Models and data used to predict the abundance and distribution of *Ixodes scapularis* (blacklegged tick) in North America: a scoping review

Yogita Sharma,[a,g] Elda K. E. Laison,[b,g] Tanya Philippsen,[a,g] Junling Ma,[a] Jude Kong,[c] Sajjad Ghaemi,[d] Juxin Liu,[e] François Hu,[f] and Bouchra Nasri[b,*]

[a]Department of Mathematics and Statistics, University of Victoria, Victoria, Canada
[b]Département de Médecine Préventive et Sociale, University of Montréal, Montréal, Canada
[c]Department of Mathematics and Statistics, York University, Toronto, Ontario, Canada
[d]Digital Technologies Research Center, National Research Council of Canada, Toronto, Canada
[e]Department of Mathematics and Statistics, University of Saskatchewan, Saskatoon, Saskatchewan, Canada
[f]Department of Mathematics and Statistics, University of Montréal, Montréal, Canada

The Lancet Regional Health - Americas
2024;32: 100706

Published Online xxx
https://doi.org/10.1016/j.lana.2024.100706

Summary

Tick-borne diseases (TBD) remain prevalent worldwide, and risk assessment of tick habitat suitability is crucial to prevent or reduce their burden. This scoping review provides a comprehensive survey of models and data used to predict *I. scapularis* distribution and abundance in North America. We identified 4661 relevant primary research articles published in English between January 1st, 2012, and July 18th, 2022, and selected 41 articles following full-text review. Models used data-driven and mechanistic modelling frameworks informed by diverse tick, hydroclimatic, and ecological variables. Predictions captured tick abundance (n = 14, 34.1%), distribution (n = 22, 53.6%) and both (n = 5, 12.1%). All studies used tick data, and many incorporated both hydroclimatic and ecological variables. Minimal host- and human-specific data were utilized. Biases related to data collection, protocols, and tick data quality affect completeness and representativeness of prediction models. Further research and collaboration are needed to improve prediction accuracy and develop effective strategies to reduce TBD.

Keywords: *Ixodes scapularis*; Models; Ticks; North America; Data-driven; Mechanistic; Mathematical

Introduction

Tick-borne diseases (TBDs) like Lyme disease (*Borrelia burgdorferi*) pose a significant public health threat in North America due to their increasing incidence. TBDs account for over half of all vector-borne illnesses in this region, and their non-specific and overlapping clinical manifestations make early diagnosis challenging.[1–3] Therefore, understanding tick population dynamics and associated supportive factors is crucial.[1,2] In North America, the ticks responsible for Lyme disease are *Ixodes scapularis* and *Ixodes pacificus*, with *I. scapularis* (blacklegged or deer tick) being the predominant and among the most medically important tick vectors[3] as it can transmit multiple pathogens simultaneously to humans in one bite.[4] *I. scapularis* has a 3-host life cycle involving transitioning from larva to nymph to adult stages and may last from 2 to 3 years.[5] Tick survival and reproduction requires available, life-stage-appropriate hosts,[6] such as white-tailed deer (*Odocoileus virginianus*) for adult ticks, and small mammals like white-footed mice (*Peromyscus leucopus*) for larvae and nymphs.[7] Most pathogens are transmitted by bites of tick nymphs. Their small size makes them difficult to detect compared to larger adult ticks, while the larval stage experiences low incidence of infection.[5,6] Humans are incidental hosts as they are not typically part of the tick life cycle.[5] TBD risk is highly correlated with infected tick abundance (absolute or relative number of ticks in a given location) and distribution (spatial extent of the species),[8] and both are influenced by intricate interplay between abiotic and biotic factors.[4,8]

Researchers use two modelling frameworks to study tick propagation: mechanistic and data-driven. Mechanistic models depend on assumptions about species characteristics and behaviours associated with various environmental drivers; however, incorporating multiple variables and dynamics lead to complex model behaviours which impacts interpretability.[8,9] Conversely, data-driven models use patterns and information from data to make predictions through computational algorithms and can easily handle complex variables; however, their predictive accuracy depends on data quality and quantity.[9]

We identified two scoping reviews: one focused on species distribution modelling (SDM) of *Amblyomma* tick

*Corresponding author.
E-mail address: bouchra.nasri@umontreal.ca (B. Nasri).
[g]Y. Sharma, E.K.E Laison and T. Philippsen contributed equally to this paper and are considered co-first authors.

species globally (studies published between 1994 and 2019)[10] and the other review prioritised SDM of several medically important tick species worldwide, including *I. scapularis* (studies published between 1998 and 2012).[11] Kopsco and colleagues[11] focus only on SDM and habitat suitability modelling in their search criteria, which may have resulted in the exclusion of tick abundance modelling from their scoping review.[11] While climatic factors are the main contributors to tick distribution, tick abundance is directly influenced by host density, both of which lead to an increase in TBDs.[12] Therefore, to better assess the risks of tick establishment and TBDs, distribution and habitat suitability along with tick abundance must be considered. Thus, our scoping review complements the work of Kopsco and collaborators[11] by providing an overview of the data and modelling approaches used for predicting *I. scapularis* distribution and abundance in North America.

