Longer study length, standardized sampling techniques, and broader geographic scope leads to higher likelihood of detecting stable abundance patterns in long term deer tick studies

Rowan Christie[1], Kaitlin Stack Whitney[1,2*], Julia Perrone[3], Christine A. Bahlai[4]

[1] Thomas H Gosnell School of Life Sciences, Rochester Institute of Technology, Rochester, NY, USA
[2] Science, Technology & Society Department, Rochester Institute of Technology, Rochester, NY, USA
[3] School of Information, Kent State University, Kent, OH, USA
[4] Department of Biological Sciences and Environmental Science and Design Research Initiative, Kent State University, Kent, OH, USA

Corresponding author:
Kaitlin Stack Whitney[1,2]
7 Lomb Memorial Drive, Rochester, NY, 14623, USA
Email address: kxwsbi@rit.edu

Abstract:

Ixodes scapularis (deer ticks) are a taxon of ecological and human health concern due to their status as primary vectors of *Borrelia burgdorferi*, the bacteria that transmits Lyme disease. Deer ticks are thought to be expanding in geographic range and population size across the eastern United States (US), leading to concern that tick-vectored illness will correspondingly rise. However, because of wide variability in deer tick monitoring strategies, synthesis efforts may be limited by the sensitivity and reliability of data produced by existing long term studies, especially to inform forecasting and proactive deer tick management. To address this, we explicitly examined the role of how study design parameters affect the likelihood of observing temporal trends in deer tick studies. We used a moving window approach to investigate the temporal stability of deer tick population trajectories across the US. We found several study factors can have an impact on the likelihood of a study reaching stability and the likelihood of tick abundance data leading to misleading results if the study does not reach stability. Our results underscore the need for longer studies of deer ticks when trying to assess long term or broad spatial patterns. Moreover, our results showcase the importance of study length, sampling technique, life stage, and geographic scope in shaping the inferences from deer tick studies. This is especially important for synthesizing across the variety of existing surveys and for potential ecological forecasting.

Introduction

Ixodes scapularis (deer ticks) are primary vectors of *Borrelia burgdorferi*, bacteria that transmits Lyme disease (Des Vignes, 1997). Perceived risk about deer ticks and Lyme disease is increasing across the United States with expanding distributions of this tick species now ranging from the southeast and northwest states in the US (Keirans, 1996). Established deer tick populations are more concentrated in upper north-central, northeastern, and west-coast states (Dennis, 1998). In addition, deer ticks (also known as blacklegged ticks) are spreading rapidly, found in more than twice the number of counties in the United States compared to twenty years ago (Eisen, 2019). Corresponding to the increasing range of this species is the concern that increased deer tick populations will lead to corresponding increases in tick vectored illness in humans and domestic animals (Leighton et al. 2012). Thus, it is critical to design effective monitoring strategies that can be synthesized into regional-level understanding of deer tick dynamics and population trajectories to better understand risks to public health.

Like for many environmental problems, understanding of how study design and monitoring strategies play out over time is critical to underpin environmental management of ticks and public policy about Lyme disease (White and Bahlai, 2021). Yet many biological and field studies on many taxa are short term and limited in scope, leading to concerns about how to interpret the many short-term studies compared to the relatively fewer long term ecological studies (Wauchope et al., 2019; Bahlai et al., 2020). For example, natural variability in a stable population might lead a 2-3 year study to appear to capture a strong upward trend in abundance for a given taxon, simply as a result of natural population cycling. Because arthropod populations are subject to strong boom bust dynamics (Wagner, 2020), as well as strong trapping and sampling biases (Fournier et al. 2019; Didham et al., 2020), short term dynamics may lead to

spurious, misleading conclusions about system dynamics (Cusser et al., 2020). Because of these underlying processes and biases, a 10 year study done in the same area could find recurring 'spikes' that occur every few years, which, if data were taken on a shorter time scale be interpreted as a steep increase (or decrease):, essentially capturing a 'snapshot' that was not representative of overall trends (Didham et al., 2020). This phenomenon is not unique to arthropods: in an analysis of no-till agricultural management, 13 years of data collection was needed in order to fully capture the benefits of no-till management - this suggests that a shorter term study would have potentially led to misleading results (Cusser et al., 2020). Indeed, temporal trends taken in variable environments may be particularly prone to spurious trends: populations observed in more variable environments are more likely to exhibit these spurious behaviors in time (Cusser et al., 2021), making understanding of this behavior critical in a changing climate. Finally, because of potential biases in study site selection, it's critical to examine whether subtle choices in experimental design led to a target taxon being sampled in a systematically biased way, potentially leading to misleading conclusions about increases or declines (Fournier et al., 2019).

Understanding population trajectories is also critical for scaling ecological data to ecological forecasting. Ecological forecasting is considered the next critical frontier in broad scale environmental science (Dietze et al., 2018). With systems such as ticks, forecasts could be used to proactively protect ecological and human health. Ecological forecasting initiatives are increasing, with an NSF-funded Research Coordination Network (NSF, 2019), conferences (EFI, 2019; EFI RCN and NEON, 2020), and ongoing ecological forecast challenge (EFI RCN NEON, 2020). In the case of deer ticks, forecasting could mean predicting which years and areas are likely to have high populations, allowing for mitigation like chemical control or behavioral change (e.g. avoidance). However, to realize meaningful forecasting, it is necessary to critically examine the barriers to effective synthesis of existing surveys, especially for species with intimate human implications.

Understanding deer tick populations patterns across ongoing surveys, let alone potential use for forecasting, is difficult due to the variability in how previous and ongoing research surveys and count them. There are several standard sampling techniques that can be used to collect data on deer ticks: dragging, flagging, and carbon dioxide baited traps, (Falco, 1992). Dragging is when a cloth is pulled behind the person sampling with a rope attached, and the pole is horizontal and perpendicular to the movement. Flagging is a similar study technique, when a cloth is attached to a pole and swept through vegetation. Baited traps lure ticks using carbon dioxide gas, which acts as an attractant. These traps can be set and then recollected. Another common tick survey method is passive surveillance, where individuals send in ticks they have found on themselves to public health departments, but do not deliberately search for ticks.

The survey techniques that are selected and used may alter which ticks are captured and counted. Existing research on deer tick sampling techniques has shown that dragging may provide the most consistent results for larval and nymphal *Ixodes dammini* (another species of deer tick) populations compared to CO2 baited traps (Falco, 1992). Comparisons of dragging and flagging have shown that these techniques show no consistent differences in the number of *Ixodes scapularis* collected (Rulison, 2013). Collection amount by life stage (larvae, nymphs, and adults) is also known to vary widely with standardized sampling techniques (Newman, 2019). In

addition, different life stages may pose different levels of risk of exposure to Lyme disease to humans. Nymph abundance, not adults, has been shown to be the primary factor in determining risk of exposure (Falco, 1999). Due to this potential variation in transmission by life stage, classification data on deer tick populations by the US Centers for Disease Control takes into account whether data is available from multiple life stages within each county reporting (CDC, undated).

To understand how variable methodology in deer tick studies may influence which population patterns researchers find, we conducted a data synthesis experiment to examine how study parameters influence deer tick population trends. Specifically, we looked at study length, sampling technique, life stage, and geographic scope. To do this, we used a nonrandom sampling algorithm (Bahlai et al., 2020) to analyze publicly available deer tick datasets to understand the impacts of study parameters on inferences about deer tick abundance and trajectory. We predicted that studies using dragging would reach stability earlier than other sampling techniques, due to existing research showing that dragging is more consistent for some life stages of deer tick collection (Falco, 1992). We also predicted that life stage could impact deer tick trends; for example, nymphs are most likely to spread *Borrelia burgdorferi* (Rulison, 2013), and could have significantly different abundance patterns than adults or larvae. This will help researchers better understand how to design deer tick studies to avoid misleading inferences, leading to improved monitoring of a public health risk.

