EV9POP·EV10ADD*. *EV6AWS·EV9POP* not significant at p = .656.

e. Model 157 Predictors: Constant, *EV1ATM*, *EV2ARN*, *EV7ELV*, *EV8USE*, *EV9POP*, *EV10ADD*, *EV1ATM·EV2ARN*, *EV1ATM·EV6AWS*, *EV1ATM·EV7ELV*, *EV1ATM·EV8USE*, *EV1ATM·EV9POP*, *EV1ATM·EV10ADD*, *EV2ARN·EV7ELV*, *EV2ARN·EV8USE*, *EV2ARN·EV9POP*, *EV6AWS·EV7ELV*, *EV6AWS·EV8USE*,

EV6AWS·EV10ADD, *EV7ELV·EV8USE*, *EV7ELV·EV9POP*, *EV7ELV·EV10ADD*, *EV8USE·EV9POP*, *EV8USE·EV10ADD*, *EV9POP·EV10ADD*. *EV7ELV* not significant at p = .528.

f. Model 158 Predictors: Constant, *EV1ATM*, *EV2ARN*, *EV8USE*, *EV9POP*, *EV10ADD*, *EV1ATM·EV2ARN*, *EV1ATM·EV6AWS*, *EV1ATM·EV7ELV*, *EV1ATM·EV8USE*, *EV1ATM·EV9POP*, *EV1ATM·EV10ADD*, *EV2ARN·EV7ELV*, *EV2ARN·EV8USE*, *EV2ARN·EV9POP*, *EV6AWS·EV7ELV*, *EV6AWS·EV8USE*, *EV6AWS·EV10ADD*, *EV7ELV·EV8USE*, *EV7ELV·EV9POP*, *EV7ELV·EV10ADD*, *EV8USE·EV9POP*, *EV8USE·EV10ADD*, *EV9POP·EV10ADD*. *EV1ATM·EV7ELV* not significant at p = .627.

g. Model 159 Predictors: Constant, *EV1ATM*, *EV2ARN*, *EV8USE*, *EV9POP*, *EV10ADD*, *EV1ATM·EV2ARN*, *EV1ATM·EV6AWS*, *EV1ATM·EV8USE*, *EV1ATM·EV9POP*, *EV1ATM·EV10ADD*, *EV2ARN·EV7ELV*, *EV2ARN·EV8USE*, *EV2ARN·EV9POP*, *EV6AWS·EV7ELV*, *EV6AWS·EV8USE*, *EV6AWS·EV10ADD*, *EV7ELV·EV8USE*, *EV7ELV·EV9POP*, *EV7ELV·EV10ADD*, *EV8USE·EV9POP*, *EV8USE·EV10ADD*, *EV9POP·EV10ADD*. *IV6AWS·EV8USE* not significant at p = .489.

h. Model 160 Predictors: Constant, *EV1ATM*, *EV2ARN*, *EV8USE*, *EV9POP*, *EV10ADD*, *EV1ATM·EV2ARN*, *EV1ATM·EV6AWS*, *EV1ATM·EV8USE*, *EV1ATM·EV9POP*, *EV1ATM·EV10ADD*, *EV2ARN·EV7ELV*, *EV2ARN·EV8USE*, *EV2ARN·EV9POP*, *EV6AWS·EV7ELV*, *EV6AWS·EV10ADD*, *EV7ELV·EV8USE*, *EV7ELV·EV9POP*, *EV7ELV·EV10ADD*, *EV8USE·EV9POP*, *EV8USE·EV10ADD*, *EV9POP·EV10ADD*. *EV1ATM·EV9POP* not significant at p = .475.

i. Model 161Predictors: Constant, *EV1ATM*, *EV2ARN*, *EV8USE*, *EV9POP*, *EV10ADD*, *EV1ATM·EV2ARN*, *EV1ATM·EV6AWS*, *EV1ATM·EV8USE*, *EV1ATM·EV10ADD*, *EV2ARN·EV7ELV*, *EV2ARN·EV8USE*, *EV2ARN·EV9POP*, *EV6AWS·EV7ELV*, *EV6AWS·EV10ADD*, *EV7ELV·EV8USE*, *EV7ELV·EV9POP*, *EV7ELV·EV10ADD*,

EV8USE·EV9POP, *EV8USE·EV10ADD*, *EV9POP·EV10ADD*. *EV7ELV·EV10ADD* not significant at p = .404.

j. Model 162 Predictors: Constant, *EV1ATM*, *EV2ARN*, *EV8USE*, *EV9POP*, *EV10ADD*, *EV1ATM·EV2ARN*, *EV1ATM·EV6AWS*, *EV1ATM·EV8USE*, *EV1ATM·EV10ADD*, *EV2ARN·EV7ELV*, *EV2ARN·EV8USE*, *EV2ARN·EV9POP*, *EV6AWS·EV7ELV*, *EV6AWS·EV10ADD*, *EV7ELV·EV8USE*, *EV7ELV·EV9POP*, *EV8USE·EV9POP*, *EV8USE·EV10ADD*, *EV9POP·EV10ADD*. *EV2ARN·EV8USE* not significant at p = .361.

k. Model 163 Predictors: Constant, *EV1ATM*, *EV2ARN*, *EV8USE*, *EV9POP*, *EV10ADD*, *EV1ATM·EV2ARN*, *EV1ATM·EV6AWS*, *EV1ATM·EV8USE*, *EV1ATM·EV10ADD*, *EV2ARN·EV7ELV*, *EV2ARN·EV8USE*, *EV2ARN·EV9POP*, *EV6AWS·EV7ELV*, *EV6AWS·EV10ADD*, *EV7ELV·EV9POP*, *EV8USE·EV9POP*, *EV8USE·EV10ADD*, *EV9POP·EV10ADD*. *EV8USE·EV9POP* not significant at p = .231.

l. Model 164 Predictors: Constant, *EV1ATM*, *EV2ARN*, *EV8USE*, *EV9POP*, *EV10ADD*, *EV1ATM·EV2ARN*, *EV1ATM·EV6AWS*, *EV1ATM·EV8USE*, *EV1ATM·EV10ADD*, *EV2ARN·EV7ELV*, *EV2ARN·EV8USE*, *EV2ARN·EV9POP*, *EV6AWS·EV7ELV*, *EV6AWS·EV10ADD*, *EV7ELV·EV9POP*, *EV8USE·EV10ADD*, *EV9POP·EV10ADD*. *EV9POP* not significant at p = .688.

