



HHS Public Access

Author manuscript

Ticks Tick Borne Dis. Author manuscript; available in PMC 2024 May 01.

Published in final edited form as:

Ticks Tick Borne Dis. 2024 May ; 15(3): 102329. doi:10.1016/j.ttbdis.2024.102329.

Adapting vector surveillance using Bayesian experimental design: An application to an ongoing tick monitoring program in the southeastern United States

B.K.M. Case^{a,b}, Kyndall C. Dye-Braumuller^b, Chris Evans^c, Huixuan Li^b, Lauren Rustin^c, Melissa S. Nolan^{b,c,*}

^aDepartment of Epidemiology and Biostatistics, University of Georgia, Athens, GA, USA

^bDepartment of Epidemiology and Biostatistics, Arnold School of Public Health, University of South Carolina, Columbia, SC, USA

^cSouth Carolina Department of Health and Environmental Control, Columbia, SC, USA

Abstract

Maps of the distribution of medically-important ticks throughout the US remain lacking in spatial and temporal resolution in many areas, leading to holes in our understanding of where and when people are at risk of tick encounters, an important baseline for informing public health response. In this work, we demonstrate the use of Bayesian Experimental Design (BED) in planning spatiotemporal surveillance of disease vectors. We frame survey planning as an optimization problem with the objective of identifying a calendar of sampling locations that maximizes the expected information regarding some goal. Here we consider the goals of understanding associations between environmental factors and tick presence and minimizing uncertainty in high risk areas. We illustrate our proposed BED workflow using an ongoing tick surveillance study in South Carolina parks. Following a model comparison study based on two years of initial data, several techniques for finding optimal surveys were compared to random sampling. Two optimization algorithms found surveys better than all replications of random sampling, while a space-filling heuristic performed favorably as well. Further, optimal surveys of just 20 visits were more effective than repeating the schedule of 111 visits used in 2021. We conclude that BED shows promise as a flexible and rigorous means of survey design for vector control, and could help alleviate pressure on local agencies by limiting the resources necessary for accurate information on arthropod distributions. We have made the code for our BED workflow publicly available on Zenodo to help promote the application of these methods to future surveillance efforts.

This is an open access article under the CC BY-NC license (<http://creativecommons.org/licenses/by-nc/4.0/>).

*Corresponding author. msnolan@mailbox.sc.edu (M.S. Nolan).

CRedit authorship contribution statement

B.K.M. Case: Conceptualization, Methodology, Software, Investigation, Visualization, Writing – original draft, Writing – review & editing. **Kyndall C. Dye-Braumuller:** Data curation, Methodology, Writing – review & editing. **Chris Evans:** Data curation, Supervision, Writing – review & editing. **Huixuan Li:** Methodology, Writing – review & editing. **Lauren Rustin:** Data curation. **Melissa S. Nolan:** Conceptualization, Investigation, Project administration, Supervision, Writing – review & editing.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:[10.1016/j.ttbdis.2024.102329](https://doi.org/10.1016/j.ttbdis.2024.102329). Code implementing the proposed Bayesian Experimental Design workflow can be found on Zenodo: doi: [10.5281/zenodo.10795595](https://doi.org/10.5281/zenodo.10795595).

Keywords

Bayesian statistics; Surveillance; Species distribution modeling; Experimental design

1. Introduction

Tickborne diseases have tripled in the last two decades and now make up more than 75% of reported vector-borne infections in the United States (Rosenberg et al., 2018). The continued geographic expansion of several medically important tick species is likely a key contributor to this sharp increase in tickborne disease incidence (Kugeler et al., 2015; Sonenshine, 2018). An accurate understanding of the spatial and temporal distribution of medically important ticks is a crucial first step to informing when and where people are at risk, and forms the basis of public health programs for the diagnosis and prevention of tickborne diseases (Eisen and Paddock, 2021). However, maps of tick distributions throughout the US are lacking in spatial and temporal resolution, and often depend on outdated sources, disparate sampling techniques, or otherwise biased data (Schulze et al., 1997; Wisely and Glass, 2019). Statistical models are therefore important tools for explaining factors associated with tick presence and filling gaps in existing distribution maps. However, the reliability of model predictions are critically dependent on the amount and quality of input data at an appropriate spatiotemporal scale (Kugeler and Eisen, 2020). An area with particularly limited knowledge of current tick distributions is the southeastern US, where local resources for monitoring and control are scarce and less than 10% of vector control agencies perform tick surveillance of any kind (Dye-Braumuller et al., 2022; Mader et al., 2021).

In addition to learning from existing data, a further use of tick distribution models is informing future surveillance and control efforts by anticipating the value of future sampling locations. For example, more fine-grained sampling might follow an initial surveillance effort focused on a subset of areas of potentially high risk (Diuk-Wasser et al., 2010). Though usually such sampling decisions are made in an ad hoc manner, a decision-theoretic alternative is to assign a score to potential sampling times and locations and attempt to find visits maximizing some objective function (Dorazio and Johnson, 2003). Deciding future visits for surveillance or control efforts may then be framed as an optimal experimental design problem, where a calendar of sampling times and/or locations is chosen to maximize their information content or increase the impact of vector control measures (Case et al., 2022; Reich et al., 2018). Because observational studies of tick distributions involve complex interactions between variable environmental conditions, population dynamics, and the collection method used, it is important to employ design criteria that are robust to uncertainty in model parameters and experimental conditions (Pronzato and Walter, 1985).

Thanks to computational advances in recent decades, Bayesian inference has become popular for model fitting in ecology and epidemiology (Parent and Rivot, 2012). The advantages of Bayesian inference include a complete treatment of model uncertainty through providing posterior distributions of plausible model parameters and predicted tick distributions, and the ability to incorporate entomological or medical knowledge through

prior beliefs. For experimental design, the Bayesian paradigm provides a framework that is robust to system uncertainty and flexible for tailoring novel design criteria to specific challenges in public health (Diggle et al., 2021; Ryan et al., 2016).