Materials & Methods

Datasets

We searched for publicly available deer tick abundance dataset that had at least 9 years of data, using keywords about deer ticks in internet search engines, literature databases, data repositories and public health websites. Specifically, we searched combinations of *tick* and *scapularis* in Data Dryad, Data One, Google Datasets, the Long Term Ecological Research data portal, and the National Ecological Observation Network data portal. We then used Google search portals to look up individual state department of health websites and state datasets where deer ticks are common for additional sets of survey data. Our complete search was done using the PRISMA workflow for systematic literature screening, review and reporting (Moher et al., 2009, Appendix 1). Following the PRISMA workflow, we found several datasets that could be used for our analysis (Table 1). We extracted and scored information about each dataset in terms of the life stage sampled, sampling techniques used, geographic scale of the study, and location.

Moving window analysis

To analyze the stability of the abundance trends of these datasets, we used the 'broken window' algorithm in R (Bahlai, 2020). This non-random resampling approach uses subsamples (i.e. '*windows*') of the time series data to gain insights into patterns of how data behaves in arbitrarily selected time periods (Bahlai et al., 2020). This algorithm defines the linear slope associated with the longest time series available as the 'true' slope (as it is calculated using the largest possible sample size) and performs calculations relative to this proxy for truth. The algorithm then returns the number of years it takes for a dataset to reach stability (defined as the number of years required for >95% of subsamples to be within one standard deviation of the 'true' slope), and the proportion of windows 'significantly wrong' (the set of windows that do not match the direction and statistical significance of the 'true' slope, which, in isolation, would likely lead a scientist to

a spurious conclusion that does not agree with the overall trend). The algorithm uses a yearly interval for its calculations, and in the case of this study, the metric used as the dependent variable was some unit of abundance or activity of ticks, as defined by the trapping method. Prior to subsetting within the algorithm, each raw dataset is subjected to normalization by converting it to a unitless Z score, allowing methods producing data on dramatically different response scales to be directly compared.The algorithm then runs a '*moving window*' function that iterates through all subsets in the data greater than 2 years, and subjects each 'window' to a linear model, which calculates the slope statistic (change of standardized density over change in year), standard error of this slope, p-value, and R^2. To calculate the number of years to reach stability, where stability is defined as being greater than the percentage of windows with slopes occuring within the standard deviation (95% CI) of the slope of the longest series, we ran the '*stability time*' function which takes in the data with the summary statistics, computes the proportion correct for each window within the standard deviation of the longest slope, and returns the number of windows whose slope are within said standard deviation as the number of years to reach stability. We also ran another function for 'proportion significantly wrong before stability' to calculate the proportion of windows that did not reach stability with a p-value that was not within 0.05 of the p-value of the true slope (slope of the longest series).

Statistical analyses

We used t-tests in base R using RStudio and R statistical software (RStudio Team, 2020; R Core Team, 2020). We compared the overall proportion wrong compared to proportion wrong before stability resulting from the moving window analysis (Bahlai, 2020) by deer tick life stage, geographic scope of study, sampling technique, and study response variable. All the R code, test results, and resulting figures are provided in the Supplemental Materials.

Results

Deer tick datasets

Our analysis included 289 datasets from 6 studies from locations in New Jersey, New York, Iowa, Connecticut, Massachusetts, ranging in length from 9 years to 24 years (Figure 1, Appendix 1). Differences between the datasets included two different sampling methods (dragging and opportunistic surveying), three different life stages (larvae, nymphs, and adults), four different geographic scopes (grid, state forest, town, and county) and two different sampling metrics (abundance and percent deer ticks infected with *B. burgdorferi*).

These datasets included a range of deer tick life stages (larvae, nymph, and adult), sampling techniques (dragging and surveys from ticks found on people), and geographic scale of study units (ranging from small experimental plots to counties). The datasets we found that met our study criteria were located in 5 states (New York, New Jersey, Massachusetts, Connecticut, and Iowa), most of which are the northeastern US (Figure 1). Using the broken window algorithm, we found that across all datasets, years to stability time ranged from 4 to 23 years (Figure 1), and the proportion significantly wrong before stability ranged from 0 to 100% (Figure 3). When comparing the overall proportion significantly wrong with proportion significantly wrong before stability calculated from abundance data,but not statistical differences were noted between the two datasets ($t = -1.21$, $df = 571.46$, p-value = 0.23; Figure 3).

Study length
None of the datasets in our study measuring tick abundance or density reached stability sooner than 4 years (Figure 1). The proportion of datasets that have reached stability by 4 years is very low, but by 10 years, the majority of datasets have reached stability (Figure 1). We found that study length is positively correlated with years to stability (Pearson product moment correlation test, $r^2 = 0.9$) and this relationship is statistically significant (p-value <0.001). Based on this correlation and the representation of this relationship shown in Figure 2, our results imply that as the number of years in a study increases, the number of years to reach stability increases.

We found that the association between proportion significantly wrong before stability time and years to stability for individual datasets was weak (Pearson product moment correlation test, r2 = 0.23, p-value <0.001, Figure 2). We also found that the association between proportion significantly wrong before stability time and overall proportion wrong were not strongly correlated (t = -1.2, df = 571.46, p-value = 0.23, Figure 3).

Sampling technique
We predicted that the dragging sampling technique would produce more consistent abundance trends than datasets associated with opportunistic surveying. We found that abundance data produced by dragging (n=90 datasets) reached stability faster than data produced by opportunistic surveying (n=198 datasets) based on t-test (t = -8.53, df = 236.23, p-value <0.001, Figure 4B). We also found that datasets with dragging sampling techniques varied less by stability time, as it had a smaller range of years to stability, with dragging data showing smaller values for stability time.

We found that datasets with dragging techniques were more likely to reach stability time in fewer years (Figure 4B). The median number of years to stability time was 7 for dragging techniques and 12 for opportunistic techniques. However, we did not find a significant difference when we compared the proportion significantly wrong before stability by sampling technique (t = 0.08, df = 155.58, p-value = 0.93, Figure 4A).

Life stage
We anticipated that datasets studying different life stages of ticks would vary in how many years it took to reach stability time. Our analysis included datasets that studied adults (oldest life stage, n=63 datasets), nymphs (middle life stage, n=66 datasets), and larvae (youngest life stage, n=8 datasets). There was evidence of variation by life stage. Datasets measuring deer tick larva reached stability significantly faster compared to adult or nymph abundance data (adults and larvae, t = -5.96, df = 10.11, p-value <0.001; nymphs and larvae, t = -5.52, df = 10.54, p-value <0.001, Figure 5B), whereas the difference between adult abundance data and nymph abundance data is insignificant (t = -0.64, df = 126.84, p-value = 0.52). The median years to stability for datasets examining larvae is 11.5, while the median for datasets examining adults and nymphs is 7.

Datasets that examine adults have a significantly higher proportion significantly wrong before stability than datasets that examine nymphs (t = 2.90, df = 113.64, p-value = 0.004, Figure 5A). The mean proportion wrong before stability for adults is 0.35, compared to nymphs (0.16) and

larvae (0.28). There was no difference between larvae and nymphs (t = -0.75, df = 7.89, p-value = 0.47) and larvae and adults (t = 0.44, df = 8.77, p-value = 0.67) This suggests that datasets that examine adults will be more likely to lead to misleading results if the study has not reached stability.

Geographic scope

The studies we found included datasets summarized by grid (n=24 datasets), state forest (n=6 datasets), town (n=186 datasets), and county (n=73 datasets). We found that the geographic scope of the study impacts the number of years it takes for a tick abundance study to reach stability (Figure 6). County level data reached stability faster than town, state forest, or grid level data (county and grid, t = -17.21, df = 33.87, p-value <0.001; county and state forest, t = -5.55, df = 5.22, p-value = 0.002; county and town, t = -17.03, df = 243.09, p-value <0.001; Figure 6B). The median number years to stability for datasets of county scope was 7 years, compared to datasets collected at the scale of town (12 years), state forest (12 years), and grid (11.5 years). There was no difference between abundance data produced by town level studies, state forest level studies, and grid level studies(town and state forest, t = 0.47, df = 6.12, p-value = 0.65; town and grid, t = 0.22, df = 86.18, p-value = 0.82; state forest and grid, t = -0.38, df = 6.03, p-value = 0.72; Figure 6B).

