Check for updates

Author Manuscript

This is the author manuscript accepted for publication and has undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process, which may lead to differences between this version and the Version of Record. Please cite this article as doi: 10.1002/ECY.2882

This article is protected by copyright. All rights reserved

Title: Inferring seasonal infection risk at population and regional scales from serology samples

Running title: Inferring infection risk across scales

Authors: Mark Q. Wilber[1,2], Colleen T. Webb[1], Fred L. Cunningham[3], Kerri Pedersen[4], Xiu-Feng Wan[5−10], Kim M. Pepin[2]

Author affiliations:
[1]Department of Biology, Colorado State University, Fort Collins, CO 80523, USA
[2]United States Department of Agriculture, Animal and Plant Health Inspection Service, Wildlife Services, National Wildlife Research Center, Fort Collins, CO 80521-2154, USA
[3]United States Department of Agriculture, Animal and Plant Health Inspection Service, Wildlife Services, National Wildlife Research Center, Mississippi Field Station, PO Box 6099, MS 39762, USA
[4]United States Department of Agriculture, Animal and Plant Health Inspection Service, Wildlife Services, 920 Main Campus Drive, Suite 200, Raleigh, NC 27606
[5]Missouri University Center for Research on Influenza Systems Biology, University of Missouri, Columbia, MO, USA
[6]Department of Molecular Microbiology and Immunology, School of Medicine, University of Missouri, Columbia, MO, USA
[7]Department of Electrical Engineering & Computer Science, College of Engineering, University of Missouri, Columbia, MO, USA
[8]Bond Life Sciences Center, University of Missouri, Columbia, MO, USA
[9]MU Informatics Institute, University of Missouri, Columbia, MO, USA
[10]Department of Pathobiology, College of Veterinary Medicine, University of Missouri, Columbia, MO, USA

Corresponding author: Mark Q. Wilber, mqwilber@gmail.com

Keywords: serology; seroprevalence; incidence; infection hazard; survival analysis; seroconversion; seasonal disease dynamics; feral swine; influenza A virus

Article Type: Statistical Report

Author Manuscript

This article is protected by copyright. All rights reserved

Abstract

Accurate estimates of seasonal infection risk can be used by animal health officials to predict future disease risk and better understand the mechanisms driving disease dynamics. It can be difficult to estimate seasonal infection risk in wildlife disease systems because surveillance assays typically target antibodies ('serosurveillance'), which are not necessarily indicative of current infection, and serosurveillance sampling is often opportunistic. Recently developed methods estimate past time of infection from serosurveillance data using quantitative serological assays that indicate the amount of antibodies in a serology sample. However, current methods do not account for common opportunistic and uneven sampling associated with serosurveillance data. We extended the framework of survival analysis to improve estimates of seasonal infection risk from serosurveillance data across population and regional scales. We found that accounting for the right-censored nature of quantitative serology samples greatly improved estimates of seasonal infection risk, even when sampling was uneven in time. Survival analysis can also be used to account for common challenges when estimating infection risk from serology data, such as biases induced by host demography and continually elevated antibodies following infection. The framework developed herein is widely applicable for estimating seasonal infection risk from serosurveillance data in humans, wildlife, and livestock.

Introduction

Identifying the timing of seasonal infections can provide information on when events of management concern, such as pathogen spillover and spatial spread (Altizer et al. 2006; Plowright et al. 2017), are likely to occur and can be used to mitigate the impacts of disease on human, livestock, and wildlife health. However, detecting and de-constructing patterns of disease seasonality in wildlife populations can be difficult because existing methods require longitudinal samples across an infection season. While longitudinal data are available for some human diseases (e.g. Rohani et al. 2002), these data are substantially

Author Manuscript

This article is protected by copyright. All rights reserved

harder to obtain for wildlife diseases, particularly over multiple years and large spatial scales (Gilbert et al. 2013). Specifically, two common features of wildlife surveillance data challenge inference of seasonal disease dynamics in wildlife populations (Gilbert et al. 2013): 1) serum samples – 'serosurveillance' detects antibodies in individuals that can persist when a pathogen is no longer present in the population, and 2) opportunistic sampling – sampling design piggybacks on other management activities and is often not representative of a population sample, but is necessarily executed due to the challenges with sampling wildlife populations.

Serological samples can provide a quantitative measure of pathogen-specific antibodies in the serum of a sampled host and can be used to determine whether a host has been previously exposed to a pathogen (i.e. is seropositive, Gilbert et al. 2013). Serological sampling is widely-used for disease surveillance in humans, livestock, and wildlife, despite challenges including pathogens that evade the immune system, cross-reactivity of antibodies, between-host heterogeneity in antibody dynamics, and seroconversion thresholds leading to misclassification of infections (Gay 1996; Gilbert et al. 2013; Pepin et al. 2017). Seroprevalence, the proportion of seropositive hosts in a sampled population, is an easy-to-understand metric of disease dynamics. If infected hosts remained permanently infected and experienced negligible mortality over the course of an epidemic, then the observed dynamics of seroprevalence provide an estimate of cumulative infection risk, from which incidence (the rate of new cases per unit time) and force of infection (the instantaneous risk of infection for a susceptible host) can be calculated (Hens et al. 2012). However, when infected hosts recover or die from infection and host demography changes the pool of susceptible hosts, seroprevalence dynamics no longer recapitulate infection risk (Appendix S1: Fig. S1). In this case, linking seasonal mechanisms to the timing of seroprevalence peaks or the shape of seroprevalence curves can be misleading as the proximate mechanisms driving increases in incidence (e.g. increases in contact during the mating season) will have occurred sometime before peak seroprevalence was observed.

Author Manuscript

This article is protected by copyright. All rights reserved

The limitations of using seroprevalence to infer seasonal infection risk have motivated a number of methodological developments that leverage quantitative serological data to infer temporal patterns in force of infection and incidence (Teunis et al. 2002; Kretzschmar et al. 2010; Teunis et al. 2012; Borremans et al. 2016; Pepin et al. 2017). These methods use within-host antibody dynamics to back-infer the time of infection (TOI) from the measured antibody level in a serological sample. Given estimates of individual-level TOI, population-level infection risk can be inferred, which can be used to understand disease dynamics and predict future risk. The current quantitative antibody methods developed for wildlife focus on estimating TOI and infection risk at the population-level under idealized sampling conditions and have yet to be extended to large-scale, opportunistically-sampled serosurveillance data (Borremans et al. 2016; Pepin et al. 2017), which are common in disease surveillance. Linking quantitative antibody methods to serosurveillance data can help answer epidemiological questions such as 1) do patterns of seasonality in infection risk change across space and time? 2) can we detect the direction of spatial spread of an epidemic through changes in the timing of peak seasonal infection risk across space? and 3) how do regional ecological and environmental factors affect the timing and magnitude of seasonal infection risk?