Implementing Bayesian Experimental Design (BED) involves three general steps (Fig. 1). First, a statistical model is chosen based on its suitability to explain any existing survey data, resulting in a posterior distribution that represents how much we have learned based on this initial survey. Here, learning indicates our level of confidence about inferred parameters in the model, as well as our confidence in predicted tick distributions. While we focus on binary presence/absence data in this work, other metrics such as tick abundance may be used with an appropriate choice of model. Next, a *utility function* is developed to score potential future survey schedules based on the expected quality of new information. How new information is quantified depends on the specific goals of surveillance, for example whether the objective is to reduce some form of uncertainty or to maximize the number of sites positive for ticks. Finally, the space of possible designs is searched for highly informative survey schedules using optimization. The output of these three steps is then a calendar of sampling times and locations that can be directly implemented as a future surveillance effort. Note that while an initial dataset is not required for BED, we frame the method for the situation where preliminary data are available from previous surveillance efforts. This information can help inform an appropriate model choice, and more effectively differentiate high quality designs (Ryan et al., 2015; Zhang, 2006).

In this work, we outline principles for how BED can be incorporated in spatiotemporal surveys to maximize the value of vector surveillance and control efforts, and we illustrate their use for an ongoing tick surveillance effort of South Carolina state parks and other public lands. We compare the ability of different search techniques to find survey schedules that maximize utility based on two design criteria tailored to different priorities of vector surveillance. In addition to informing future data collection efforts, we demonstrate how high utility designs can be further analyzed to provide novel insight into sources of uncertainty in tick distributions.

2. Methods

2.1. Data collection and preparation

Data originated from an existing state-wide tick surveillance project in South Carolina city and state parks, beginning in March 2020. The project also included submissions from South Carolina animal shelters and citizen scientists, though these data were not used in this study. Here we used data from 2020 to 2021 state and city parks, with observations in 30 distinct parks spanning 26 counties from March to December. Sampling was relatively opportunistic; more visits took place between March and August compared to later months, and 4 parks were visited at least 10 times over the two years while 3 parks were visited just once. In total, 59 distinct visits occurred in 2020, and 111 visits occurred in 2021.

A scientific collecting permit from the SC Department of Parks, Recreation, & Tourism was secured for both years, and written permission was granted from the appropriate municipal government for city parks. The coordinates for each site were selected near the

entrance to each park in a forested area for consistency. Tick collections were performed following recommended CDC Ixodidae surveillance guidelines (Centers for Disease Control and Prevention, 2021). In brief, tick traps consisting of a 0.61m² muslin cloth baited with 1.5-lb dry ice each were placed in parks along hiking and nature trails and left in the park for 1.5–2 h. Additionally, tick drags were performed along hiking and nature trails. Tick drags were constructed with a 1.22-m x 1.52-m white duck canvas attached to a 1.22-m wooden dowel, with zinc washers as weights on the bottom. Each collection visit consisted of ten tick traps and a 30 min timed tick drag (sixty 30-second segments to regularly check for ticks) to ensure that the recommended surface area of at least 750 m for host-seeking ticks was surveyed (Centers for Disease Control and Prevention, 2021).

Ticks were processed at the Laboratory of Vector-Borne and Zoonotic Diseases at the University of South Carolina, where they were identified to species, sex, and life stage. Morphological identifications were conducted with multiple dichotomous keys (Clifford and Anastos, 1960; Keirans and Durden, 1998; Keirans and Litwak, 1989). The response outcome for each visit was then recorded as a binary variable of presence/absence for each species, where presence indicates nymphs were found with at least one of the two sampling methods. *Amblyomma maculatum*, *Ixodes brunneus*, and *I. kearnsi* (formerly *I. affinis* (Nava et al., 2023)) were present in only 3, 5, and 12 of these visits, respectively. Therefore to improve the quality of model predictions, *A. maculatum* was removed from all further analyses, and the *Ixodes* species were aggregated into a single *Ixodes* spp. group. *Amblyomma americanum* was found in 45% of the 170 visits, *Ixodes* spp. was found in 24%, and *Dermacentor variabilis* in 8%.

2.2. Environmental risk factors

Several meteorological and geographic variables were selected as potential covariates of tick occurrence based on tick ecology and previous modeling studies (Hahn et al., 2016; Lippi et al., 2021). Land cover and forest canopy data were obtained from the USGS 2019 National Land Cover Database (Dewitz, 2021), while elevation and meteorological variables were obtained from PRISM (Oregon State University, 2014). Included meteorological variables were monthly total precipitation, monthly average of the daily maximum temperature, monthly average of the daily minimum temperature, average daily minimum temperature in January, and monthly average of the daily average humidity, calculated from average daily temperature and average dew point temperature (Alduchov and Eskridge, 1996). Since we are interested in making predictions for future years, 30-year meteorological monthly normals were used, defined as the value for each month on average over the last 30 years. The continuous covariates were centered and scaled prior to all statistical analyses, and the monthly minimum temperature was removed due to multicollinearity.

2.3. Modeling tick distributions using Bayesian regression

Complete mathematical definitions for the model specification and experimental design procedure are given in the Supplemental Methods. We used a hierarchical, mixed-effects framework, where correlations between the observed tick distributions and various environmental, spatial, and temporal effects are captured (Parent and Rivot, 2012). The probability of encountering a tick of species j , in a visit to site i during month t , is therefore a

function of the 7 environmental covariates, the survey location, and the month the visit took place.

To find a model parsimonious with the collections data, 28 candidate models were constructed from simplifying different components of the full model. Each of the environmental, spatial, and temporal model components were considered either shared or different between tick groups, and linear or spline-based functions were considered for the environmental effects. Models were compared based on the Deviance Information Criterion, which measures a model's goodness-of-fit to the data and robustness, while penalizing model complexity. The best performing model was used in all subsequent analyses.

2.4. Experimental design for vector surveillance

Considering the limited capacity of vector control agencies and researchers, a reasonable surveillance strategy should consider the feasibility and convenience of sampling sites while allowing sufficient diversity to realistically be able to extrapolate to the region of interest. To maintain this balance, we restricted future sampling to a set of 57 sites on public land across South Carolina, of which the 30 sites visited in the initial data were a subset. These candidate sites included all 47 South Carolina state parks and historic sites, 6 locations within national parks and wildlife refuges, and 4 other locations present in the initial data. Collection visits were delineated monthly and could take place in any month, resulting in a space of 684 possible visits (i.e. month-location pairs) which may be added to a candidate survey.