Data produced by state forest level studies has a significantly smaller proportion significantly wrong before stability versus data produced by county and town level studies (state forest and county, t = 5.24, df = 74.29, p-value <0.001; state forest and town, t = 7.42, df = 46.44, p-value = p-value <0.001; Figure 6A). The mean proportion significantly wrong for data at the scale of state forest is 0.077, compared to datasets collected at the scale of county (0.35), town (0.31), and grid (0.23). Data produced by county level and town level studies (t = 0.72, df = 115.03, p-value = 0.48), county level and grid level studies (t = 1.33, df = 43.07, p-value = 0.19), town level and grid level studies (t = 1.00, df = 28.52, p-value = 0.32), and state forest level and grid level studies (t = -1.91, df = 25.10, p-value = 0.068) did not differ.

Sampling metric

The datasets that focus on different sampling metrics, abundance of deer ticks (n=175 datasets) and percent tested for *B. burgdorferi* (n=114 datasets), reached stability in a similar number of years and had similar levels of proportion significantly wrong before stability (Figure 7). The abundance data reached stability slightly faster, but the difference was not statistically significant (t = -1.29, df = 283.9, p-value = 0.20; Figure 7B). The abundance data also had a slightly higher proportion significantly wrong before stability, but this difference is insignificant (t = -1.18, df = 232.98, p-value = 0.24; Figure 7A).

Discussion

In our sampling of 289 unique location datasets from 6 projects, we found several study factors can have an impact on the likelihood of a study reaching stability and the likelihood of tick abundance data leading to misleading results if the study does not reach stability. Our findings are consistent with recent research that indicates resampling existing data can give insights into analyzing existing data and designing future monitoring programs (White and Bahlai, 2021).

Study length is important to consider in any deer tick study if the data is to reach stability. Our findings indicate that short term studies that survey for less than at least 5 years are very unlikely to reach stability and thus are insufficient to explain long term dynamics. In addition, our findings indicate there is a greater proportion of studies that reach stability in a longer period of time, implying that studies that are longer in length are more likely to reach stability. Several other studies provide support for using long term studies to reach stability (White, 2019; Cusser et al., 2020). White found that at least 72% of the vertebrate populations they studied required at least 10 years of study to detect a significant trend in abundance, while Cusser determined that it would take at least 16-19 years to detect a significant effect from no-till management on crop yield and soil moisture, emphasizing the importance of long term studies. However, we did not find a strong correlation between study length and the proportion of windows significantly wrong before stability was reached.

Sampling technique used was shown to have an impact on the number of years it took to reach stability - more specifically, using the standardized method dragging for measuring abundance allows the study to reach stability quicker, and thus those studies are more likely to reach stability with a lower number of years of study. This may imply that standardized sampling techniques are more reliable in terms of reaching stability. The data produced between the sampling methods has not been shown to be significantly different (Rulison, 2013). Another study compared the effectiveness of dragging, mouse trapping, and CO_2 baited traps to measure the abundance of immature deer ticks, and found that dragging was the most effective sampling method, providing further support for dragging (Falco, 1992).

We found differences in the amount of time it took for populations to reach stability based on deer tick life stage. Specifically, time to stability was faster for larvae, the youngest life stage, compared to adults and nymphs. This finding is critical for researchers trying to glean insights into broad scale deer tick trends across disparate study methods. We also found that the geographic scope of the study impacts the number of years it takes for a tick abundance study to reach stability. Specifically, this means that the largest geographic unit in our study reached stability faster (compared to town, forest, and grids). This indicates that sampling a larger geographic area leads to more consistent patterns in deer tick studies than studies done on smaller areas. This may indicate that variables influencing population trends for deer ticks do not uniformly scale across space and time. Nonlinearity in ecological phenomena means data may require transformation before findings are applied to different scales (Wessman, 1992).

Given that ecological monitoring requires tradeoffs in effort, cost, and insight (Bennett et al., 2018), this finding may indicate that studies at larger spatial scales may be able to give indication of population trajectories faster than studies at smaller scales. While in theory more data is always better, researchers and public policymakers will never have 'complete information' about deer ticks to make decisions about Lyme disease and health. Our results can help inform when applied ecology researchers have "enough" data to make informed, rigorous claims about deer tick population trends to make good decisions (Canessa et al., 2015).

Conclusions

We used 6 long term deer tick studies with 289 underlying datasets to reveal that variation in study design and implementation is associated with variation in inferences in deer tick

population trajectories and how many years it takes to find stable patterns. We make clear that ecological forecasting initiatives based on existing deer tick datasets must take into account this variability and resulting differences in stability patterns. Specifically, our results show the importance of study length, sampling technique, life stage, and geographic scope for monitoring deer tick populations. Our findings are consistent with previous research supporting the importance of long term ecological monitoring for understanding populations in field biology.

Acknowledgments
RMC was supported by the National Institute of General Medical Sciences of the National Institutes of the Health under Award Number R25GM122672. CAB, JP, and KSW were supported by the Office of Advanced Cyberinfrastructure in the National Science Foundation under Award Number #1838807. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health or the National Science Foundation.

References

Bahlai, CA. 2019. cbahlai/bad_breakup: The bad breakup Algorithm (Version v1.0). Zenodo. doi.org/10.5281/zenodo.2561051

Bahlai CA, White ER, Perrone JD, Cusser S, Stack Whitney K. 2020. bioRxiv 2020.07.07.192211; doi: https://doi.org/10.1101/2020.07.07.192211

Bennett JR, Maxwell SL, Martin AE, Chadés I, Fahrig L, Gilbert B. 2018. When to monitor and when to act: Value of information theory for multiple management units and limited budgets. J. Appl. Ecol. 55, 2102–2113. doi: 10.1111/1365-2664.13132

Centers for Disease Control (CDC). Undated. Surveillance for Ixodes scapularis and pathogens found in this tick species in the United States. Available at <https://www.cdc.gov/ticks/resources/TickSurveillance_Iscapularis-P.pdf> (accessed 15 November 2021)

Centers for Disease Control (CDC). 2018. Signs and Symptoms of Untreated Lyme Disease | Lyme Disease | CDC. Available at https://www.cdc.gov/lyme/signs_symptoms/index.html (accessed 15 November 2021)

Canessa S, Guillera-Arroita G, Lahoz-Monfort JJ, Southwell DM, Armstrong DP, Chadés I, Lacy RC, Converse SJ. 2015. When do we need more data? A primer on calculating the value of information for applied ecologists. Methods Ecol. Evol. 6, 1219–1228. doi: 10.1111/2041-210X.12423

Cusser S, Helms IV J, Bahlai C, Hadded NM. 2021. How long do population level field experiments need to be? A meta-analysis across the 40-year old LTER network. Ecology Letters https://doi.org/10.1111/ele.13710

Cusser S, Bahlai C, Swinton SM, Robertson GP, Haddad NM. 2020. Long□term research avoids spurious and misleading trends in sustainability attributes of no□till. Glob Change Biol. 26: 3715– 3725. https://doi.org/10.1111/gcb.15080

Dennis DT, Nekomoto TS, Victor JC, Paul WS, Piesman J. 1998. Reported distribution of Ixodes scapularis and Ixodes pacificus (Acari: Ixodidae) in the United States. Journal of Medical Entomology 35(5):629-38. doi: 10.1093/jmedent/35.5.629. PMID: 9775584.