m. Model 165 Predictors: Constant, *EV1ATM*, *EV2ARN*, *EV8USE*, *EV10ADD*, *EV1ATM·EV2ARN*, *EV1ATM·EV6AWS*, *EV1ATM·EV8USE*, *EV1ATM·EV10ADD*, *EV2ARN·EV7ELV*, *EV2ARN·EV8USE*, *EV2ARN·EV9POP*, *EV6AWS·EV7ELV*, *EV6AWS·EV10ADD*, *EV7ELV·EV9POP*, *EV8USE·EV10ADD*, *EV9POP·EV10ADD*. All predictors significant ($p < \alpha = .20$).

n. Model 166 Predictors: Constant, *EV1ATM*, *EV2ARN*, *EV8USE*, *EV10ADD*, *EV1ATM·EV2ARN*, *EV1ATM·EV6AWS*, *EV1ATM·EV8USE*, *EV1ATM·EV10ADD*, *EV2ARN·EV7ELV*, *EV2ARN·EV8USE*, *EV2ARN·EV9POP*, *EV6AWS·EV7ELV*,

EV6AWS·EV10ADD, *EV7ELV·EV9POP*, *EV8USE·EV10ADD*, *EV9POP·EV10ADD*. Tweedie MVP adjusted to 1.45. All predictors significant ($p < \alpha = .20$).

o. Model 167 Predictors: Constant, *EV1ATM*, *EV2ARN*, *EV8USE*, *EV10ADD*, *EV1ATM·EV2ARN*, *EV1ATM·EV6AWS*, *EV1ATM·EV8USE*, *EV1ATM·EV10ADD*, *EV2ARN·EV7ELV*, *EV2ARN·EV8USE*, *EV2ARN·EV9POP*, *EV6AWS·EV7ELV*, *EV6AWS·EV10ADD*, *EV7ELV·EV9POP*, *EV8USE·EV10ADD*, *EV9POP·EV10ADD*. Tweedie MVP adjusted to 1.40. All predictors significant ($p < \alpha = .20$).

p. Model 168 Predictors: Constant, *EV1ATM*, *EV2ARN*, *EV8USE*, *EV10ADD*, *EV1ATM·EV2ARN*, *EV1ATM·EV6AWS*, *EV1ATM·EV8USE*, *EV1ATM·EV10ADD*, *EV2ARN·EV7ELV*, *EV2ARN·EV8USE*, *EV2ARN·EV9POP*, *EV6AWS·EV7ELV*, *EV6AWS·EV10ADD*, *EV7ELV·EV9POP*, *EV8USE·EV10ADD*, *EV9POP·EV10ADD*. Tweedie MVP adjusted to 1.43. All predictors significant ($p < \alpha = .20$).

q. Model 169 Predictors: Constant, *EV1ATM*, *EV2ARN*, *EV8USE*, *EV10ADD*, *EV1ATM·EV2ARN*, *EV1ATM·EV6AWS*, *EV1ATM·EV8USE*, *EV1ATM·EV10ADD*, *EV2ARN·EV7ELV*, *EV2ARN·EV8USE*, *EV2ARN·EV9POP*, *EV6AWS·EV7ELV*, *EV6AWS·EV10ADD*, *EV7ELV·EV9POP*, *EV8USE·EV10ADD*, *EV9POP·EV10ADD*. Tweedie MVP adjusted to 1.44. All predictors significant ($p < \alpha = .20$).

Table C7

Stage B5 DV_{COUNT} Results for DV5, All EVs

Model	Tweedie MPV	*D*	χ^2_P	*AIC*	*LL2 Ratio* χ^2	*df*	*.sig*
170	1.44	.984	1482.918	2475.447	551.190	27	.000
171	1.44	.984	1482.918	2475.447	551.190	27	.000
172	1.44	.983	1482.900	2474.448	551.190	26	.000
173	1.44	.982	1484.133	2471.659	550.979	25	.000
174	1.44	.981	1485.264	2469.775	550.862	24	.000
175	1.44	.980	1485.510	2467.936	550.701	23	.000
176	1.44	.979	1487.497	2466.097	550.540	22	.000
177	1.44	.979	1492.014	2464.508	550.129	21	.000
178	1.44	.978	1491.176	2462.855	549.782	20	.000
179	1.44	.978	1492.774	2462.272	548.365	19	.000
180	1.44	.969	1481.610	2437.651	541.793	19	.000
181	1.40	.997	1515.908	2512.804	563.609	19	.000
182	1.40	.997	1523.859	2512.378	562.035	18	.000

a. Model 170 Predictors: Constant, *EV1ATM*, *EV2ARN*, *EV8USE*, *EV10ADD*, *EV1ATM·EV2ARN*, *EV1ATM·EV6AWS*, *EV1ATM·EV8USE*, *EV1ATM·EV10ADD*, *EV2ARN·EV7ELV*, *EV2ARN·EV8USE*, *EV2ARN·EV9POP*, *EV6AWS·EV7ELV*, *EV6AWS·EV10ADD*, *EV7ELV·EV9POP*, *EV8USE·EV10ADD*, *EV9POP·EV10ADD*, *January*, *February*, *March*, *April*, *May*, *June*, *July*, *August*, *September*, *October*, *November*, *December*. *January* not significant at p = .997.

b. Model 171 Predictors: Constant, *EV1ATM*, *EV2ARN*, *EV8USE*, *EV10ADD*, *EV1ATM·EV2ARN*, *EV1ATM·EV6AWS*, *EV1ATM·EV8USE*, *EV1ATM·EV10ADD*, *EV2ARN·EV7ELV*, *EV2ARN·EV8USE*, *EV2ARN·EV9POP*, *EV6AWS·EV7ELV*, *EV6AWS·EV10ADD*, *EV7ELV·EV9POP*, *EV8USE·EV10ADD*, *EV9POP·EV10ADD*, *February*, *March*, *April*, *May*, *June*, *July*, *August*, *September*, *October*, *November*, *December*. *December* not significant at $p = .997$.