In BED, potential outcomes (i.e. presence/absence of ticks) resulting from a proposed design are assigned a *utility*, or a score based on the quality of new information provided by that outcome. Since the outcome of a visit is uncertain prior to sampling, the utility of the design is then averaged over the posterior predictive distribution of possible outcomes, i.e. predictions given the initial survey data (Fig. 1, step 2). To score potential outcomes, two design criteria were considered. First was a form of Bayesian d-optimality, which quantifies the volume of uncertainty in the posterior covariance matrix for the environmental effects (Chaloner and Verdinelli, 1995). This criterion measures how much has been learned about the environmental effects as a whole, which grants a fuller representation of uncertainty compared to examining the variance of each effect individually. Broadly speaking, this criterion will reward surveillance schedules that maintain a diverse range of environmental conditions. A second criterion was then designed to improve the reliability of prediction maps in regions where risk of exposure is highest. Here we assigned utility based on the maximum reduction in standard deviation of risk from the initial dataset among high risk areas, where high risk areas were defined as any point in the spatiotemporal prediction map with expected risk greater than 0.75 for at least one tick species. These high risk points were concentrated along the southern shoreline of the state in the months of March through May. The maximum rather than mean reduction in uncertainty was chosen for greater variety compared to the first criterion, and represents an “all-in” approach where visits are chosen to target uncertainty at a specific part of the prediction map. In contrast, the d-optimality criterion is a more “big picture” goal that reflects learning across all visits.

Equipped with a predictive model and utility function, the space of possible designs can be searched to optimize utility (Fig. 1, step 3). Because Bayesian design criteria are available in closed form only in the simplest cases, mathematical formulae for the utility function are not available and numerical methods are typically used. This design space is far too vast to test the utility of most designs, therefore specialized search techniques are needed to find designs of high quality (Ryan et al., 2016). We implemented two optimization algorithms – first was a simulated annealing algorithm, a common stochastic search technique which incrementally adds and removes visits to a design while avoiding local minima, and second was an exchange algorithm which rotates through neighboring months and sites until no neighbors improve utility (Reich et al., 2018). We also considered two *heuristics* to choose designs based on characteristics suspected to lead to high utility, while requiring fewer computational resources than optimization. The first chose visits with the highest predicted variance given the existing collections data, and the second was a space-filling design that spread visits evenly over time and prohibited visiting any two parks less than 25 km away from each other (Chipeta et al., 2017). We then compared these four search strategies to random sampling for an increasing sample size from 5 to 20 visits, as well as to several repeated sampling scenarios which were constructed by repeating aspects of the initial collections data. The repeated sampling schemes were to visit the 30 distinct sites from the initial collections in June, to visit these 30 sites in December, and to repeat the exact schedule performed in 2021.

3. Results

In the model comparison study, the model with lowest DIC included linear environmental effects shared between species, and both spatial and temporal effects separate for each species, although several alternative models performed nearly as well (Figure S1). The posterior environmental and temporal effects from the best performing model are summarized in Fig. 2. Daily maximum temperature and January minimum temperature both had a strong positive association with tick presence, while precipitation had a strong negative association, as coefficients for these variables had over 95% of the posterior mass above/below zero. Relative humidity, elevation, and a Mixed Forest land cover also appear negatively associated with tick presence. The coefficients for land cover had particularly high variance, likely due to less data being available for any single land cover class. Spatiotemporal prediction maps throughout South Carolina for each tick group are shown in Fig. 3, while the average predicted risk of visits in the initial collections data are compared to the true presence/absence in Figure S2. There appears to be reasonably good agreement between the expected risk and the observed data. The average predicted risk for both *A. americanum* and *Ixodes* spp. was high in the southeast of the state, though in different months, while average risk was consistently low for all species in the northwest. Ultimately, Fig. 3 shows considerable model uncertainty in tick presence remains throughout much of the state, illustrating the importance of continued surveillance.

Fig. 4 summarizes the results from the simulation study, with utilities of the designs produced by each search method compared to random sampling. For the d-optimality criterion, both of the optimization algorithms and the space-filling heuristic were able to find designs better than any random sample, and had very similar performance for all sample

sizes, although the space-filling strategy was less computationally expensive. The variance heuristic performed comparably or worse than random sampling. The repeated sampling designs based on the initial schedule also had much lower utility relative to their larger sampling budget – revisiting all 30 previously visited sites in June was generally worse than choosing 20 visits uniformly at random, as well as a schedule of just 10 visits found using exchange. Repeating all 111 visits from 2021 had lower utility than designs of 20 visits. Results were similar for the second criterion of maximum variance reduction in high risk areas, although space-filling was less effective than exchange and simulated annealing. Here, repeating the 111 visits from 2021 only slightly outperformed the best designs of just 20 visits. For this criterion, the best design of 20 locations had an expected utility of 0.1, which amounts to a 30–50% reduction in uncertainty for a particular future visit.

To more concretely visualize the output of our experimental design pipeline, panels C and D of Fig. 4 show the spatiotemporal distribution of proposed visits for both design criteria, laid over the design of the initial surveillance data for comparison. There was almost no overlap between the initial visits and the proposed designs, with just three such occurrences across all design criteria and search algorithms. Similarly, identical visits were proposed by the different search techniques in only two instances. For both criteria, successful designs spread visits fairly evenly across months, although interestingly few visits were proposed in the late summer/fall for either criteria. In total over both criteria, January and December were the most sampled months while August had the least proposed visits. Over the course of the year, each strategy also proposed a balanced spread of locations across the state, although the middle of the state appears under-sampled relative to the number of parks that are located there.

To further gain a sense of what makes certain surveys better than others, the environmental conditions of each possible visit were then projected using Factor Analysis for Mixed Data (FAMD), a dimension reduction technique for continuous and categorical data (Pagès, 2004). The associated covariates are embedded within the four most important dimensions and plotted in Figure S3, with the covariates selected by each of the search strategies highlighted. For the first criterion, the three successful strategies chose visits that spread covariates throughout the center of the two most important dimensions, while the variance heuristic placed points on the edges and far away from the observed data. For the second criterion the successful designs instead placed many more points in the bottom-right quadrant of the first and second dimensions, suggesting that a more specific combination of survey conditions were critical for maximizing information regarding certain high risk points, although the spread of designs in Fig. 4D makes clear that some sampling variability remains beneficial.

4. Discussion

Accurate information regarding the time and place of probable tick encounters is an essential first step to reducing the burden of tick-borne pathogens. Statistical modeling allows extrapolating available information to a wider scale, which in turn enables local vector control agencies to better direct critical resources. However, the reliability of such model predictions are critically dependent on the nature of available data. Combining a Bayesian

workflow and design of experiments is a principled approach to getting more out of data from existing surveillance efforts, and directing future efforts for the greatest effect. Thanks to advances in software and computing throughout the last decade, optimal Bayesian survey design is feasible to implement for a diverse array of researchers throughout epidemiology.