Des Vignes F, Fish D. 1997. Transmission of the agent of human granulocytic ehrlichiosis by host-seeking Ixodus scapularis (Acari:Ixodidae) in southern New York state. Journal of Medical Entomology, 34(4), 379–382. https://doi.org/10.1093/jmedent/34.4.379

Didham RK, Basset Y, Collins CM, Leather SM, Littlewood NA, Menz MHM, Müller J, Packer L, Saunders ME, Schönrogge K, Stewart AJA, Yanoviak SP, Hassall C. 2020. Interpreting insect declines: seven challenges and a way forward. Insect Conservation and Diversity Volume 13 Issue 2:103-114. https://doi.org/10.1111/icad.12408

Dietze, MC, Fox A, Beck-Johnson LM, Betancourt JL, Hooten MB, Jarnevich CS, Keitt TH, Kenney MA, Laney CM, Larsen LG, Loescher HW, Lunch CK, Pijanowski BC, Randerson JT, Read EK, Tredennick AT, Vargas R, Weathers KC, White EP. Iterative near-term ecological forecasting. Proceedings of the National Academy of Sciences Feb 2018, 115 (7) 1424-1432; DOI: 10.1073/pnas.1710231115

Ecological Forecasting Initiative (EFI). 2019. EFI 2019 Conference May 13-15, 2019. Available at <https://ecoforecast.org/efi2019/> (accessed 15 November 2021)

Ecological Forecasting Initiative (EFI) and National Ecological Observation Network (NEON). 2020. NSF-sponsored EFI-Research Coordination Network and NEON. Dates: May 12-13, 2020. Available at <https://ecoforecast.org/efi-rcn-2020-conference/> (accessed 15 November 2021)

Ecological Forecasting Initiative (EFI) Research Coordination Network (RCN) and National Ecological Observation Network (NEON). 2020. EFI RCN NEON Ecological Forecast Challenge. Available at <https://ecoforecast.org/efi-rcn-forecast-challenges/> (accessed 15 November 2021)

Eisen RJ, Eisen L, Graham CB, Hojgaard A, Mead P, Kersh G, Neitzel D. 2019. Surveillance for Ixodes scapularis and pathogens found in this tick species in the United States. *CDC Stacks*. Available at <https://stacks.cdc.gov/view/cdc/78432> (accessed 15 November 2021)

Falco RC, Fish D. 1992. A comparison of methods for sampling the deer tick, Ixodes dammini, in a Lyme disease endemic area. Experimental & applied acarology, 14(2), 165–173. https://doi.org/10.1007/BF01219108

Falco RC, McKenna DF, Daniels TJ, Nadelman RB, Nowakowski J, Fish D, Wormser GP. 1999. Temporal relation between Ixodes scapularis abundance and risk for Lyme disease associated with erythema migrans. Am J Epidemiol. 15;149(8):771-6. doi: 10.1093/oxfordjournals.aje.a009886. PMID: 10206627.

Fournier AM, White ER, Heard SB. 2019. Site-selection bias and apparent population declines in long-term studies. Conserv. Biol. 33, 1–10. doi: 10.1111/cobi.13371

Keirans JE, Hutcheson HJ, Durden LA, Klompen JS. 1996. Ixodes (Ixodes) scapularis (Acari:Ixodidae): redescription of all active stages, distribution, hosts, geographical variation, and medical and veterinary importance. Journal of Medical Entomology 33(3):297-318. doi: 10.1093/jmedent/33.3.297. PMID: 8667375.

Leighton PA, Koffi JK, Pelcat Y, Lindsay LR, Ogden NH. 2012. Predicting the speed of tick invasion: an empirical model of range expansion for the Lyme disease vector Ixodes scapularis in Canada. Journal of Applied Ecology 49(2): 457-464. https://doi.org/10.1111/j.1365-2664.2012.02112.x

Moher D, Liberati A, Tetzlaff J, Altman DG, The PRISMA Group. 2009. Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement. PLoS Med 6(7): e1000097. doi:10.1371/journal.pmed1000097

National Science Foundation (NSF). 2019. NEON RCN: The Ecological Forecasting Initiative RCN: Using NEON-enabled near-term forecasting to synthesize our understanding of predictability across ecological systems and scales. NSF Division of Environmental Biology Award Abstract #1926388. Available at <https://www.nsf.gov/awardsearch/showAward?AWD_ID=1926388> (accessed 15 November 2021)

Newman BC, Sutton WB, Wang Y, Schweitzer CJ, Moncayo AC, Miller BT. 2019. A standardized method for the construction of a tick drag/flag sampling approach and evaluation of sampling efficacy. Exp Appl Acarol 79, 433–446 2019. https://doi.org/10.1007/s10493-019-00429-6

R Core Team. 2020. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org/. (accessed 15 November 2021)

RStudio Team. 2020. RStudio: Integrated Development for R. RStudio, PBC, Boston, MA URL http://www.rstudio.com/. (accessed 15 November 2021)

Rulison EL, Kuczaj I, Pang G, Hickling GJ, Tsao JI, Ginsberg HS. 2013. Flagging versus dragging as sampling methods for nymphalIxodes scapularis(Acari: Ixodidae). Journal of Vector Ecology, 38(1), 163-167. doi:10.1111/j.1948-7134.2013.12022.x

Wagner DL. 2020. Insect Declines in the Anthropocene. Annual Review of Entomology 2020 65:1, 457-480.

Wauchope HS, Amano T, Sutherland WJ, Johnston A. 2019. When can we trust population trends? A method for quantifying the effects of sampling interval and duration. Methods in Ecology and Evolution 10: 2067–2078. doi:10.1111/2041-210X.13302

Wessman CA. 1992. “Spatial Scales and Global Change: Bridging the Gap from Plots to GCM Grid Cells.” Annual Review of Ecology and Systematics, vol. 23, 1992, pp. 175–200. JSTOR, www.jstor.org/stable/2097286.

White ER. 2019. Minimum Time Required to Detect Population Trends: The Need for Long-Term Monitoring Programs, BioScience, Volume 69, Issue 1, January 2019, Pages 40–46, https://doi.org/10.1093/biosci/biy144

White ER, Bahlai CA. 2021. Experimenting With the Past to Improve Environmental Monitoring. Frontiers in Ecology and Evolution 8 (391). DOI=10.3389/fevo.2020.572979

Supplemental Materials

The data and code for this study are located here - https://github.com/SMCCoder/deer_tick_bad_breakup_algorithms

Appendix 1. PRISMA workflow for finding and screening deer tick datasets

Screening criteria: Datasets that are at least 9 years and include data (either density or total number ticks found - any sampling method) of ticks species that are disease vectors

6/26/2019 I used the LTER data portal using the keyword “scapularis” and “tick”
N = 12 results
Screening process
After screening using criteria, n = 1

6/26/2019 I used the Data One portal using the key words “tick” and “scapularis”, and selected the checkbox under the tab “Data files” to only select results with data
N = 19 results
After reviewing duplicates n = 18
After screening using criteria, n = 2

6/26/2019 I used the Data Dryad repository using the key words “tick” and “scapularis”
N = 9 results
After reviewing duplicates n = 7
After screening using criteria, n = 0

6/26/2019 I used the NEON data portal using the keywords “tick”
N = 2 results
After reviewing duplicates n = 2
After screening using criteria, n = 0

6/26/2019 I used the Google Datasets repository using the keywords “scapularis”
N = 75
After reviewing duplicates n = 70
After screening using criteria, n = 4

6/28/2019 I searched through the NJ Department of Health and tabbed to the Communicable Disease Service page, accessed the Statistics, Reports, & Publication dropdown, and clicked on Vector borne surveillance reports
After screening n = 0

6/28/2019 I searched the CT Department of Health using the keyword “tick”

N = 2,920
After screening using criteria, n = 1

6/28/2019 I searched the PA Department of Health using the keyword “tick”
N = 54
After screening using criteria, n = 0

6/28/2019 I searched Google using the key phrase “pennsylvania tick datasets”
Did not screen every single page of results but found one article with extensive tick data
N = 447,000
After screening using criteria, n = 1

6/28/2019 I searched the DE Department of Health using the keyword “tick”
N = 61
After screening using criteria, n = 0

6/28/2019 I searched Google using the key phrase “delaware tick datasets”
N = 18
After screening using criteria, n = 0

6/28/2019 I searched the RI Department of Health using the keyword “tick”
N = 23
After screening using criteria, n = 0