c. Model 172 Predictors: Constant, *EV1ATM*, *EV2ARN*, *EV8USE*, *EV10ADD*, *EV1ATM·EV2ARN*, *EV1ATM·EV6AWS*, *EV1ATM·EV8USE*, *EV1ATM·EV10ADD*, *EV2ARN·EV7ELV*, *EV2ARN·EV8USE*, *EV2ARN·EV9POP*, *EV6AWS·EV7ELV*, *EV6AWS·EV10ADD*, *EV7ELV·EV9POP*, *EV8USE·EV10ADD*, *EV9POP·EV10ADD*, *February*, *March*, *April*, *May*, *June*, *July*, *August*, *September*, *October*, *November*. *November* not significant at $p = .646$.

d. Model 173 Predictors: Constant, *EV1ATM*, *EV2ARN*, *EV8USE*, *EV10ADD*, *EV1ATM·EV2ARN*, *EV1ATM·EV6AWS*, *EV1ATM·EV8USE*, *EV1ATM·EV10ADD*, *EV2ARN·EV7ELV*, *EV2ARN·EV8USE*, *EV2ARN·EV9POP*, *EV6AWS·EV7ELV*, *EV6AWS·EV10ADD*, *EV7ELV·EV9POP*, *EV8USE·EV10ADD*, *EV9POP·EV10ADD*, *February*, *March*, *April*, *May*, *June*, *July*, *August*, *September*, *October*. *March* not significant at $p = .733$.

e. Model 174 Predictors: Constant, *EV1ATM*, *EV2ARN*, *EV8USE*, *EV10ADD*, *EV1ATM·EV2ARN*, *EV1ATM·EV6AWS*, *EV1ATM·EV8USE*, *EV1ATM·EV10ADD*, *EV2ARN·EV7ELV*, *EV2ARN·EV8USE*, *EV2ARN·EV9POP*, *EV6AWS·EV7ELV*, *EV6AWS·EV10ADD*, *EV7ELV·EV9POP*, *EV8USE·EV10ADD*, *EV9POP·EV10ADD*, *February*, *April*, *May*, *June*, *July*, *August*, *September*, *October*. *February* not significant at $p = .687$.

f. Model 175 Predictors: Constant, *EV1ATM*, *EV2ARN*, *EV8USE*, *EV10ADD*, *EV1ATM·EV2ARN*, *EV1ATM·EV6AWS*, *EV1ATM·EV8USE*, *EV1ATM·EV10ADD*, *EV2ARN·EV7ELV*, *EV2ARN·EV8USE*, *EV2ARN·EV9POP*, *EV6AWS·EV7ELV*, *EV6AWS·EV10ADD*, *EV7ELV·EV9POP*, *EV8USE·EV10ADD*, *EV9POP·EV10ADD*, *April*, *May*, *June*, *July*, *August*, *September*, *October*. *October* not significant at $p = .689$.

g. Model 176 Predictors: Constant, *EV1ATM*, *EV2ARN*, *EV8USE*, *EV10ADD*, *EV1ATM·EV2ARN*, *EV1ATM·EV6AWS*, *EV1ATM·EV8USE*, *EV1ATM·EV10ADD*,

EV2ARN·EV7ELV, *EV2ARN·EV8USE*, *EV2ARN·EV9POP*, *EV6AWS·EV7ELV*, *EV6AWS·EV10ADD*, *EV7ELV·EV9POP*, *EV8USE·EV10ADD*, *EV9POP·EV10ADD*, *April*, *May*, *June*, *July*, *August*, *September*. *September* not significant at $p = .522$.

h. Model 177 Predictors: Constant, *EV1ATM*, *EV2ARN*, *EV8USE*, *EV10ADD*, *EV1ATM·EV2ARN*, *EV1ATM·EV6AWS*, *EV1ATM·EV8USE*, *EV1ATM·EV10ADD*, *EV2ARN·EV7ELV*, *EV2ARN·EV8USE*, *EV2ARN·EV9POP*, *EV6AWS·EV7ELV*, *EV6AWS·EV10ADD*, *EV7ELV·EV9POP*, *EV8USE·EV10ADD*, *EV9POP·EV10ADD*, *April*, *May*, *June*, *July*, *August*. *August* not significant at $p = .556$.

i. Model 178 Predictors: Constant, *EV1ATM*, *EV2ARN*, *EV8USE*, *EV10ADD*, *EV1ATM·EV2ARN*, *EV1ATM·EV6AWS*, *EV1ATM·EV8USE*, *EV1ATM·EV10ADD*, *EV2ARN·EV7ELV*, *EV2ARN·EV8USE*, *EV2ARN·EV9POP*, *EV6AWS·EV7ELV*, *EV6AWS·EV10ADD*, *EV7ELV·EV9POP*, *EV8USE·EV10ADD*, *EV9POP·EV10ADD*, *April*, *May*, *June*, *July*. *July* not significant at $p = .234$.

j. Model 179 Predictors: Constant, *EV1ATM*, *EV2ARN*, *EV8USE*, *EV10ADD*, *EV1ATM·EV2ARN*, *EV1ATM·EV6AWS*, *EV1ATM·EV8USE*, *EV1ATM·EV10ADD*, *EV2ARN·EV7ELV*, *EV2ARN·EV8USE*, *EV2ARN·EV9POP*, *EV6AWS·EV7ELV*, *EV6AWS·EV10ADD*, *EV7ELV·EV9POP*, *EV8USE·EV10ADD*, *EV9POP·EV10ADD*, *April*, *May*, *June*. All predictors significant ($p < \alpha = .20$).

k. Model 180 Predictors: Constant, *EV1ATM*, *EV2ARN*, *EV8USE*, *EV10ADD*, *EV1ATM·EV2ARN*, *EV1ATM·EV6AWS*, *EV1ATM·EV8USE*, *EV1ATM·EV10ADD*, *EV2ARN·EV7ELV*, *EV2ARN·EV8USE*, *EV2ARN·EV9POP*, *EV6AWS·EV7ELV*, *EV6AWS·EV10ADD*, *EV7ELV·EV9POP*, *EV8USE·EV10ADD*, *EV9POP·EV10ADD*, *April*, *May*, *June*. Tweedie MVP adjusted to 1.46. All predictors significant ($p < \alpha = .20$).