Our results for the application of scheduling monthly tick surveillance in public natural areas demonstrate large gains in information are possible through carefully chosen surveys. Even when restricting sampling to a limited number of locations, efficient survey design can make the difference for learning critical information and improving reliability of tick distribution maps. Although implementing an optimal surveillance design may require additional planning prior to sampling, the higher quality of information provided by such designs means that ultimately fewer visits are required to reach a certain level of confidence. Optimal sampling can therefore serve to reduce the overall resources necessary for effective surveillance.

Successful designs can also inform general practices for surveillance. For example, for the first design criterion based on environmental covariates, the success of a space-filling strategy shows that spreading future visits across time and space is more valuable than other intuitive options such as focusing sampling on specific months which were previously under-sampled. Because similar environmental conditions will tend to be clustered in time and space, spreading visits in this way will tend to spread design points evenly across covariate space as well, which is theoretically optimal with respect to the d-optimality criterion for simple logistic regression (Chaloner and Larntz, 1989). A similar takeaway regarding the second criterion of variance reduction among high risk points on the prediction map is that focusing sampling on months or parks also with high suspected risk is not an efficient strategy. Instead, effective surveys should avoid sampling redundancy and maintain a diversity of visits to identify remaining sources of uncertainty for these high risk predictions.

While we have emphasized finding efficient designs using optimization, a secondary takeaway from our analysis is that random sampling generally outperforms some common forms of convenience sampling, such as repeated sampling or sampling based on current uncertainty. This reinforces random sampling as a gold standard catch-all technique that is independent of the design criterion (Diekmann et al., 2007). Thus, although we have demonstrated that the most effective surveys for minimizing uncertainty in tick distribution maps must be found using optimization, random sampling could still be a reasonable choice when the design goals of future analyses are unclear, or for a preliminary round of sampling when little existing information is available.

Analysis of the initial survey data during model comparison provides insight into the current tick patterns in natural areas throughout the southeastern US, while also demonstrating further data are needed. The top performing models all included a temporal trend for each tick species (Figure S1, and the posterior marginals for each trend show strong seasonal patterns (Fig. 2B). These residual trends could indicate contributions from variables not included in the analysis such as isothermality (Hahn et al., 2016), or from seasonality in ecological factors such as host availability. Another explanation for this temporal trend is our

use of 30-year normals data, which ignores climatic differences between the two years in the initial data that could lead to a temporal offset in risk between years.

The best suited models all included a term for spatial variability for each tick species, which has previously been deemed important for modeling *I. scapularis* density (Diuk-Wasser et al., 2010). Overall, the large variance in both spatial and temporal effects among top models suggests uncertainty in tick presence is due to a combination of unmeasured environmental effects, population dynamics, and variability inherent to all methods of tick collection in a natural environment (Rynkiewicz and Clay, 2014). It is also interesting that the top performing model did not include species-dependent nor non-linear effects with the environment, as the importance of such non-linear environmental effects are frequently stressed when modeling tick distributions (Elias et al., 2021; Hahn et al., 2016). This is likely due in part to the relatively limited geographic range of our data.

Our prediction map of expected probability of tick presence in South Carolina generally agrees with previously published results, although data at a similar spatial and temporal scale are limited. For *I. scapularis*, county level data show 30% of South Carolina counties had an established tick presence by 2015, most of which were in the southeast region of the state (Eisen et al., 2016), while models calibrated to the same data predicted suitability in the center of the state as well (Hahn et al., 2016). County-level establishment of *A. americanum* follows a similar pattern as *I. scapularis* (Springer et al., 2014), although predictions based on that data indicated all counties were highly suitable (Springer et al., 2015). This previous prediction of *A. americanum* in the northwest of the state is in contrast with our findings, as the species was never encountered during initial data collection in the region, and our model predicted low risk there in all months.

The application of BED for vector surveillance used in this work focused on establishing tick presence in public natural areas, although we note that the framework used here can be applied to other metrics such as abundance with minimal changes. While measuring tick presence or abundance in outdoor recreational areas such as state parks is a widely used method for establishing human exposure risk (Falco and Fish, 1989; Hassett et al., 2022), and for detecting expanding ranges of ticks and tick-borne pathogens (Johnson et al., 2017), the reliability of such data for predicting individualized risk of infection is unclear. For Lyme disease, it has been suggested that private property is the main source of exposure to host-seeking nymphs (Eisen and Eisen, 2016; Mead et al., 2018). The infection status of ticks is also a critical source of information, although the importance of measuring density of infected ticks compared to tick prevalence likely depends on the study area. At the county level, nymph density has been found ineffective for predicting Lyme disease incidence in low incidence areas, but is comparable to density of infected nymphs in high incidence counties (Pepin et al., 2012). Another limitation with our assumed data collection method is that patterns of tick presence and abundance will vary greatly over the span of a state park. Thus, establishing the distribution of ticks is ultimately just a single step to any comprehensive strategy for vector surveillance and control.

The BED procedure illustrated here suggests several avenues for future statistical and computational development. First, additional work is needed to better understand optimal

designs for the types of mixed-effects models used in this and other studies of species distributions, as research combining BED and mixed-effects models is scarce (Reich et al., 2018; Ryan et al., 2015). Second, specialized optimization strategies for finding optimal survey schedules should be developed, as spatiotemporal survey design presents distinct challenges such as incomplete control over the environmental conditions available among possible visits. Another possibility is to employ adaptive sampling, where locations are visited in smaller batches and the data collected from each iteration are able to inform sampling in future batches, although updating data sequentially leads to additional logistical constraints during surveillance (Case et al., 2022; Chipeta et al., 2016). A final avenue for future work is in the choice of design criteria, which may change depending on the specific goals of the analysis. While we simply restricted designs to a certain number of visits, researchers with different goals or specific resource limitations could employ criteria that account for distance traveled, availability of materials, or density of human traffic at a particular park and month. More broadly, a general procedure for design criteria that adapt to the current needs of local vector control agencies would allow widespread application of experimental design strategies at a fine-grained spatiotemporal scale.

5. Conclusions

In this work, we have outlined Bayesian Experimental Design as a formal approach to the surveillance of disease vectors. Compared to classical methods of experimental design, a Bayesian framework provides a natural way to incorporate initial survey data, while rigorously accounting for remaining uncertainty in model predictions. We applied a BED workflow to an ongoing tick surveillance study in South Carolina state parks, and found that surveys optimized to satisfy specific goals were universally more efficient than simple random sampling. These results demonstrate the promise of optimal survey design for researchers and vector control agencies to maximize the impact of the data they collect.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

The authors would like to thank Drs. Banky Olatosi and Xiaoming Li for their programmatic support of BC during the T35 summer training program at the University of South Carolina.