Figure Titles and Captions

Figures

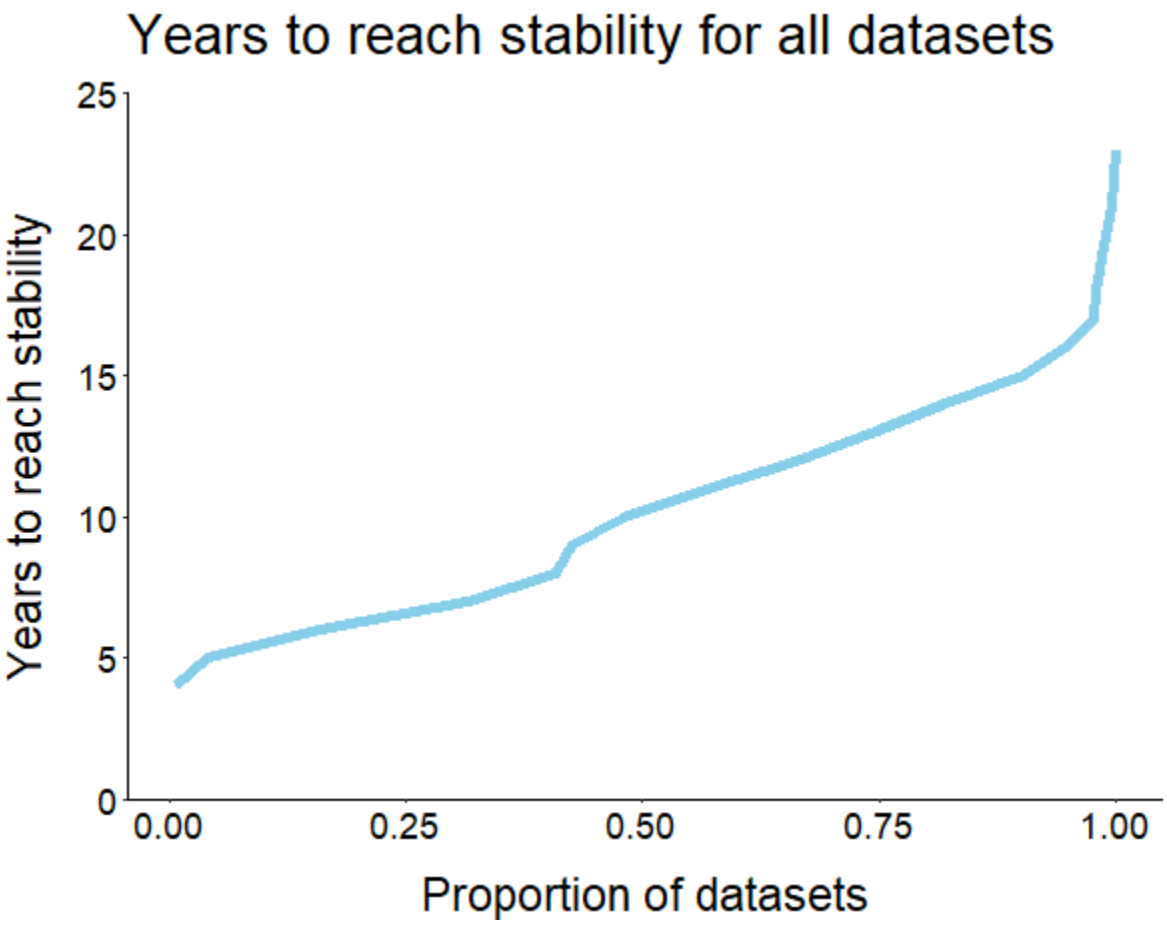

Figure 1. The fraction of datasets that take *y* years to reach stability.

The proportion of datasets within our study that reach stability within a given number of years. The majority of the datasets (n = 289) reach stability after 10 years, and none of the datasets reach stability before 4 years.

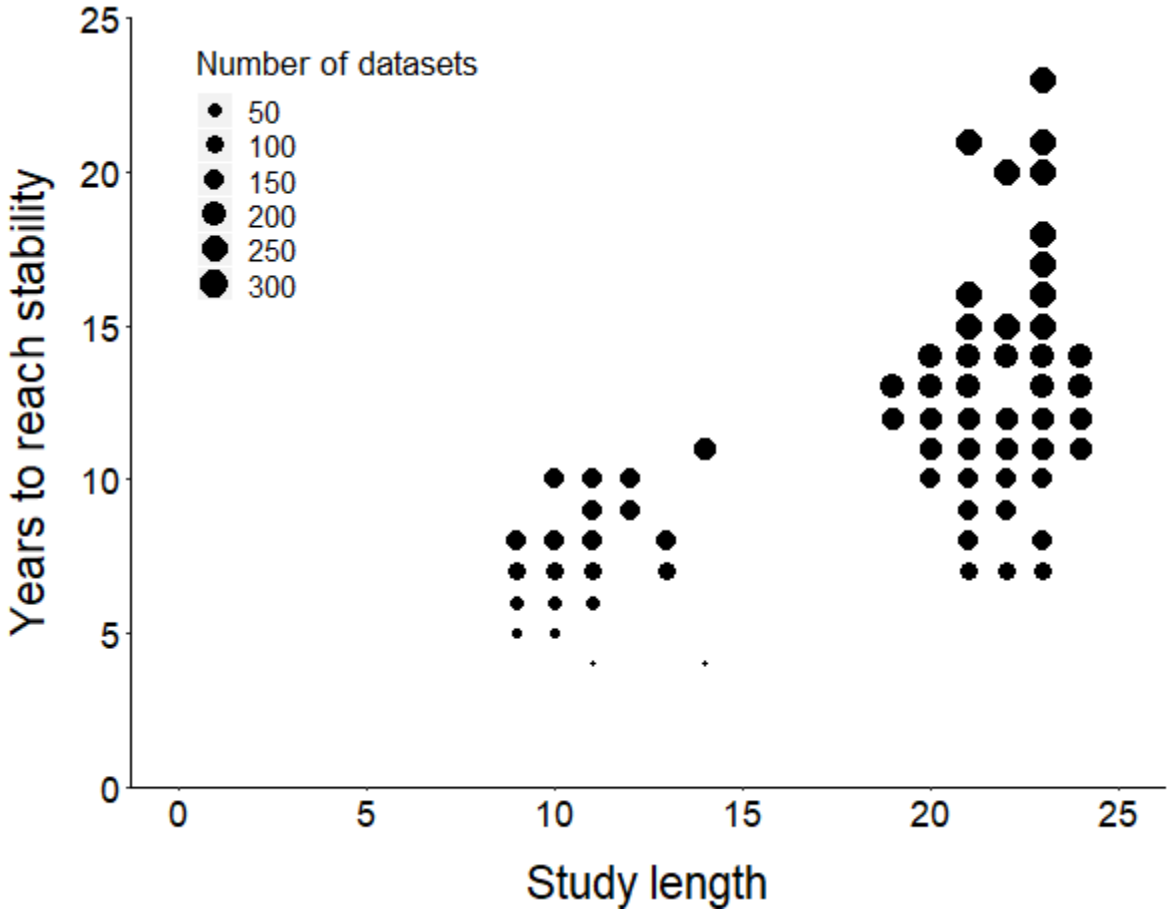

Figure 2. Comparison of study length, years to reach stability and the number of datasets.

The number of datasets is represented by point size, and the location of the points shows the datasets that have x study length and have reached stability by y years (n= 289 datasets). Overall, years to reach stability time increased with study length.