We developed a novel framework that links quantitative, individual-level serology data to population- or regional-level patterns of infection risk (Fig. 1). We extended the framework of survival analysis to the emerging field of quantitative antibody methods. We used survival analysis to infer true incidence from serological data obtained through opportunistic sampling at either population or regional scales. We demonstrate our approach by using it to examine seasonal patterns of influenza A virus (IAV) incidence in feral swine from national-scale serosurveillance data collected in the United States of America (USA).

Author Manuscript

This article is protected by copyright. All rights reserved

Methods

Figure 1 describes four steps that are required to estimate seasonal infection risk from serological data. We focused on the fourth step – how to improve our inference on retrospective patterns of infection risk, given TOI estimates. Previous studies have addressed different epidemiological challenges associated with obtaining unbiased TOI estimates such as dose-dependent sources of antibody titer variation or differences in variation among experimental and field settings (Borremans et al. 2016; Pepin et al. 2017). Our goal was, given TOI estimates (biased or unbiased), how to best estimate the incidence function $f(t)$ that defines the rate of infection at time or date t. When we refer to seasonal infection risk we are jointly referring to the incidence function $f(t)$, the survival function $S(t)$ (the probability that a host has not been infected by t), and the cumulative distribution function $F(t)$ (the probability that a host has been infected before t; Appendix S2: Table S1). We do not consider disease-induced mortality in our current framework, which can lead to underestimates of $f(t)$ (Heisey et al. 2006). However, previous studies have accounted for disease-induced mortality when estimating TOI and our framework could be extended to include these approaches (Pepin et al. 2017).

Using survival analysis to estimate infection risk from TOI data

Consider N serology samples collected longitudinally or cross-sectionally and assume we have estimated the TOI for each seropositive host and that seropositive hosts remain seropositive over an infection season. Each sample comes from a unique host. The ith sample is associated with two pieces of information: (δ_i, t_i). δ_i is one if a sampled host is seropositive and zero if the host is seronegative. If $\delta_i = 1$, t_i is the TOI relative to a starting date of interest (e.g. the date at which sampling began, the date at which a pathogen was suspected to have invaded). Note that t_i is still time of infection, just relative to a starting date (e.g. $t_i = 24$ days since March 1). This contrasts with time since infection that is the time between the sampling date and the date of infection (e.g. 5 days between a sampling date of March 30 and an infection date of March 25). If $\delta_i = 0$, an uninfected, seronegative

Author Manuscript

This article is protected by copyright. All rights reserved

host is right-censored, meaning that infection has not occurred before the sampling time s_i (Appendix S1: Fig. S2). When $\delta_i = 0$, $t_i = s_i$ (Klein and Moeschberger 2003). The likelihood for the N samples is given by (Klein and Moeschberger 2003)

$$L = \prod_{i=1}^{N} f(t_i)^{\delta_i} S(t_i)^{1-\delta_i} \tag{1}$$

Equation 1 assumes that sampling and infection times are independent. Importantly, accounting for censoring in the TOI data ameliorates the effects of uneven seasonal sampling on estimates of infection risk. This is because patterns of censoring reflect sampling patterns, such that accounting for censoring also accounts for sampling design (Klein and Moeschberger 2003). We used a flexible, parametric framework to estimate $f(t)$ (Appendix S3). Note that $S(t)$ does not necessarily have to start at 1 (no infection risk) or end at 0 (all hosts infected) over the time interval when infection risk is being inferred.

We tested the ability of equation 1 to correctly infer seasonal infection risk, given different sampling designs and different known incidence functions. Using simulation, we sampled 1000 hosts over time following either a uniform, unimodal, or bimodal sampling distribution from a population of hosts that was experiencing either uniform, unimodal, or bimodal incidence. We considered all nine combinations of sampling and incidence functions. We considered a host seropositive if it was sampled after its true infection date and we assumed we knew TOI for seropositive hosts. We made this assumption to test equation 1 independent of any uncertainty or biases that might arise when estimating TOI (Appendix S2; Pepin et al. 2017).

Extending survival analysis for common challenges

Host demography: When serosurveillance data are sampled over multiple infection seasons, host demography needs to be accounted for to prevent biased estimates of infection risk as hosts are not necessarily at risk for an entire sampling period. Host age, or host age class, is a commonly collected host characteristic in human and wildlife serology studies (Borremans et al. 2016) and can be used to account for a component of host demography

Author Manuscript

This article is protected by copyright. All rights reserved

when inferring seasonal infection risk. Let age_i be the age of host i at sampling s_i and $\text{dob}_i = s_i - \text{age}_i$ be the approximate date of birth of host i. The approximate date of birth can be used to update equation 1

$$L = \prod_{i=1}^{N} f(t_i)^{\delta_i} [F(\text{dob}_i) + S(t_i)]^{1-\delta_i} \tag{2}$$

The term in square brackets states that an uninfected host could be observed because it remained susceptible after birth ($F(\text{dob}_i)$ defines the "amount" of infection risk a host missed by being born at dob_i) or because it was at risk of infection after birth, but was censored before infection occurred conditional on being at risk, $[1 - F(\text{dob}_i)] \left(\frac{S(t_i)}{S(\text{dob}_i)} \right) = S(t_i)$. More detailed aspects of age-based immunity, such as maternal antibodies, could also be incorporated into the framework when estimating seasonal infection risk. For example, if maternal antibodies prevent seroconversion for an average of τ_{maternal} time units after birth, $F(\text{dob}_i)$ could be updated to $F(\text{dob}_i + \tau_{\text{maternal}})$. We examined how host births during the sampling season could bias estimates of infection risk and tested the ability of equation 2 to recover true seasonal infection risk dynamics.

Elevated antibodies: We have thus far assumed that the TOI for a seropositive host was known. In reality we never know the true TOI and quantitative antibody methods are needed to back-infer the TOI using the dynamics of within-host antibody curves (Fig. 1C; Simonsen et al. 2009; Borremans et al. 2016; Pepin et al. 2017). However, the nature of the antibody curve can affect how TOI estimates are used in the survival analysis framework. For example, consider the antibody curve in Appendix S1: Fig. S2A. After exposure to a pathogen, host antibodies rise and then fall to the level y_{end}, but the host remains seropositive. If this host was sampled when its antibody level was equal to y_{end}, a unique TOI could not be inferred, as there are many times since infection with an antibody level of y_{end}. We refer to this phenomenon as "elevated antibodies".