Funding support

BC received financial support from National Institutes for Health T35 grant 5T35AI165252-02 and 1P20 GM125498-01 Centers of Biomedical Research Excellence Award. This work was also partially supported by the Centers for Disease Control and Prevention Epidemiology and Laboratory Capacity for Prevention and Control of Emerging Infectious Diseases NU50CK000542-04-00.

Data availability

All code and data necessary to reproduce our analyses are available on Zenodo, doi: [10.5281/zenodo.10795595](https://doi.org/10.5281/zenodo.10795595)

References

- Alduchov OA, Eskridge RE, 1996. Improved Magnus Form Approximation of Saturation Vapor Pressure. *J. Appl. Meteorol. Clim* 35, 601–609. 10.1175/1520-0450(1996)035<0601.IMFAOS>2.0.CO;2.
- Case BKM, Young J-G, Penados D, Monroy C, Hébert-Dufresne L, Stevens L, 2022. Spatial epidemiology and adaptive targeted sampling to manage the Chagas disease vector *Triatoma dimidiata*. *PLOS Negl. Trop. Dis* 16, e0010436 10.1371/journal.pntd.0010436. [PubMed: 35653307]
- Centers for Disease Control and Prevention, 2021. Guide to the surveillance of Metastriate Ticks (Acari: ixodidae) and their Pathogens in the United States [WWW Document]. URL <https://www.cdc.gov/ticks/surveillance/MetastriateTicks.html> (accessed 9.19.23).
- Chaloner K, Larntz K, 1989. Optimal Bayesian design applied to logistic regression experiments. *J. Statist. Plann. Inference* 21, 191–208. 10.1016/0378-3758(89)90004-9.
- Chaloner K, Verdinelli I, 1995. Bayesian experimental design: a review. *Inst. Math. Stat* 10, 273–304.
- Chipeta M, Terlouw D, Phiri K, Diggle P, 2017. Inhibitory geostatistical designs for spatial prediction taking account of uncertain covariance structure. *Environmetrics* 28, 1–11. 10.1002/env.2425.
- Chipeta M, Terlouw D, Phiri K, Diggle P, 2016. Adaptive geostatistical design and analysis for prevalence surveys. *Spat. Stat* 15, 70–84. 10.1016/j.spasta.2015.12.004.
- Clifford CM, Anastos G, 1960. The use of chaetotaxy in the identification of larval ticks (Acarina: ixodidae). *J. Parasitol* 46, 567–578. 10.2307/3274939. [PubMed: 13694092]
- Dewitz J., 2021. National land cover database (NLCD) 2019 products. 10.5066/P9KZCM54.
- Diekmann M, Kühne A, Isermann M, 2007. Random vs non-random sampling: effects on patterns of species abundance, species richness and vegetation-environment relationships. *Folia Geobot* 42, 179–190. 10.1007/BF02893884.
- Diggle P, Amoah B, Fronterre C, Giorgi E, Johnson O, 2021. Rethinking neglected tropical disease prevalence survey design and analysis: a geospatial paradigm. *Trans. R. Soc. Trop. Med. Hyg* 115, 208–210. 10.1093/trstmh/trab020. [PubMed: 33587142]
- Diuk-Wasser MA, Vourc'h G, Cislo P, Hoen AG, Melton F, Hamer SA, Rowland M, Cortinas R, Hickling GJ, Tsao JI, Barbour AG, Kitron U, Piesman J, Fish D, 2010. Field and climate-based model for predicting the density of host-seeking nymphal *Ixodes scapularis*, an important vector of tick-borne disease agents in the eastern United States. *Glob. Ecol. Biogeogr* 19, 504–514. 10.1111/j.1466-8238.2010.00526.x.
- Dorazio RM, Johnson FA, 2003. Bayesian inference and decision theory—a framework for decision making in natural resource management. *Ecol. Appl* 13, 556–563. 10.1890/1051-0761(2003)013[0556:BIADTA]2.0.CO;2.
- Dye-Braumuller KC, Gordon JR, Johnson D, Morrissey J, McCoy K, Dinglasan RR, Nolan MS, 2022. Needs assessment of southeastern United States vector control agencies: capacity improvement is greatly needed to prevent the next vector-borne disease outbreak. *Trop. Med. Infect. Dis* 7, 73. 10.3390/tropicalmed7050073. [PubMed: 35622700]
- Eisen L, Eisen RJ, 2016. Critical evaluation of the linkage between tick-based risk measures and the occurrence of lyme disease cases. *J. Med. Entomol* 53, 1050–1062. 10.1093/jme/tjw092. [PubMed: 27330093]
- Eisen RJ, Eisen L, Beard CB, 2016. County-Scale Distribution of *ixodes scapularis* and *ixodes pacificus* (Acari: ixodidae) in the continental United States. *J. Med. Entomol* 53, 349–386. 10.1093/jme/tjv237. [PubMed: 26783367]
- Eisen RJ, Paddock CD, 2021. Tick and tickborne pathogen surveillance as a public health tool in the United States. *J. Med. Entomol* 58, 1490–1502. 10.1093/jme/tjaa087. [PubMed: 32440679]
- Elias SP, Gardner AM, Maasch KA, Birkel SD, Anderson NT, Rand PW, Lubelczyk CB, Smith RP Jr., 2021. A generalized additive model correlating blacklegged ticks with white-tailed deer density, temperature, and humidity in Maine, USA, 1990–2013. *J. Med. Entomol* 58, 125–138. 10.1093/jme/tjaa180. [PubMed: 32901284]
- Falco RC, Fish D, 1989. Potential for exposure to tick bites in recreational parks in a Lyme disease endemic area. *Am. J. Public Health* 79, 12–15. 10.2105/AJPH.79.1.12.