l. Model 181 Predictors: Constant, *EV1ATM*, *EV2ARN*, *EV8USE*, *EV10ADD*, *EV1ATM·EV2ARN*, *EV1ATM·EV6AWS*, *EV1ATM·EV8USE*, *EV1ATM·EV10ADD*, *EV2ARN·EV7ELV*, *EV2ARN·EV8USE*, *EV2ARN·EV9POP*, *EV6AWS·EV7ELV*, *EV6AWS·EV10ADD*, *EV7ELV·EV9POP*, *EV8USE·EV10ADD*, *EV9POP·EV10ADD*, *April*, *May*, *June*. Tweedie MVP adjusted to 1.40. *EV2ARN·EV7ELV* not significant at p = .210.

m. Model 182 Predictors: Constant, *EV1ATM*, *EV2ARN*, *EV8USE*, *EV10ADD*, *EV1ATM·EV2ARN*, *EV1ATM·EV6AWS*, *EV1ATM·EV8USE*, *EV1ATM·EV10ADD*, *EV2ARN·EV7ELV*, *EV2ARN·EV8USE*, *EV2ARN·EV9POP*, *EV6AWS·EV7ELV*, *EV6AWS·EV10ADD*, *EV7ELV·EV9POP*, *EV8USE·EV10ADD*, *EV9POP·EV10ADD*, *April*, *May*, *June*. All predictors significant ($p < \alpha = .20$).

Stage B6 DV_{COUNT} Analysis with *DV5*, All EVs, α = .05

Table C8

Stage B6 DV_{COUNT} Analysis with DV5, EVs, 2FIs, and Months, α = .05

Model	Tweedie MPV	*D*	χ^2_P	*AIC*	*LL2* Ratio χ^2	*df*	*.sig*
183	1.400	.998	1533.391	2513.754	558.659	17	.000
184	1.390	1.003	1539.083	2526.590	563.058	17	.000
185	1.395	1.000	1536.228	2520.159	560.828	17	.000

a. Model 183 Predictors: Constant, *EV1ATM*, *EV2ARN*, *EV8USE*, *EV10ADD*, *EV1ATM·EV2ARN*, *EV1ATM·EV6AWS*, *EV1ATM·EV10ADD*, *EV2ARN·EV7ELV*, *EV2ARN·EV8USE*, *EV2ARN·EV9POP*, *EV6AWS·EV7ELV*, *EV6AWS·EV10ADD*, *EV7ELV·EV9POP*, *EV8USE·EV10ADD*, *EV9POP·EV10ADD*, *April*, *May*, *June*. All predictors remained significant at $p < .05$.

b. Model 184 Predictors: Constant, *EV1ATM*, *EV2ARN*, *EV8USE*, *EV10ADD*, *EV1ATM·EV2ARN*, *EV1ATM·EV6AWS*, *EV1ATM·EV10ADD*, *EV2ARN·EV7ELV*, *EV2ARN·EV8USE*, *EV2ARN·EV9POP*, *EV6AWS·EV7ELV*, *EV6AWS·EV10ADD*, *EV7ELV·EV9POP*, *EV8USE·EV10ADD*, *EV9POP·EV10ADD*, *April*, *May*, *June*. Tweedie MVP adjusted to 1.39. All predictors remained significant at $p < .05$.

c. Model 185 Predictors: Constant, *EV1ATM*, *EV2ARN*, *EV8USE*, *EV10ADD*, *EV1ATM·EV2ARN*, *EV1ATM·EV6AWS*, *EV1ATM·EV10ADD*, *EV2ARN·EV7ELV*, *EV2ARN·EV8USE*, *EV2ARN·EV9POP*, *EV6AWS·EV7ELV*, *EV6AWS·EV10ADD*, *EV7ELV·EV9POP*, *EV8USE·EV10ADD*, *EV9POP·EV10ADD*, *April*, *May*, *June*. Tweedie MVP adjusted to 1.395. All predictors remained significant at $p < .05$.

Appendix D: Stage C Analysis

Stage C DV_{COUNT} Analysis using $DV_{PRESENCE}$ Final Model Terms, α = .05

Table D1

Stage C DV_{COUNT} with $DV_{PRESENCE}$ Final Model EVs

Model	*Tweedie MPV*	*D*	χ^2_P	*AIC*	*LL2 Ratio* χ^2	*df*	*.sig*
186	1.5	1.004	1559.274	2458.608	469.392	23	.000
187	1.5	1.003	1559.212	2456.608	469.391	22	.000
188	1.5	1.002	1561.148	2454.658	469.341	21	.000
189	1.5	1.000	1563.837	2452.930	469.070	20	.000
190	1.5	1.000	1562.524	2451.161	468.833	19	.000
191	1.5	.999	1564.192	2449.426	468.574	18	.000
192	1.5	.998	1564.231	2447.783	468.216	17	.000
193	1.5	.998	1563.929	2445.957	468.043	16	.000
194	1.5	.997	1574.935	2444.426	467.574	15	.000
195	1.5	.996	1575.527	2442.851	467.148	14	.000
196	1.5	.996	1580.569	2442.352	466.151	13	.000
197	1.5	.996	1584.252	2440.697	465.503	12	.000
198	1.5	.995	1583.150	2439.738	464.262	11	.000
199	1.4	1.041	1640.38	2565.422	494.991	11	.000
200	1.395	1.044	1643.344	2572.033	496.955	11	.000
201	1.45	1.017	1610.892	2501.028	477.882	11	.000

