SEQUENTIAL SAMPLING SCHEMES FOR WEST NILE VIRUS INFECTION IN *CULEX* MOSQUITOES AND ANTICIPATION OF EPIDEMIC THRESHOLDS

By

Danielle J. Donovan

A THESIS

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

MASTER OF SCIENCE

Entomology

2011

ABSTRACT

SEQUENTIAL SAMPLING SCHEMES FOR WEST NILE VIRUS INFECTION IN *CULEX* MOSQUITOES AND ANTICIPATION OF EPIDEMIC THRESHOLDS

By

Danielle J. Donovan

Sequential sampling of insects offers the benefits of increased efficiency of sampling and support of decision making. However, its applications to public health entomology are rare, even though it could be a powerful tool in guiding sampling programs for vectors, and in forecasting risk of epidemics of mosquito-borne, viral disease. In this study, I chose to analyze two dependent variables for sequential sampling schemes, each representing a surrogate of infection rate in *Culex* mosquito populations for West Nile virus (WNv); namely, the number and percentage of mosquito pools that tested positive for the virus in real time PCR. I obtained longitudinal data of WNv infection in mosquitoes in Illinois from local study site, county, and state spatial scales, as well as, the dates of onset of confirmed human cases of WNv infection, from a public source, occurring within these spatial units through a completely anonymized process. Sampling and epidemic threshold regression lines were successfully developed with aggregation analysis for each spatial scale and dependent variable. Sequential sampling schemes were found to be most sensitive to prediction of epidemic outbreaks at the local study site scale, and relatively less sensitive at the county and state spatial scales, likely due to the effects of local transmission dynamics that were obscured at broader geographic scales by sampling error. Furthermore, the two dependent variables showed little difference in their ability to forecast West Nile virus human cases at the study site and county scales, making either a possibility for use in a sequential sampling plan.

This thesis is dedicated to my family and friends.

ACKNOWLEDGEMENTS

First I would like to thank my advisor Dr. Edward Walker for all his support, guidance, and expertise throughout the course of my master's research. In addition, I would like to thank the rest of my committee, Dr. James Miller and Dr. Gabriel Ording. I would like to make a special thanks to Dr. Gabriel Ording for convincing me to go to graduate school and being an excellent mentor in my teaching experience.

In addition, I would like to thank the many other collaborators on the project Dr. Gabriel Hamer, Dr. Tony Goldberg, Dr. Marilyn Ruiz, and Bill Brown. I would like thank Dr. Gabriel Hamer especially for his help and guidance on the project and his willingness to answer and discuss my many questions throughout.

I would like to thank the Michigan State University College of Agriculture and Natural Resources Statistical Counseling Center for all their statistical help, especially to Wenzhao Yang who was very patient and knowledgeable through my numerous statistics questions.

There were so many people that contributed to the project both in the field and lab. I would like to thank Lisa Abernathy, Zach Allison, Rachel Atkins, Melanie Bender, Garrett Berry, Erica Brown, Blair Bullard, Seth Dallmann, Milica Dimic, Diane Ghode, Mike Goshorn, Patrick Kelly, Megan Machiniak, Alvin Makohon-Moore, Jonathon McClain, Mike Neville, Beth Pultorak, Eric Secker, Jennifer Sidge, Timothy Thompson, Liza Watson, and Amy Wechsler.

I would also like to thank my family and friends for their support. A special thanks to my mom for being just plain awesome, and always giving me encouragement and guidance.

Finally, I would like to thank my Ultimate crew. Those games provided some great memories and were always good for a little stress relief.

This work was supported by National Science Foundation/National Institutes of Health program in Ecology of Infectious Diseases and the Michigan Agricultural Experimental Station Multistate Project NE 507.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

SEQUENTIAL SAMPLING SCHEMES FOR WEST NILE VIRUS INFECTION IN *CULEX* MOSQUITOES AND ANTICIPATION OF EPIDEMIC THRESHOLDS

Introduction

Emergence of West Nile Virus

West Nile virus (WNv) is classified within the genus *Flavivirus*, and is a neuro-invasive virus transmitted to humans through mosquitoes, usually certain species of the genus *Culex* in the United States and elsewhere (Hayes & Gubler 2006, Kramer et al. 2008, Reisen & Brault 2007). It belongs to a complex of closely related but geographically widespread viruses – the Japanese encephalitis virus complex – all of which have affiliation with *Culex* mosquitoes and birds as vertebrate hosts, including Japanese encephalitis, Murray Valley encephalitis, St. Louis encephalitis, Kunjin, and several other viruses (Davis et al. 2006, Hayes & Gubler 2006, Kramer et al 2008). Many of these viruses exhibit enzootic, epizootic and epidemic behavior in nature; Japanese encephalitis virus is a major cause of morbidity and mortality in humans in rice-growing regions of Asia (Misra & Kalita 2010). While humans are "dead end hosts" in the sense that virus infections in the human blood stream are not infectious to mosquitoes even when humans become infected and severely ill, a variety of birds serve as so-called vertebrate amplifier and reservoir hosts of West Nile and other viruses in the Japanese encephalitis virus complex (Hayes and Gubler 2006, Kramer et al. 2008, Reisen & Brault 2007).

West Nile virus was first detected in the United States in 1999. Before that time, it was restricted to the Old World and likely originated in central Africa, from where it emerged into the Middle East, Europe, and parts of Asia sometime after the 1940s and continuing into the

1990s (Davis et al. 2006, Kramer et al. 2008, Samuel & Diamond 2006). The appearance of WNv in New York City in 1999 represented a large geographic range expansion for the virus, into the western hemisphere. Most human infections with West Nile virus are asymptomatic (80%), however for the remaining 20% infection may progress to West Nile fever, West Nile meningitis, West Nile encephalitis, or acute flaccid paralysis (Davis et al. 2006, Hayes & Gubler 2006, Samuel & Diamond 2006). From 1999-2009, there were 29,681 cases and 1,163 deaths due to WNv infection reported in the United States through public health surveillance systems (CDC 2011).

One region of continuously intense seasonal WNv activity in the United States is suburban Chicago, Illinois, including Cook County. 2002 was the first year human West Nile virus was detected in Illinois. Since then and until 2009 there were about 1,591 cases in the state, with 942 cases in Cook County and 624 of those in suburban Cook County excluding the city of Chicago, making this region a true "hot spot" (i.e., a place with particularly and consistently high West Nile virus transmission) (CDC 2011, IDPH 2011). Studies initiated there in 2005 and continued through 2010 quantified WNv infection prevalence in mosquitoes and birds (Ruiz et al. 2007, Hamer et al. 2008, Hamer et al. 2009). Data from this site and these published studies were utilized in the present study. In addition, data of WNv infection in mosquito populations from Cook County, Illinois; and the state of Illinois; were obtained from the Illinois Department of Public Health, as were data on onset of human cases of infection reported from these areas.

Several previous studies investigated the potential for an early warning system for West Nile virus human cases. Most of these studies used the spatial analysis of dead birds for the predicative variable. Bird deaths are important because WNv shows epizootic activity before epidemic activity, meaning that the virus infection rate increases in mosquitoes and in birds temporally before humans become exposed to infectious mosquito bites, during the course of the summer season (Hamer et al. 2008, Ruiz et al. 2010). Eidson et al. (2001) identified the temporal appearance of dead birds of the family Corvidae (particularly crows and blue jays) as a useful sign for predicting human cases of West Nile virus, noting that an increase in bird deaths occurred before human cases occurred. Theophilides et al. (2003) went on to assess the possibility of using the Dynamic Continuous-Area Space-Time (DYCAST) system to create a risk analysis for human West Nile virus for New York, New York. The DYCAST system employs a statistical approach incorporating both spatial and temporal factors of bird deaths to predict risk for West Nile virus. The model showed success in predicting high risk areas for 71.4% of the locations where human cases occurred in 2001 with a minimum of 13 days before the date of onset. Theophilides et al. (2006) continued to test the DYCAST system in the Chicago, Illinois for the year of 2002. The results showed similar success with identifying high risk areas for 76.2% of the locations where human cases occurred with a minimum of 15 days before the date of onset. Additionally, Mostashari et al. (2003) developed an early warning system for West Nile virus for New York, New York utilizing a spatial scan statistic analyzing clusters of bird deaths. For the year 2001 the study found the warning system predicted high risk areas before a majority of human cases occurred within them. The model called for control measures to be taken 4 weeks before any positive laboratory results for West Nile virus. Despite the initial

success of using dead birds in early warnings systems there are several limitations. Some important limitations being that identifying bird deaths is highly dependent on variables such as public reporting, media coverage and support of governing agencies (Eidson et al. 2001 and Mostashari et al. 2003). These factors can result in the consistency and accuracy in reporting bird deaths among different locations to be highly variable, which could limit the effectiveness of models relying on bird deaths to predict human cases. Supporting the limitations in warning systems using bird deaths, Brownstein et al. (2004) developed a West Nile virus human surveillance model for all counties in the continental United States that incorporated both dead birds and mosquito infection. The investigators found that when comparing the predictor variable of bird deaths versus the number of confirmed West Nile virus mosquito pools, both were significant but differed in regard to their ability to account for the variation of human cases. Mosquito pools accounted for 38% of the variation in human cases, while bird deaths accounted for 2.5%. This would indicate future warning system models using mosquito infection would have the most promise in forecasting human West Nile virus. However, this model suffers from coarse resolution of spatial scale and inclusion of data from a wide variety of landscapes.

Sequential Sampling History

Sequential sampling was first developed by Wald (1947) for economic and military applications. Since that time, sequential sampling has been adapted to the field of entomology and pest management (Waters 1955, Onsager 1976). In particular, sequential sampling has been developed to determine whether pest population densities have risen to levels at which economic damage is about to occur, such that pest management actions can be taken to reduce

the population densities and obviate the damage (Fowler and Lynch 1987). Effective utilization of sequential sampling in this way requires knowledge of the sampling distribution of the individuals comprising the pest population in their various stages, sample size estimation and statistical power, and the risks presented by so-called alpha and beta error. The former error arises when probabilistically the statistical analysis leads to the inference, wrongly, that a pest outbreak is about to occur, leading to wasted resources for controlling the pests when in fact there is no pending outbreak. This error is equivalent to incorrectly rejecting the null hypothesis of "no pending pest outbreak." By contrast, beta error arises probabilistically when the statistical analysis leads to the inference, wrongly, that no pest outbreak is about to occur when in fact one does occur, leading to no pest management action and to economic loss due to pest damage. The primary advantages of sequential sampling are as follows. It allows the determination of critical population size from actual field sampling during natural population density changes, which in turn allows sampling efforts to be terminated once a decision has been reached; and it permits the ability to set the values of alpha and beta error terms prospectively. Both processes (i.e., sampling and setting error rates) are under the control of the investigator, making sequential sampling an appealing process for pest managers. Two drawbacks to sequential sampling are as follows. Firstly, one must pre-determine the pest population densities at which economic injury and loss do not, and do, occur. This is problematic if the relationship between pest density and damage is variable and strongly nonlinear, if there is a lag time between changes in pest density and the appearance of damage, or if the relationship between density and damage is poorly researched. Secondly, the relationship between damage and economic loss is not a fixed relationship but rather subject to

market forces and commodity valuation, which vary continuously due to such factors as weather and economic scarcity. Under such circumstances, pest management decisions in one season may provide the desired protection from economic loss, but in another season may result in loss, simply if the value of the crop has declined relative to the cost of the pest management action from one season to the next. Indeed, a grower's lack of knowledge of the value of their crop at harvest is a psychological and managerial impediment to the implementation of sequential sampling for this reason.

It is clear that sequential sampling is a tool useful in classification of arthropod pest populations, and relating those classifications to economic injury and damage. The system typically reduces the classification to three categories: low population densities with no relevance to economic injury; to those intermediate population densities which may be growing towards densities at which injury may occur; to those higher densities at which economic injury and economic loss is a certainty. The transition from the first to the second categories requires establishment of a "lower threshold" or boundary separating the qualities of the two populations meaningfully by definition; and an "upper threshold" or boundary defining the transition from the population in the second category to the third category. These categories and boundaries are conceptual, empirical, statistical, and operational in nature. A large literature has developed around these themes, encompassing both the theoretical and applied areas and with sampling theory forming the framework (Fowler & Lynch 1987, Nyrop et al. 1986, Onsager 1976, Pedigo & Zeiss 1996, Waters 1955). More generally, sequential sampling involves one of three approaches (Lindblade et al. 2000). The direct approach requires selection of a critical population density below which risk of economic injury is low or

nil, and above which injury is imminent and risk is high. This binomial approach requires foreknowledge of the critical population density at which damage will occur and knowledge of the sampling and dispersion properties of the pest population in its various stages. The precision for various sample sizes is established by the method of the error as a percentage of the mean, and this method then provides an estimate of the accuracy of the classification of the population density at any given time point. The minimum sample size approach considers the number of samples containing at least one individual of the pest population and is useful at low pest densities, that is, when damage occurs when pest populations are low. Here, it is not the density of the pest population that is used as the assessment of the critical density, but rather the number of sampling units containing pests (vs. "zero" counts) that is related to the potential for economic damage. This system is also called "presence/absence" sampling and acquires a binomial distribution which greatly simplifies sampling and calculation of outbreak thresholds (Gerrard & Chiang 1970, Mogi et al. 1990, Wilson & Room 1983). Thirdly, the system developed by Wald (1947) involves the tripartite hierarchical classification described above, with low, intermediate, and critical pest population densities delineated by normal and outbreak threshold or decision lines.

Applications of Sequential Sampling

Applications of sequential sampling to arthropod pests in plant agriculture are legion but few in medical entomology. Knight (1964) identified the possibility that sequential sampling may be useful to classify mosquito larvae densities for control purposes. The first sequential sampling plan for mosquitoes was created by Wada (1965). This study explored the possibility for sequential sampling to classify mosquito larval density into 3 density levels for

control purposes. It elaborated sequential sampling plans for larvae of the Japanese encephalitis virus vector, *Culex tritaeniorhynchus*. However, there have been relatively few applications of sequential sampling to predict human disease in a variety of mosquito vectored diseases, such as encephalitis, dengue and malaria. For example, one of the first uses of sequential sampling to help prevent human disease, specifically Japanese encephalitis, was in the early 1970's. Wada et al. (1971) classified densities of larval *Culex tritaeniorhynchus summorsus* mosquitoes in rice fields. Also working with *Culex* larval densities, Mackey and Hoy (1977) explored the possibility of using sequential sampling to target pesticide applications in California rice fields. Lindblade et al. (2000) developed sequential sampling of adult *Anopheles gambiae* s.l. females, collected indoors, to predict human malaria epidemics in a highland region of southwestern Uganda. This was one of the initial studies to use adult mosquitoes as the monitoring variable, compared to larvae. The direct approach, the minimal sample size method, and the classic sequential sampling method were all compared using operating characteristic curves. They found that increased indoor resting densities of this mosquito species predicted epidemics about one month ahead of their occurrence, and suggested basic sampling plans for public health agencies.

Mogi et al. (1990) provided one of the first efforts to apply sequential sampling theory to predict epidemics of dengue, a mosquito-borne viral disease prevalent in the tropics and subtropics and involving only human beings as the vertebrate hosts. The mosquito stage sampled was the egg and ovitraps were the sampling device; presence/absence sampling was found to be adequate as predictor variable, obviating the need to count eggs thus saving time. This analysis resembles the minimum sample size approach discussed above. Lee and Singh

(1993) developed predictive models to classify the risk of increased prevalence of human dengue virus infection using adult *Aedes aegypti* mosquito density as the predictor variable. More recently, Barrera et al. (2006) used the density of *Aedes aegypti* pupae in Puerto Rico to predict risk of dengue epidemics (2006). They found that sequential sampling would require less sampling to reach a decision compared to other sampling plans. Barrera (2009) showed that sequential sampling may be a beneficial tool for confirming established epidemic thresholds to access their accuracy, using dengue as a model.