- Hahn MB, Jarnevich CS, Monaghan AJ, Eisen RJ, 2016. Modeling the geographic distribution of *Ixodes scapularis* and *Ixodes pacificus* (Acari: ixodidae) in the contiguous United States. *J. Med. Entomol* 53, 1176–1191. 10.1093/jme/tjw076. [PubMed: 27282813]
- Hassett E, Diuk-Wasser M, Harrington L, Fernandez P, 2022. Integrating tick density and park visitor behaviors to assess the risk of tick exposure in urban parks on Staten Island, New York. *BMC Public Health* 22, 1602. 10.1186/s12889-022-13989-x. [PubMed: 35999523]
- Johnson TL, Graham CB, Boegler KA, Cherry CC, Maes SE, Pilgard MA, Hojgaard A, Buttke DE, Eisen RJ, 2017. Prevalence and diversity of tick-borne pathogens in nymphal *Ixodes scapularis* (Acari: ixodidae) in Eastern National Parks. *J. Med. Entomol* 54, 742–751. 10.1093/jme/tjw213. [PubMed: 28028138]
- Keirans JE, Durden LA, 1998. Illustrated key to nymphs of the tick genus *Amblyomma* (Acari: ixodidae) found in the United States. *J. Med. Entomol* 35, 489–495. 10.1093/jmedent/35.4.489. [PubMed: 9701933]
- Keirans JE, Litwak TR, 1989. Pictorial key to the adults of hard ticks, family ixodidae (Ixodida: ixodoidea), East of the Mississippi River. *J. Med. Entomol* 26, 435–448. 10.1093/jmedent/26.5.435. [PubMed: 2795615]
- Kugeler KJ, Eisen RJ, 2020. Challenges in predicting lyme disease risk. *JAMA Netw. Open* 3, e200328. 10.1001/jamanetworkopen.2020.0328. [PubMed: 32125424]
- Kugeler KJ, Farley GM, Forrester JD, Mead PS, 2015. Geographic distribution and expansion of human lyme disease, United States. *Emerg. Infect. Dis* 21, 1455–1457. 10.3201/eid2108.141878. [PubMed: 26196670]
- Lippi CA, Gaff HD, White AL, Ryan SJ, 2021. Scoping review of distribution models for selected *Amblyomma* ticks and rickettsial group pathogens. *PeerJ* 9, e10596. 10.7717/peerj.10596. [PubMed: 33643699]
- Mader EM, Ganser C, Geiger A, Harrington LC, Foley J, Smith RL, Mateus-Pinilla N, Teel PD, Eisen RJ, 2021. A survey of tick surveillance and control practices in the United States. *J. Med. Entomol* 58, 1503–1512. 10.1093/jme/tjaa094. [PubMed: 34270770]
- Mead P, Hook S, Niesobecki S, Ray J, Meek J, Delorey M, Prue C, Hinckley A, 2018. Risk factors for tick exposure in suburban settings in the Northeastern United States. *Ticks Tick-borne Dis* 9, 319–324. 10.1016/j.tbd.2017.11.006. [PubMed: 29174449]
- Nava S, Beati L, Venzal JM, Durden LA, Bermudez SE, Tarragona EL, Mangold AJ, Gleason D, Mastropaolo M, Guglielmone AA, 2023. Description of two new species in the *Ixodes ricinus* complex from the New World (Acari: ixodidae), and redescription of *Ixodes affinis* Neumann, 1899. *Zootaxa* 5361, 53–73. 10.11646/zootaxa.5361.1.2. [PubMed: 38220777]
- Oregon State University, 2014. PRISM Climate Group [WWW Document]. URL <https://prism.oregonstate.edu/normals/> (accessed 9.19.23).
- Pagès J., 2004. Analyse factorielle multiple de données mixtes: principe et exemple d'application. *Rev. Stat. Appl* 52, 93–111.
- Parent E, Rivot E, 2012. Introduction to Hierarchical Bayesian Modeling for Ecological Data. CRC Press, Boca Raton.
- Pepin KM, Eisen RJ, Mead PS, Piesman J, Fish D, Hoen AG, Barbour AG, Hamer S, Diuk-Wasser MA, 2012. Geographic variation in the relationship between human lyme disease incidence and density of infected host-seeking *Ixodes scapularis* nymphs in the eastern United States. *Am. J. Trop. Med. Hyg* 86, 1062–1071. 10.4269/ajtmh.2012.11-0630. [PubMed: 22665620]
- Prinzato L, Walter E, 1985. Robust experiment design via stochastic approximation. *Math. Biosci* 75, 103–120. 10.1016/0025-5564(85)90068-9.
- Reich BJ, Pacifici K, Stallings JW, 2018. Integrating auxiliary data in optimal spatial design for species distribution modelling. *Methods Ecol. Evol* 9, 1626–1637. 10.1111/2041-210X.13002.
- Rosenberg R, Lindsey NP, Fischer M, Gregory CJ, Hinckley AF, Mead PS, Paz-Bailey G, Waterman SH, Drexler NA, Kersh GJ, Hooks H, Partridge SK, Visser SN, Beard CB, Petersen LR, 2018. Vital signs: trends in reported vectorborne disease cases — United States and territories, 2004–2016. *Morb. Mortal. Wkly. Rep* 67, 496–501. 10.15585/mmwr.mm6717e1.
- Ryan EG, Drovandi CC, McGree JM, Pettitt AN, 2016. A review of modern computational algorithms for bayesian optimal design. *Int. Stat. Rev* 84, 128–154 10.1111/insr.12107.

- Ryan EG, Drovandi CC, Pettitt AN, 2015. Simulation-based fully Bayesian experimental design for mixed effects models. *Comput. Statist. Data Anal* 92, 26–39. 10.1016/j.csda.2015.06.007.
- Rynkiewicz EC, Clay K, 2014. Tick community composition in Midwestern US habitats in relation to sampling method and environmental conditions. *Exp. Appl. Acarol* 64, 109–119. 10.1007/s10493-014-9798-7. [PubMed: 24705853]
- Schulze TL, Jordan RA, Hung RW, 1997. Biases associated with several sampling methods used to estimate abundance of ixodes scapularis and Amblyomma americanum (Acari: ixodidae). *J. Med. Entomol* 34, 615–623. 10.1093/jmedent/34.6.615. [PubMed: 9439115]
- Sonenshine DE, 2018. Range Expansion of Tick Disease Vectors in North America: implications for spread of tick-borne disease. *Int. J. Environ. Res. Public. Health* 15, 478. 10.3390/ijerph15030478. [PubMed: 29522469]
- Springer YP, Eisen L, Beati L, James AM, Eisen RJ, 2014. Spatial distribution of counties in the continental united states with records of occurrence of Amblyomma americanum (Ixodida: ixodidae). *J. Med. Entomol* 51, 342–351. 10.1603/ME13115. [PubMed: 24724282]
- Springer YP, Jarnevich CS, Barnett DT, Monaghan AJ, Eisen RJ, 2015. Modeling the Present and Future Geographic Distribution of the Lone Star Tick, Amblyomma americanum (Ixodida: ixodidae), in the Continental United States. *Am. J. Trop. Med. Hyg* 93, 875–890. 10.4269/ajtmh.15-0330. [PubMed: 26217042]
- Wisely SM, Glass GE, 2019. Advancing the science of tick and tick-borne disease surveillance in the united states. *insects* 10, 361. 10.3390/insects10100361. [PubMed: 31635108]
- Zhang Y., 2006. Bayesian D-Optimal Design for Generalized Linear Models. Virginia Polytechnic Institute and State University.