Appendix E: Stage A–C Final Model Estimates

Table E1

Stage A3 DV1, All EVs SPSS Parameter Estimates, α = .20

Variables in the Equation							
		B	S.E.	Wald	df	Sig.	Exp(B)
Step 1[a]	IV1ATM	-4.371	1.695	6.645	1	0.010	0.013
	IV2ARN	234.861	163.438	2.065	1	0.151	9.97E+101
	IV5ABP	-0.145	0.117	1.552	1	0.213	0.865
	IV8USE(1)	-67.000	60.045	1.245	1	0.264	0.000
	IV9POP	0.015	0.008	3.481	1	0.062	1.015
	IV10ADD	-0.151	0.038	15.432	1	0.000	0.860
	IV1_IV2	0.192	0.084	5.177	1	0.023	1.211
	IV1_IV5	0.004	0.002	6.433	1	0.011	1.004
	IV1_IV7	0.000	0.000	10.094	1	0.001	1.000
	IV2_IV5	-0.247	0.159	2.396	1	0.122	0.781
	IV2_IV7	0.003	0.002	1.779	1	0.182	1.003
	IV2_IV8	-1.618	1.500	1.164	1	0.281	0.198
	IV5_IV7	0.000	0.000	28.581	1	0.000	1.000
	IV5_IV8	-0.064	0.059	1.189	1	0.276	0.938
	IV6_IV10	-0.016	0.004	12.801	1	0.000	0.984
	IV7_IV10	0.000	0.000	18.452	1	0.000	1.000
	IV8_IV9	-0.012	0.008	1.957	1	0.162	0.989
	Apr	1.386	0.644	4.629	1	0.031	3.999
	May	1.597	0.711	5.046	1	0.025	4.938
	Jun	1.652	0.857	3.718	1	0.054	5.216
	Jul	3.361	0.821	16.750	1	0.000	28.811
	Aug	4.158	0.799	27.079	1	0.000	63.950
	Sep	4.360	0.690	39.883	1	0.000	78.254
	Oct	3.401	0.502	45.859	1	0.000	29.988
	Nov	2.577	0.362	50.717	1	0.000	13.159
	Constant	214.881	100.505	4.571	1	0.033	2.09E+93

a. Variable(s) entered on step 1: IV1ATM, IV2ARN, IV5ABP, IV8USE, IV9POP, IV10ADD, IV1_IV2, IV1_IV5, IV1_IV7, IV2_IV5, IV2_IV7, IV2_IV8, IV5_IV7, IV5_IV8, IV6_IV10, IV7_IV10, IV8_IV9, Apr, May, Jun, Jul, Aug, Sep, Oct, Nov.

Note: Model 36, $R^2_{\text{Nagelkerke}} = .285$, $\chi^2(25) = 739.250$, $p < .001$.

Table E2

Stage A5 DV2, All EVs SPSS Parameter Estimates, α = .20

Variables in the Equation									
								95% C.I.for EXP(B)	
		B	S.E.	Wald	df	Sig.	Exp(B)	Lower	Upper
Step 1[a]	IV1ATM	-8.380146	1.526	30.143	1	0.000		0.000	0.005
	IV6AWS	78.978139	17.674	19.969	1	0.000	1.99421E+34	1.80213E+19	2.20675E+49
	IV9POP	0.396894	0.239	2.753	1	0.097	1.487	0.931	2.377
	IV1_IV5	0.008365	0.002	30.898	1	0.000	1.008	1.005	1.011
	IV1_IV6	-0.005448	0.006	0.967	1	0.325	0.995	0.984	1.005
	IV1_IV10	0.001304	0.002	0.612	1	0.434	1.001	0.998	1.005
	IV2_IV6	-0.332218	0.197	2.838	1	0.092	0.717	0.487	1.056
	IV2_IV10	0.182601	0.133	1.872	1	0.171	1.200	0.924	1.559
	IV5_IV6	-0.076861	0.017	19.765	1	0.000	0.926	0.895	0.958
	IV5_IV8	0.003656	0.001	20.475	1	0.000	1.004	1.002	1.005
	IV5_IV9	-0.000374	0.000	2.545	1	0.111	1.000	0.999	1.000
	IV5_IV10	-0.000175	0.000	1.290	1	0.256	1.000	1.000	1.000
	IV6_IV8	-0.419742	0.147	8.153	1	0.004	0.657	0.493	0.877
	IV6_IV9	0.000640	0.000	3.535	1	0.060	1.001	1.000	1.001
	IV6_IV10	-0.033318	0.015	5.016	1	0.025	0.967	0.939	0.996
	IV8_IV9	-0.018761	0.008	5.128	1	0.024	0.981	0.966	0.997
	IV9_IV10	0.000245	0.000	3.401	1	0.065	1.000	1.000	1.001
	May	-1.628505	0.430	14.348	1	0.000	0.196	0.084	0.456
	Jun	-0.449369	0.216	4.328	1	0.037	0.638	0.418	0.974
	Aug	0.199673	0.169	1.402	1	0.236	1.221	0.877	1.699
	Sep	0.163101	0.176	0.861	1	0.353	1.177	0.834	1.661
	Oct	0.659320	0.225	8.590	1	0.003	1.933	1.244	3.005
	Nov	-0.354533	0.514	0.475	1	0.491	0.702	0.256	1.922
	Constant	-13.277192	2.823	22.118	1	0.000	0.000		

a. Variable(s) entered on step 1: IV1ATM, IV6AWS, IV9POP, IV1_IV5, IV1_IV6, IV1_IV10, IV2_IV6, IV2_IV10, IV5_IV6, IV5_IV8, IV5_IV9, IV5_IV10, IV6_IV8, IV6_IV9, IV6_IV10, IV8_IV9, IV9_IV10, May, Jun, Aug, Sep, Oct, Nov.

Note: Model 84, $R^2_{\text{Nagelkerke}} = .293$, $\chi^2\,(23) = 762.344$, $p < .001$.