These studies demonstrate the possibilities to forecast human disease and predict when control would be optimal with sequential sampling. Furthermore, compared to other sampling plans, sequential sampling offers the possibilities for reducing sampling and targeted disease prevention efforts, which offer economic savings (Barrera et al 2006, Fowler & Lynch 1987, Kinght 1964, Mackey & Hoy 1977). In addition, it presents an opportunity to better understand the point when the pest population and virus pose the highest epidemic threat to humans. This raises a critical issue with regard to the application of sequential sampling systems to vector-borne diseases: what variable should be used to make the forecasts? From the above review it is evident that larval, pupal, and adult mosquitoes have been used, but never has pathogen infection in adult, female mosquitoes been used as the variable. Yet, the most proximate factor in risk of vector-borne infectious diseases in humans is the number of *infected* vectors in the total vector population. Abundant but uninfected vectors present no risk of infection to humans, thus abundance (or density, the variable most commonly used in sequential sampling of insects) is much less relevant. One of the innovative elements of this research is indeed the identification and utilization of the number of infected vectors as the monitoring variable. I

have been unable to find an example of where infected insect vectors of plant pathogens has been used in this way, although one study did consider the prevalence of infection of aster leafhoppers with aster yellows phytoplasma as a means of establishing the damage thresholds for carrots (Burkness et al. 1999). However, the number of leafhoppers was used as the sequentially sampled variable in that study and not the number of infected leafhoppers.

Objectives

The aim of this study expands the use of sequential sampling and its benefits to developing an early warning scheme for West Nile virus. This is the first study to examine the possibility of WNv prevention with sequential sampling. In addition, this study explores the possibilities of using virus infection in mosquito populations as the monitoring variable in sequential sampling compared to many past studies using mosquito population density estimates for mosquitoes (Barrera 2006, Lindblade 2000, Wada 1971). It will provide a direct assessment of the virus threat to humans. Two virus prevalence indices were tested for their potential as monitoring variables to forecast West Nile virus human cases. These two indices were the weekly number of positive West Nile virus tested *Culex* mosquito pools, as well as the percent of positive pools. Furthermore, in most entomological sequential sampling plans to date, sampling is done for successive spatial units (Barrera et al. 2006, Mackey & Hoy 1977, Onsager 1976, Wada 1965). However, for this study instead of having a spatial unit as a variable, a temporal unit (with one constant spatial scale) was used due to mosquitoes and virus presence having dynamic, fast paced changes in prevalence, as well as highly variable spatial aggregation that would best be detected in a temporal scale with a constant spatial unit. Data were fit to sequential sampling models for 3 different spatial scales (study site, county and

state) and the two different monitoring variables. After this, it was determined if sequential sampling would be a useful tool to predict West Nile virus in humans, and gain a better understanding of the point when the mosquito population and virus abundance pose a risk for an epidemic in humans. In addition, it was determined the spatial scale and monitoring variable that would be most appropriate for a sequential sampling model.

Materials and Methods

Field Site Data

In the southwest suburban Chicago, Illinois, study site mosquitoes were sampled from May to October since 2005. Mosquitoes were collected by 3 methods: CDC miniature light traps baited with CO_2, CDC gravid traps baited with rabbit pellet infusion, and battery powered backpack aspirators. Trap locations included a variety of natural and residential sites, and the number varied by year (Hamer et al. 2009). In 2005 there were 15 sampling sites, in 2006 there were 26 (with 14 sites carried over from 2005), and in 2007 there were 19 sites (with 14 sites from previous years) (Hamer et al. 2009).Trap locations were sampled every two to three weeks.

Once collected, mosquitoes were identified to species morphologically (Andreadis et al. 2005) and separated into bloodfed and non-bloodfed categories. Due to difficulty in observing the subtly distinguishing morphological features after rough handling in trap samples, *Culex pipiens* and *Culex restuans* were pooled together (Harrington & Poulson 2008). For this study only *Culex* mosquito results were utilized due to the fact they are the main vectors of West Nile

virus. The non-bloodfed mosquitoes were pooled in groups of up to 25 individuals in 2.0 mL centrifuge tubes by trap location and date. The bloodfed mosquitoes were stored individually.

The following methodology summarizes the virus testing procedure as previously described in Hamer et al. (2008). In the lab, reverse transcription, real time polymerase chain reaction (PCR) was utilized to test for the presence of West Nile virus RNA in each of the non-bloodfed pools. To homogenize the mosquitoes 1 mL of a 50:50 mixture of phosphate-buffered saline (PBS) and 2X lysis buffer (Applied Biosystems, Foster City, CA) and three #7 steel shots were added. Each pool was then subjected to a high-speed mechanical homogenizer (Retsch MM 300) for 4 minutes at 20 cycles/second. After pools were homogenized, each was centrifuged for 2 minutes at 13,000 rpm. Using an ABI Prism 6100 Nucleic Acid Prep Station and following the Tissue RNA Isolation Protocol (Applied Biosystems; P/N 4330252) RNA was extracted from mosquito pools, and the extracted RNA was eluted in a final volume of 60 µL of elution solution. Using real-time, reverse transcription polymerase chain reaction (RT-PCR), a region of the WNv RNA envelope gene was detected (Lanciotti et al. 2000). An ABI Prism 9700HT sequence detector at the Research Technology Support Facility at Michigan State University was used for thermocycling utilizing the Taq- Man One-Step RT-PCR Master Mix Protocol (Applied Biosystems; P/N 04310299). For bloodfed mosquitoes the abdomens were removed for blood meal analysis (Hamer et al. 2009) and the remaining carcass for each individual was tested as described above.

County & State Data

Data of the number of virus tested *Culex* mosquito pools for the State of Illinois and Cook County were available from the Illinois Department of Public Health for the years 2004 to 2006 and we provided by research collaborator Dr. Marilyn O. Ruiz of the University of Illinois.

Human Data

The number of human cases by week within each of the spatial units under study were obtained from records compiled by the Illinois Department of Public Health, and were completely anonymous with regard to name, address, physician, locality, and course of disease. Such data are publicly available and exempt from review for human subjects research by institutional review boards for this reason. In addition, for the study site human cases were obtained by creating a 5-km buffer around the field sites using ArcGIS 9.2 software (Environmental Systems Research Institute, Redland, CA).

Monitoring Variables

Two different monitoring variables were analyzed to determine the variable that would be best suited for forecasting human disease. These variables were the number of WNv tested positive *Culex* mosquito pools per week for each spatial scale; and the percent of pools that tested positive for WNv at each spatial scale. The variable best suited would be the one that would cross both decision lines first in years where human cases occurred, leading to an earlier warning and more opportunity to react. For all years and spatial scales sequential sampling

plans were created for both monitoring variables to determine the most successful variable for forecasting WNv human cases.

To facilitate comparing the two potential monitoring variables, a time series cross correlation analysis was conducted between the number or percent of positive pools per week and the number of human cases per week. This was done for each year, spatial scale and monitoring variable. The cross correlation function would indicate if there was a different significant lag time between the individual monitoring variables and human cases. This would be important in a forecasting tool. Ideally, the best monitoring variable would have the longest significant time lag correlation with human cases. This would potentially result in earlier warning times.

Aggregation Analysis

All monitoring variables were considered to follow an aggregated distribution with a likely fit to the negative binomial distribution, where the sample variance to sample mean ratio is greater than one for all monitoring variables, years and spatial scales. In addition, the distribution of the monitoring variables for each spatial scale were fit to a negative binomial distribution using PROC GENMOD procedure in Statistical Analysis Software (SAS), version 9.2 (SAS Institute, Inc., Cary, North Carolina, USA). The negative binomial is characterized by the data being clumped compared to being randomly dispersed. This would be logical for virus prevalence in mosquitoes over time, being that there is a peak time of virus activity that occurs in the season with little virus being seen before or after the season is done.

Furthermore, all sampling plans were determined if a common dispersion parameter existed among years at each spatial scale. To be able to create sequential sampling plans in the future for predicting disease epidemics before they occur, establishing that the population dispersion parameter (k) values don't differ significantly between years is critical. If k values are significantly different between years, it is much more difficult to construct a plan when the dispersion parameter fluctuates significantly. The k values were analyzed between years for each spatial scale and monitoring variable to determine if there was a common k, using the PROC GENMOD procedure in SAS, version 9.2. In this manner a common k was estimated for each spatial scale and monitoring variable.

Error Level Determination

Next, the error levels were determined. In this case, the alpha error (α) occurs when a WNv human epidemic would be predicted when in fact one never transpires. The beta error (β) would occur when no human epidemic is predicted, and one arises. These values should be considered for each individual sampling plan. Utilizing sequential sampling to predict human disease, results in the beta error being the more crucial of the two. A false prediction of human disease would be better received over failure to predict an epidemic.

To determine the most suitable error levels, we assessed how changes in the values would affect the decision lines. As alpha error is increased, all values in decision line 1 (lower threshold) increase, and values in decision line 2 (upper threshold) decrease. The decrease in values for decision line 2 is of a greater magnitude than for the increase in decision line 1. The lower decision line 2 becomes the greater the chances for forecasting an epidemic. Likewise, as

beta error increases, the values of the decision line 1 increase, and values of the decision line 2 decrease. In this instance, the magnitude is greatest for the increase in decision line 1 values. This means that as beta increases, the bottom line rises. This would result in a greater chance in slowing sampling throughout the season because more values would fall below both decision lines. From this assessment the α and β were set to 0.05 for all spatial scales and years.

Threshold Determination

The next parameters that had to be determined were the threshold parameters. These are the values that determine the value of the y-intercepts of the two decision lines. They establish the amount of *Culex* WNv positive pools per week (raw and percent) that would be needed to predict high and low risk of an epidemic occurring, and must be determined empirically. Based on the data available, the lower threshold for the number of positive pools was set to 1, 3, and 6 and the upper threshold to 10, 30 and 60 for the study site, county and state scales respectively. For the percent positive pools the lower threshold for all scales was set to 2 and the upper threshold to 20. For the percent positive pools all scales had the same threshold because it was assumed that the percentage of positive pools to predict an epidemic would be the same among spatial scales. However, with the number of positive pools there is overall more testing done as the spatial scale increases, so the number of positive pools to predict human disease would increase with spatial scale.

These thresholds help to form two hypotheses for each spatial scale and monitoring variable. The null hypothesis is no risk of an epidemic, or more precisely the number or percent of positive pools per week is below the lower threshold (H_0: $\bar{X}' \leq \bar{X}_1$). The alternative

hypothesis is a risk of an epidemic, meaning the number or percent of positive pools per week is above the upper threshold (H_1: $\bar{X}' \geq \bar{X}_2$).

Operating Characteristic & Average Sample Number Curves

To assess the functionality of the plan, operating characteristic (OC) and average sample number (ASN) curves were created for each monitoring variable and spatial scale. The formulas used for both the OC (Tables 1 & 2) and ASN (Tables 3 & 4) curves were derived from Onsager 1976. The ASN curves permits assessment of the number of sample-weeks that would be required to make a decision if the epidemic threat is a low or high probability. In principle, the ASN most suitable would be one that has the fewest number of weeks to make a decision at low and high average number or percent of positive pools. Additionally, for the average number or percent of positive pools in between the low and high values, the number of weeks to make a decision should be practical. For example, if the number of weeks needed to make a decision exceeded the virus amplification season, then obviously it would be an unsuitable plan.

The OC curve determines the probability of accepting the first hypothesis, and concluding there is no epidemic for any given number or percent of positive pools per week, and the inverse determines the probability of accepting the alternative hypothesis, that there is an epidemic. The OC curve that would be most appropriate would have a high probability of accepting the first hypothesis at very low averages of number or percent of positive pools per week, and a decrease in probability of acceptance as the average number or percent of positive pools per week increased.

Table 1. Operating Characteristic Curve Formulae-Number of Positive Pools

X'				LP Result		
Study Site	***County***	***State***	**LP Formula**	***Study Site***	***County***	***State***
0	0	0	1*	1	1	1
$\bar{X}_1 = 1$	$\bar{X}_1 = 3$	$\bar{X}_1 = 6$	$1 - \alpha$	0.95	0.95	0.95
$b = 2.71$	$b = 7.75$	$b = 15.47$	$\frac{h_2}{(h_2 - h_1)}$	0.50	0.50	0.50
$\bar{X}_2 = 10$	$\bar{X}_2 = 30$	$\bar{X}_2 = 60$	β	0.05	0.05	0.05

*Formulae derived from Onsager 1976

Table 2. Operating Characteristic Curve Formulae-Percent Positive Pools

X'				LP Result		
Study Site	***County***	***State***	**P Formula**	***Study Site***	***County***	***State***
0	0	0	1*	1	1	1
$\bar{X}_1 = 1$	$\bar{X}_1 = 1$	$\bar{X}_1 = 1$	$1 - \alpha$	0.95	0.95	0.95
$b = 3.38$	$b = 3.29$	$b = 3.23$	$\frac{h_2}{(h_2 - h_1)}$	0.50	0.50	0.50
$\bar{X}_2 = 20$	$\bar{X}_2 = 20$	$\bar{X}_2 = 20$	β	0.05	0.05	0.05

*Formulae derived from Onsager 1976

Table 3. Average Sample Number Formulae-Number of Positive Pools

X'					ASN Result		
Study Site	***County***	***State***	**LP**	**ASN Formula**	**Study Site**	**County**	**State**
0	0	0	1	$\frac{h_1}{-b}$	3.43	7.27	4.47
$\bar{X}_1 = 1$	$\bar{X}_1 = 3$	$\bar{X}_1 = 6$	0.95	$\frac{h_2 + (h_1 - h_2)(LP)}{\bar{X}' - b}$	4.89	10.68	6.57
$b = 2.71$	$b = 7.75$	$b = 15.47$	0.50	$\frac{h_1 * h_2}{-\left(\frac{b^2}{k} + b\right)}$	4.38	9.30	5.72
$\bar{X}_2 = 10$	$\bar{X}_2 = 30$	$\bar{X}_2 = 60$	0.05	$\frac{h_2 + (h_1 - h_2)(LP)}{\bar{X}' - b}$	1.15	2.28	1.40

*Formulae derived from Onsager 1976

Table 4. Average Sample Number Formulae-Percent Positive Pools

X'					ASN Result		
Study Site	***County***	***State***	**LP**	**ASN Formula**	**Study Site**	**County**	**State**
0	0	0	1	$\frac{h_1}{-b}$	2.83	4.77	7.93
$\bar{X}_1 = 1$	$\bar{X}_1 = 1$	$\bar{X}_1 = 1$	0.95	$\frac{h_2 + (h_1 - h_2)(LP)}{\bar{X}' - b}$	3.62	6.17	10.33
$b = 3.38$	$b = 3.29$	$b = 3.23$	0.50	$\frac{h_1 * h_2}{-\left(\frac{b^2}{k} + b\right)}$	2.79	4.69	7.79
$\bar{X}_2 = 20$	$\bar{X}_2 = 20$	$\bar{X}_2 = 20$	0.05	$\frac{h_2 + (h_1 - h_2)(LP)}{\bar{X}' - b}$	0.52	0.84	1.38

*Formulae derived from Onsager 1976

Sequential Sampling Plans

Equations utilized by Pedigo and Zeiss (1996) for negative binomial distributions were used to develop the decision lines for each sampling plan (Table 5). After generating these lines, for every unit of time (week, number of days, etc.) the monitoring variable was cumulated and plotted with the decision lines. If the cumulative point falls below both decision lines then sampling could be slowed or possibly stopped because the variable in question is low enough that virus amplification in the near future would be very unlikely (Figure 1-Area C). If the point falls in between the two lines, not enough information is known to determine if the insect vector could or could not pose a threat to the human population (Figure 1-Area B). At this point sampling may be intensified due to possible amplification. When the point falls above both lines, this would indicate that the virus is amplifying in the mosquito population and could pose a threat to the human population (Figure 1-Area A). This would be the point to implement preventive measures to prevent human disease from occurring. The most suitable sequential sampling plans would forecast risk of human West Nile virus well before the data of onset of any human cases.