This scoping review aims to answer the following research questions: (1) What are the current modelling frameworks and data used to predict *I. scapularis* population distribution, abundance, and factors contributing to its spread in North America? (2) What are the limitations and suitability of these approaches and data sources, and how do they impact model predictive accuracy?

Methodology

This scoping review followed the Preferred Reporting Items for Systematic Review and Meta-Analysis Extension for Scoping Reviews (PRISMA-ScR) guidelines.[13] Literature searches were conducted in Web of Science and PubMed online databases selected for their open access policy, range of health-related publications, and article update frequency.[11,14] Keywords related to *I. scapularis* modelling were identified for the search (Appendix A).

YS performed the literature search, and YS and EL performed the first round of title and abstract screening in Covidence (Covidence systematic review software, Veritas Health Innovation, Melbourne, Australia; available at www.covidence.org). To examine temporal publication trends for this topic, date ranges were not restricted. The initial screening included peer-reviewed primary research articles with full-text available and published in English, focusing on predictive models for *I. scapularis* distribution, abundance, and Lyme disease risk. Exclusion criteria omitted review articles, experimental studies, short communications, articles published in languages other than English, and studies focusing on non-*I. scapularis* tick species.

After initial screening, article eligibility disagreements were resolved (BN). Notably, most relevant literature was published between January 1, 2012 and July 18, 2022, which was set as the study inclusion timeframe. A data extraction spreadsheet was developed to collect information about study objectives, model types, data sources, and predictor variables, among others (Appendix A). The full-text screening of all retained articles underwent two phases. First, articles that did not meet the above inclusion criteria were excluded (YS, EL). Relevant data were independently extracted from the retained articles (YS, EL) and results were further discussed (YS, EL, BN) with disagreements resolved by BN. As many articles focused on Lyme disease risk or regions other than North America, the inclusion criteria were further refined (Fig. 1).

The second full-text screening phase included studies focusing on *I. scapularis* distribution and abundance modelling in North America and excluded studies prioritizing Lyme disease risk modelling or focusing on other geographic areas. BN supervised and approved the entire extraction process. Extracted data was then further categorized.

Models were identified as data-driven or mechanistic. Tick data were grouped by collection method: active, passive, and proxy. In active surveillance, researchers or health officials directly sample field areas, by dragging or trapping hosts (deer, birds, rodents) to inspect them for ticks. In passive surveillance, civilians report ticks to health officials after finding ticks on themselves or their pets.[15] Proxy data includes citizen science (public participation in scientific research and data collection activities)[16] or historical data (tick surveillance data from previous studies). Other variables used in model development were categorised as hydroclimatic, ecological, human behavioural, or TBD cases data (Table 1). Model response variables were categorized as tick abundance (including tick density and population counts) and tick distribution (including habitat suitability, tick occurrence, presence/absence, invasion, or expansion risk).[8]

Results

A total of 4661 articles (PubMed: 1890, Web of Science: 2771) were identified. Based on the initial inclusion and exclusion criteria, 4074 articles were excluded after title and abstract screening and duplicate removal. The remaining 587 articles also underwent full-text screening, and 431 additional articles were excluded. Additionally, articles that did not focus on North America (n = 54) or were primarily focused on Lyme disease risk assessment (n = 61) were also excluded (n = 115). No additional studies beyond this search were included. The final data extraction included 41 articles (Figs. 1 and 2, Table 2, Appendix B). Appendix C provides summaries of the parameter, variable, and model selection methods and model validation approaches used in each study, if reported.

Studies were geographically focused on the US (n = 18, 44%), Canada (n = 20, 49%), Mexico and the US (n = 2, 5%) and the US and Canada (n = 1, 2%), and