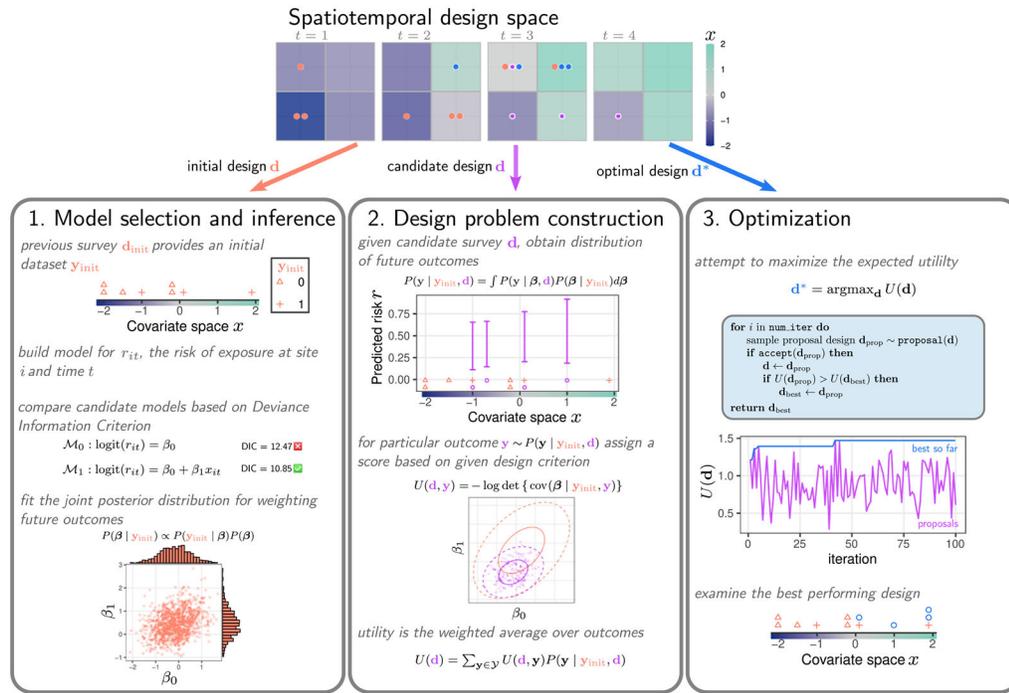


Fig. 1. Implementing Bayesian Experimental Design in spatiotemporal surveillance. A motivating example with a single environmental covariate is shown, with the goal of establishing environmental factors associated with tick presence. Top: a small design space consisting of four possible survey locations (e.g. parks) and four timepoints (e.g. months). A surveillance schedule is a collection of visits (time-location pairs) and amounts to arranging points in design space. The changing values of the environmental covariate x are shown for each survey point, and the values of x corresponding to a design d are mapped to the 1-dimensional covariate space, indicated by arrows for each d . Bottom: in step 1, a response y_{init} and associated x values from an initial survey d_{init} are shown in covariate space, two candidate models are compared, and a posterior distribution for the selected model is fit. In step 2, the utility of some candidate design is defined as an average over future outcomes and the amount of new information that would be provided by each outcome. Here a Bayesian d-optimality criterion is used, which scores outcomes based on the volume of confidence ellipsoids approximating the updated posterior distribution. In step 3, finding an effective design is treated as an optimization problem over the space of candidate designs. A generic stepwise procedure is shown, and the best design found after 100 iterations is then examined. In accordance with BED theory, this design spread additional points throughout the middle of covariate space, while putting special attention at the extreme $x \approx 2$ which was under-sampled in the initial survey.

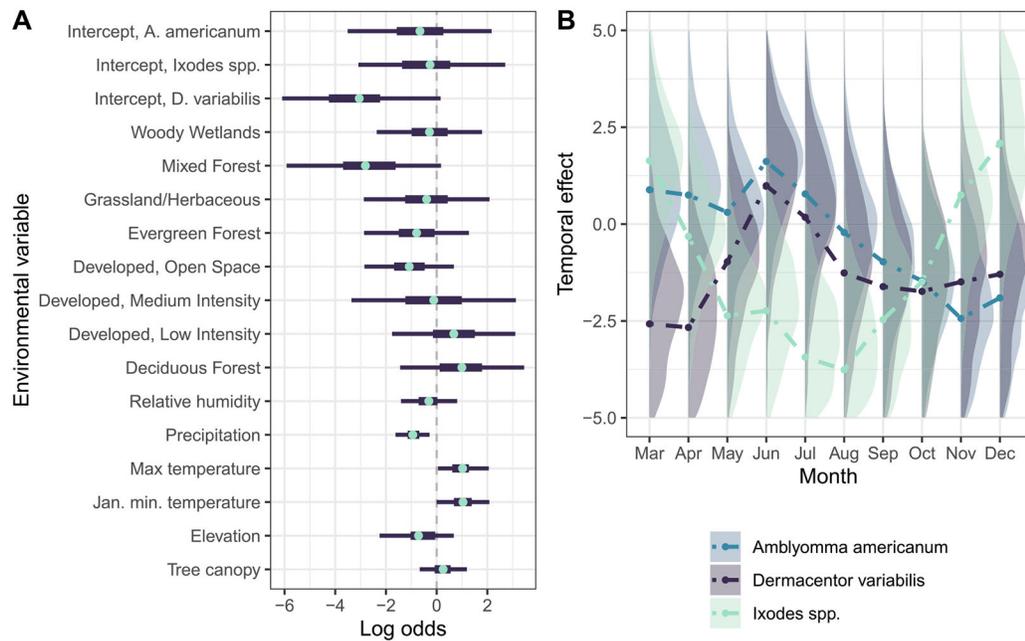


Fig. 2. Posterior environmental and temporal effects from the initial survey. Results are shown for the best-performing model fit to the initial survey data from 2020 to 2021. All results are in log-odds scale. (A) Marginal posterior means for each species intercept and coefficient for the environmental variables are shown as points, while 50% and 95% Highest Posterior Density Intervals are shown as purple bars. (B) Mean temporal trend for each tick group is given by dashed lines, along with full marginal posterior densities for each month/tick group.