Multiple years' data were fit to sequential sampling plans at 3 different spatial scales (study site, county and state) and for two different monitoring variables (number and percent positive pools). Sequential sampling plans were generated for each spatial scale based on data availability for each year. Chicago study site plans were created for years 2005-2007 and Cook County, IL and Illinois plans for years 2004-2006. In addition available human case data was plotted with sequential sampling plans to determine if sequential sampling plans would have

Table 5. Sequential Sampling Formulae for Negative Binomial Distributions

Sequential Sampling Parameters	
Formula	***Definition***
m_1	Lower threshold
m_2	Upper threshold
$p_1 = \frac{m_1}{k}$	Contrived value, k=dispersion parameter
$p_2 = \frac{m_2}{k}$	Contrived value, k=dispersion parameter
$q_1 = 1 + p_1$	Contrived value
$q_2 = 1 + p_2$	Contrived value
$b=k^* \frac{\ln\left(\frac{q_2}{q_1}\right)}{\ln\left(\frac{p_2{}^*q_1}{p_1{}^*q_2}\right)}$	Slope
$h_1= \frac{\ln\left(\frac{\beta}{1-\alpha}\right)}{\ln\left(\frac{p_2{}^*q_1}{p_1{}^*q_2}\right)}$	Lower decision line y-intercept
$h_2= \frac{\ln\left(\frac{1-\beta}{\alpha}\right)}{\ln\left(\frac{p_2{}^*q_1}{p_1{}^*q_2}\right)}$	Upper decision line y-intercept
$d_1 = bx + h_1$	Lower decision line equation, x = sample number in time
$d_2 = bx + h_2$	Upper decision line equation, x=sample number in time

*Formulae derived from Pedigo & Zeiss 1996

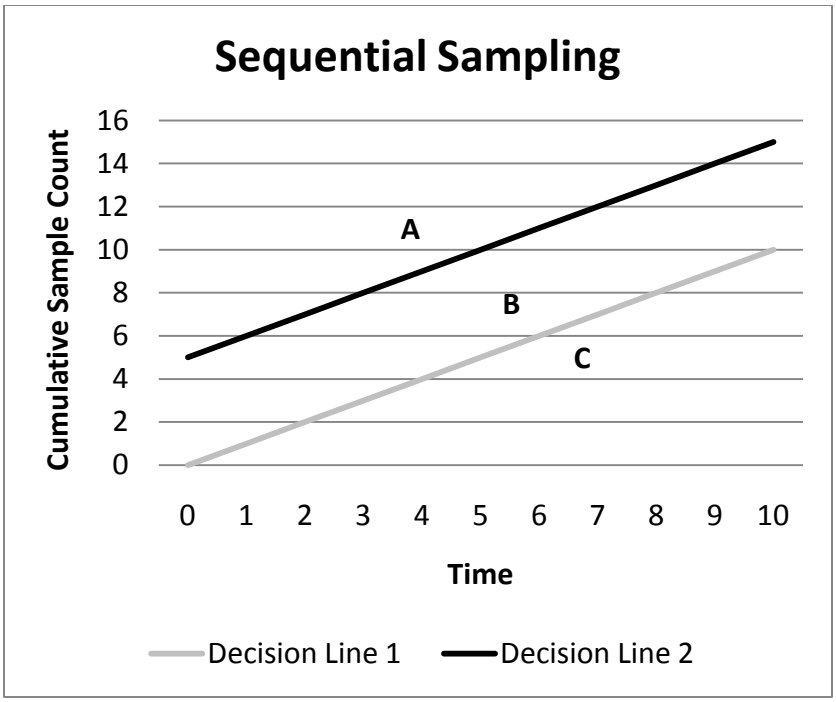

Figure 1. Example of a sequential sampling graph showing the 3 possible outcomes. For disease forecasting, if the cumulative count of virus-positive samples or pools falls in zone C, the statistical inference would indicate that the likelihood of human cases is very low and sampling could be slowed or even stopped, depending on the plan. If the cumulative count falls in zone B, the inference is that a larger sample and continued sampling is required before the risk of human cases can be assumed. If the count falls within zone A, then the inference is that mosquito infection has risen to an infection prevalence where the risk of human infection is high and human cases are highly likely to occur (unless some management action is taken to interdict transmission).

forecasted a high probability of human cases before they occur. Human case data was available for years 2004-2006 for the county and state spatial scales and years 2005-2007 study site scale.

Results

Common Dispersion Parameter (k)

All spatial scales and monitoring variables had a common dispersion parameter. For the number of positive pools the common k was 0.4327, 0.1799 and 0.2914 for study site, county and state scales. For the percent positive pools the common k was 0.3864, 0.2197, and 0.1289 for the study site, county and state scales.

Cross Correlation Function Analyses

For the study site spatial scale in years 2005-2007, the cross correlation time series analysis showed that the number of positive pools was positively correlated with the distribution of human cases in a time lag in weeks of 1, 5, and 5 for each year respectively (Figure 2 A-C). The percent positive pools became positively correlated in a time lag in weeks of 1, 4, and 5 for each year respectively (Figure 2 D-F). For the county spatial scale in the years 2004-2006, the cross correlation time series analysis showed that the number of positive pools was positively correlated in a time lag of 4 weeks for all years (Figure 3 A-C). The percent positive pools became positively correlated in a time lag of 3, 3, and 4 weeks (Figure 3 D-F). For the state spatial scale in the years 2004-2006, the cross correlation analysis resulted in significant time lags of 4, 4, and 5 weeks respectively for each year (Figure 4 A-C). For the

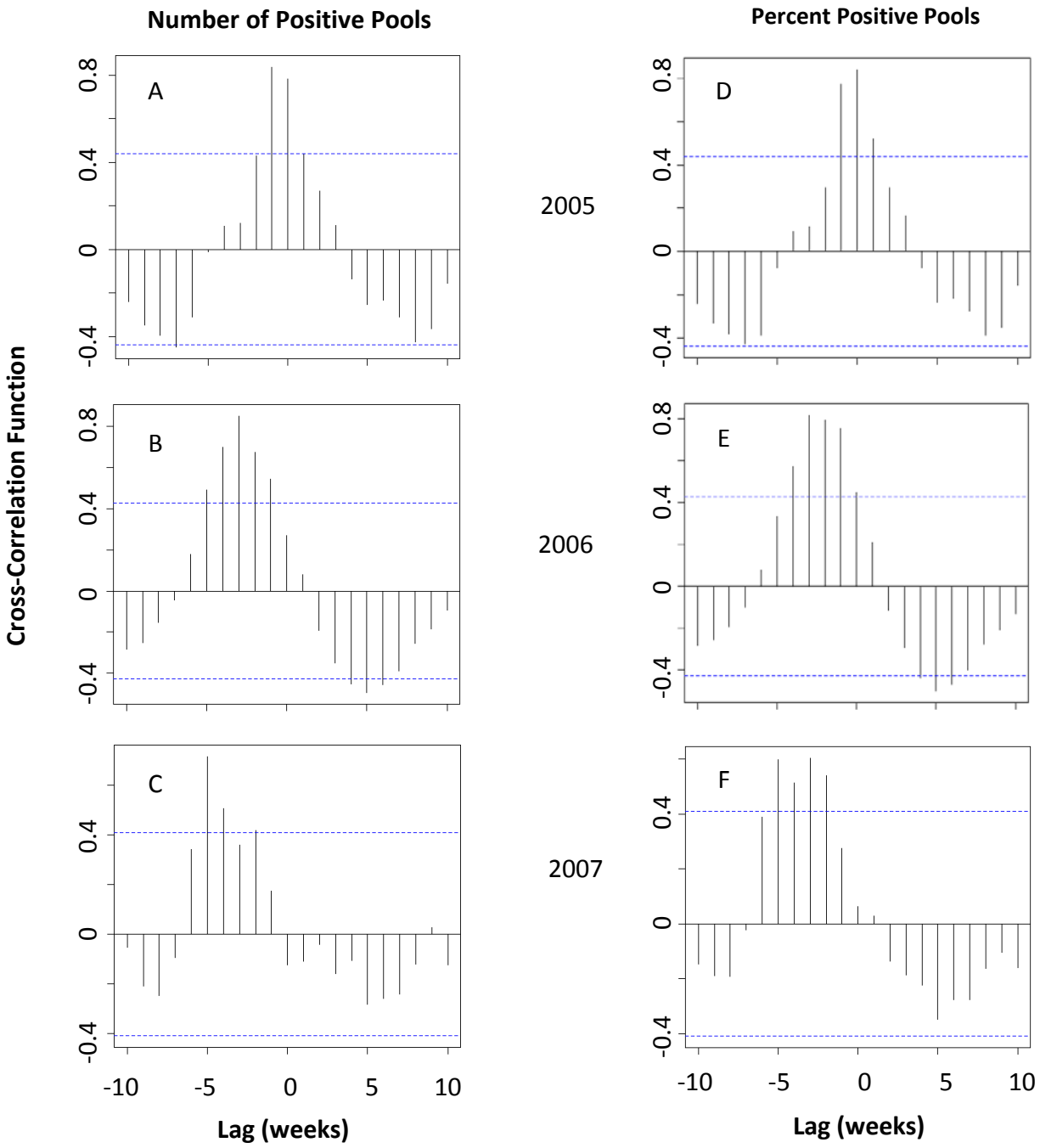

Figure 2. Time series, cross correlation analyses between the Suburban Cook County, Illinois human cases and the monitoring variables, both grouped by week, for the study site spatial scale. A, B and C show the cross correlation analysis between the number of positive pools and human cases for the years 2005, 2006 and 2007, respectively. D, E and F show the cross correlation analysis between the percent positive pools and human cases for the years 2005, 2006 and 2007, respectively.

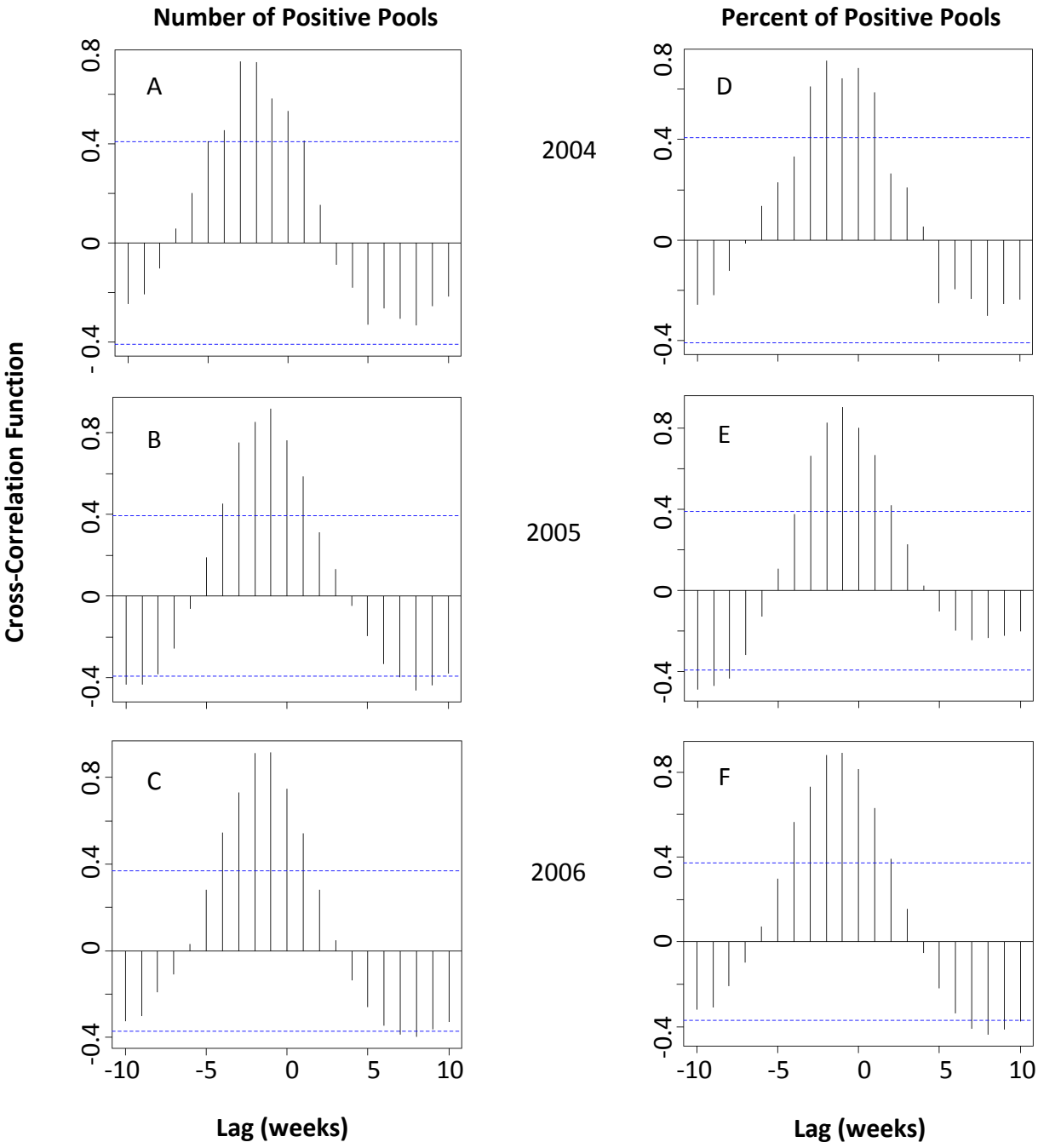

Figure 3. Time series, cross correlation analyses between Cook County, Illinois human cases and the monitoring variables, both grouped by week, for the county spatial scale. A, B and C show the cross correlation analysis between the number of positive pools and human cases for the years 2004, 2005 and 2006, respectively. D, E and F show the cross correlation analysis between the percent positive pools and human cases for the years 2004, 2005 and 2006, respectively.

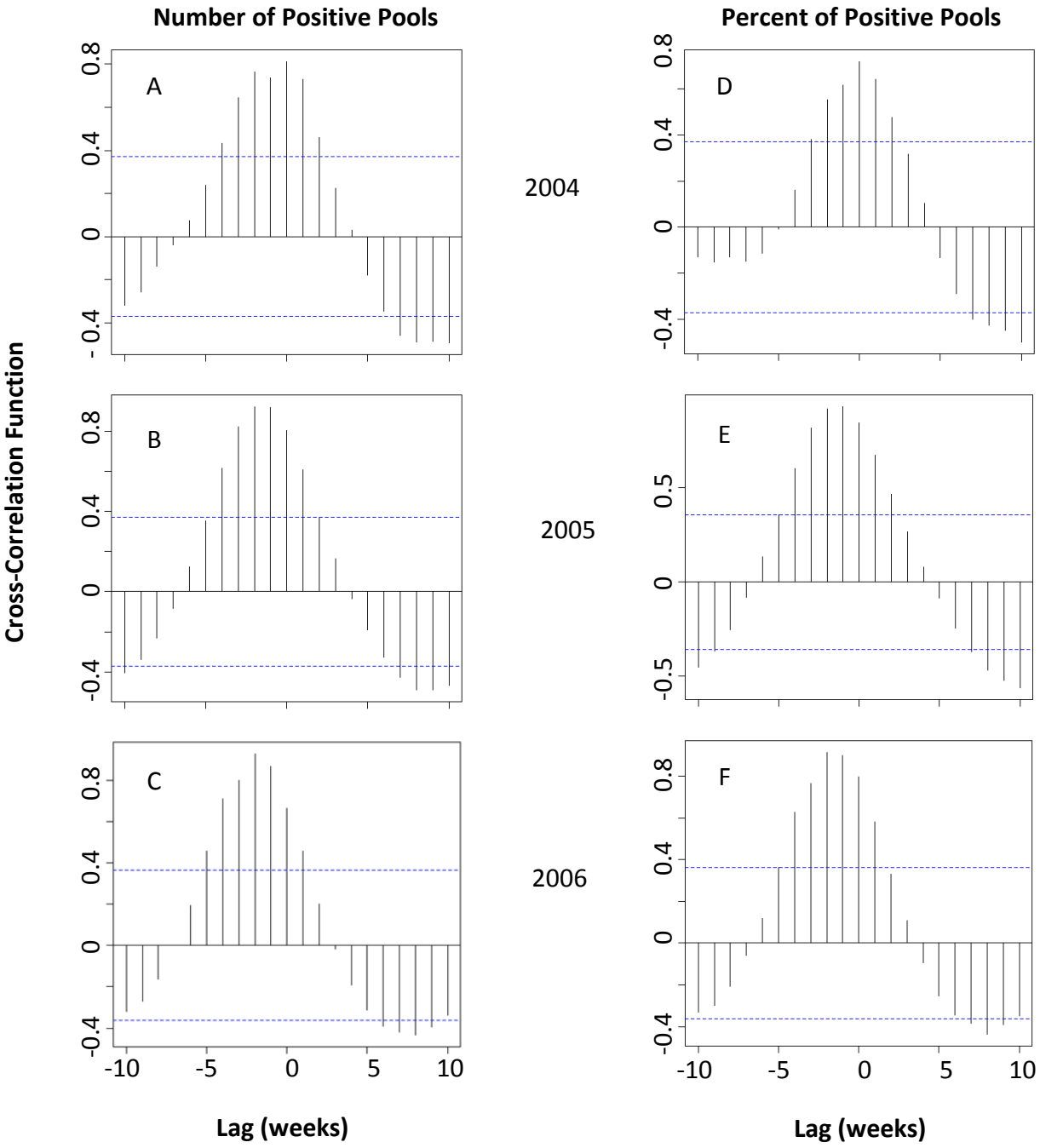

Figure 4. Time series, cross correlation analyses between Illinois human cases and the monitoring variables, both grouped by week, for the state spatial scale. A, B and C show the cross correlation analysis between the number of positive pools and human cases for the years 2004, 2005 and 2006, respectively. D, E and F show the cross correlation analysis between the percent positive pools and human cases for the years 2004, 2005 and 2006, respectively.

percent positive pools the cross correlation analysis resulted in significant time lags of 3, 4, and 4 weeks respectively for each year (Figure 4 D-F).