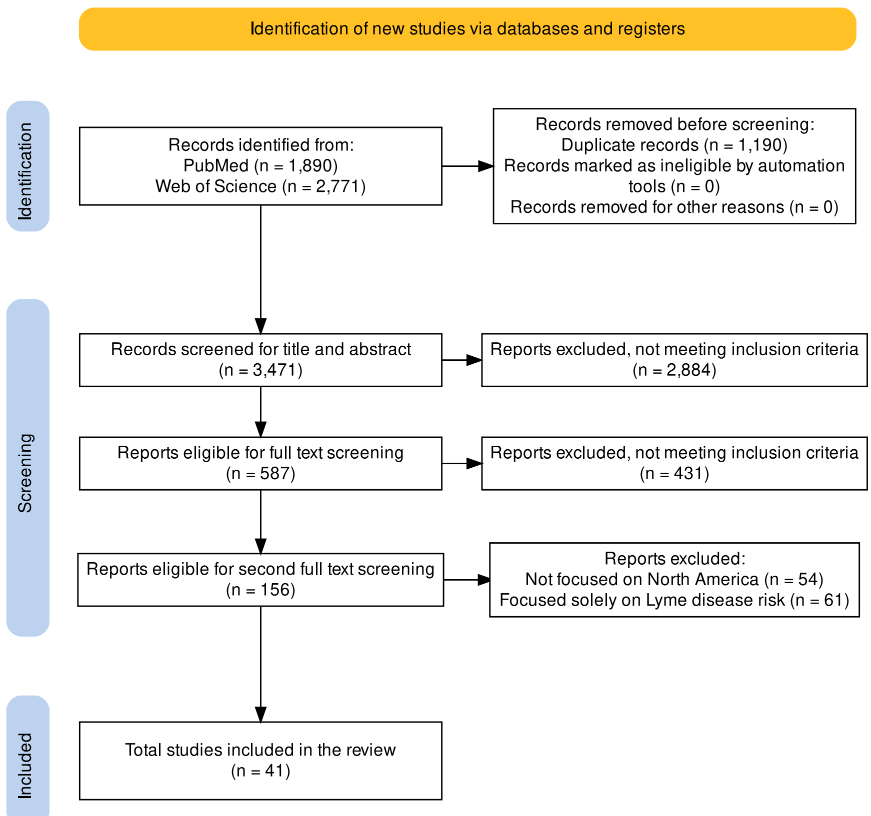

Fig 1: PRISMA Flowchart Illustrating the Scoping Review screening and selection process. The figure was generated by the PRISMA flow chart online tool (https://estech.shinyapps.io/prisma_flowdiagram/).

consistent research interest in *I. scapularis* over time is evident (Fig. 3). Data-driven (n = 33, 80%) and mechanistic (n = 8, 20%) models were identified (Fig. 4 and 5). Overall, model predictions focused on tick abundance (n = 14, 34.1%), tick distribution (n = 22, 53.7%) and both tick abundance and distribution (n = 5, 12.2%) (Figs. 4 and 5A).

For tick abundance, data-driven (n = 13, 92.8%) and mechanistic models (n = 1, 7.2%) were implemented. Data-driven studies focusing on tick abundance (n = 13, 100%) employed count data regression models (Fig. 6), comprising simple Poisson and negative binomial (NB) (n = 1, 7.7%),[17] multivariable NB (n = 1, 7.7%),[18] simple and multivariable NB (n = 1, 7.7%),[19] simple NB models (n = 1, 7.7%),[20] multivariable zero-inflated Poisson (n = 1, 7.7%) and multivariable zero-inflated NB (n = 1,7.7%) in cases of excess zeros.[21,22] NB was preferred over Poisson regression when handling count data demonstrating overdispersion.[17] Additionally, Generalized Estimating Equation (GEE) NB and Poisson regression models (n = 1, 7.7%) were employed to accommodate potential clustering arising from repeated measures.[51] Other approaches included simple and multivariable linear regression (n = 5, 38.4%)[23–27] and count generalized linear models (GLM) (n = 1, 7.7%).[28] For mechanistic modelling, tick abundance was addressed using temperature-driven ordinary differential equations (ODEs) model to simulate population dynamics of *I. scapularis* (n = 1, 100%) (Appendix D, Figure D1).[29]

For tick distribution, both data-driven (n = 16, 72.7%) and mechanistic models (n = 6, 27.3%) were employed. Among the data-driven tick distribution studies (n = 16, 100%), logistic regression (LgR) models (n = 7, 43.7%)[7,30–33,35,36] were used to predict binary outcomes (i.e., tick presence or absence), while MaxEnt was used to model *I. scapularis* habitat suitability (n = 6, 37.5%). Two studies included survival regression (n = 2, 12.5%) with one also utilizing simple LgR.[1,34] These approaches integrated environmental and dispersal factors and provide a comprehensive analysis and temporal projections of *I. scapularis* establishment.[1,34] A model was also identified (n = 1, 6.3%) employing multivariable regression and machine learning techniques like boosted regression tree, generalized linear model (GLM) multivariate adaptive regression spline, MaxEnt, and random forest models.[37] Overall, LgR was most utilised, including simple and multivariable logistic (n = 1,

Hydroclimatic variables		Ecological variables		Human components
Temperature	Temperature during warmest month	**Location**	Latitude	**Human behaviour**
	Temperature during coldest month		Longitude	**TBD cases**
	Temperature during warmest quarter		Geographic location	
	Temperature during coldest quarter	**Vegetation**	Forest cover	
	Annual maximum temperature		Vegetation index	
	Daily mean temperature		Canopy cover	
	Minimum temperature during coldest month		Soil type	
	Maximum temperature during warmest month		Soil moisture	
	Monthly mean temperature		Landcover (e.g., shrub coverage, grassland, agricultural area)	
			Litter depth	
	Annual cumulative degree days (DD > 0 °C)	**Landscape**	Presence of river	
	Mean diurnal temperature		Landscape fragmentation	
	Isothermality		Elevation	
	Annual mean temperature	**Host density**	Host density	
Precipitation	Total annual rainfall			
	Precipitation during driest quarter			
	Precipitation during wettest quarter			
	Precipitation during warmest quarter			
	Precipitation during coldest quarter			
	Annual mean precipitation			
Humidity	Relative humidity			
	Index of atmospheric moisture			
	Vapour pressure			

Table 1: **Description of the three main classes of predictor variables used in the models.**