Elevated antibodies have been thoroughly addressed when estimating age profiles of

Author Manuscript

This article is protected by copyright. All rights reserved

force of infection, where a host's immunity to a pathogen is assumed to be life-long (Heisey et al. 2006; Hens et al. 2012). Elevated antibodies present different challenges when inferring seasonal infection risk because the goal is to obtain estimates of infection risk at time points during the season, rather than for particular host ages. For hosts with an antibody quantity equal to y_{end}, all we know is that infection occurred sometime before the date of sampling, but far enough away from the date of sampling such that the host's antibody level could rise and then fall back to y_{end} (which takes τ_{end} time units; Appendix S1: Fig. S2). In survival analysis parlance these hosts are left-censored and we can update equation 1 as

$$L = \prod_{i=1}^{N} [f(t_i)\mathbb{1}_{y_i \neq y_{\text{end}}} + F(s_i - \tau_{\text{end}})\mathbb{1}_{y_i = y_{\text{end}}}]^{\delta_i} [S(t_i)]^{1-\delta_i} \tag{3}$$

where s_i gives the time at which a seropositive host was sampled, y_i is the antibody quantity of host i, and $\mathbb{1}_{y_i = y_{\text{end}}}$ is an indicator variable that evaluates to one if true and zero otherwise. In practice, seropositive hosts do not need to exactly satisfy the equality $y_i = y_{\text{end}}$ to be considered left-censored and the user can make a decision as to what criteria defines a left-censored host. Equation 3 reverts to a standard left-censored analysis if we considered all seropositive hosts to be left-censored (i.e. no TOI estimates were available; Hens et al. 2012).

Equation 3 can be updated to simultaneously account for host age and elevated antibodies in the survival analysis framework.

$$L = \prod_{i=1}^{N} [f(t_i)\mathbb{1}_{y_i \neq y_{\text{end}}} + (F(s_i - \tau_{\text{end}}) - F(\text{dob}_i))\mathbb{1}_{y_i = y_{\text{end}}}]^{\delta_i} [F(\text{dob}_i) + S(t_i)]^{(1-\delta_i)} \tag{4}$$

Equation 4 states that left-censored hosts can only be infected between their date of birth and $s_i - \tau_{\text{end}}$.

We used simulations to test whether equation 3 correctly accounted for hosts with

Author Manuscript

This article is protected by copyright. All rights reserved

elevated antibodies when estimating seasonal infection risk. We sampled hosts for one year. If a host was infected prior to sampling, we assigned it an antibody quantity based on a known antibody curve that reached y_{end} 132 days after initial exposure. Any seropositive host with a TOI greater than 132 days was considered left-censored with a TOI that could not be estimated, as would be the case in practice. We assumed that we knew whether a host came from the rising or falling arm of the antibody curve (Appendix S4). We compared estimates of seasonal infection risk from equation 3 to the special case where we had no information on TOI and thus all seropositive hosts were considered left-censored.

Application: Seasonal dynamics of influenza A virus in feral swine

We applied our approach to explore the seasonal dynamics of influenza A virus (IAV) infections in feral swine across the United States of America. IAV dynamics can be highly seasonal, with incidence in humans peaking in the winter months in the northern and southern hemispheres (Earn et al. 2002; Finkelman et al. 2007). Feral swine are susceptible to multiple IAV subtypes, including human and avian subtypes, making them important vessels for the mixing and transmission of influenza viruses (Feng et al. 2014; Martin et al. 2017). Here we compared seasonal patterns of observed seroprevalence and estimated incidence with the goal of understanding the implications of using incidence compared with seroprevalence for inferring the timing and magnitude of seasonal infection risk of IAV in feral swine.

We used IAV serosurveillance samples from feral swine that were sampled across the USA from 2010-2017 as part of the U.S. Department of Agriculture, Wildlife Services' National Wildlife Disease Program (Fig. 1A; Feng et al. 2014). The methods we develop here can be used to extract novel information out of this common type of surveillance data, such as seasonal patterns of infection risk at large spatial scales. We applied our approach to examine seasonal IAV infection risk in feral swine sampled in five watershed regions (Hydrologic Unit Code 2) within the USA where greater than 60 seropositive samples were detected over the seven years of sampling (Appendix S5; Fig. 1A). We used previously

Author Manuscript

This article is protected by copyright. All rights reserved

published experimental, longitudinal data on IAV infections in feral swine to estimate the within-host antibody dynamics (Appendix S5; Sun et al. 2015). We considered seropositive hosts left-censored if estimated TOI was greater than 200 days. We fit all models using Python (3.6.8) and the probabilistic programming language Stan (PyStan 2.18.1.0, Carpenter et al. 2017). Scripts are available at https://zenodo.org/badge/latestdoi/173839768.

Results

The survival analysis framework allowed for inference on seasonal infection risk even when sampling was highly uneven in time (Fig. 1D, Fig. 2). Not accounting for sampling design and right-censoring led to serious biases in the estimated seasonal infection risk (Fig. 2, see Appendix S1: Fig. S3 for estimated incidence functions). Extending the survival analysis framework to account for host age significantly improved estimates of seasonal infection risk compared to when host age was ignored (Appendix S1: Fig. S4). Ignoring host age resulted in underestimates of peak infection risk and a more dispersed estimate of infection risk throughout the sampling period (Appendix S1: Fig. S4). Similarly, excluding hosts with elevated antibodies underestimated infection risk (Appendix S1: Fig. S5). While excluding these hosts was a naive approach, it was necessary under equation 1 as TOI cannot be estimated for these hosts. By appropriately treating hosts with elevated antibodies as left-censored, equation 3 recovered the true patterns of seasonal infection risk (Appendix S1: Fig. S5). When we discarded the information on TOI and treated all seropositive hosts as left-censored, equation 3 could still recover true seasonal infection risk, but there was increased uncertainty about the estimate relative to when information on TOI was included (Appendix S1: Fig. S6).

The seasonal patterns of IAV infection risk in feral swine varied across watershed regions in the USA (Fig. 3). For all five watershed regions, there were notable peak(s) in infection risk in the 2013-2014 and 2014-2015 infection seasons (Fig. 3). Peaks in seroprevalence lagged peaks in incidence and the amount of lag depended on the antibody quantities of

Author Manuscript

This article is protected by copyright. All rights reserved

seropositive hosts. For example, seropositive feral swine identified in the Hawaii Region with an estimated TOI in 2013-2014 had higher median antibody titers than seropositive feral swine in the South Atlantic-Gulf Region with TOIs in 2013-2014 (Hawaii Region median antibody level, $n = 39$: 0.54, South Atlantic-Gulf Region median, $n = 118$: 0.21; 95% bootstrapped confidence interval in the difference between medians: [0.19, 0.41]), suggesting a more recent infection relative to sampling time. The estimated incidence peak in the Hawaii Region in 2013 led the observed seroprevalence peak by one month, while the estimated incidence peak in the South Atlantic-Gulf Region in 2014 led the observed seroprevalence peak by eight months (Fig. 3).

Discussion

Linking survival analysis and quantitative antibody methods improved estimates of seasonal infection risk. Importantly, the framework of survival analysis could estimate seasonal infection risk even when sampling was uneven through time. As temporally even sampling of serosurveillance data for either a population or region is logistically impossible for most wildlife hosts, this property of survival analysis is useful because it allows information on seasonal infection risk to be obtained from a wide-range of serosurveillance sampling designs. Moreover, we demonstrated that our framework could be used to improve estimates of infection risk by accounting for common challenges associated with serological data such as host age and left-censoring through elevated antibodies.