Operating Characteristic Curves

The operating characteristic (OC) curve for all spatial scales and each monitoring variable displayed the ideal form (Figure 5). Results for the OC curves for all scales and monitoring variables demonstrated that increasing the number of positive pools per week resulted in a higher probability of rejecting H_0 and concluding there is a risk for an epidemic.

Average Sample Number Curves

The average sample number (ASN) curve for all spatial scales and monitoring variables confirm the lowest number of weeks to determine a decision would be for values nearing either of the two thresholds levels (Figure 6). The estimated maximum number of weeks to make a decision for study site, county, and state for number of positive pools was 4.9, 10.7, and 6.6 respectively. For the percent positive pools the estimated maximum number of weeks to make a decision was 3.6, 6.2 and 10.2 for study site, county and state. All values were practical in regards to the time to reach a decision within the West Nile virus amplification season, which occurs between May and October.

Sequential Sampling Plans: Study Site

For the suburban Chicago, Illinois study site scale in all years and for both monitoring variables the cumulative positive pool line crossed the upper decision line. Furthermore, for all

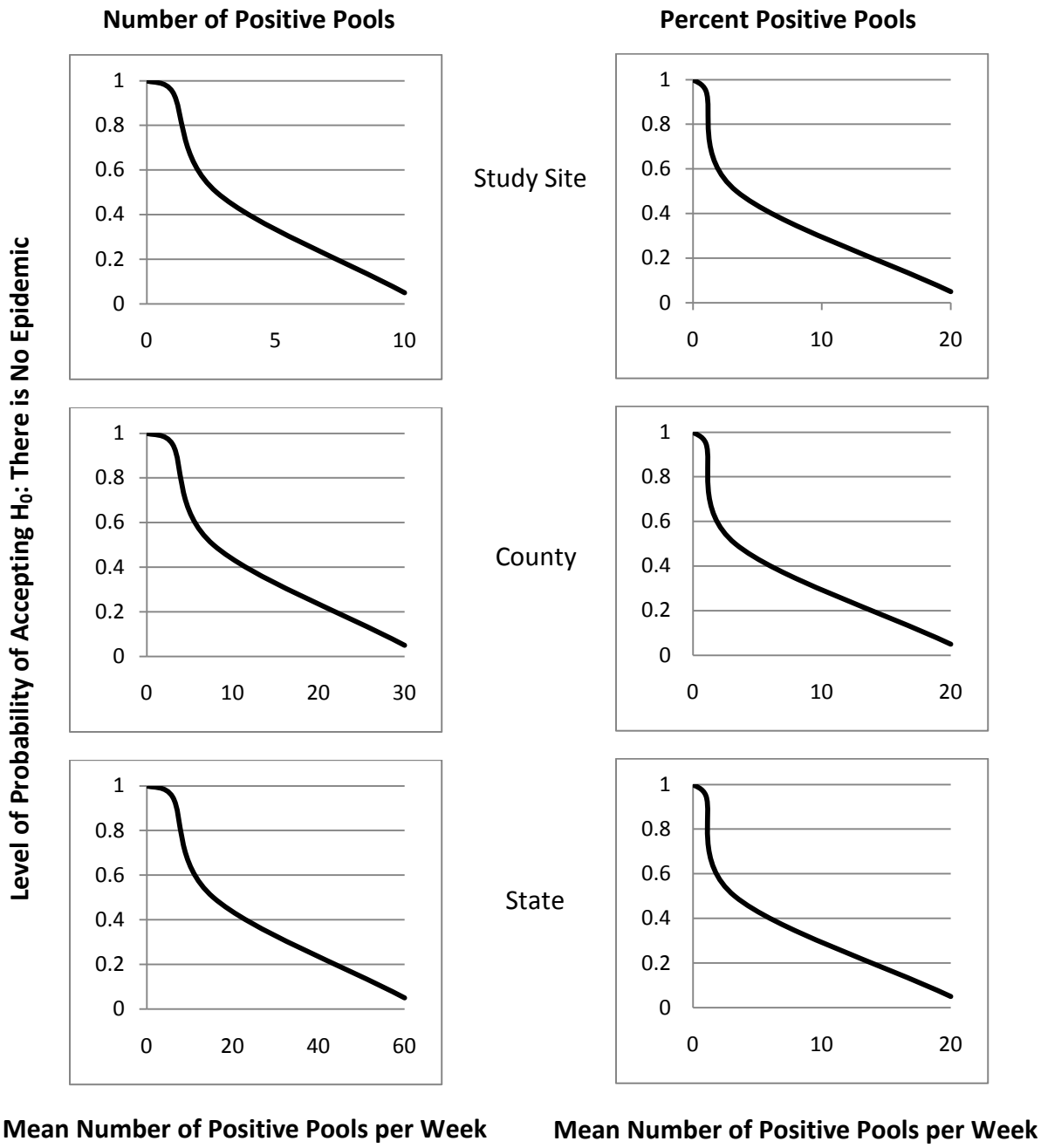

Figure 5. Operating Characteristic Curves for each spatial scale and monitoring variable.

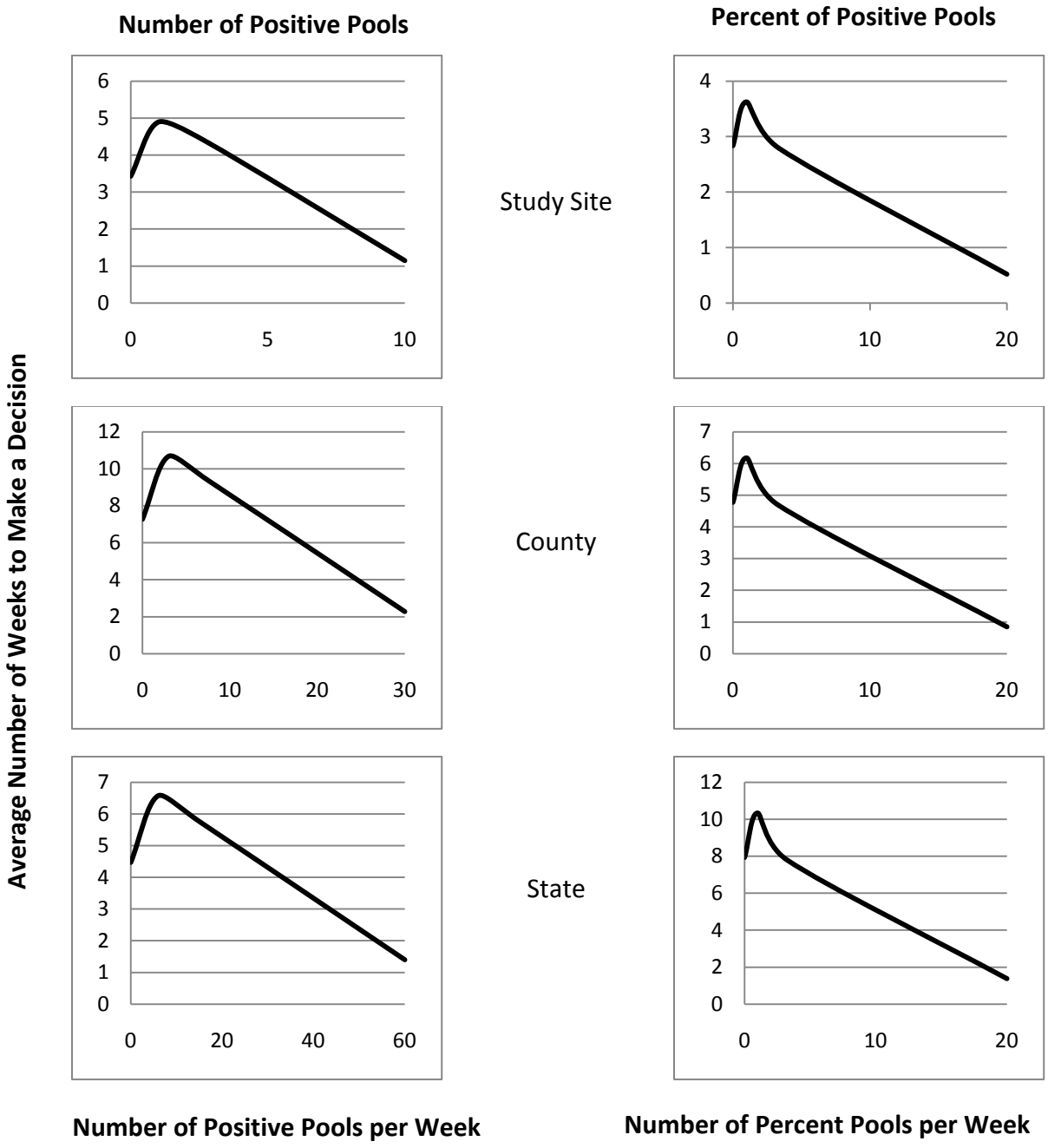

Figure 6. Average sample number curves for each spatial scale and monitoring variable.

years and monitoring variables, the cumulative positive pool line crossed the upper decision line before a majority of the human cases started to appear.

In 2005, there were 223 WNv positive pools out of 665 tested, and 20 human cases. This year was noted for a high number of West Nile virus cases that occurred in a very short amount of time. The cumulative number and percent of positive pools crossed the upper decision line in week 30, the same week human cases started to appear (Figure 7).

In 2006, there were 200 WNv positive pools out of 1063 tested, and 42 human cases. It should be noted that the increase in overall pools and human cases was due to study site increasing in mosquito trapping area. From 2005 to 2006 the study site increased from 856 hectares to 1218.1 hectares. The cumulative number and percent of positive pools crossed the upper decision line in week 29, 2 weeks prior to human cases occurring (Figure 8).

In 2007, there were 95 WNv positive pools out of 715 tested, and 14 human cases. This year had an unusual human case distribution, with one human case occurring very early in the season, week 23, and several single instances a few weeks after (28 & 30), before seeing amplification. The cumulative number and percent of positive pools line crossed the upper decision line in week 32, 9 weeks after the first human case, and one week before amplification (Figure 9).

Sequential Sampling Plans: County

At the Cook County, Illinois spatial scale in all years the cumulative number and percent of positive pools line crossed the upper decision line.

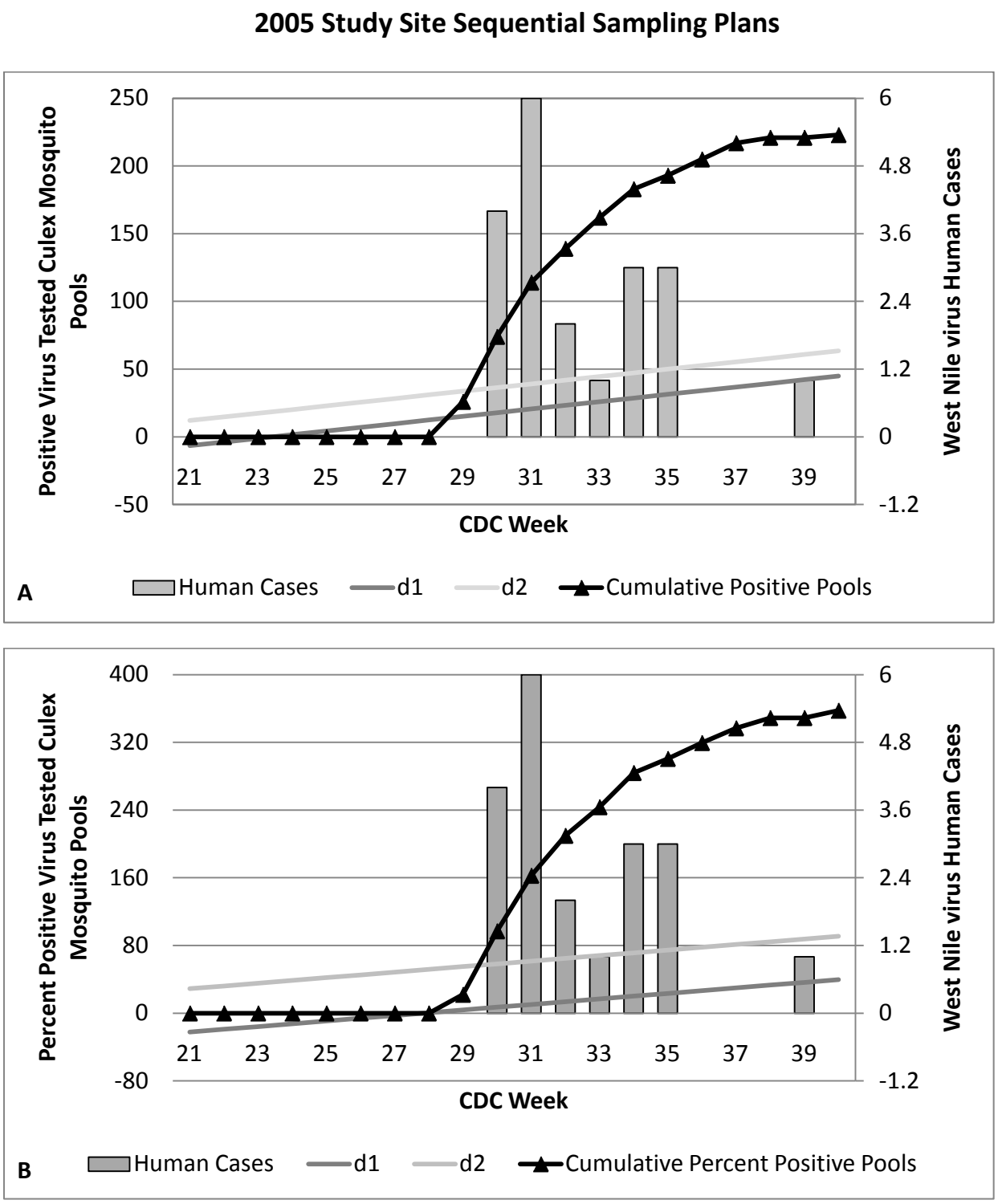

Figure 7. Sequential sampling plans for the study site in Suburban Chicago, Illinois for 2005. (A) The plan using the number of West Nile virus tested *Culex* mosquito pools. (B) The plan using the percent of West Nile virus tested *Culex* mosquito pools.