Table E3

Stage A6 DV2, All EVs SPSS Parameter Estimates, α = .05

Variables in the Equation									
								95% C.I.for EXP(B)	
		B	S.E.	Wald	df	Sig.	Exp(B)	Lower	Upper
Step 1[a]	IV1ATM	-8.35886	1.500	31.073	1	0.000	0.000	0.000	0.004
	IV6AWS	79.44608	17.215	21.297	1	0.000	3.18416E+34	7.06698E+19	1.43468E+49
	IV9POP	0.42669	0.235	3.288	1	0.070	1.532	0.966	2.430
	IV1_IV5	0.00836	0.001	31.922	1	0.000	1.008	1.005	1.011
	IV1_IV6	-0.00588	0.005	1.205	1	0.272	0.994	0.984	1.005
	IV2_IV6	-0.31451	0.193	2.668	1	0.102	0.730	0.501	1.065
	IV2_IV10	0.16639	0.132	1.594	1	0.207	1.181	0.912	1.529
	IV5_IV6	-0.07717	0.017	21.009	1	0.000	0.926	0.896	0.957
	IV5_IV8	0.00369	0.001	20.967	1	0.000	1.004	1.002	1.005
	IV5_IV9	-0.00040	0.000	3.054	1	0.081	1.000	0.999	1.000
	IV6_IV8	-0.42944	0.147	8.545	1	0.003	0.651	0.488	0.868
	IV6_IV9	0.00067	0.000	3.908	1	0.048	1.001	1.000	1.001
	IV6_IV10	-0.04770	0.007	42.046	1	0.000	0.953	0.940	0.967
	IV8_IV9	-0.01866	0.008	5.076	1	0.024	0.982	0.966	0.998
	IV9_IV10	0.00020	0.000	2.340	1	0.126	1.000	1.000	1.000
	May(1)	1.60548	0.428	14.055	1	0.000	4.980	2.151	11.529
	Jun(1)	0.43281	0.216	4.014	1	0.045	1.542	1.009	2.354
	Aug(1)	-0.20634	0.169	1.499	1	0.221	0.814	0.585	1.132
	Sep(1)	-0.17622	0.173	1.034	1	0.309	0.838	0.597	1.177
	Oct(1)	-0.70619	0.212	11.138	1	0.001	0.494	0.326	0.747
	Constant	-16.14772	2.319	48.469	1	0.000	0.000		

a. Variable(s) entered on step 1: IV1ATM, IV6AWS, IV9POP, IV1_IV5, IV1_IV6, IV2_IV6, IV2_IV10, IV5_IV6, IV5_IV8, IV5_IV9, IV6_IV8, IV6_IV9, IV6_IV10, IV8_IV9, IV9_IV10, May, Jun, Aug, Sep, Oct.

Note: Model 93, $R^2_{\text{Nagelkerke}} = .292$, $\chi^2\,(20) = 760.095$, $p < .001$.

Table E4

Stage B3 DV_{COUNT} DV1, All EVs SPSS Parameter Estimates, α = .05

Parameter Estimates

			Interval		Hypothesis Test				Interval for Exp(B)	
Parameter	B	Std. Error	Lower	Upper	Wald Chi-Square	df	Sig.	Exp(B)	Lower	Upper
(Intercept)	-4.936	2.1071	-9.065	-0.806	5.487	1	0.019	0.007	0.000	0.447
IV1ATM	-0.029	0.0169	-0.062	0.004	2.887	1	0.089	0.972	0.940	1.004
IV7ELV	0.003	0.0012	0.001	0.006	7.509	1	0.006	1.003	1.001	1.006
[IV8USE=0]	2.929	1.1384	0.698	5.161	6.622	1	0.010	18.717	2.010	174.263
[IV8USE=1]	0[a]							1		
IV1_IV2	0.332	0.0808	0.174	0.491	16.909	1	0.000	1.394	1.190	1.633
IV1_IV7	-6.564E-05	1.6972E-05	-9.891E-05	-3.238E-05	14.959	1	0.000	1.000	1.000	1.000
IV1_IV8	0.030	0.0144	0.002	0.059	4.472	1	0.034	1.031	1.002	1.060
IV1_IV9	3.222E-05	1.9349E-05	-5.708E-06	7.014E-05	2.772	1	0.096	1.000	1.000	1.000
IV2_IV5	-0.028	0.0062	-0.040	-0.016	20.480	1	0.000	0.972	0.961	0.984
IV2_IV9	0.009	0.0035	0.002	0.016	6.695	1	0.010	1.009	1.002	1.016
IV5_IV10	0.000	4.2568E-05	0.000	-4.215E-05	8.703	1	0.003	1.000	1.000	1.000
IV7_IV10	7.685E-05	3.6785E-05	4.750E-06	0.000	4.364	1	0.037	1.000	1.000	1.000
IV8_IV10	-0.123	0.0510	-0.223	-0.023	5.856	1	0.016	0.884	0.800	0.977
IV9_IV10	0.000	0.0001	0.000	5.993E-05	2.068	1	0.150	1.000	1.000	1.000
[Feb=0]	1.853	1.1033	-0.310	4.015	2.820	1	0.093	6.377	0.734	55.430
[Feb=1]	0[a]							1		
[Mar=0]	1.218	0.8264	-0.402	2.838	2.172	1	0.141	3.380	0.669	17.075
[Mar=1]	0[a]							1		
[Apr=0]	1.328	0.5343	0.280	2.375	6.174	1	0.013	3.772	1.324	10.749
[Apr=1]	0[a]							1		
[May=0]	1.940	0.4267	1.104	2.777	20.676	1	0.000	6.960	3.016	16.063
[May=1]	0[a]							1		
[Jun=0]	2.202	0.3755	1.466	2.938	34.381	1	0.000	9.040	4.331	18.871
[Jun=1]	0[a]							1		
[Aug=0]	-0.930	0.1385	-1.201	-0.658	45.017	1	0.000	0.395	0.301	0.518
[Aug=1]	0[a]							1		
[Sep=0]	-1.157	0.1438	-1.439	-0.875	64.775	1	0.000	0.314	0.237	0.417
[Sep=1]	0[a]							1		
[Oct=0]	-0.952	0.1981	-1.341	-0.564	23.126	1	0.000	0.386	0.262	0.569
[Oct=1]	0[a]							1		
[Nov=0]	-0.905	0.2886	-1.471	-0.339	9.831	1	0.002	0.405	0.230	0.712
[Nov=1]	0[a]							1		
(Scale)	1.493[b]	0.0349	1.426	1.563						

Dependent Variable: DV1

a. Set to zero because this parameter is redundant.

b. Maximum likelihood estimate.