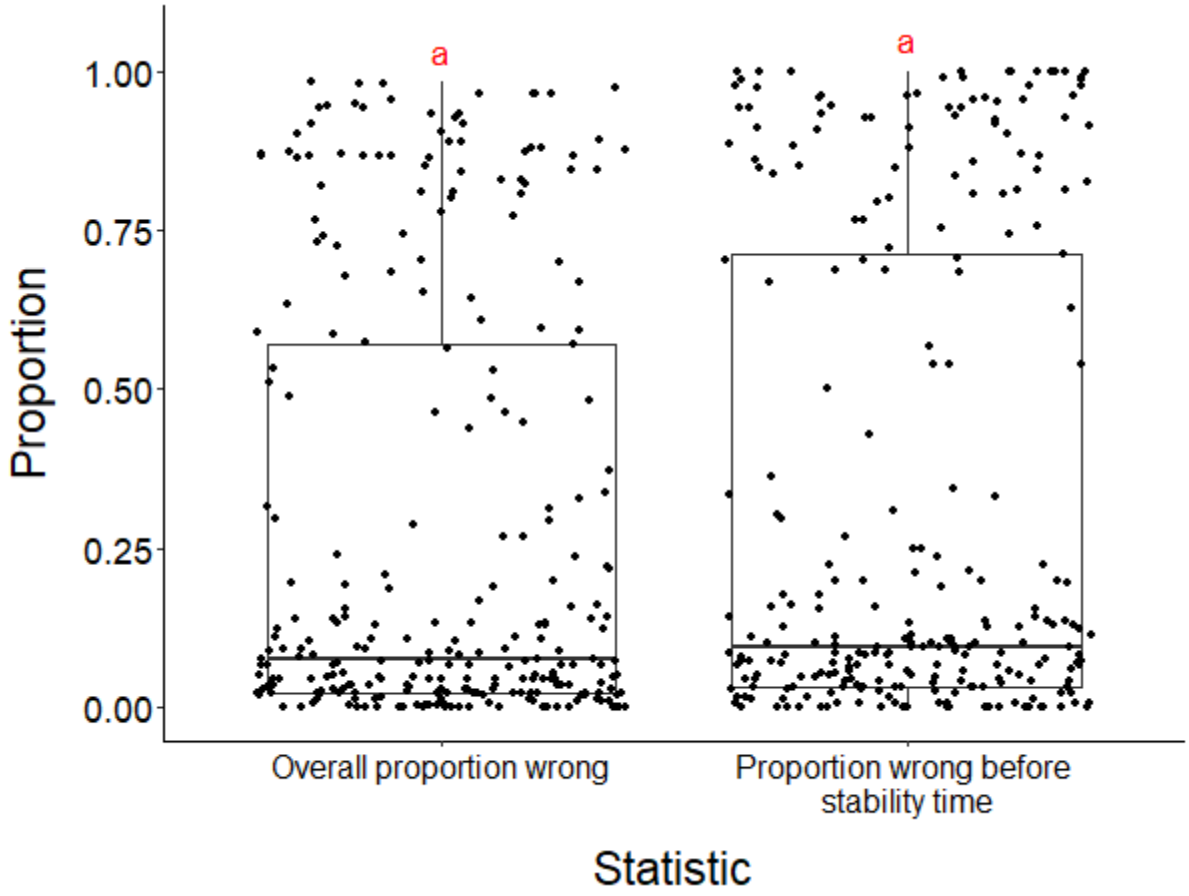

Figure 3. Overall proportion wrong compared to proportion wrong before stability.

The ends of the whiskers represent minimum and maximum values, the box represents the interquartile range, and the line in the box represents the median. The dots represent the data points for either overall proportion wrong or proportion wrong before stability time, and are spread over the x-axis to more clearly visualize them. The letters show which pair of factors are significantly different - if the factors share the same letter, they are not significantly different, but if not, then the factors are significantly different. The difference between overall proportion wrong and proportion wrong before stability time is insignificant ($t = -1.2$, $df = 571.46$, p-value = 0.23).

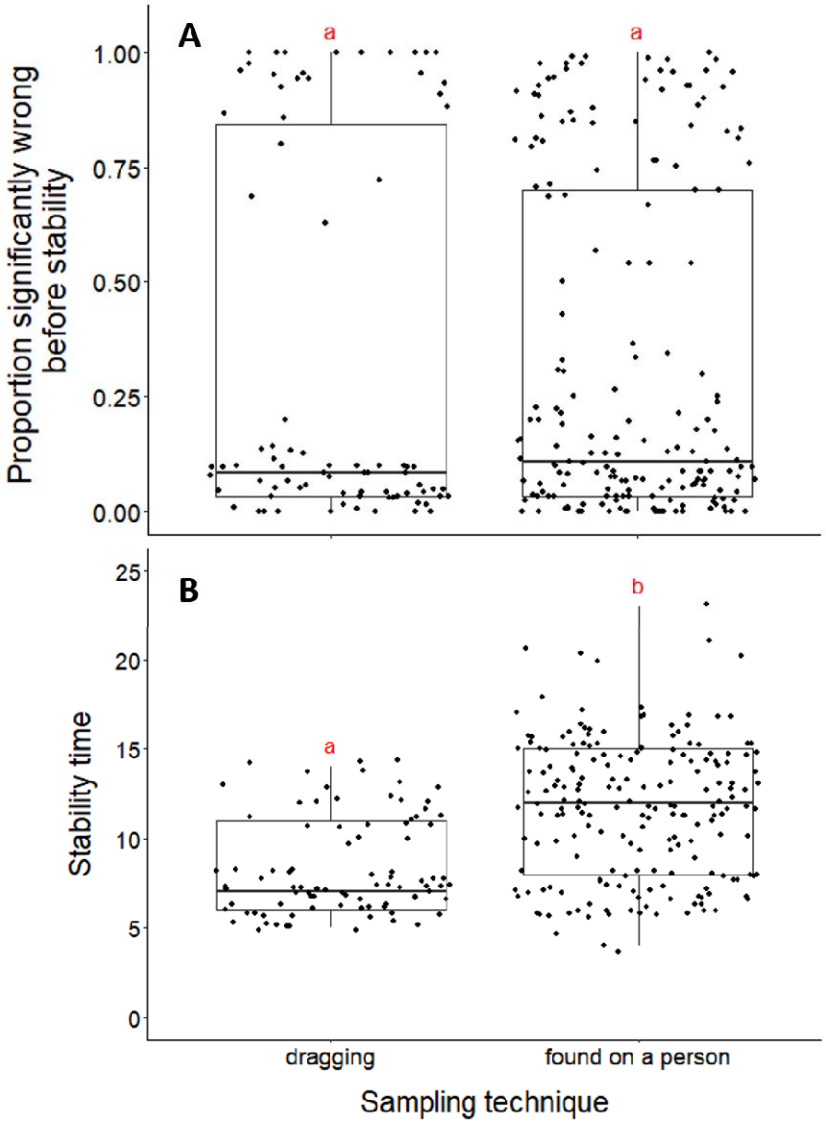

Figure 4.

Figure 4A. Comparison of proportion significantly wrong before stability by sampling techniques. The ends of the whiskers represent minimum and maximum values, the box represents the interquartile range, and the line in the box represents the median. The dots represent the data points for data produced by each sampling technique, and are spread over the x-axis to more clearly visualize them. The letters show which pair of factors are significantly different - if the factors share the same letter, they are not significantly different, but if not, then the factors are significantly different. Shows the difference between studies that use a standardized sampling technique (dragging, n=90 datasets) and studies that use an opportunistic public survey method based on reports of ticks found on people (n=198 datasets), in proportion significantly wrong before stability. The difference between datasets produced by different sampling techniques is insignificant (t = 0.08, df = 155.58, p-value = 0.93).

Figure 4B. Comparison of years to stability by sampling technique. The ends of the whiskers represent minimum and maximum values, the box represents the interquartile range, and the line in the box represents the median. The dots represent the data points for data produced by each sampling technique, and are spread over the x-axis to more clearly visualize them. The letters show which pair of factors are

significantly different - if the factors share the same letter, they are not significantly different, but if not, then the factors are significantly different. Shows the difference between studies that use a standardized sampling technique (dragging, n=90 datasets) and studies that use an opportunistic public survey method based on reports of ticks found on people (n=198 datasets), in years to reach stability. Data produced by dragging varies less by stability time ($t = -8.53$, $df = 236.23$, p-value <0.001).