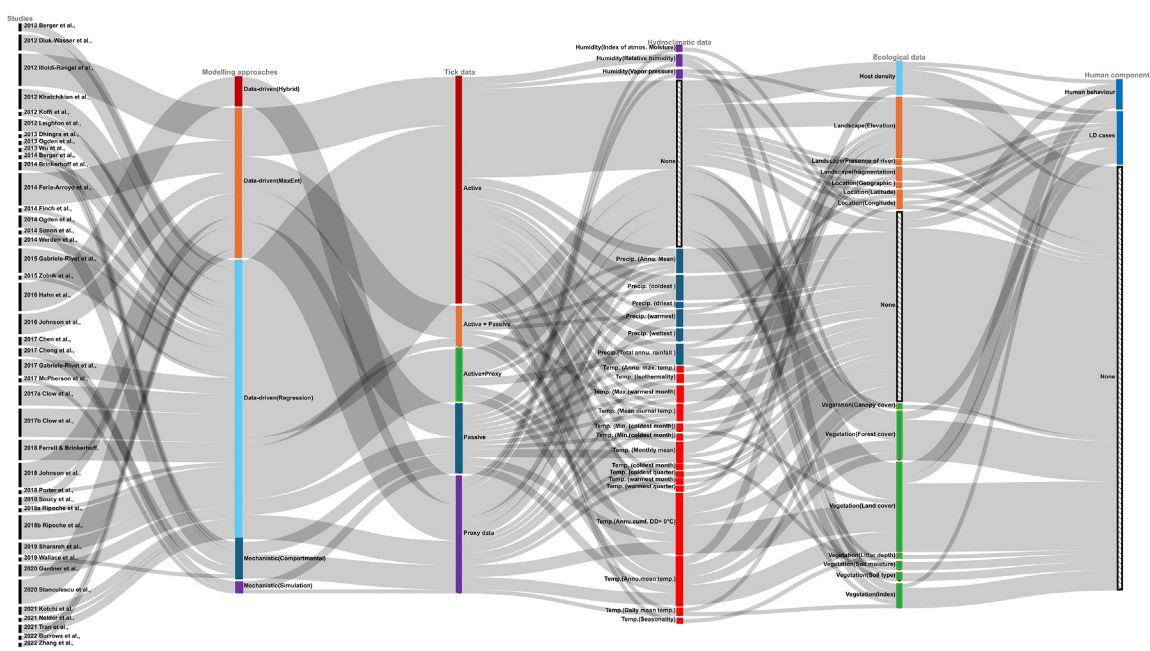

Fig 2: Alluvial Diagram depicting relationships among studies, modelling methods, tick data, hydroclimatic variables, and ecological variables.

Response variable focus: tick abundance (n = 14)							
Studies	Location & time frame	Model framework (Type) & specific models	Predictor variable categories	Study objective(s)	Significant predictor(s)	Limitations	Statistical methods to assess outcomes, validation approaches & model predictive accuracy[a] (if provided)
Nelder et al., 2021[17]	Ontario (Canada) 2011–2017	Data-driven (Regression): Simple Negative binomial regression Simple Poisson regression	Tick data (Passive)	Examine whether there is a spatial and temporal relationship between submission rates of *I. scapularis* ticks found on humans compared to those identified on pets at the public health unit (PHU) geographic level in Ontario.	None reported	Data reflects the prioritization of tick submissions from humans over pets in PHUs, as well as the absence of tick collection in some PHUs. Lack of veterinary awareness of pet-associated tick submission options potentially underestimates pet-tick encounters. Absence of pet populations required estimation by proxy using human population data. Model may underestimate tick exposure risk because: 1) tick establishment in Ontario is recent; and 2) passive tick surveillance data have some predictive limitations.	**Statistical Methods** Incidence Rate Ratio (IRR) **Validation Methods** None reported **Predictive Accuracy** None reported
Porter et al., 2019[18]	Connecticut, Maine, Massachusetts, Vermont, New York, New Hampshire, New Jersey, New York, Rhode Island, Pennsylvania (USA) January 2016 and August 2017	Data-driven (Regression): Multivariable Negative binomial regression	Tick data (Proxy) Human data	Assess the influence of seasonality and human behaviour on tick encounters in the northeastern United States and the association with Lyme disease risk through the use of citizen science (CS) tick submission data	Human behaviours (yard work, outdoor activities, recreation in forests)	CS data collection processes often lack systematic structure and can be influenced by location biases and participant representation. Nymph detection is more difficult thus CS tick data may be skewed towards adult ticks resulting in possible underestimation of nymph exposure.	**Statistical Methods** Spearman Rank Correlation **Validation Methods** Mean Absolute Error (MAE) Root Mean Square Error (RMSE) Normalized Root Mean Square Error (NRMSE) **Predictive Accuracy** MAE (Model 3): 60.7 RMSE (Model 3): 101.7
Finch et al., 2014[19]	Block Island, Rhode Island (USA) May–August, 2012	Data-driven (Regression): Simple Negative binomial regression Multivariable Negative binomial regression	Tick data (Active) Ecological data	Examine the relationship between landscape measures, nymph density, and risk of human–tick interaction and Lyme disease on Block Island, Rhode Island.	Positive association: Shrub metrics (Class Area, Largest Patch Index, Total Edge, Edge Density, Landscape Shape Index) Negative association: Lawn metrics (Largest Patch Index, Edge Density)	Exact location of exposure was difficult to identify, and details regarding the varying use of preventive measures were not collected.	**Statistical Methods** Difference in means (estimates, p-values) **Validation Methods** Not reported. **Predictive Accuracy** Land cover classification (errors of omission and commission): 83.6%
Zolnik et al., 2015[20]	Northeastern USA June and July 2013	Data-driven (Regression): Simple Negative binomial regression	Tick data (Active) Ecological data	Use landscape fragmentation characteristics to represent host biodiversity and examine the "dilution effect hypothesis" on the prevalence and density of infected nymphs in tick-endemic areas of the northeastern United States.	None reported	Detached forest patches were absent in the majority of the sample sites so forest patch size was excluded from the models, preventing the assessment of their association on disease risk.	**Statistical Methods** Difference in means (estimates, p-values) **Validation Methods** Not reported **Predictive Accuracy** None reported