Left-censored survival analysis is the standard approach for estimating age-based force of infection from serology data (Heisey et al. 2006; Hens et al. 2012). Our approach differs from these past techniques as it estimates incidence in calendar time rather than incidence by age. Moreover, our approach directly incorporates information on antibody dynamics to improve precision when estimating seasonal infection risk. Instead of considering all seropositive hosts as left-censored, our approach uses antibody dynamics to inform estimates of TOI for non-left-censored seropositive hosts and to provide additional information on the bounds of the TOI interval for left-censored hosts. Using the antibody

Author Manuscript

This article is protected by copyright. All rights reserved

curve to narrow the TOI interval for left-censored hosts reduces uncertainty and bias in seasonal infection risk estimates.

There are many ecological and methodological challenges associated with estimating seasonal infection risk that we do not address in this study (see Appendix S6, Appendix S7 and Pepin et al. 2017). The framework we present here is a single component of a larger process that can be used to address some, but not all, of these challenges. When estimating infection risk from serology data, sources of bias can enter at four different stages of the analysis: sampling design, laboratory techniques for estimating antibody quantities, inferring time since infection, and inferring seasonal infection risk (Fig. 1).

For example, an important methodological challenge that we did not address in this study was the seroconversion threshold (Gay 1996; Garnier et al. 2017). The seroconversion threshold is important because hosts that are left-censored with one seroconversion threshold may appear to be seronegative with a different seroconversion threshold. Not accounting for this apparent recovery can lead to underestimates in the magnitude of infection risk. Our framework can be extended to account for apparent recovery by modeling two potential conditions of uninfected hosts: a host that has not yet experienced infection by time t or that has experienced infection prior to sampling but has since decreased below the detectable seroconversion threshold (Appendix S6).

An important biological challenge facing quantitative antibody methods is how to best account for reinfection. In vertebrates, antibody levels tend to rise and fall more quickly after the primary infection (Pepin et al. 2017). If covariate data can uniquely distinguish between a primary and anamnestic response, then TOI estimates could be adjusted accordingly. If not, survival analysis could account for the probability of a seropositive host experiencing either a primary or anamnestic response. For example, an anamnestic response might be more likely for older hosts sampled as seropositive later in the infection season. Of course, as hosts experience more unobserved exposures these will be increasingly difficult to uniquely identify without some additional covariate data.

Author Manuscript

This article is protected by copyright. All rights reserved

We applied our approach to an extensive database of IAV serology samples in feral swine sampled since 2010 in the USA (Feng et al. 2014; Martin et al. 2017). Peaks in IAV seroprevalence necessarily lagged peaks in estimated incidence. Importantly, the amount of lag inferred was based on the dynamics of the antibody curves, such that across a single watershed region, seroprevalence did not always lag incidence by the same length of time. When identifying seasonal mechanisms driving disease dynamics, the timing of infection risk can be correlated with seasonal factors such as host behavior and environmental variables (Altizer et al. 2006). This requires an understanding of when infection actually occurred, which is not obvious when using seroprevalence data as a metric of infection risk. In feral swine exposed to IAV, seroprevalence peaks lagged incidence from one month to eight months, depending on the region. Our case study illustrates that correlating seasonal factors, such as temperature or seasonal animal behaviors, with peaks in seroprevalence could miss the true mechanisms driving seasonal infection risk, as peak infection risk can occur in a different season than peak seroprevalence, and the lag time is inconsistent across seasons and years.

Our method can be applied to understand the patterns and mechanisms of seasonal infection risk in pathogens of concern for wildlife and humans including plague, avian influenza, West Nile virus, and myxomatosis in rabbits (Kerr 1997; McKee et al. 2015; Pepin et al. 2017), where longitudinal antibody and serosurveillance data already exist. However, high-resolution longitudinal antibody data is not a necessary prerequisite to use the survival analysis framework we present here, as our method is a generalization of serology methods that only use seroprevalence data to estimate infection risk, with no information on TOI (Heisey et al. 2006; Hens et al. 2012). Even incomplete knowledge of the antibody curve (e.g. the average τ_{end}) or highly uncertain estimates of TOI could be included to improve the precision of seasonal infection risk compared to seroprevalence alone. Our framework is an important step for leveraging individual-level quantitative antibody data to understand and predict seasonal disease dynamics at population and

Author Manuscript

This article is protected by copyright. All rights reserved

regional scales, with implications for better managing disease emergence and spread.

Acknowledgments

Thanks to Thomas Deliberto, Brandon Schmit, and Ted Anderson for help with the IAV serosurveillance and laboratory data used in this analysis. Funding for XFW was provided by NIH NIAID R01AI116744. Funding provided by the U.S. Department of Agriculture, Animal and Plant Health Inspection Services' National Feral Swine Damage Management Program Animal and National Wildlife Disease Program.

Literature Cited

Altizer, S., A. Dobson, P. Hosseini, P. Hudson, M. Pascual, and P. Rohani. 2006. Seasonality and the dynamics of infectious diseases. Ecology Letters **9**:467–84.

Borremans, B., N. Hens, P. Beutels, H. Leirs, and J. Reijniers. 2016. Estimating time of infection using prior serological and individual information can greatly improve incidence estimation of human and wildlife infections. PLoS Computational Biology **12**:e1004882.

Carpenter, B., A. Gelman, M. D. Hoffman, D. Lee, B. Goodrich, M. Betancourt, M. Brubaker, J. Guo, P. Li, and A. Riddell. 2017. Stan : A Probabilistic Programming Language. Journal of Statistical Software **76**:1–32.

Earn, D. J., J. Dushoff, and S. A. Levin. 2002. Ecology and evolution of the flu. Trends in Ecology and Evolution **17**:334–340.

Feng, Z., J. A. Baroch, L. P. Long, Y. Xu, F. L. Cunningham, K. Pedersen, M. W. Lutman, B. S. Schmit, A. S. Bowman, T. J. DeLiberto, and X. F. Wan. 2014. Influenza A subtype H3 viruses in feral swine, United States, 2011-2012. Emerging Infectious Diseases **20**:843–846.

Finkelman, B. S., C. Viboud, K. Koelle, M. J. Ferrari, N. Bharti, and B. T. Grenfell. 2007. Global patterns in seasonal activity of influenza A/H3N2, A/H1N1, and B from 1997 to 2005: Viral coexistence and latitudinal gradients. PLoS ONE **2**.

Garnier, R., R. Ramos, A. Sanz-Aguilar, M. Poisbleau, H. Weimerskirch, S. Burthe,

Author Manuscript

This article is protected by copyright. All rights reserved

J. Tornos, and T. Boulinier. 2017. Interpreting ELISA analyses from wild animal samples: Some recurrent issues and solutions. Functional Ecology **31**:2255–2262.

Gay, N. J. 1996. Analysis of serological surveys using mixture models: application to a survey of parvovirus B19. Statistics in Medicine **15**:1567–73.