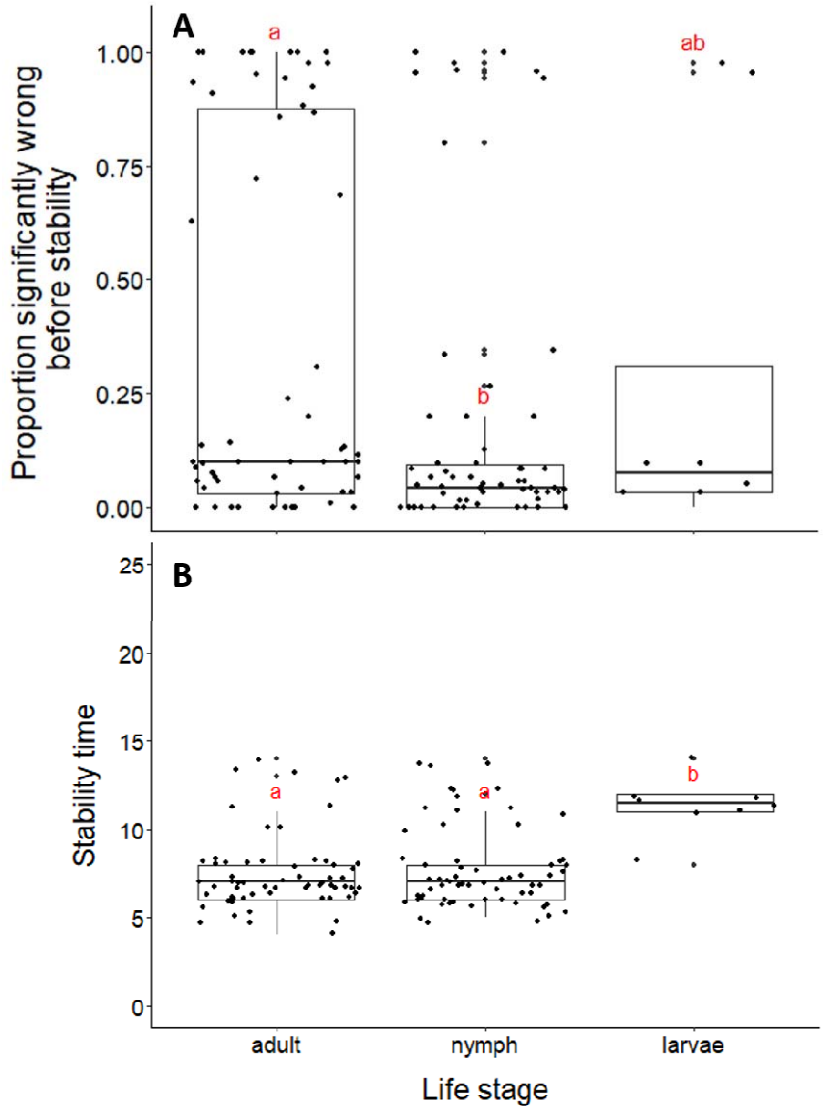

Figure 5.

Figure 5A. Comparison of proportion significantly wrong before stability by life stage.
The ends of the whiskers represent minimum and maximum values, the box represents the interquartile range, and the line in the box represents the median. The dots represent the data points for either overall proportion wrong or proportion wrong before stability time, and are spread over the x-axis to more clearly visualize them. The letters show which pair of factors are significantly different - if the factors share the same letter, they are not significantly different, but if not, then the factors are significantly different. Shows the difference between studies that measure adults (oldest, n=63 datasets), nymphs (mid-stage, n=66 datasets), and larvae (youngest, n=8 datasets), in proportion significantly wrong before stability. The difference between nymphs and larvae($t = -0.75$, $df = 7.89$, p-value = 0.47) and adults and larvae ($t = 0.44$, $df = 8.77$, p-value = 0.67) is insignificant, but there is a significant difference between adults and nymphs ($t = 2.90$, $df = 113.64$, p-value = 0.004).

Figure 5B Comparison of years to stability by life stage.
The ends of the whiskers represent minimum and maximum values, the box represents the interquartile range, and the line in the box represents the median. The dots represent the data points for either overall

proportion wrong or proportion wrong before stability time, and are spread over the x-axis to more clearly visualize them. The letters show which pair of factors are significantly different - if the factors share the same letter, they are not significantly different, but if not, then the factors are significantly different. Shows the difference between studies that measure adults (oldest, n=63 datasets), nymphs (mid-stage, n=66 datasets), and larvae (youngest, n=8 datasets), in years to reach stability. The difference between larvae and adults ($t = -5.96$, $df = 10.11$, p-value <0.001) and larvae and nymphs ($t = -5.52$, $df = 10.54$, p-value <0.001) is significant, while there is no significant difference between nymphs and adults ($t = -0.64$, $df = 126.84$, p-value $= 0.52$).

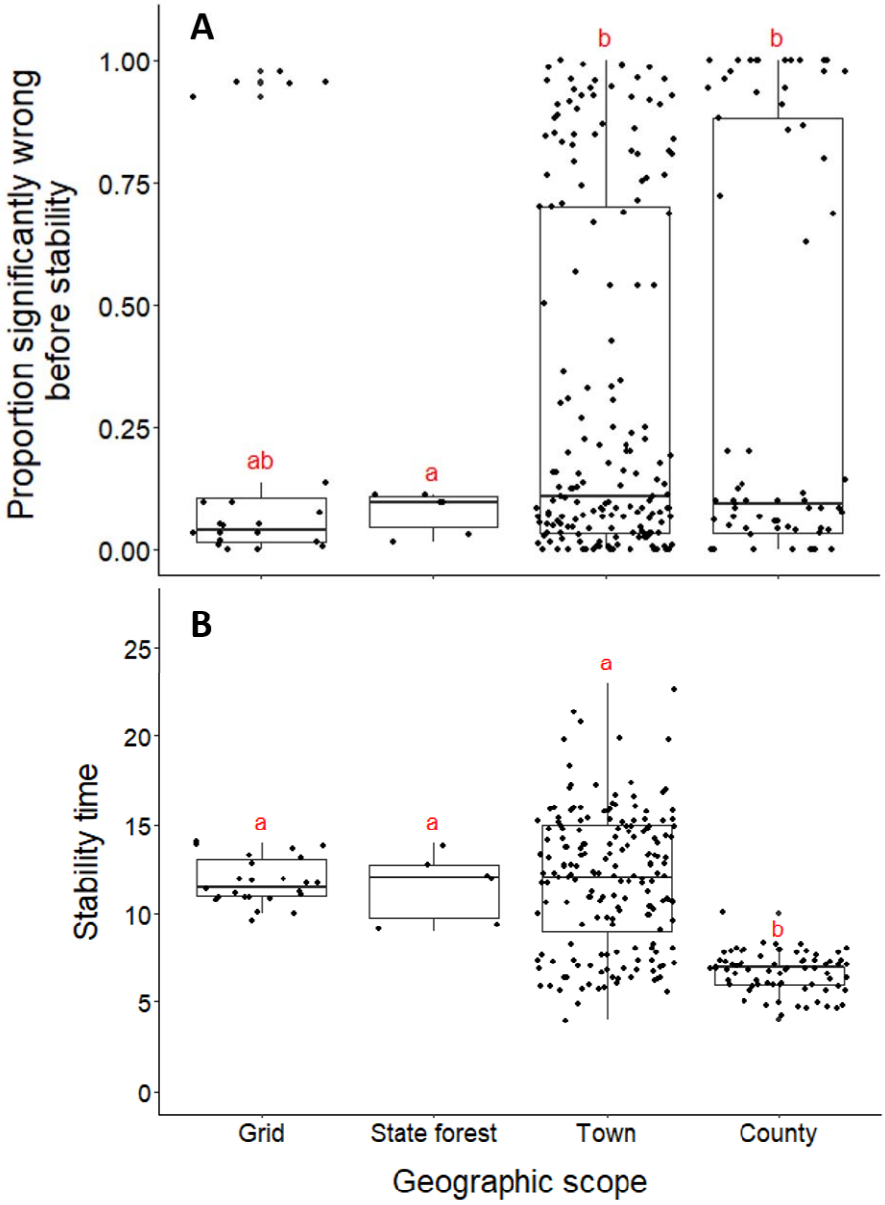

Figure 6. Comparison of proportion significantly wrong before stability by geographic scope.