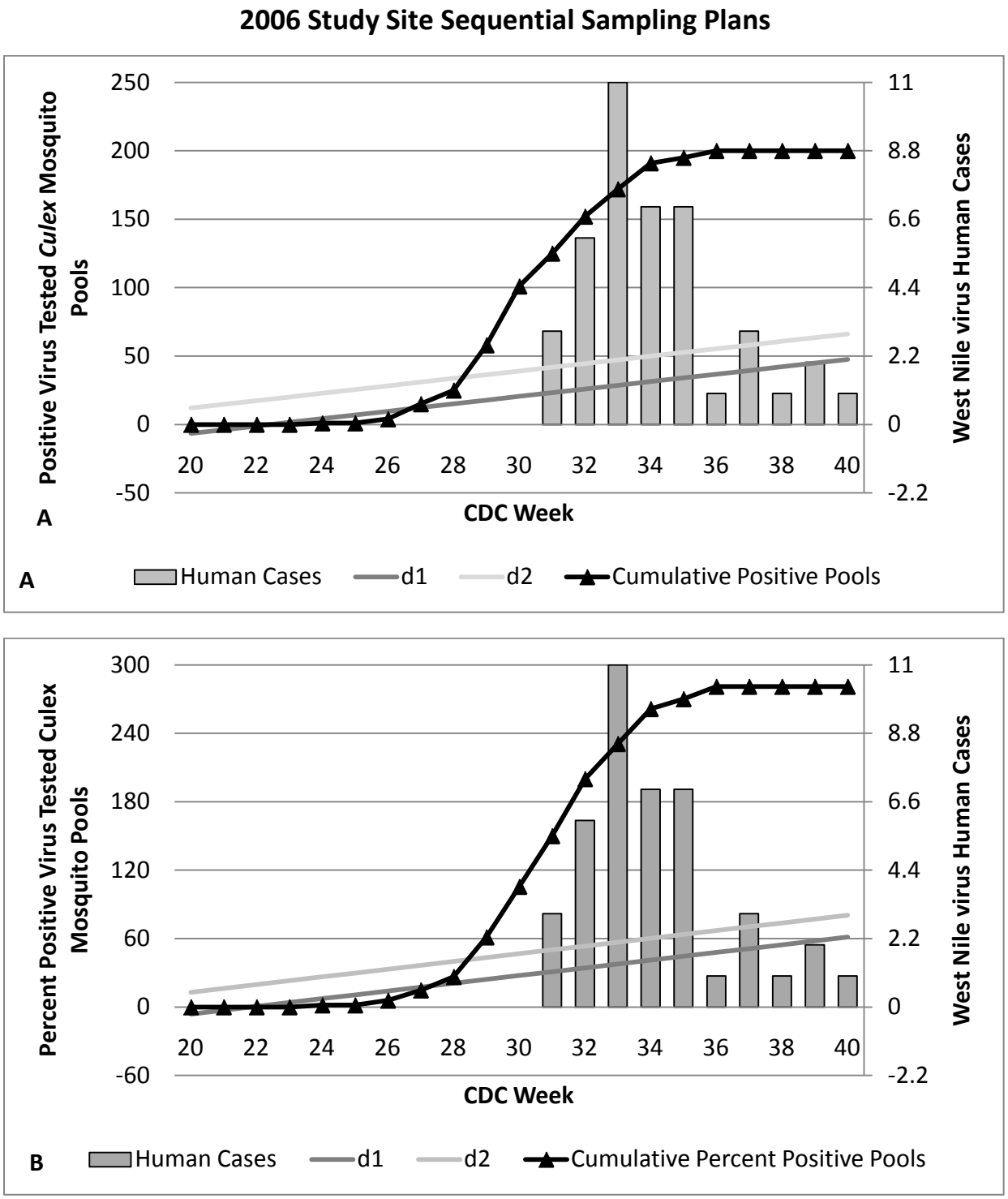

Figure 8. Sequential sampling plans for the study site in Suburban Chicago, Illinois for 2006. (A) The plan using the number of West Nile virus tested *Culex* mosquito pools. (B) The plan using the percent of West Nile virus tested *Culex* mosquito pools.

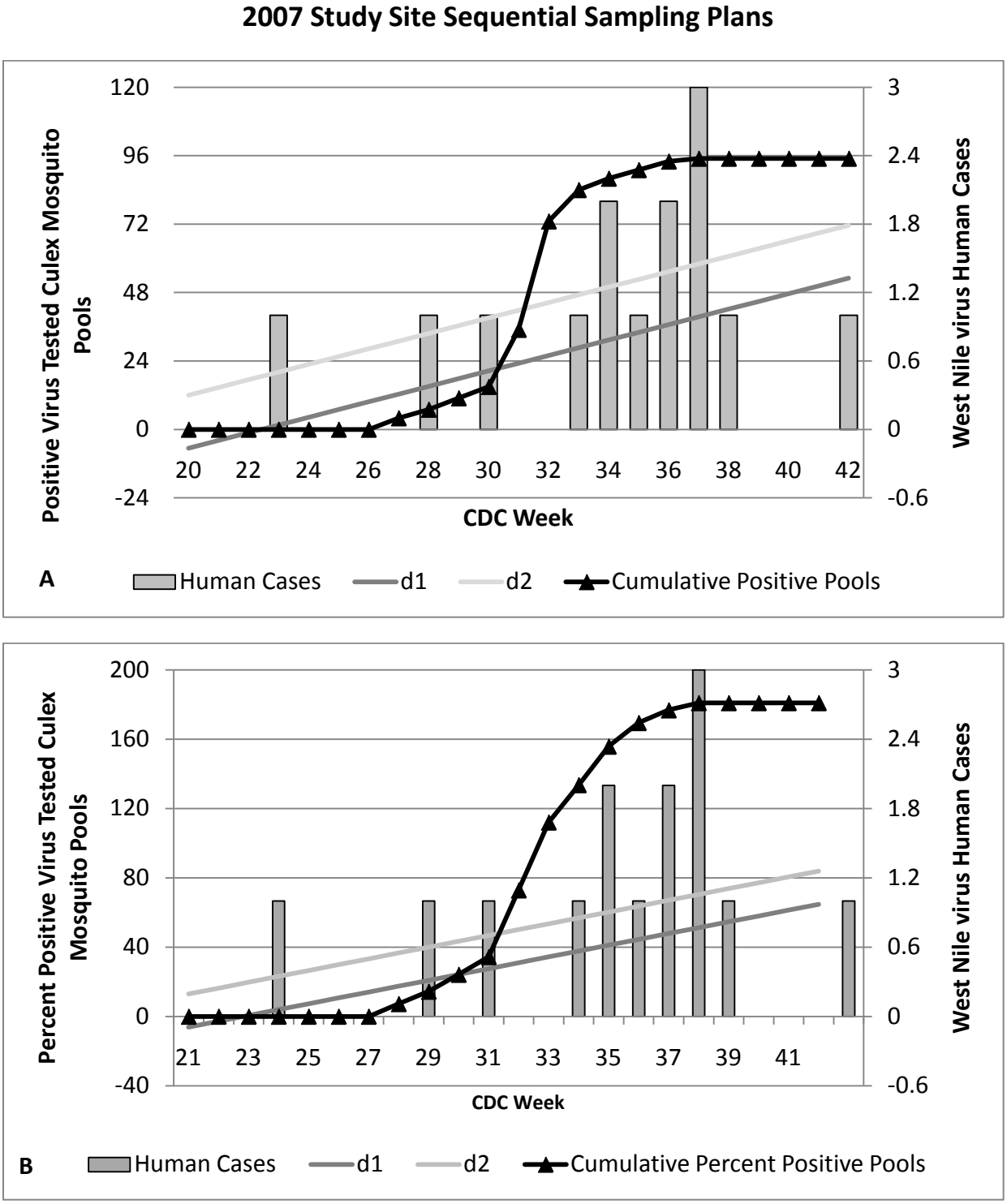

Figure 9. Sequential sampling plans for the study site in Suburban Chicago, Illinois for 2007. (A) The plan using the number of West Nile virus tested *Culex* mosquito pools. (B) The plan using the percent of West Nile virus tested *Culex* mosquito pools.

In 2004, there were 652 positive pools out of 4,829 tested, and there were 23 human West Nile virus cases. This year the cumulative number, as well as the percent, of positive pools crossed the upper decision line in week 31, the week human cases started to occur (Figure 10).

In 2005, there were 1,557 positive pools out of 4,977 tested, and there were 134 human cases. This year the cumulative number and percent of positive pools crossed the upper decision line in week 29, 3 weeks after the first human case occurred, and the week when human cases started to amplify (Figure 11).

In 2006, there were 1,633 positive pools out of 7,041 tested, and there were 86 human cases. This year the cumulative number of positive pools crossed the upper decision line in week 29, 2 weeks after the first human case occurred, and the same week human cases started to amplify (Figure 12A). The percent positive pools crossed the upper decision line one week later in week 30 (Figure 12B).

Sequential Sampling Plans: State

At the Illinois spatial scale in all years the cumulative number and percent of positive pools line crossed the upper decision line. In 2004, there were 1,600 positive pools out of 13,258 tested, and there were 60 human WNv cases. This year the cumulative number of positive pools crossed the upper decision line in week 24, one week after the first human case, but 3 weeks before we started to see amplification (Figure 13A). The cumulative percent positive pools crossed the upper decision line in week 23, the week the first human case

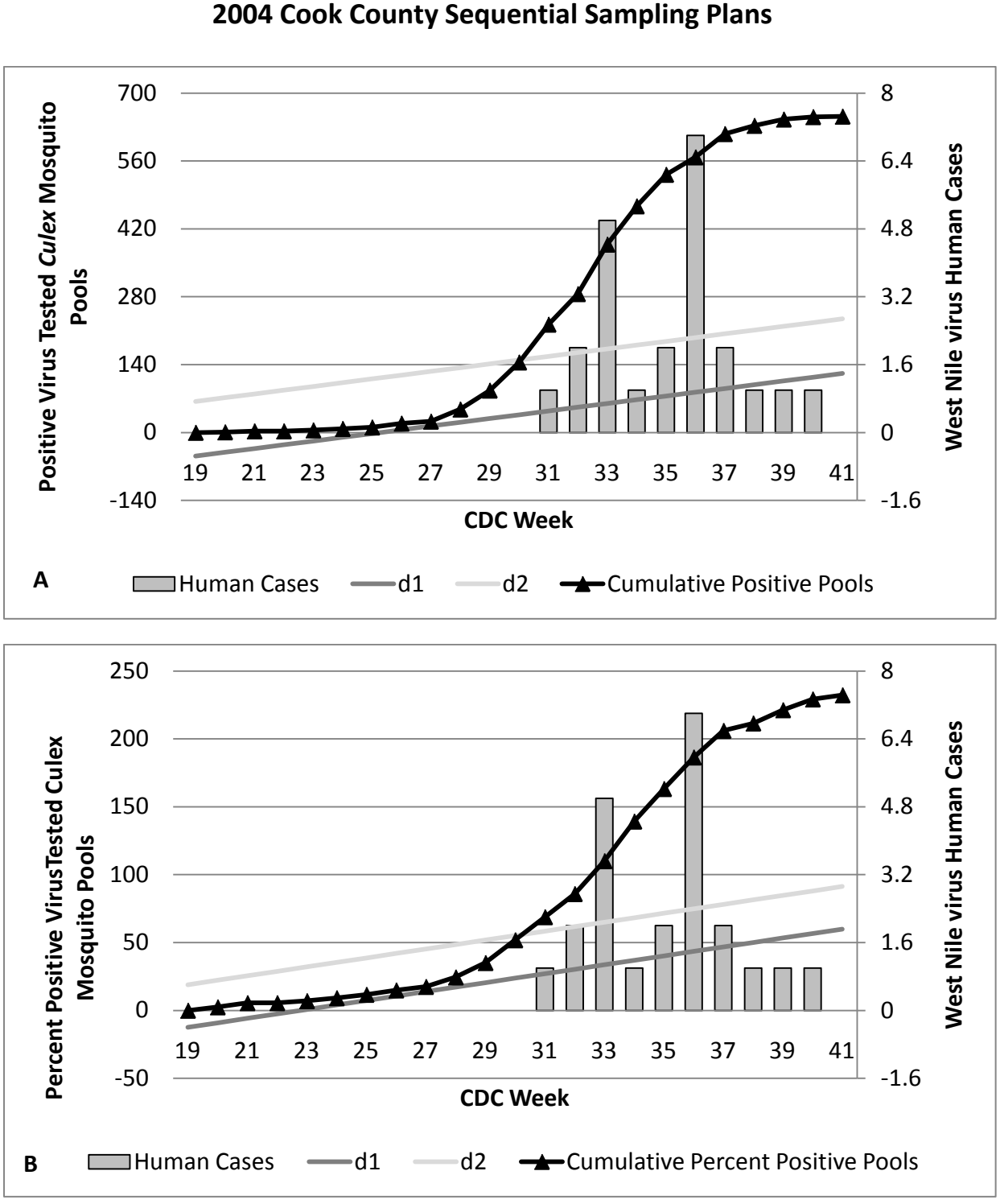

Figure 10: Sequential sampling plans for the county scale in Cook County, Illinois for 2004. (A) The plan using the number of West Nile virus tested *Culex* mosquito pools. (B) The plan using the percent of West Nile virus tested *Culex* mosquito pools.

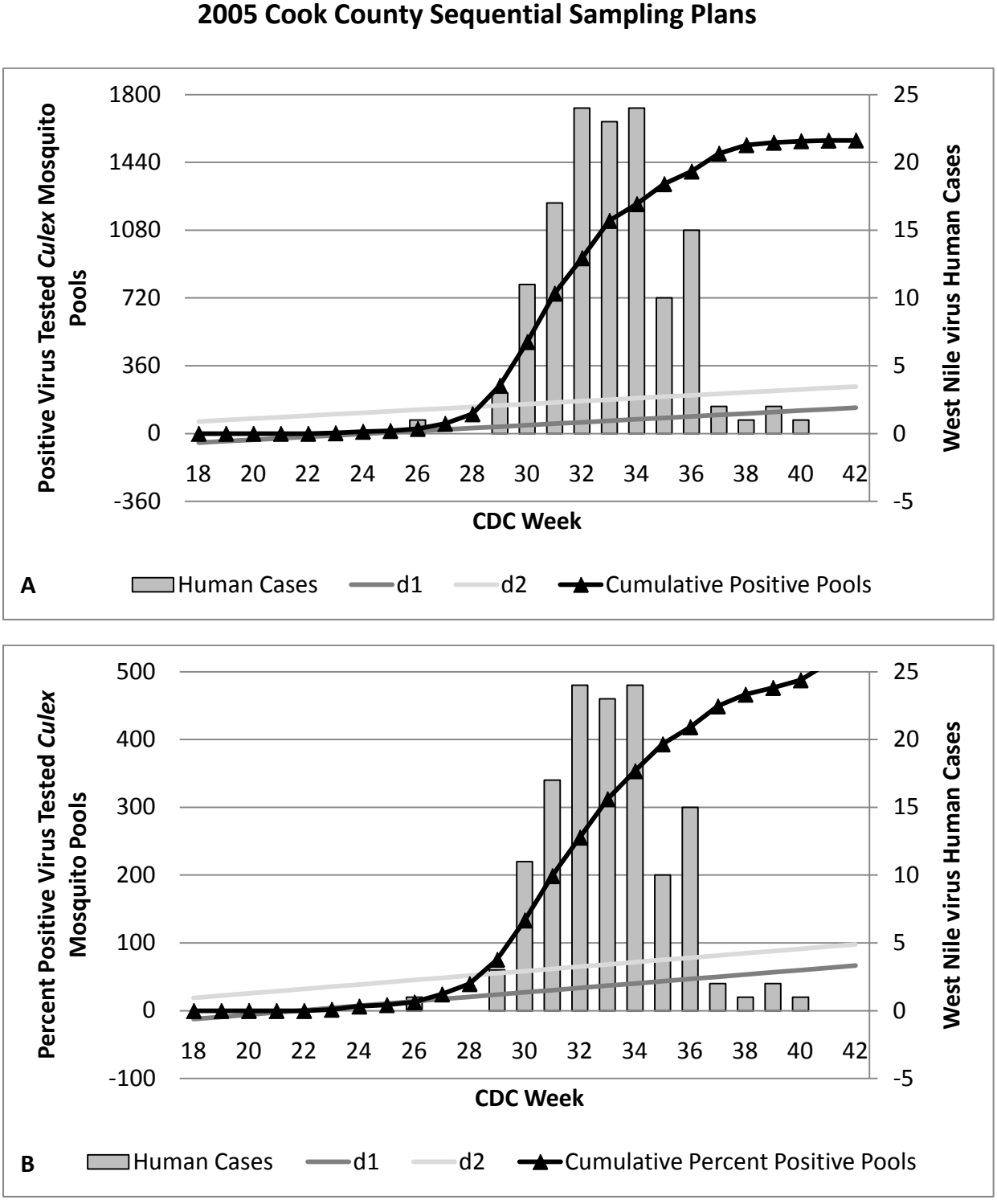

Figure 11: Sequential sampling plans for the county scale in Cook County, Illinois for 2005. (A) The plan using the number of West Nile virus tested *Culex* mosquito pools. (B) The plan using the percent of West Nile virus tested *Culex* mosquito pools