(Table 2 continues on next page)

Review

Response variable focus: tick abundance (n = 14)

Studies	Location & time frame	Model framework (Type) & specific models	Predictor variable categories	Study objective(s)	Significant predictor(s)	Limitations	Statistical methods to assess outcomes, validation approaches & model predictive accuracy[a] (if provided)
(Continued from previous page)							
Khatchikian et al., 2012[21]	Hudson River Valley, New York State (USA) 2004–2010	Data-driven (Regression): Multivariable linear regression Multivariable Poisson regression Multivariable Zero-inflated Poisson regression	Tick data (Active) Hydroclimatic data Ecological data	Report empirically detailed ten-year data for *I. scapularis* population growth and identify associated environmental factors in Hudson River Valley, New York State.	Nymph model: location, year, sampling week (season), precipitation (summer), temperature in winter (minimum), amount of forest coverage Adult model: location, year, sampling week of sampling (season), precipitation (winter), temperature in winter (minimum), amount of forest coverage urbanization, and interaction between forest coverage and urbanization)	Annual visits did not cover all collection sites, potentially leading to overestimates of tick density.	**Statistical Methods** Differences in means (estimates, p-values) **Validation Methods** Data **Predictive Accuracy** R^2 (nymphal tick density): 0.642 R^2 (adult tick density): 0.622
Diuk-Wasser et al., 2012[22]	37 states east of 100th meridian (USA) 2004–2007	Data-driven (Regression): Multivariable Zero-inflated negative binomial regression	Tick data (Active) Hydroclimatic data Ecological data	Create a continuous predictive map for human risk of Lyme disease by analyzing the influence of environmental factors on nymphal tick infections.	For infected nymph absence: elevation, vapour pressure deficit (mean), yearly temperature cycle phase (minimum) For infected nymph occurrence: largest mixed forest cover patch (+ spatial autocovariate)	Density of infected nymphs are underestimated in specific areas (e.g., Parker Dam state Park) due to their separation from areas with high density, low autocovariate values, and the impact of high elevation. The sampling strategy covered only natural regions which were used for continent-wide standardization, thus the predictions are regionally limited due to the small sample size.	**Statistical Methods** Differences in means (estimates, p-value) **Validation Methods** Not reported **Predictive Accuracy** Accuracy: 0.91 Sensitivity: 0.93 Specitivity: 0.90 Positive predictive value (PPV): 0.63 Negative predictive value (NPV): 0.99
Ripoche et al., 2018a[23]	Quebec (Canada) 2009–2014	Data-driven (Regression): Multivariable Generalized Estimating Equation (GEE) negative binomial regression Multivariable Generalized Estimating Equation (GEE) Poisson regression	Tick data (Active, Passive) Human data	Develop new Lyme disease risk indicators by analysing the connections between tick and human surveillance data and assessing their capability to differentiate between defined risk levels implemented in Quebec.	# ticks provided through passive surveillance approaches during the previous 2 years, adjusted according to the population size (humans)	Predictive accuracy may have been reduced due to the exclusion of tick submissions from pets and environmental variables from the model. Nymph density may be underestimated due to single site samples and few nymphs obtained by passive surveillance. Human-tick encounters associated with passive surveillance data do not always signify well-established tick populations (i.e., adventitious ticks).	**Statistical Methods** Difference in means (estimates, p-values) **Validation Methods** Pearson correlation **Predictive Accuracy** (Based on Criterion 3): AUC: 0.961 Sensitivity: 0.889 Specificity: 0.908 NPV: 99% PPV: 13%
Berger et al., 2014[23]	Rhode Island (USA) 1997–2010	Data-driven (Regression): Simple linear regression Multivariable linear regression	Tick data (Active) Ecological data Hydroclimatic data	Estimate "tick adverse moisture events (TAMEs)" using atmospheric moisture, relative humidity, and leaf litter, and evaluate their effect on *I. scapularis* survival and the risk of human–tick interactions.	# TAMEs occurring in June every year lasting over 8 h	Host density and temperature were not considered in this study but may be responsible for impacting nymph occurrence when collecting samples and nymph questing activities, respectively.	**Statistical Methods** Difference in means (estimates, p-value) **Validation Methods** Not reported **Predictive Accuracy** Not reported

(Table 2 continues on next page)

Response variable focus: tick abundance (n = 14)