Note: Model 129, Tweedie MVP = 1.5, D = 1.898, $\chi^2(22)$ = 582.309, p < .001.

Table E5

Stage B6 DV_{COUNT} DV5, All EVs SPSS Parameter Estimates, α = .05

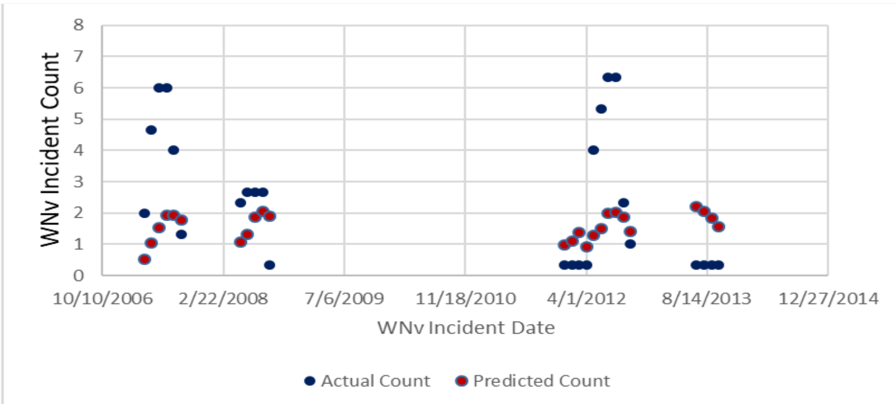
Parameter Estimates										
			Interval		Hypothesis Test				Interval for Exp(B)	
Parameter	B	Std. Error	Lower	Upper	Wald Chi-Square	df	Sig.	Exp(B)	Lower	Upper
(Intercept)	-4.368	0.8942	-6.120	-2.615	23.856	1	0.000	0.013	0.002	0.073
IV1ATM	0.051	0.0122	0.028	0.075	17.832	1	0.000	1.053	1.028	1.078
IV2ARN	-12.111	3.9636	-19.880	-4.343	9.337	1	0.002	5.496E-06	2.324E-09	0.013
[IV8USE=0]	0.629	0.2434	0.152	1.106	6.681	1	0.010	1.876	1.164	3.022
[IV8USE=1]	0[a]							1		
IV10ADD	0.184	0.0782	0.031	0.338	5.548	1	0.019	1.202	1.031	1.402
IV1_IV2	0.142	0.0511	0.042	0.242	7.752	1	0.005	1.153	1.043	1.274
IV1_IV6	0.003	0.0007	0.002	0.005	24.927	1	0.000	1.003	1.002	1.005
IV1_IV10	-0.004	0.0009	-0.005	-0.002	17.688	1	0.000	0.996	0.995	0.998
IV2_IV9	0.007	0.0019	0.003	0.010	13.539	1	0.000	1.007	1.003	1.011
IV6_IV7	0.000	5.8973E-05	0.000	-7.922E-05	10.911	1	0.001	1.000	1.000	1.000
IV6_IV10	-0.013	0.0056	-0.024	-0.002	5.620	1	0.018	0.987	0.976	0.998
IV7_IV8	0.001	0.0003	0.001	0.002	21.585	1	0.000	1.001	1.001	1.002
IV7_IV9	-2.264E-06	7.9127E-07	-3.815E-06	-7.131E-07	8.186	1	0.004	1.000	1.000	1.000
IV8_IV10	-0.070	0.0260	-0.121	-0.019	7.165	1	0.007	0.933	0.886	0.982
IV9_IV10	0.000	3.9912E-05	4.499E-05	0.000	9.530	1	0.002	1.000	1.000	1.000
Apr	0.489	0.1545	0.186	0.792	10.006	1	0.002	1.630	1.204	2.207
May	0.404	0.0916	0.224	0.583	19.418	1	0.000	1.497	1.251	1.792
Jun	0.282	0.0740	0.137	0.427	14.518	1	0.000	1.326	1.147	1.532
(Scale)	.728[b]	0.0233	0.684	0.775						
Dependent Variable: DV5										
a. Set to zero because this parameter is redundant.										
b. Maximum likelihood estimate.										

Note: Model 185, Tweedie = 1.395, D = *1.000*, *LL2 Ratio* $\chi^2(17) = 560.828$, $p < .001$.

Appendix F: GZLM Count Data

Figure F1

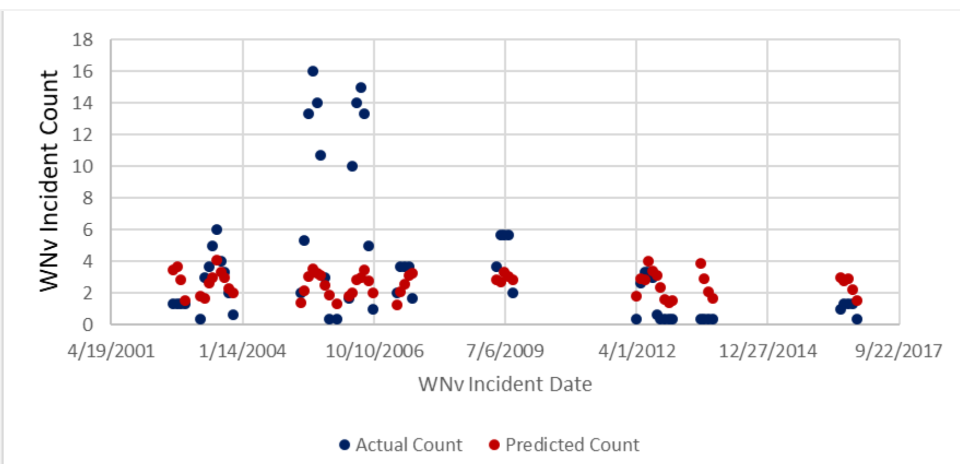
Actual versus Predicted WNv Count, Aiken County 2002-2016