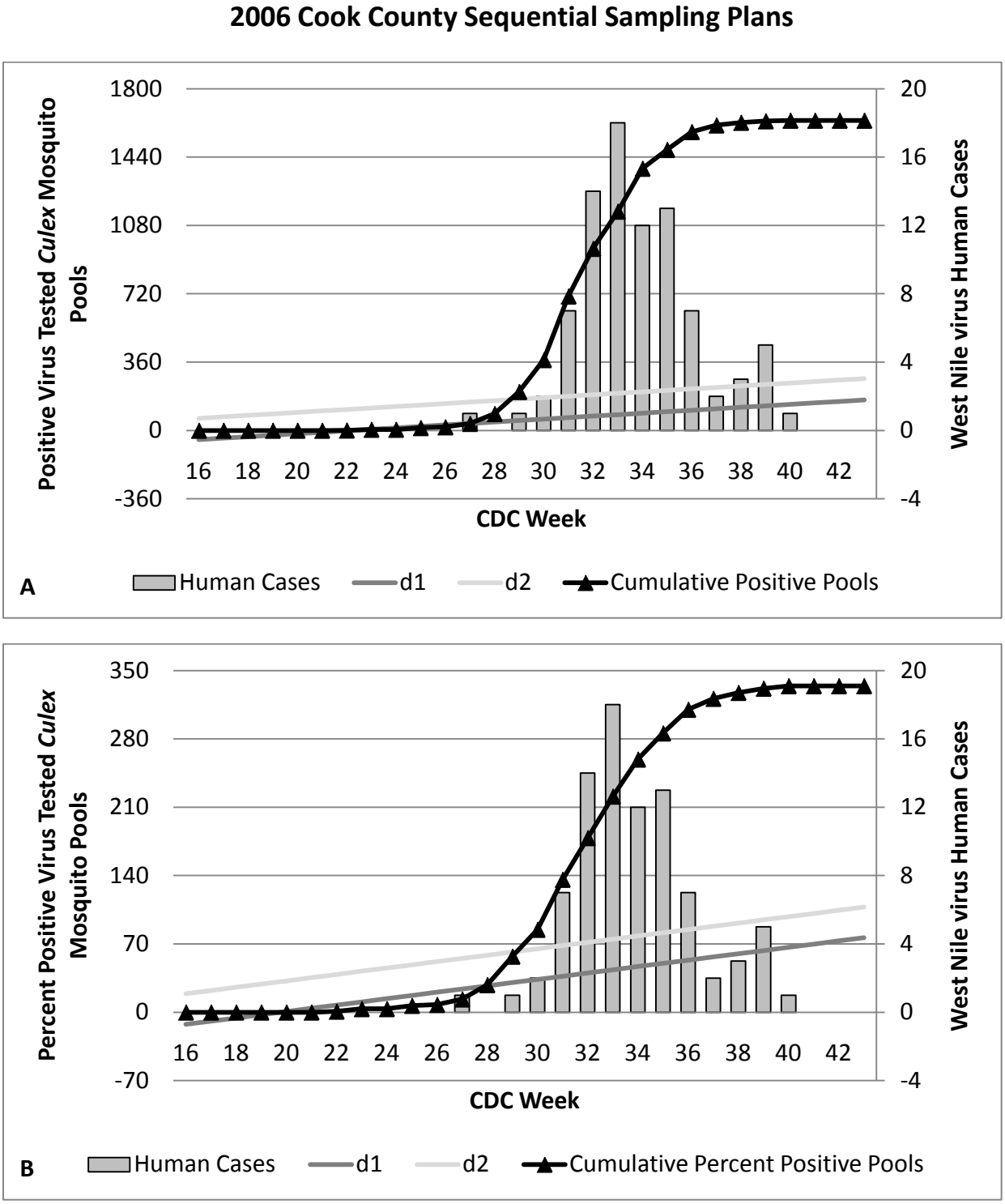

Figure 12. Sequential sampling plans for the county scale in Cook County, Illinois for 2006. (A) The plan using the number of West Nile virus tested *Culex* mosquito pools. (B) The plan using the percent of West Nile virus tested *Culex* mosquito pools.

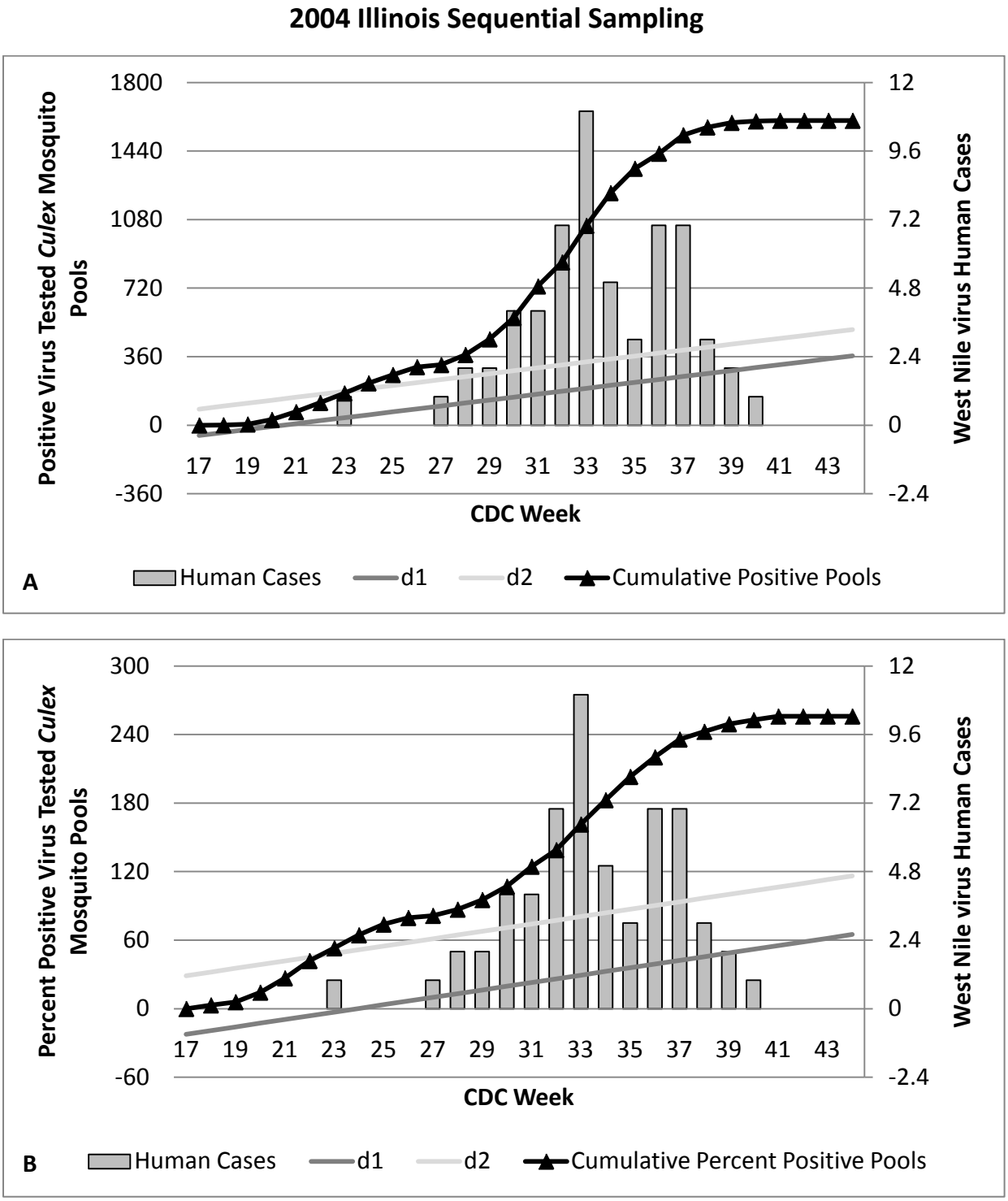

Figure 13: Sequential sampling plans for the state scale in Illinois for 2004. (A) The plan using the number of West Nile virus tested *Culex* mosquito pools. (B) The plan using the percent of West Nile virus tested *Culex* mosquito pools.

occurred (Figure 13B). In 2005, there were 2,480 positive pools out of 13,271 and 252 human cases. This year the cumulative number of positive pools crossed the upper decision line in week 29, 3 weeks after the first human case appeared, and one week after amplification (Figure 14A). The cumulative percent positive pools crossed the upper decision line in week 30 (Figure 14B). In 2006, there were 2,626 positive pools out of 15,390 and 215 human cases. This year the cumulative number of positive pools crossed the upper decision line in week 29, 7 weeks after the first human case appeared and the week amplification started to occur (Figure 15A). The cumulative percent positive pools crossed the upper decision line, 2 weeks later, in week 31 (Figure 15B).

Discussion

The research identified two variables that might be useful as the dependent variables in the sequential sampling scheme, i.e., the percent of pools that tested positive for West Nile virus, and the absolute number of pools that tested positive for the virus. The former variable is normalized as a percentage but does not lend itself well to statistical analysis of the kind presented here, where counts of events are considered. The latter variable does lend itself well to the analysis because it represents counts of positive tests per sample-week, where sample-week represents the aggregate of all pools of mosquitoes sampled and tested from all traps deployed and checked during that week. This variable does not require a consideration of infection rate or prevalence of infection (Biggerstaff 2006) because the sample-week, and not the mosquito population, provides the sampling framework. Ginsberg et al. (2010) showed that counts of positive pools per week was positively correlated with prevalence of infection in the

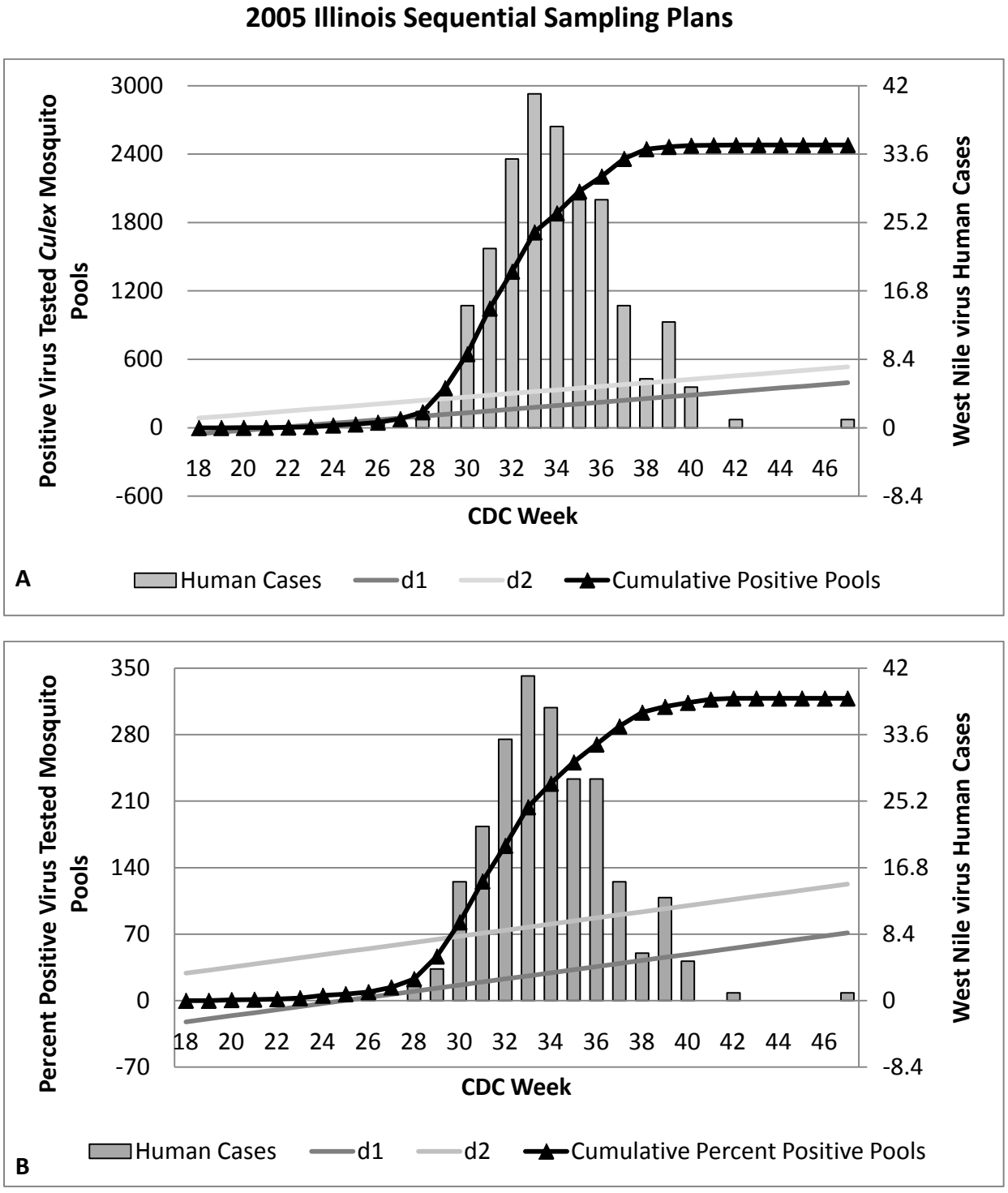

Figure 14. Sequential sampling plans for the state scale in Illinois for 2005. (A) The plan using the number of West Nile virus tested *Culex* mosquito pools. (B) The plan using the percent of West Nile virus tested *Culex* mosquito pools.

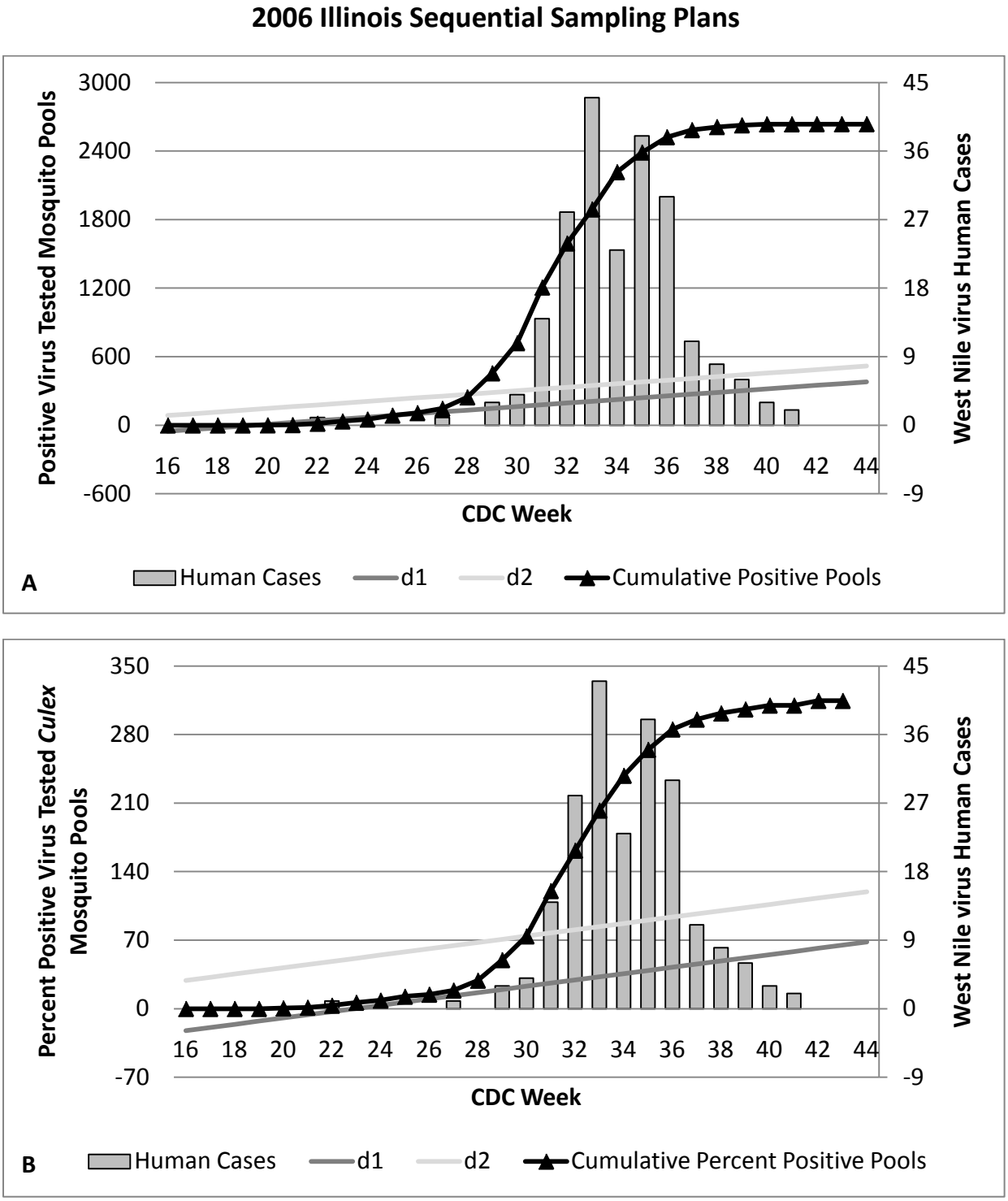

Figure 15. Sequential sampling plans for the state scale in Illinois for 2006. (A) The plan using the number of West Nile virus tested *Culex* mosquito pools. (B) The plan using the percent of West Nile virus tested *Culex* mosquito pools.

mosquito population as well as with the percentage of positive pools, supporting the use of the count data solely as a useful monitoring variable.