Studies	Location & time frame	Model framework (Type) & specific models	Predictor variable categories	Study objective(s)	Significant predictor(s)	Limitations	Statistical methods to assess outcomes, validation approaches & model predictive accuracy[a] (if provided)
(Continued from previous page)							
Brinkerhoff et al., 2014[24]	Virginia (USA) 2008–2011	Data-driven (Regression): Weighted multivariable linear regression	Tick data (Active) Ecological data Human data	To investigate, under the assumption of a positive correlation, the link between infected tick density and the county risk level for Lyme disease in Virginia.	Longitude	Tick data is spatially and temporally incomplete which makes it difficult to accurately establish relationships between county-level Lyme disease and tick density variation. Timing of tick collection (daytime) may have excluded ticks with other questing behaviour.	**Statistical Methods** Difference in means (p-values) **Validation Methods** Field observations **Predictive Accuracy** Not reported
Tran et al., 2021[25]	New York State (USA) 2016–2017	Data-driven (Regression): Simple linear regression Multivariable linear regression	Tick data (Active, Proxy) Ecological data Human data	Examine whether citizen science (CS) datasets can be used to precisely describe *Ixodes scapularis* populations when compared with traditional tick surveillance data over a broad range of locations and time periods.	CS submitted ticks (+ all collector associated variables)	CS data can be impacted by underlying issues associated with its collection, as well as participation bias limitations resulting in inaccurate tick abundance predictions.	**Statistical Methods** NA **Validation Methods** 10-fold cross-validation Root Mean Square Error (RMSE) Field observations **Predictive Accuracy** Full model without Lyme disease: RMSE: 0.45 R^2: 0.63
Berger et al., 2013[26]	Southeastern New England (USA) 2009–2010	Data-driven (Regression): Simple linear regression	Tick data (Proxy) Hydroclimatic data Ecological data	Evaluate the use of remotely sensed surface moisture conditions and the Temperature Vegetation Dryness Index (TVDI) to describe the factors impacting tick habitats and potential risk of Lyme disease in New England.	Surface moisture Temperature Forest cover	Limitation not given.	**Statistical Methods** Differences in means (estimates, p-value) **Validation Methods** Field observations **Predictive Accuracy** Accuracy of model with dry/wet-edge parameters (May–Aug 2009): Adjusted R^2 range: 0.67–0.78 ($p < 0.001$) Accuracy of model with dry/wet-edge parameters (May–Aug 2010): Adjusted R^2 range: 0.57–0.76 ($p < 0.001$)

(Table 2 continues on next page)

Response variable focus: tick abundance (n = 14)

Studies	Location & time frame	Model framework (Type) & specific models	Predictor variable categories	Study objective(s)	Significant predictor(s)	Limitations	Statistical methods to assess outcomes, validation approaches & model predictive accuracy[a] (if provided)
(Continued from previous page)							
Werden et al., 2014[27]	Ontario (Canada) 2009–2010	Data-driven (Regression): Multivariable linear regression Multivariable logistic regression	Tick data (Active) Hydroclimatic data Ecological data	Explore the effects of climatic, environmental, spatial location, and host biodiversity factors on the risk of Lyme disease in the Thousand Islands area of Ontario.	Avg. Daily minimal temperature (Jul–Sept), avg. # deer pellet groups/hectare, relative mice abundance, small mammal species diversity, % canopy cover, distance away from US mainland, interaction between mice and richness	Focusing on small animal host populations only rather than all possible hosts may hinder the development of a more holistic understanding of the impact of biodiversity on Lyme disease risk.	**Statistical Methods** Difference in means (estimates, p-values) **Validation Methods** None reported. **Predictive Accuracy** R^2 (Model 1–Number of nymphs (NON)): 0.51 R^2 (Model 1–Number of infected nymphs (NIN)): 0.45
Ferrell & Brinkerhoff., 2018[28]	Central Virginia (USA) 2014	Data-driven (Regression): Multivariable Count GLM with combined forward and backward stepwise selection	Tick data (Active) Ecological data Human data	Identify land cover features linked with *I. scapularis* abundance and the utilization of the habitat by hosts.	Model 1: Total linear forest edge (5 km), Shannon's diversity index (10 km), Elevation Model 2: total linear forest edge (5 km), Shannon's diversity index (10 km), distance to nearest forest patch	Association between landscape and host density is likely contingent on contextual conditions and structures.	**Statistical Methods** Difference in means (estimates, p-values) **Validation Methods** Not reported. **Predictive Accuracy** Not reported
Wallace et al., 2019[29]	Hanover, New Hampshire (USA) 1990–2015	Mechanistic (Compartmental): Ordinary differential equations (ODEs)	Tick data (Proxy) Hydroclimatic data Ecological data	Develop a model to predict the effect of host densities and rising mean annual temperature on *I. scapularis* abundance and Lyme disease risk in Hanover, New Hampshire.	Mean annual/seasonal temperature (based on daily max/min temperature data)	Model assumptions for host-tick encounter probabilities are based on relative host density assumptions which may be inaccurate. Human behavior data associated with possible tick exposure was not included.	**Statistical Methods** NA **Validation Methods** Data Field observations **Predictive Accuracy** Not reported

Response variable focus: tick distribution (n = 22)