Note. Stage B Model 185, Tweedie 1.395

Figure F2

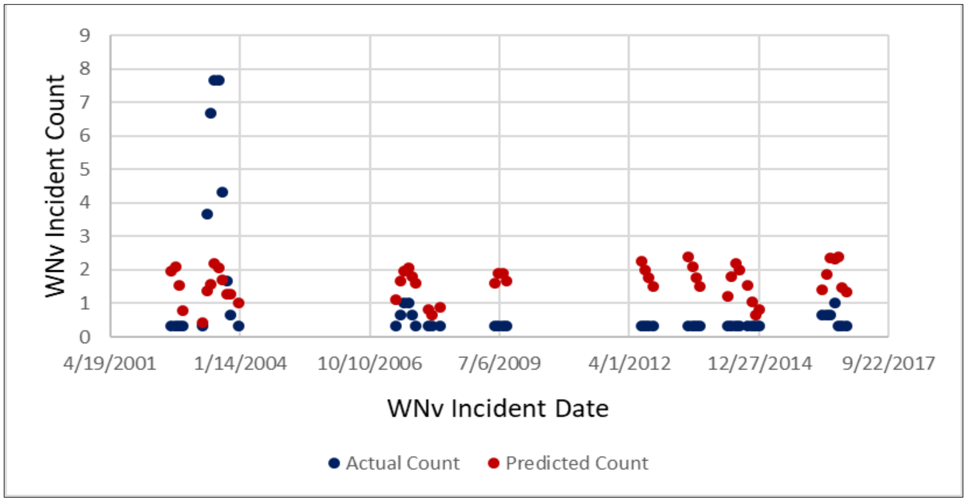
Actual versus Predicted WNv Count, Charleston County 2002-2016



Note. Stage B Model 185, Tweedie 1.395

Figure F3

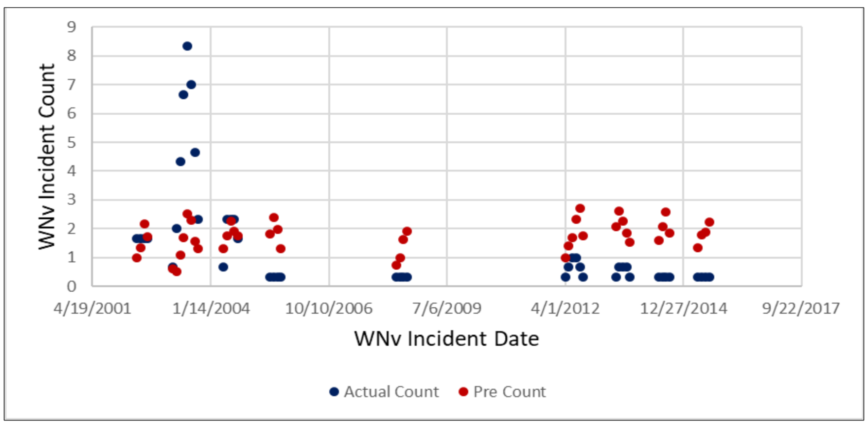
Actual versus Predicted WNv Count, Dorchester County 2002-2016



Note. Stage B Model 185, Tweedie 1.395

Figure F4

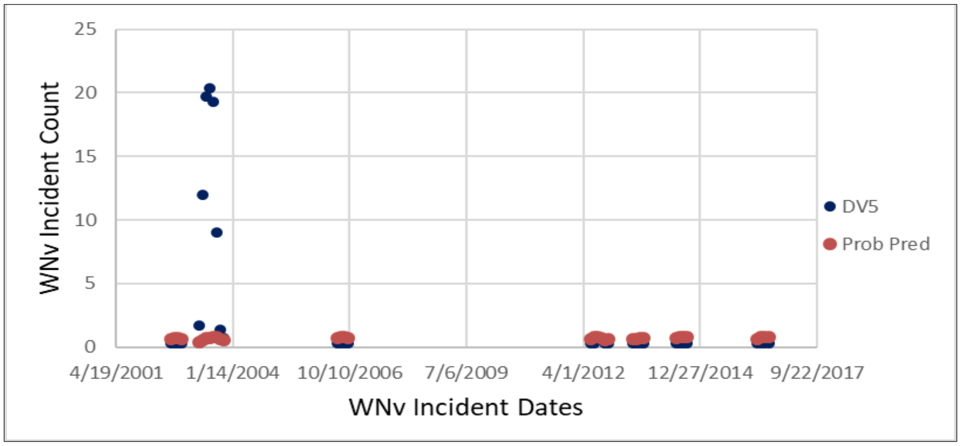
Actual versus Predicted WNv Count, Greenville County 2002-2016



Note. Stage B Model 185, Tweedie 1.395

Figure F5

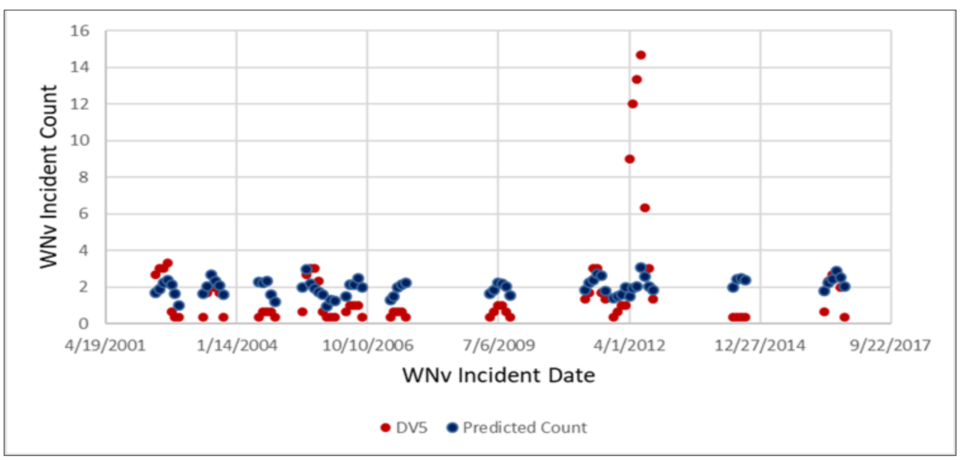
Actual versus Predicted WNv Count, Horry County 2002-2016



Note. Stage B Model 185, Tweedie 1.395

Figure F6

Actual versus Predicted WNv Count, Richland County 2002-2016



Note. Stage B Model 185, Tweedie 1.395