Figure 6A. The ends of the whiskers represent minimum and maximum values, the box represents the interquartile range, and the line in the box represents the median. The dots represent the data points for either overall proportion wrong or proportion wrong before stability time, and are spread over the x-axis to more clearly visualize them. Shows the difference between studies that summarize data by county (n=73 datasets), town (n=186 datasets), state forest (n=6 datasets) and grid (n=24 datasets), in proportion significantly wrong before stability. The letters show which pair of factors are significantly different - if the factors share the same letter, they are not significantly different, but if not, then the factors are significantly different. For proportion significantly wrong before stability, the difference between data produced by state forest level studies versus data produced by county and town level studies is significantly different (state forest and county, t = 5.24, df = 74.29, p-value <0.001; state forest and town, t = 7.42, df = 46.44, p-value <0.001), whereas the difference between data produced by county level and town level studies (t = 0.72, df = 115.03, p-value = 0.48), county level and grid level studies (t = 1.33, df

= 43.07, p-value = 0.19), town level and grid level studies (t = 1.00, df = 28.52, p-value = 0.32), and state forest level and grid level studies (t = -1.91, df = 25.10, p-value = 0.07) is insignificant.

Figure 6B. Comparison of years to stability by geographic scope.
The ends of the whiskers represent minimum and maximum values, the box represents the interquartile range, and the line in the box represents the median. The dots represent the data points for either overall proportion wrong or proportion wrong before stability time, and are spread over the x-axis to more clearly visualize them. The letters show which pair of factors are significantly different - if the factors share the same letter, they are not significantly different, but if not, then the factors are significantly different. Shows the difference between studies that summarize data by county (n=73 datasets), town (n=186 datasets), state forest (n=6 datasets) and grid (n=24 datasets), in years to reach stability. For years to stability, the difference between data produced by county level studies versus data produced by town, state forest, and grid level studies is significantly different (county and town, t = -17.03, df = 243.09, p-value <0.001; county and state forest, t = -5.55, df = 5.22, p-value = 0.002; county and grid, t = -17.21, df = 33.87, p-value <0.001) whereas the difference between data produced by town level studies, state forest level studies, and grid level studies is insignificant (town and state forest, t = 0.47, df = 6.12, p-value = 0.65; town and grid, t = 0.22, df = 86.18, p-value = 0.8; state forest and grid, t = -0.38, df = 6.03, p-value = 0.72).

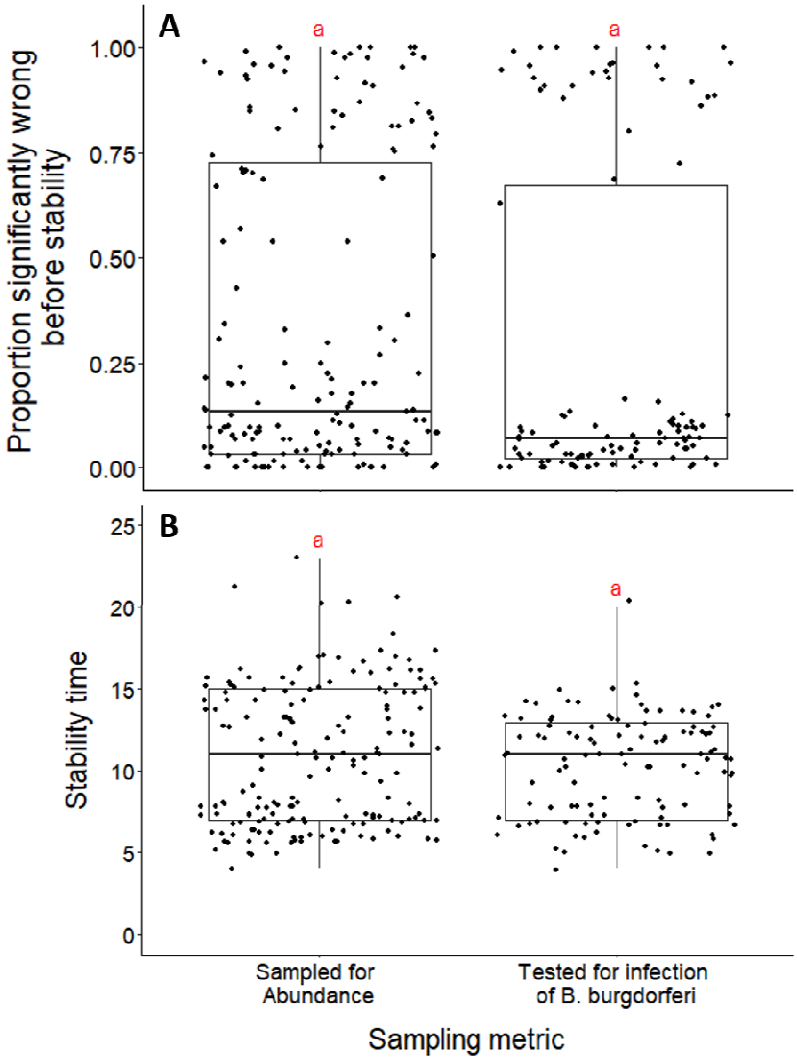

Figure 7.
Figure 7A. Comparison of proportion significantly wrong before stability by study response variable. The ends of the whiskers represent minimum and maximum values, the box represents the interquartile range, and the line in the box represents the median. The dots represent the data points for either overall proportion wrong or proportion wrong before stability time, and are spread over the x-axis to more clearly visualize them.The letters show which pair of factors are significantly different - if the factors share the same letter, they are not significantly different, but if not, then the factors are significantly different. Shows the difference between studies that sample deer ticks for abundance (n=175) and studies that test deer ticks for *B. burgdorferi* (n=114), in proportion significantly wrong before stability. The difference between the data produced by the response variables in proportion significantly wrong before stability is insignificant (t = -1.18, df = 232.98, p-value = 0.24).

Figure 7B Comparison of stability time by sampling metric.
The ends of the whiskers represent minimum and maximum values, the box represents the interquartile range, and the line in the box represents the median. The dots represent the data points for either overall proportion wrong or proportion wrong before stability time, and are spread over the x-axis to more clearly visualize them.The letters show which pair of factors are significantly different - if the factors share the same letter, they are not significantly different, but if not, then the factors are significantly different.

Shows the difference between studies that sample deer ticks for abundance (n=175) and studies that test deer ticks for *B. burgdorferi* (n=114), in stability time. The difference between the data produced by the response variables in years to stability is insignificant (t = -1.29, df = 283.9, p-value = 0.20).

Tables

Dataset name	Data repository	Location	Geographic scope	Life stage	Sampling technique
Summaries of Tick Testing	The Connecticut Agricultural Experiment Station	Connecticut, US	Town	Unspecified	Found on a person
Incidence of Ticks and Tick Bites at Harvard Forest since 2006	LTER Data Portal	Harvard Forest, Massachusetts, US	State forest	Unspecified	Found on a person
The growing importance of lone star ticks in a Lyme disease endemic county: Passive tick surveillance in Monmouth County, NJ, 2006 – 2016	NCBI	Monmouth County, New Jersey, US	Town	Adults, nymphs	Found on a person
Deer Tick Surveillance: Adults (Oct to Dec) excluding Powassan virus: Beginning 2008	NY Department of Health	New York, US	County	Adults	Dragging
Deer Tick Surveillance: Nymphs (May to Sept) excluding Powassan virus: Beginning 2008	NY Department of Health	New York, US	County	Nymphs	Dragging
Tick-borne disease risk in a forest food web	Data Dryad	Cary Forest, New York, US	Grid	Adults, larvae, nymphs	Dragging
Range expansion and increasing Borrelia burgdorferi infection of the tick Ixodes scapularis (Acari: Ixodidae) in Iowa, 1990-2013	Data Dryad	Iowa, US	County	Adults, nymphs	Found on a person

Table 1. Location, geographic scope, life stage, and sampling technique of deer tick datasets found and included in our analysis.