Studies	Location & time frame	Model framework (Type) & specific models	Predictor variable categories	Study objective(s)	Significant predictor(s)	Limitations	Statistical methods to assess outcomes, validation approaches & model predictive accuracy[a] (if provided)
Gardner et al., 2020[1]	Midwestern USA 1962–2011 1992–2007	Data-driven (Regression): Multivariable Survival analysis (Cox regression) Simple logistic regression	Tick data (Proxy) Ecological data Human data	Examine the impact of landscape characteristics on the spatiotemporal trends of *I. scapularis* propagation and Lyme disease prevalence in the Midwestern United States from 1967 to 2018 and forecast regions that have risk of tick invasion.	Forest coverage, river proximity, adjacency to regions with historical tick establishment (prior 5–10 years)	Some ecological variables corresponding with tick presence (e.g., forest category, altitude, soil class) and climate variables are not included. Possible bias in the model due to the use of historical data and inconsistencies in surveillance activities.	**Statistical Methods** Risk (Hazard) ratio **Validation Methods** Data AUC **Predictive Accuracy** AUC: 0.95 Sensitivity: 90.6% Specificity: 98.5%

(Table 2 continues on next page)

Response variable focus: tick distribution (n = 22)							
Studies	Location & time frame	Model framework (Type) & specific models	Predictor variable categories	Study objective(s)	Significant predictor(s)	Limitations	Statistical methods to assess outcomes, validation approaches & model predictive accuracy[a] (if provided)
(Continued from previous page)							
Clow et al., 2017a[7]	Ontario (Canada) May–October (2014, 2015)	Data-driven (Regression): Mixed-effect Multivariable logistic regression	Tick data (Active) Hydroclimatic data Ecological data	Identify key ecological and hydroclimatic factors responsible for the spread of *I. scapularis* in Ontario.	Density of understory, relative shrub abundance (and interaction), longitude, cumulative Degree-Days Above 0 °C (DD > 0 °C), tick presence at location	Temporal tick density and life stage variability impacts sensitivity and specificity. Tick density cannot be normalized, limiting the model type selection. Rudimentary measurement approaches of ecological variables may impact precision. Missing data and singular observations resulted in small sampling sizes.	**Statistical Methods** Odds ratio **Validation Methods** Not reported **Predictive Accuracy** Not reported
Soucy et al., 2018[15]	Ottawa, Ontario (Canada) 2013–2016	Data-driven (MaxEnt)	Tick data (Active, Passive) Ecological data	Create an ecological niche model for the spread of *Ixodes scapularis* in the Ottawa region using environmental (land cover, elevation) and passive tick data and develop a habitat suitability map.	Proximity to agricultural land, proximity to land with trees and hedges	Model predictions may not fully match true tick presence due to unconsidered factors like host movement impediments, seasonal fluctuations, and human utilization of and changes to habitats.	**Statistical Methods** NA **Validation Methods** AUC **Predictive Accuracy** AUC: 0.878 ± 0.019 Classification accuracy: 0.835 ± 0.020 Sensitivity: 0.956 ± 0.026 Specificity: 0.769 ± 0.028 NPV: 0.972 ± 0.015 PPV: 0.705 ± 0.026
Koffi et al., 2012[30]	Southern Quebec (Canada) 1.2007–2008 2. June–October, 2010	Data-driven (Regression): Multivariable logistic regression	Tick data (Active, Passive) Hydroclimatic data Ecological data	Determine *I. scapularis* emergence risk in Quebec using an environmental suitability index and passive tick data.	Passive tick data, invasion risk index (i.e., considering cumulative annual degree days >0 °C and adventitious tick counts from migratory birds)	The specificity of passive tick surveillance may be impacted by the presence of adventitious ticks (false positive results).	**Statistical Methods** Odds ratio **Validation Methods** ROC/AUC Field data **Predictive Accuracy** AUC: 0.816 PPV: 60.42%
Gabriele-Rivet et al., 2015[31]	New Brunswick (Canada) May–September 2014	Data-driven (Regression): Multivariable logistic regression	Tick data (Active) Hydroclimatic data Ecological data	Assess the spread of the established tick populations andstudy the environmental factors affecting their habitat in New Brunswick.	Active tick data, # years in 2009–2014 where degree-days above 0 °C (DD > 0 °C) was >2800	Temperatures in many sites were at or near the *I. scapularis* survival range. Provincial deer density estimates found to be lower than required to support the survival of *I. scapularis*.	**Statistical Methods** Odds ratio **Validation Methods** Hosmer–Lemeshow statistic test **Predictive Accuracy** ROC/AUC: 0.78 (95% CI 0.74–0.82) Sensitivity: 97.84% Specificity: 17.65% PPV: 33.8 NPV: 95

(Table 2 continues on next page)

Review

Response variable focus: tick distribution (n = 22)