Gilbert, A. T., A. R. Fooks, D. T. Hayman, D. L. Horton, T. Müller, R. Plowright, A. J. Peel, R. Bowen, J. L. Wood, J. Mills, A. A. Cunningham, and C. E. Rupprecht. 2013. Deciphering serology to understand the ecology of infectious diseases in wildlife. EcoHealth **10**:298–313.

Heisey, D. M., D. O. Joly, and F. Messier. 2006. The fitting of general force-of-infection models to wildlife disease prevalence data. Ecology **87**:2356–65.

Hens, N., Z. Shkedy, M. Aerts, C. Faes, P. Van Damme, and P. Beutels. 2012. Modeling Infectious Disease Parameters Based on Serological and Social Contact Data. Springer, New York.

Kerr, P. J. 1997. An ELISA for epidemiological studies of myxomatosis: Persistence of antibodies to myxoma virus in European rabbits (*Oryctolagus cuniculus*). Wildlife Research **24**:53–65.

Klein, J. P., and M. L. Moeschberger. 2003. Survival Analysis: Techniques for Censored and Truncated Data. Springer, New York.

Kretzschmar, M., P. F. Teunis, and R. G. Pebody. 2010. Incidence and reproduction numbers of pertussis: estimates from serological and social contact data in five European countries. PLoS Medicine **7**.

Martin, B. E., H. Sun, M. Carrel, F. L. Cunningham, J. A. Baroch, K. C. Hanson-Dorr, S. G. Young, B. Schmit, J. M. Nolting, K. J. Yoon, M. W. Lutman, K. Pedersen, K. Lager, A. S. Bowman, R. D. Slemons, D. R. Smith, T. DeLiberto, and X. F. Wan. 2017. Feral swine in the United States have been exposed to both avian and swine influenza A viruses. Applied and Environmental Microbiology **83**:e01346–17.

McKee, E. M., E. D. Walker, T. K. Anderson, U. D. Kitron, J. D. Brawn, B. L. Krebs,

Author Manuscript

This article is protected by copyright. All rights reserved

C. Newman, M. O. Ruiz, R. S. Levine, M. E. Carrington, R. G. McLean, T. L. Goldberg, and G. L. Hamer. 2015. West Nile Virus antibody decay rate in free-ranging birds. Journal of Wildlife Diseases **51**:601–608.

Pepin, K. M., S. L. Kay, B. D. Golas, S. S. Shriner, A. T. Gilbert, R. S. Miller, A. L. Graham, S. Riley, P. C. Cross, M. D. Samuel, M. B. Hooten, J. A. Hoeting, J. O. Lloyd-Smith, C. T. Webb, and M. G. Buhnerkempe. 2017. Inferring infection hazard in wildlife populations by linking data across individual and population scales. Ecology Letters **20**:275–292.

Plowright, R. K., C. R. Parrish, H. McCallum, P. J. Hudson, A. I. Ko, A. L. Graham, and J. O. Lloyd-Smith. 2017. Pathways to zoonotic spillover. Nature Reviews Microbiology **15**:502–510.

Rohani, P., M. Keeling, and B. Grenfell. 2002. The interplay between determinism and stochasticity in childhood diseases. The American Naturalist **159**:469–481.

Simonsen, J., K. Molbak, G. Falkenhorst, K. A. Krogfelt, A. Linneberg, and P. F. M. Teunis. 2009. Estimation of incidences of infectious diseases based on antibody measurements. Statistics in Medicine **28**:1882–1895.

Sun, H., F. L. Cunningham, J. Harris, Y. Xu, L. P. Long, K. Hanson-Dorr, J. A. Baroch, P. Fioranelli, M. W. Lutman, T. Li, K. Pedersen, B. S. Schmit, J. Cooley, X. Lin, R. G. Jarman, T. J. Deliberto, and X. F. Wan. 2015. Dynamics of virus shedding and antibody responses in influenza A virus-infected feral swine. Journal of General Virology **96**:2569–2578.

Teunis, P. F. M., O. G. van der Heijden, H. E. de Melker, J. F. P. Schellekens, F. G. A. Versteegh, and M. E. E. Kretzschmar. 2002. Kinetics of the IgG antibody response to pertussis toxin after infection with *B. pertussis.* Epidemiology and Infection **129**:479–89.

Teunis, P. F. M., J. C. H. van Eijkeren, C. W. Ang, Y. T. H. P. van Duynhoven, J. B. Simonsen, M. A. Strid, and W. van Pelt. 2012. Biomarker dynamics: estimating infection rates from serological data. Statistics in Medicine **31**:2240–2248.

Author Manuscript

This article is protected by copyright. All rights reserved

Figure 1: There are four steps when estimating seasonal infection risk from serological samples at population or regional scales: **A.** Hosts are first serologically sampled over space and time. The map shows serological samples collected from feral swine (gray dots) from 2010-2017 that were then screened for influenza A virus (IAV) antibodies. The labels for each region give the sample size N and seroprevalence. **B.** Serological samples are tested using an antibody assay that provides a measure of the antibody quantity in a sample. For example, an enzyme-linked immunosorbent assays (ELISA) quantifies the antibody titer in a sample by measuring the fluorescence of a sample, relative to a known control. **C.** Quantitative antibody methods use the assay data and an estimated antibody curve to back infer the time of infection for seropositive hosts (Appendix S2). **D.** Finally, these time of infection estimates are used to infer past seasonal infection risk, accounting for the design of serological sampling. The data shown in D. are simulated data.

Figure 2: A comparison of how different sampling efforts interact with the underlying true incidence function to affect inference on the seasonality of infection risk. We used equation 1 to estimate infection risk from 1000 sampled hosts (solid green line, shaded region gives the 95% credible interval). The black lines are the true infection risk. The dashed orange lines give the estimated infection risk if uninfected (right-censored) hosts were ignored. The x-axis refers to "Days since start of infection season" that we define as the difference between TOI date and the date of the start of the infection season.

Figure 3: Regional incidence dynamics of IAV in feral swine across five regions in the USA. The x-axis gives the date in terms of month-day for a given infection season. The colored lines give the median estimated incidence, accounting for antibody dynamics, host age, and left-censoring with different criteria for determining left-censoring (80-200 days). The light black lines are 25 realizations of the incidence function from the posterior distribution of the thick, solid, 200 day left-censoring criteria line. The black lines are observed monthly seroprevalence for a given region and infection season, smoothed with a four month rolling mean. Seroprevalence and incidence are plotted on the same scale.