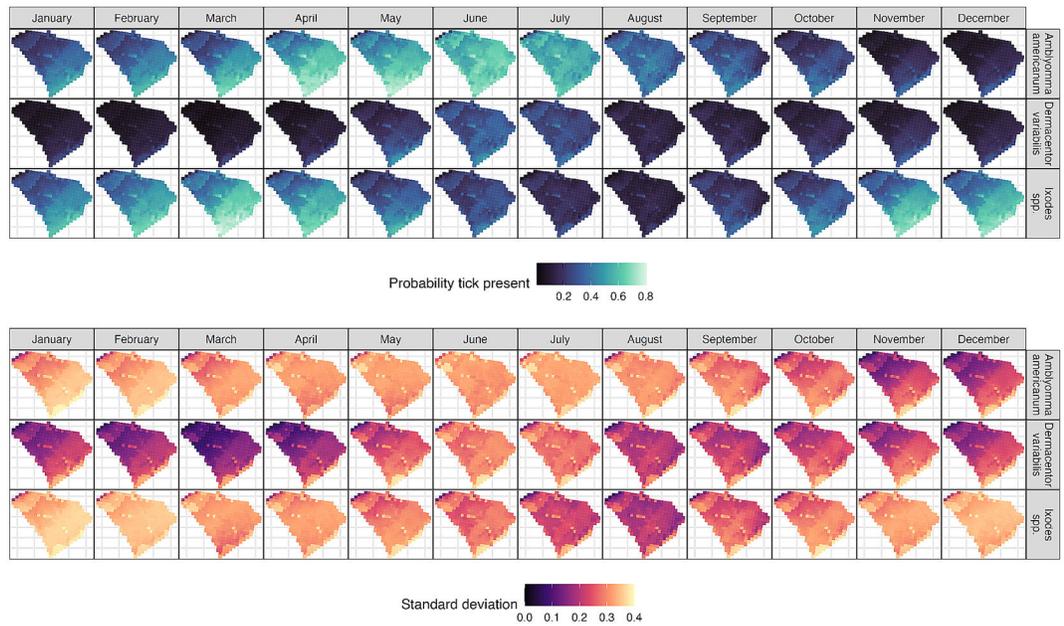


Fig. 3. Spatiotemporal mean and standard deviation of risk. Results are shown for the best-performing model fit to the initial survey data. Posterior marginals for the probability of tick presence were computed along a 16 km grid of locations across South Carolina, and summarized by the mean (top) and standard deviation (bottom).

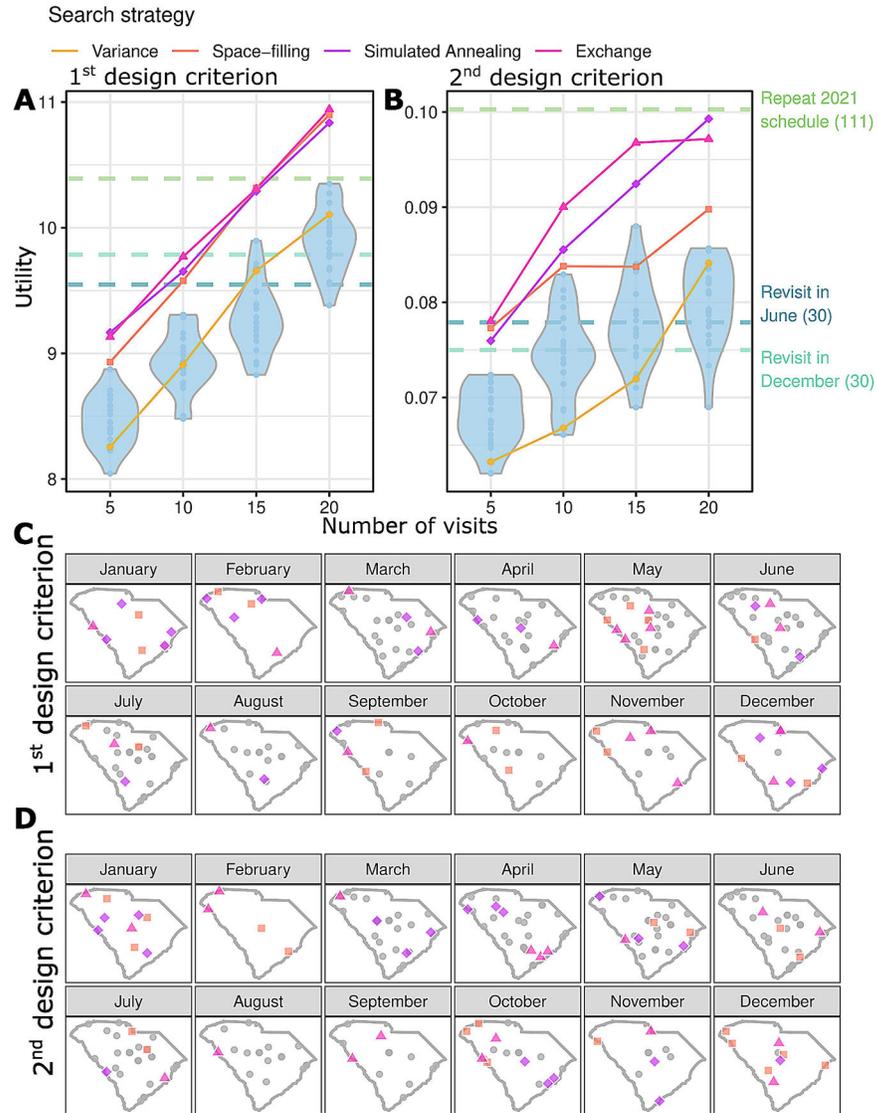


Fig. 4. Comparing search methods for effective designs of tick surveillance. The utility of designs found using different search strategies are compared to 20 replications of random sampling (light blue distributions) for different sample sizes, and to different convenience/repeated sampling schemes (dashed lines; number of visits in parentheses). (A) Results from optimizing the d-optimality criterion for the environmental covariates. (B) Results using the maximum variance reduction criterion among high risk areas. (C-D) Proposed schedules of 20 visits for the 3 effective search strategies are shown across space and time. The schedule of 170 visits used for the initial survey data is shown as gray points.