A comparison of the cross correlation analyses for the monitoring variables indicates a general trend showing that the percent positive pools have a smaller statistically significant lag time to human cases compared to the number of positive pools, however, both generally precede the first appearance of human cases temporally by one to several weeks. This result implies that the number of positive pools has a greater lag time and would therefore be a better predictor of human cases compared to the percent positive pools. In addition, comparing the two variables for each year at the study site spatial scale, the curves for the two responses were very similar (Figure 16). However, at the county and state spatial scales the number of positive pools showed an increase of a greater magnitude compared to the percent positive pools (Figures 17 & 18). This would also indicate that the number of positive pools would be a better forecasting tool compared to percent positive pools for sequential sampling plans, especially at larger scales.

When comparing the sequential sampling plans for the two variables, they have extremely similar results. For the study site and county spatial scales, both monitoring variables crossed the upper decision line at the same week, except for 2006. The cumulative number of positive pools crossed the upper decision line in week 29 and the percent positive pools in week 30 for the county scale. However, for the state spatial scale each of the monitoring variables crossed the upper decision line in different weeks. In 2004 the number of positive pools crossed in week 24 and the percent positive pools in week 23. This was an unusual year because there

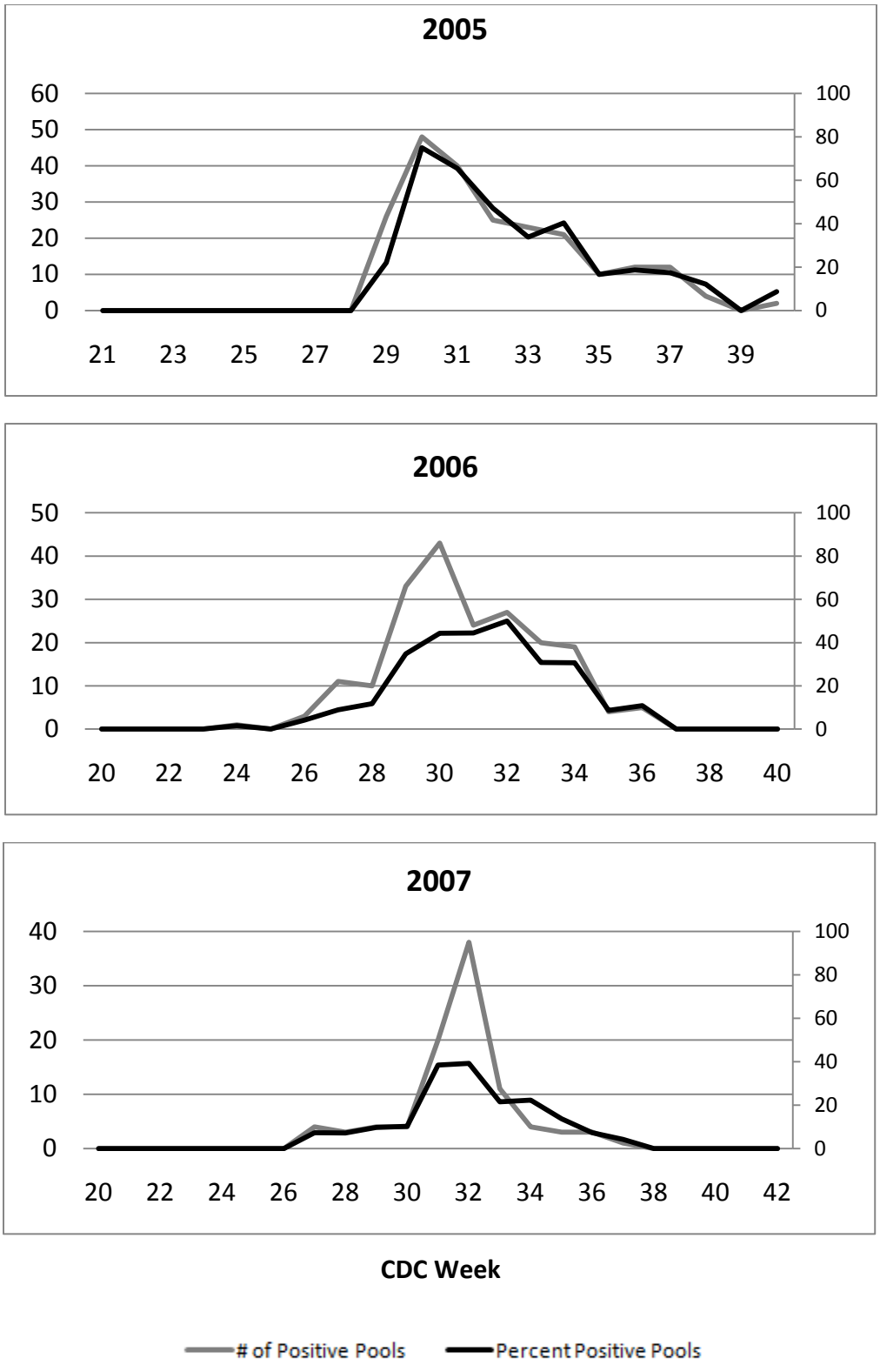

Figure 16. Suburban study site: comparison between number and percent positive pools for each year.

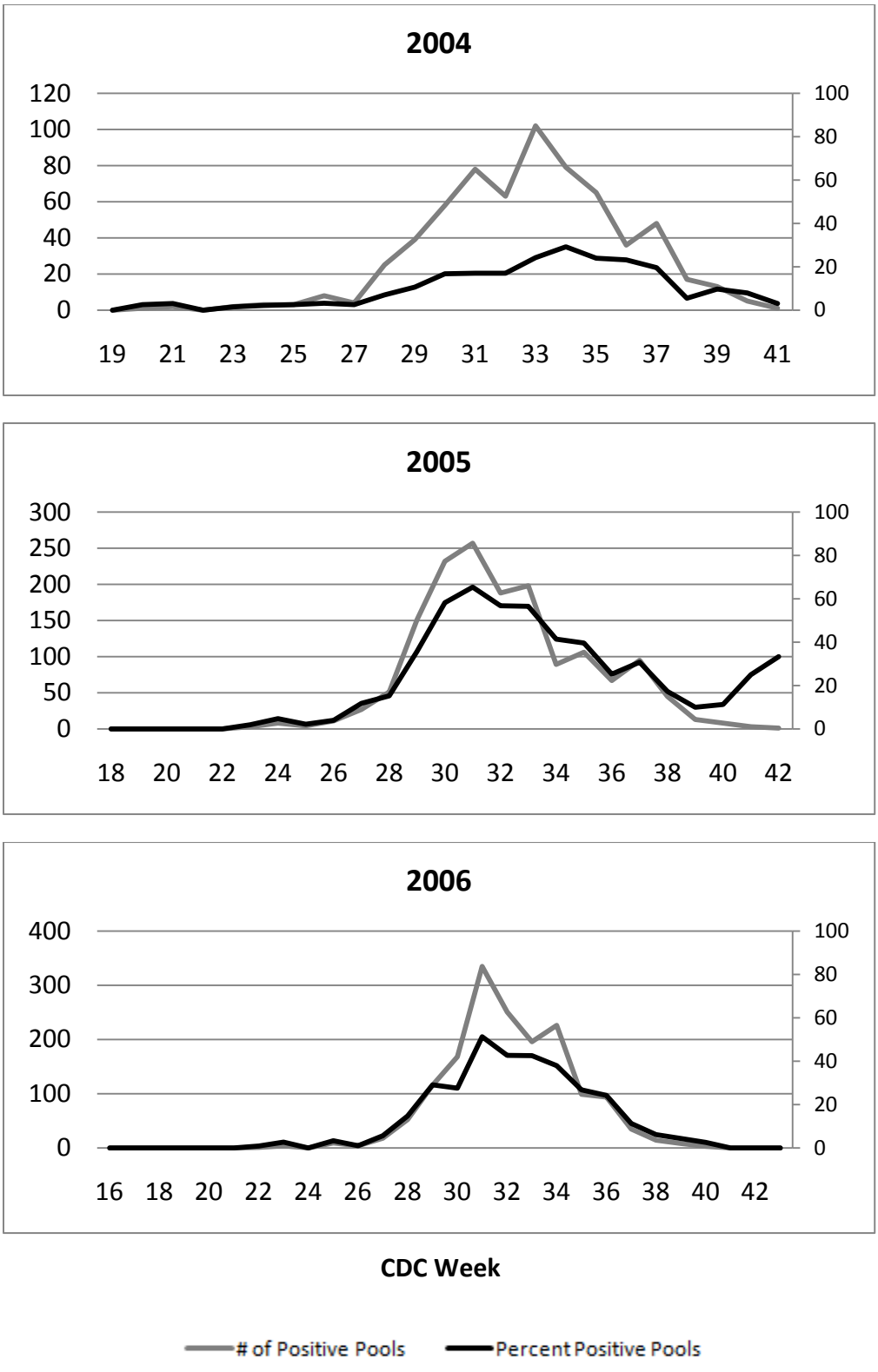

Figure 17. County: comparison between number and percent positive pools for each year.

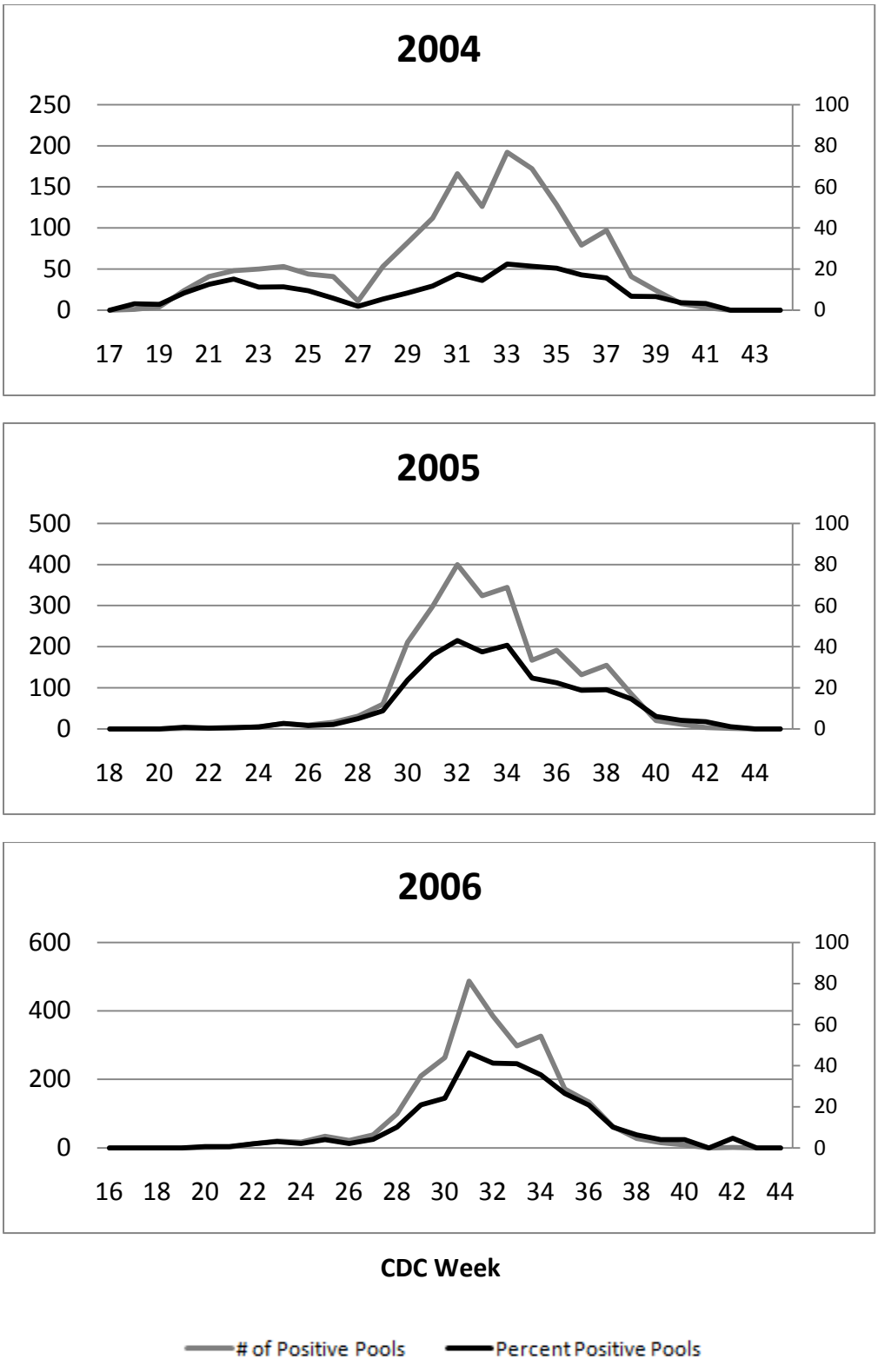

Figure 18. Illinois: comparison between number and percent positive pools for each year.

was an increase in virus earlier in the season, as well a human case in week 23. Regardless, both monitoring variables crossed the upper decision line well before human cases started to amplify. For 2005 and 2006 the number of positive pools crossed the upper decision line in week 29 and the percent positive pools in week 30 and 31 respectively. This may be an indication that the number of positive pools would be a more appropriate forecasting tool for human West Nile virus at larger spatial scales.

A variety of variables can be used for the monitoring or sampling plan. Past studies utilizing sequential sampling for disease prevention have used population density estimates of various life stages of the mosquito vector species under consideration separately from infection rate for the pathogen of interest (Barrera et al. 2006, Lindblade et al. 2000, Mogi 1990, Wada 1971). Rote density estimators obviate the need for testing for the pathogen in pooled samples, however, the possibility that high mosquito populations could develop but without the occurrence of an epidemic exists, elevating the risk of alpha error.

In addition to identifying the most suitable monitoring variable, the most suitable spatial scale for sequential sampling was assessed in this study. Examination of the data for years that were common across each spatial scale (i.e., 2005 & 2006) resulted in the indication that smaller spatial scales provided the most appropriate context for accurate forecasts of epidemics of human cases. At larger spatial scales, sensitivity was reduced.

For 2005 at the study site spatial scale both monitoring variables forecasted human cases the week they started to occur. For the county and state scales the number of positive pools forecasted human cases 3 weeks after the first human case appeared. For the percent

positive pools the county scale forecasted human cases 3 weeks after the first human cases and the state scale 4 weeks after. This year had very rapid amplification of virus in the mosquito population, probably because of the high summer temperatures (Ruiz et al. 2010). Regardless, the study site would offer an earlier reaction time compared to the county and state scales, therefore offering the potential of more human cases prevented. It should be noted that in 2005 at both the county and state spatial scale the human case that first occurred in week 26 was 2 weeks prior to when human cases began to increase in a continuous epidemic curve. Both spatial scales and monitoring variables crossed the second decision line before a majority of the human cases occurred, but the study site would offer the possibility to prevent a higher percentage of human cases occurring at that spatial scale compared to the county and state scales.