Author Manuscript

This article is protected by copyright. All rights reserved

Author Manuscript

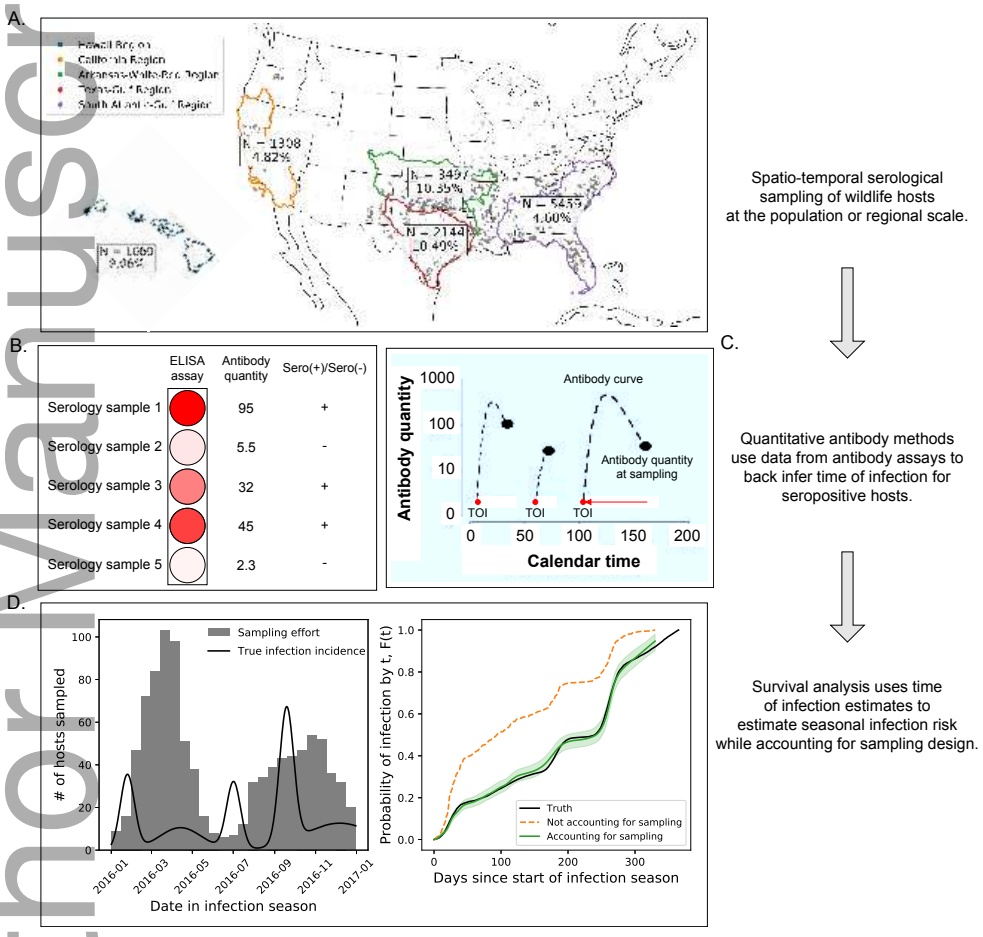


Figure 1:

This article is protected by copyright. All rights reserved

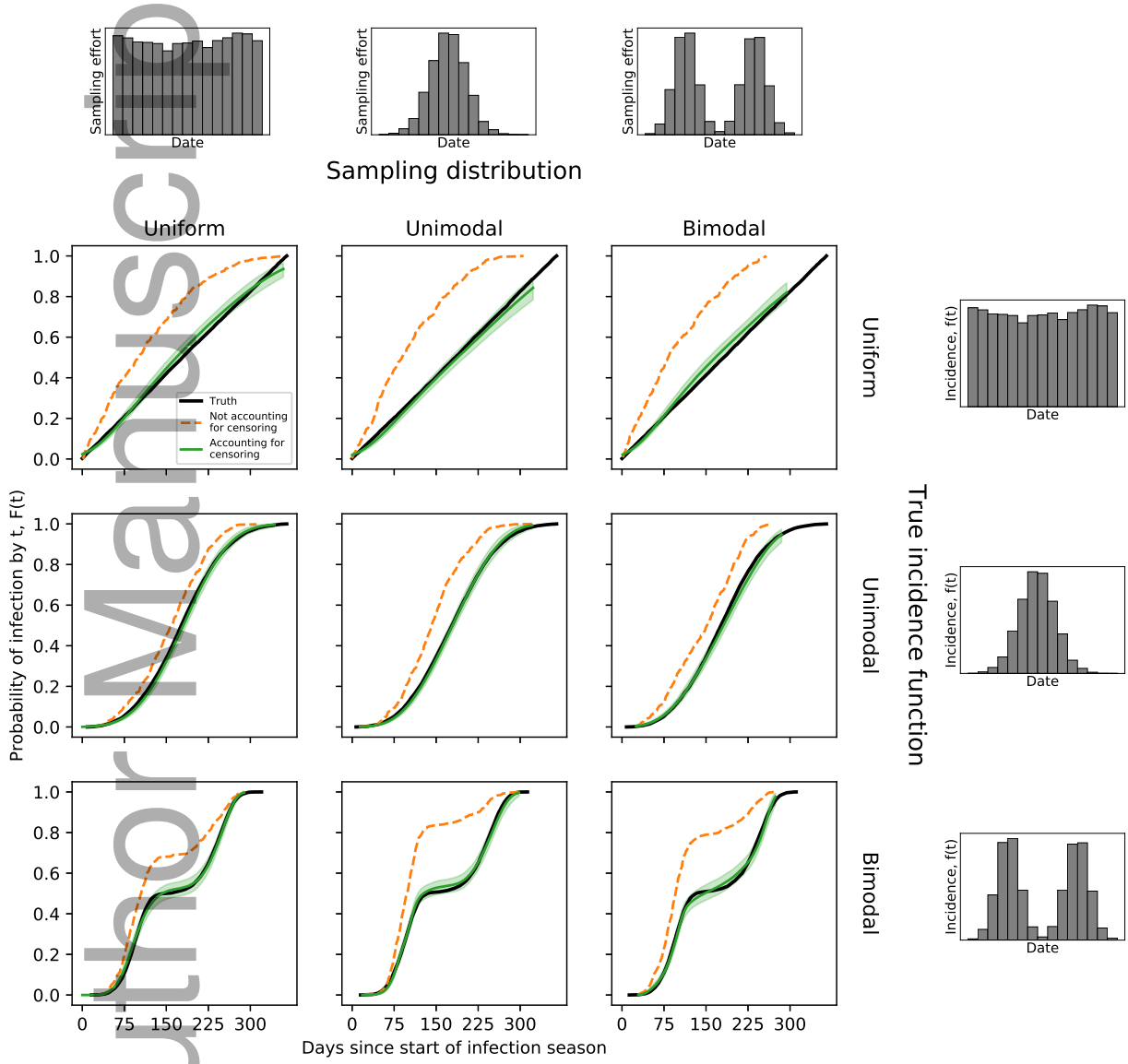


Figure 2:

Author Manuscript

This article is protected by copyright. All rights reserved

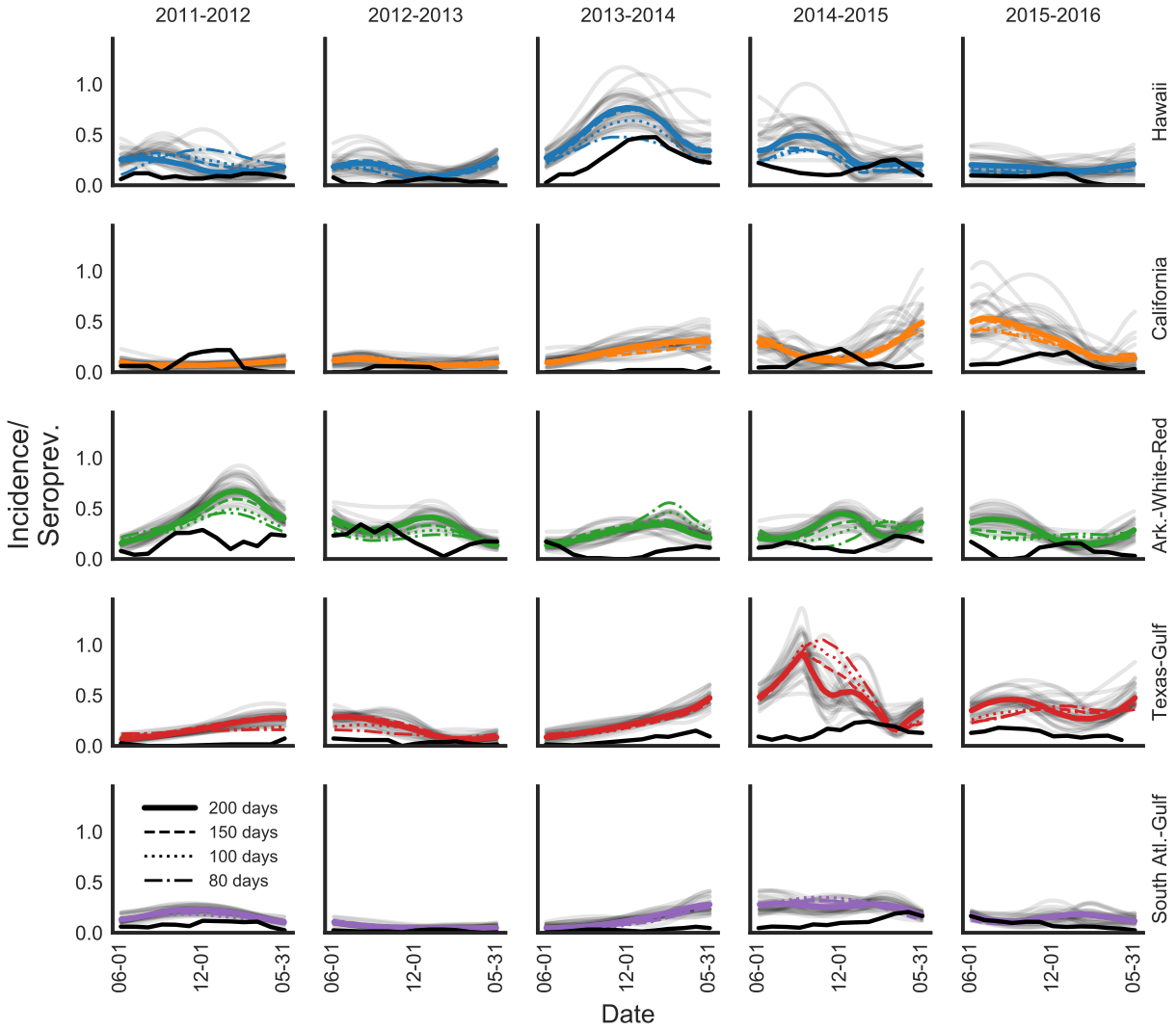


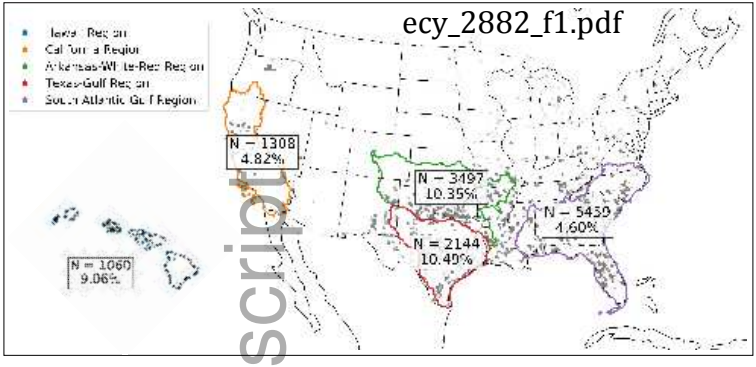
Figure 3:

Author Manuscript

This article is protected by copyright. All rights reserved

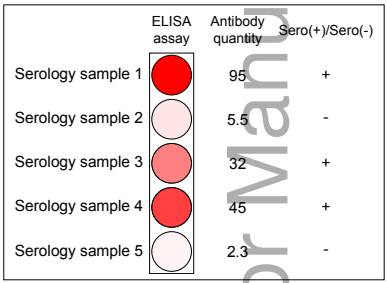
ecy_2882_f1.pdf

A.

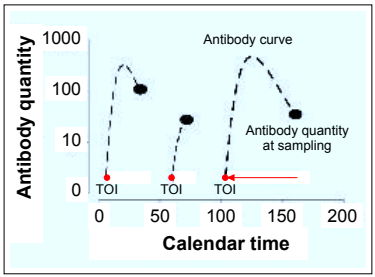


Spatio-temporal serological sampling of wildlife hosts at the population or regional scale.

B.

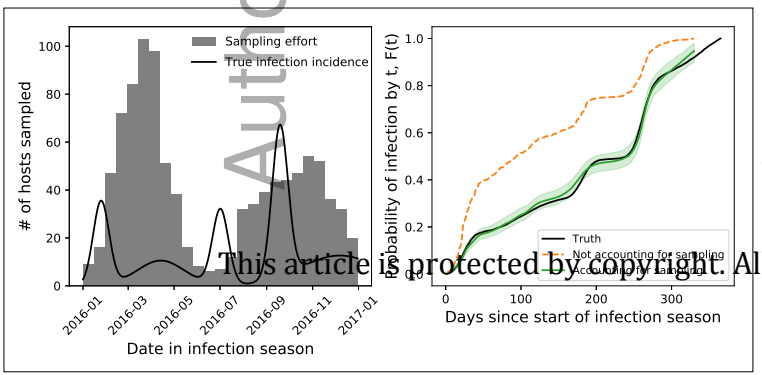


C.



Quantitative antibody methods use data from antibody assays to back infer time of infection for seropositive hosts.

D.