In 2006, the number of positive pools for the study site plan forecasted human cases 2 weeks before human cases occurred, the county scale 2 weeks after and the state 7 weeks after. For the percent positive pools the study site forecasted human cases two weeks prior to any human cases occurring, the county 2 weeks after and the state 9 weeks after. Once again during this year the study site plan would offer the most reaction time and prevention of more human cases. It should be noted again that the human case in week 27 for the county spatial scale occurred 2 weeks prior to when human cases began to increase in a continuous epidemic curve, so once again at the county scale the second decision line was crossed before a majority of the human cases occurred. For the state spatial scale, the case that occurred in week 22 was 5 weeks prior to any other human case occurring and 7 weeks before human cases began to

increase in a continuous epidemic curve. In any case, once again the study site spatial scale would offer the most reaction time before any human cases occurred.

One reason why smaller spatial scales may be better suited for sequential sampling is weather patterns and habitat is more consistent across this scale compared to the county or state scales. Weather patterns have been noted to drive West Nile virus amplification (Ruiz et al. 2010). It has been found for the Chicago, Illinois area that increased temperatures and decreased precipitation are significantly correlated with higher mosquito infection rates (Ruiz et al. 2010). For the state and county spatial scale there is different weather patterns, including temperature and precipitation, and habitat type throughout, which could create smaller areas of West Nile virus activity. These areas may be overlooked at the larger scales, however, with the smaller study site scale it is less likely for these 'hot spots' to be overlooked (Eisen and Eisen 2007). This can be seen in the county and state spatial scales for the human cases that occur in the weeks prior to amplification in human cases occurring. It would be more likely at a smaller spatial scale to forecast these human cases due to smaller pockets of virus amplification due to differences in local ecosystem dynamics.

An additional component that may contribute to the smaller spatial scale leading to more successful sequential sampling plans is the combination of competent vectors and reservoirs in the environment. Just as temperature and precipitation can vary at larger spatial scales the prevalence of vectors and reservoirs can demonstrate the same variability. In the Chicago area it has been found that there are several bird species that have a high reservoir competence, such as American robins, that contribute to a large percentage of mosquito

infections (Hamer et al. 2009). These bird species can differ in prevalence spatially and fluctuate in presence temporally throughout the year, thus driving West Nile virus activity where they are present. Not only does the species of bird play a significant role in WNv activity, but the age of the bird does as well. It has been found that hatch year birds enhance WNv amplification. If areas have higher numbers of hatch year birds the possibility for WNv amplification may be higher (Hamer et al 2008). All of these factors lead to smaller spatial scales capturing differences in bird populations and WNv activity better than larger spatial scales.

In addition to reservoirs, there can be difference in vectors and their tendency to feed on humans as demonstrated by Hamer et al. 2009. The investigators found that overuse of some bird species, such as American robins, for bloodmeals by the main amplifiying vector in the area, *Culex pipiens*. This results in the amplification of West Nile virus due to American robins having a high reservoir competence. Additionally, *Culex pipiens* demonstrated spatial variation in their tendency to bite humans (Hamer et al. 2009). This can be driven by habitat type, host availability and genetic variation in the mosquito (Hamer et al. 2009). Furthermore, in another study it has also been demonstrated that *Culex pipiens* vector competence fluctuates in both a temporal and spatial aspect (Kilpatrick et al 2010). The implications of these studies are that the main vectors of WNv have many factors that vary both spatially and temporally and this results, once again, in spatial differences in vectors that drive WNv amplification that would best be detected at smaller spatial scales.

Another important aspect of the study site sequential sampling plans, compared to the larger spatial scales, is an increased time lag when the cumulative count crosses the second

decision line. It is critical to have a time lag between when the cumulative pools cross the upper decision line and when human cases begin to occur to allow time for testing and reaction. Once the samples are collected for a given week, they will need to be tested in a lab, which would take time. In addition, once the pools are tested and can be added to the plan, it would take time to implement preventive measures. Therefore, the earlier a human epidemic is forecasted the more time allowance for testing and reaction, which results in the potential for more human cases to be prevented.

One advantage often noted for sequential sampling is the decrease in sampling effort, money and labor (Barrera et al 2006, Fowler & Lynch 1987, Knight 1964, Mackey & Hoy 1977). Once a decision is made to classify the epidemic threat then sampling could be slowed or stopped. This would offer decreased sampling and labor efforts, as well as the costs associated with them (Barrera et al 2006, Fowler & Lynch 1987, Knight 1964, Mackey & Hoy 1977). In addition, preventive efforts can be targeted to times when the insect population poses the greatest threat to humans instead of throughout the season. This would reduce costs once again in labor and effort, but would also potentially reduce costs in prevention methods. Furthermore, this would target prevention efforts that incorporate pesticide use. This would avoid multiple applications throughout the season and lower the environmental impact.

An issue that arises in this study relates to adequacy of sample size. Here, I utilized sample-week with aggregation of all samples from all trap deployments within the spatial unit of interest. Each of the plans proved robust when considering the overall number of mosquito pools tested which exceeded the number required to reach a decision point (Figure 6). In

overall number of pools tested, the study site ranged from 665 to 1,063, the county ranged from 4,829 to 7,041 and the state ranged from 13,258 to 15,390. These sample sizes far exceed the minimum number required by the sequential sampling plans developed here.

Sequential sampling has the potential to be an important forecasting tool for West Nile virus. This system offers flexibility in sampling, simplicity in design and cost and labor savings. In order for a particular locality to utilize a sequential sampling plan previous data on virus prevalence in mosquitoes and humans would be required to generate the plans. Therefore areas that don't have previous data collected wouldn't be able to efficiently create a sequential sampling plan. In addition, because of differences in weather, habitat, hosts and vectors that would affect the occurrence and amplification of WNv, plans from one location could not be expected to work for a different location.

One area that could use further study would be the threshold determination. Currently there is a lack of a consistent quantitative approach for threshold determination. The values in this study were arbitrarily picked based on past data and the occurrence of human cases. As more years and data become available further analysis would be useful. A possible approach would be to use the time series cross correlation analysis to determine if there is a certain amount of positive pools (percent or number) that causes the first significant correlation with human cases to constitute a 'critical mass' of virus that would likely cause human cases. For this study, with three years being available for each spatial scale, doing this threshold determination procedure proved to be difficult. As more years of data become available this method has great promise and potential in threshold determination.

In conclusion, sequential sampling plans that consistently predict annual West Nile virus outbreaks for a given area, before a majority of the human cases occur can be created. Every spatial scale and monitoring variable forecasted when a majority of the human cases are going to occur. However, the study site spatial scale was most successful at forecasting human cases before any take place. In addition, the number and percent of positive pools were both successful monitoring variables, with the number of positive pools possibly being more efficient at larger spatial scales compared to the percent. Sequential sampling has the capability to be a competent forecasting tool for West Nile virus. Sequential sampling plans bring the benefits of decreased sampling & targeted treatments, and better understanding of when the pest population poses a threat to humans compared to other sampling plans. Additional study in determining the best threshold determination and applicability for different locations would be beneficial in fine tuning and implementing sequential sampling plans.

REFERENCES

REFERENCES

Andreadis, TG, Thomas, MC, Shepard, JJ. Identification Guide to the Mosquitoes of Connecticut. New Haven, CT: The Connecticut Agricultural Experiment Station; 2005.

Barrera, R, Amador, M, Clark, GG. Sample-size requirements for developing strategies, based on the pupal/demographic survey, for the targeted control of dengue. Annals of Tropical Medicine & Parasitology. 2006; 100(1): S33-S43.

Barrera, R. Simplified pupal surveys of *Aedes aeypti* (L.) for entomologic surveillance and dengue control. American Journal of Tropical Medicine & Hygiene. 2009;81(1): 100-107.

Biggerstaff, BJ. PooledInfRate, Version 3.0: a Microsoft Excel add-in to compute prevalence estimates from pooled samples. Centers for Disease Control and Prevention, Fort Collins, Colorado, USA; 2006.

Brownstein, JS, Holford, TR, Fish, D. Enhancing West Nile virus surveillance, United States. Emerg Infect Dis. 2004;10(6):1129-1133.

Burkness, EC, Venette, RC, O'Rourke, PK, Hutchison, WD. Binomial sequential sampling for management of aster leafhopper (Homoptera: Cicadellidae) and aster yellows phytoplasma in carrot: impact of tally threshold on the accuracy of treatment decisions. Environ. Ent. 1999;28(5): 851-857.

CDC (Centers for Disease Control and Prevention). West Nile virus home page. Available at: http://www.cdc.gov/ncidod/dvbid/westnile/index.htm. Accessed April 15, 2011.

Davis, LE, DeBiasi, R, Goade, DE, Haaland, KY, Harrington, JA, Harnar, JB, Pergam, SA, King, MK, DeMasters, BK, Tyler, KL. West Nile virus neuroinvasive disease. Annals of Neurology. 2006;60(3):286-300.

Eidson, M, Komar, N, Sohage, F, Melson, R, Talbot, T, Mostashari, F. Crow deaths as a sentinel surveillance system for West Nile virus in the northeastern United States. Emerg Infect Dis 2001;7: 615-620.

Eisen, L, Eisen, RJ. Need for improved methods to collect and present spatial epidemiologic data for vectorborne diseases. Emerg. Infect. Dis. 2007;13(12):1816-1820.

Fowler, GW, Lynch, AM. Sampling plans in insect pest management based on Wald's sequential probability ratio test. Environmental Entomology 1987;16: 345-354.

Gerrard, DJ, Chiang, HC. Density estimation of corn rootworm egg populations based upon frequency of occurrence. Ecology 1970;51(2): 237-245.

Ginsberg, HS, Rochlin, I, Campbell, SR. The use of early summer mosquito surveillance to predict late summer West Nile virus activity. J Vector Ecol. 2010;35(1): 35-42.

Hamer G, Walker ED, Brawn JD, Loss SR, Ruiz MO, Goldberg TL, Schotthoefer AM, Brown WM, Wheeler E, Kitron, UD. Rapid amplification of West Nile virus: the role of hatch year birds. Vector Borne Zoonotic Dis. 2008;8: 57-67.

Hamer G, Kitron U, Goldberg TL, Brawn JD, Loss SR, Ruiz MO, Hayes DB, and Walker ED. Host selection by *Culex pipiens* and West Nile virus amplification. Amer. J. Trop. Med. Hyg. 2009;80: 268-278.

Harrington LC, Poulson RL. Considerations for accurate identification of adult *Culex restuans* (Diptera: Culicidae) in field studies. J Med Entomol. 2008;45(1): 1-8.

Hayes, EB, Gubler, DJ. West Nile virus: epidemiology and clinical features of an emerging epidemic in the United States. Annu. Rev. Med. 2006;57: 181-194.

Illinois Department of Public Health (IDPH). West Nile virus. Available at: http://www.idph.state.il.us/envhealth/wnv.htm. Accessed April 15, 2011.

Kilpatrick, AM, Fonseca, DM, Ebel, GD, Reddy, MR, Kramer, LD. Spatial and temporal variation in vector competence of *Culex pipiens* and *Cx. restuans* mosquitoes for West Nile virus. Am. J. Trop. Med. Hyg. 2010;83(30): 607-613.

Knight, KL. Quantitative methods for mosquito larval surveys. J. Med. Entomol. 1964;1(1): 109-115.

Kramer, LD, Styer, LM, Ebel, GD. A global perspective on the epidemiology of West Nile virus. Annu. Rev. Entomol. 2008;53: 61-81.

Lanciotti, RS, Kerst, AJ, Nasci, RS, Godsey, MS, Mitchell, CJ, et al. Rapid detection of West Nile virus from human clinical specimens, field-collected mosquitoes, and avian samples by a TaqMan reverse transcriptase-PCR assay. J Clin Microbiol. 2000;38(11): 4066-4071.

Lee, HL, Inder Singh, K. Sequential analysis of adult *Aedes aegypti* and *Aedes albopictus* in Kuala Lumpur City--its potential use in dengue epidemics prediction. Tropical Biomedicine 1993;10: 117-123.

Lindblade, KA, Walker, ED, Wilson, ML. Early warning of malaria epidemics in African highlands using *Anopheles* (Diptera: Culicidae) indoor resting density. J. Med. Entomol. 2000;37(5): 664-674.

Mackey, BE, Hoy, JB. *Culex tarsalis*: Sequential sampling as a means of estimating populations in California rice fields. J. Economic Entomology 1977;71(2): 329-334.

McLaughlin, RE, Brown, MA, Vidrine, MF. The sequential sampling of *Psorophora columbiae* larvae in rice fields. J. American Mosquito Control Association 1987;3(3): 423-428.

Misra, UK, Kalita, J. Overview: Japanese encephalitis. Progress in Neurobiology 2010;91: 108-120.

Mogi, M, Wada, Y. Spatial distribution of larvae of the mosquito *Culex tritaeniorhynchus summorosus* in a rice field area. Department of Medical Zoology, Nagaski University School of Medicine 1973;207: 69-83.

Mogi, M, Choochote, W, Khamboonruang, C, Suwanpanit, P. Applicability of presence-absence and sequential sampling for ovitrap surveillance of *Aedes* (Diptera: Culicidae) in Chiang Mai, northern Thailand. J. Med. Entomol. 1990;27(4): 509-514.

Mostashari, F, Kulldorff, M, Hartman, JJ, Miller, JR, Kulasekera, V. Dead bird clusters as an early warning system for West Nile virus activity. Emerg. Infect. Dis. 2003;9(6): 641-646.

Nyrop, JP, Foster RE, Onstad DW. Value of sample information in pest control decision making. J. Econ. Entomol. 1986;79: 1421-1429.

Onsager, JA. The rationale of sequential sampling, with emphasis on its use in pest management. Technical Bulletin 1526. Washington, D.C.: Agricultural Research Service, U.S. Department of Agriculture; 1976

Pedigo, LP, Zeiss, MR. Analyses in Insect Ecology and Management. Ames, Iowa State University Press; 1996.

Reisen, W, Brault AC. West Nile virus in North America: perspectives on epidemiology and intervention. Pest Manag. Sci. 2007;63:641-646.

Ruiz MO, Walker ED, Foster E, Haramis L, Kitron, UD. Association of West Nile virus illness and urban landscapes in Chicago and Detroit. International J. Health Geographics 2007;6:10.

Ruiz, MO, Chaves, LF, Hamer, GL, Sun, T, Brown, WM, Walker, ED, Haramis, L, Goldberg, TL, Kitron, UD. Local impact of temperature and precipitation on West Nile virus infection in *Culex* species mosquitoes in northeast Illinois, U.S.A. Parasites & Vectors 2010;3:19.

Samuel, MA, Diamond, MS. Pathogenesis of West Nile virus infection: a balance between virulence, innate and adaptive Immunity and viral evasion. Journal of Virology 2006;80(19): 9349-9360.

Theophilides, CN, Ahern, SC, Grady, S, Merlino, M. Identifying West Nile virus risk areas: the Dynamic Continuous-Area Space-Time system. Am. J. Epidemiol. 2003;157: 843-854.

Theophilides, CN, Ahern, SC,Binkowski, ES, Paul, WS, Gibbs, K. First evidence of West Nile virus amplification and relationship to human infections. International Journal of Geographical Information Science 2006;20: 103-115.

Wada, Y. Population studies on edmonton mosquitoes. Quaestiones entomologicae 1965; 1: 187-222.

Wada, Y, Mogi, M, Nishigaki, J. Studies on the population estimation for insects of medical importance: III. sequential sampling technique for *Culex tritaeniorhynchus summorosus* larvae in the paddy-field. Tropical Medicine 1971;13(1): 16-25.

Wald, A. Sequential tests of statistical hypotheses. Annals of Mathematical Statistics 1947;16: 117-186.

Waters, WE. Sequential sampling in forest insect surveys. For. Sci. 1955;1:68-79.

Wilson, LT, Room, PM. Clumping patterns of fruits and arthropods in cotton, with implications for binomial sampling. Environ. Entomol. 1983;12:50-54.