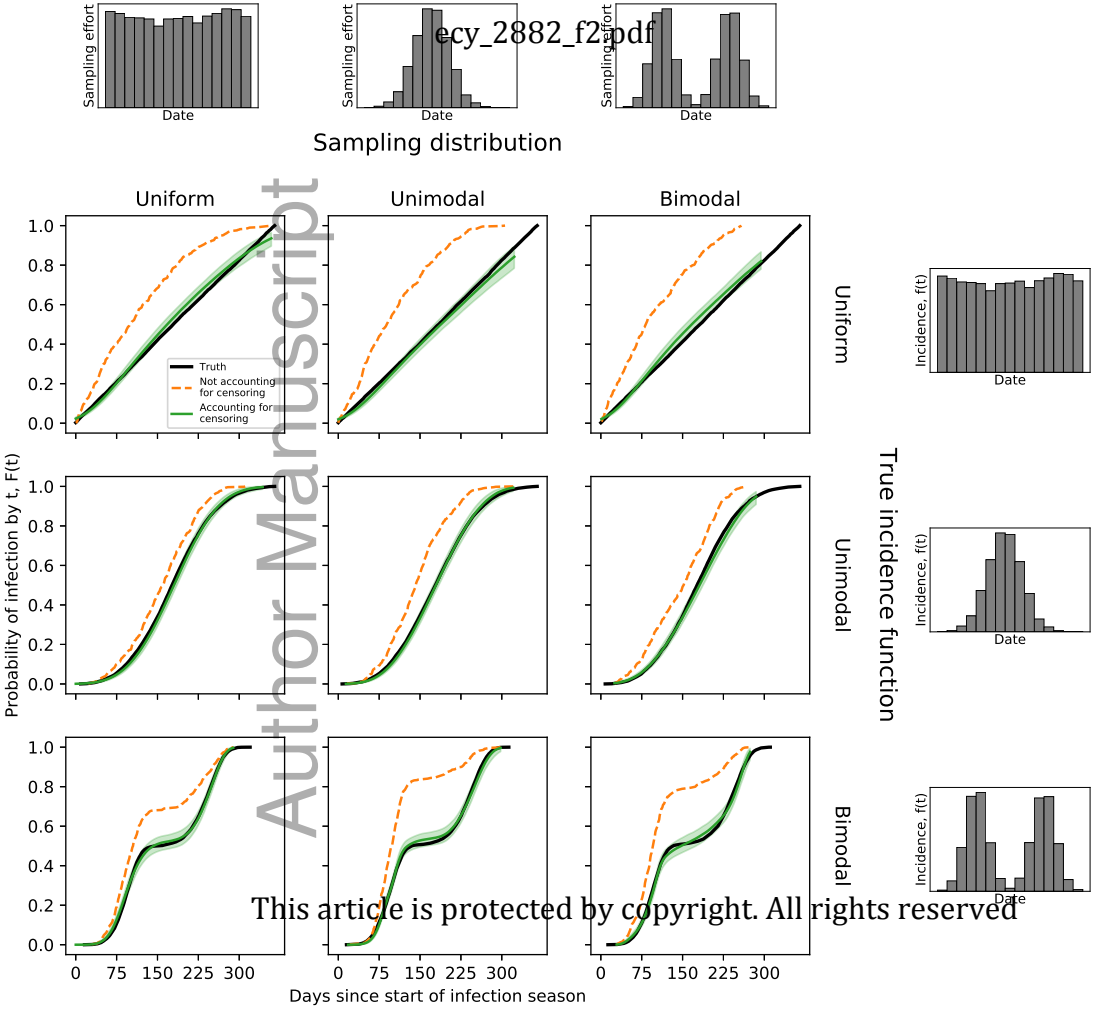


Survival analysis uses time of infection estimates to estimate seasonal infection risk while accounting for sampling design.

Author Manuscript

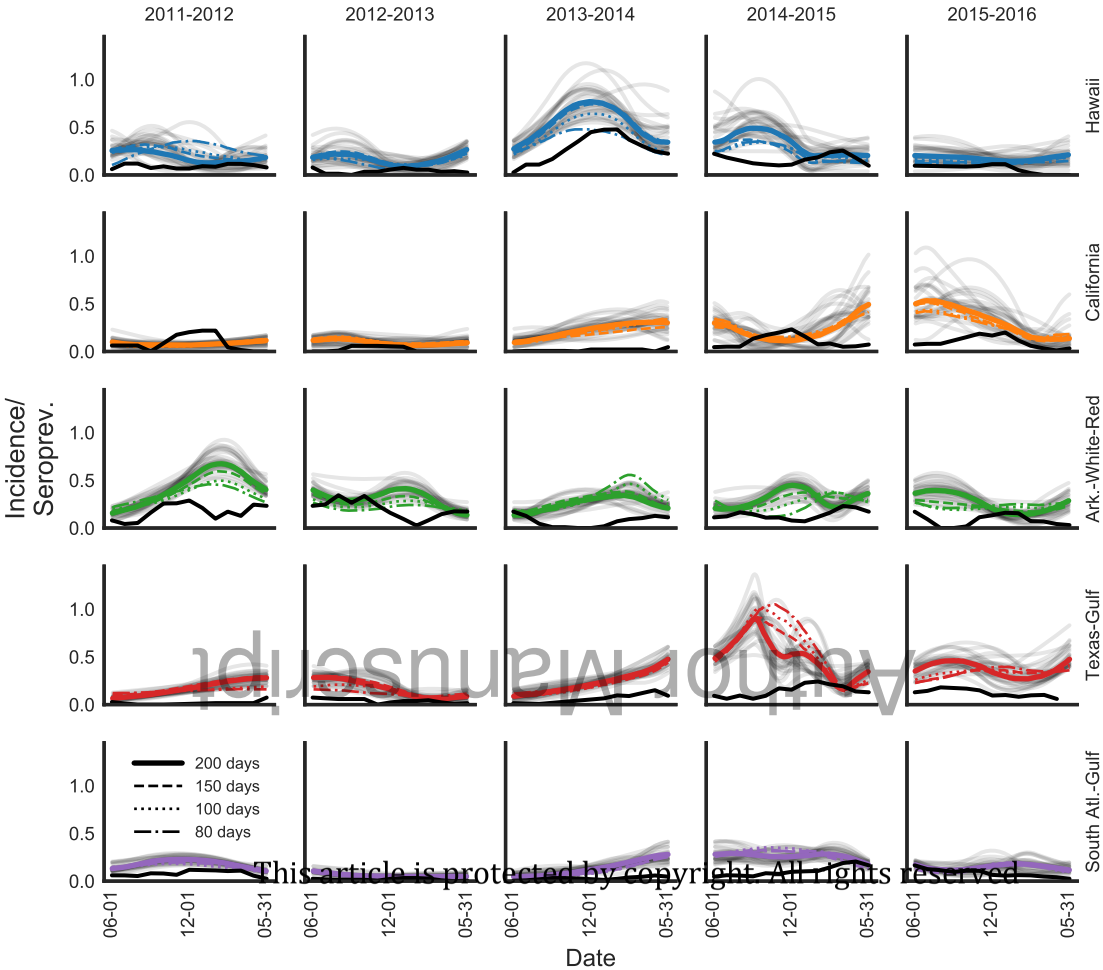
This article is protected by copyright. All rights reserved

ecy_2882_f2.pdf



Author Manuscript

This article is protected by copyright. All rights reserved



Author Manuscript

This article is protected by copyright. All rights reserved.