Studies	Location & time frame	Model framework (Type) & specific models	Predictor variable categories	Study objective(s)	Significant predictor(s)	Limitations	Statistical methods to assess outcomes, validation approaches & model predictive accuracy[a] (if provided)
(Continued from previous page)							
Gabriele-Rivet et al., 2017[32]	Alberta, Saskatchewan, Manitoba (Canada) 2005–2015	Data-driven (Regression): Multivariable logistic regression	Tick data (Active, Proxy) Hydroclimatic data Ecological data	Create spatial *I. scapularis* establishment risk maps in the Canadian prairies using appropriate temperature, precipitation and land cover variables.	Landcover (forested and non-forested), mean annual cumulative degree days DD > 0 °C (2009–2014), rainfall.	Surveillance data inconsistencies between sampling sites hinder effective calibration of the model. Detection of adventitious ticks might lead to misclassification and erroneously low sensitivity results.	**Statistical Methods** Odds ratio **Validation Methods** AUC/ROC **Predictive Accuracy** AUC (all 6 Risk Map algorithms): 0.71–0.74 AUC (Risk Map algorithm 5): 0.74 Sensitivity (risk level >0): 83.1% Specificity (risk level >0): 50.9% NPV (risk level >0): 94.5% PPV (risk level >0): 22%
Clow et al., 2017b[33]	Ontario (Canada) May–October 2015	Data-driven (Regression): Simple logistic regression Multivariable Multinomial regression	Tick data (Active, Proxy) Hydroclimatic data Ecological data	Assess *I. scapularis* distribution, shifts in risk parameters, and if the expansion rate of ticks in Ontario aligns with the previously estimated rate (~46 km/year)[34]	Number of years to tick population establishment (based on speed of expansion.[34])	Small sample size and a short sampling period decrease statistical power and impact ability to demonstrate temporal invasion patterns. In low-density emergence areas, tick dragging is less sensitive and can lead to inaccurate results.	**Statistical Methods** Relative risk ratio **Validation Methods** Fagerland, Hosmer and Bofin goodness of fit test **Predictive Accuracy** Not reported
Sharareh et al., 2019[35]	Northeastern USA 2014	Data-driven (Regression): Multivariable Multinomial regression	Tick data (Active) Ecological data Human data	Examine the relationship between tick encounter risk and factors such as vegetation, pathways, tick and rodent host data, and human behaviours at a New York state university	Understory (%) Human risk Rodents (#)	The data collection time frame (summer), the absence of nymph tick collection during this period, and the limited sample size likely affected model outcomes.	**Statistical Methods** Difference in means (estimates, p-values) **Validation Methods** Pseudo R-square (Negelkerke method) Pseudo R-square (McFadden method) **Predictive Accuracy** 90.9% (method not stated)

(Table 2 continues on next page)

Response variable focus: tick distribution (n = 22)

(Continued from previous page)

Studies	Location & time frame	Model framework (Type) & specific models	Predictor variable categories	Study objective(s)	Significant predictor(s)	Limitations	Statistical methods to assess outcomes, validation approaches & model predictive accuracy[a] (if provided)
Kotchi et al., 2021[36]	Manitoba Ontario Quebec New Brunswick Nova Scotia (Canada) 2007–2013 2014–2018	Data-driven (Regression): Simple logistic regression Multivariable logistic regression	Tick data (Active) Hydroclimatic data Ecological data	Use remote sensing data to examine spatial and temporal trends related to tick habitat suitability and create a map of *I. scapularis* risk for eastern and central regions of Canada.	Cumulative annual surface degree days >0 °C of forest covers Percent broadleaf forest Geographical region (Province)	Low specificity of risk maps likely due to absence of identified tick populations in habitable regions. Some tick data may represent adventitious ticks rather than established populations in particular locations.	**Statistical Methods** Odds ratio **Validation Methods** Field data ROC/AUC **Predictive Accuracy** All provinces (summary): ROC AUC (SE): 0.604 (0.016) Sensitivity (%): 99.3 Specificity (%): 5.31 Individual provinces (range): ROC AUC: 0.401–0.700 Sensitivity (%): 94.4–100 Specificity (%): 0.34–9.80
Leighton et al., 2012[34]	Southern Quebec Ontario Maritime provinces (Canada) 1990–2008	Data-driven (Regression): Multivariable Survival analysis (Parametric)	Tick data (Passive) Hydroclimatic data Ecological data	Examine *I. scapularis* northward range expansion and its implications for Lyme disease by analyzing two decades of Canadian tick data, identifying factors contributing to tick spread, and projecting future distribution.	Temperature (Degree-Days Above 0 °C (DD > 0 °C) Annual Precipitation (positive) Elevation (negative)	Passive tick surveillance data may not represent actual tick burden thus the accuracy of expansion rate forecasts may be impacted.	**Statistical Methods** Risk (Hazard) ratio **Validation Methods** AUC/ROC **Predictive Accuracy** AUC: 0.897
Hahn et al., 2016[37]	Midwestern and Eastern USA 1950–2000 1980–2000	Data-driven (Hybrid): Multivariable Boosted regression tree Multivariable Generalized linear model (GLM) Multivariable adaptive spline regression MaxEnt Random forest	Tick data (Proxy) Hydroclimatic data Ecological data	Forecast *I. scapularis* geographic spread in the United States by implementing an ensemble modeling approach that incorporates climatic and ecological data along with tick data captured at the local county level.	Maximum Temperature of the Warmest Month (Bio 5) Precipitation in the Warmest Quarter (Bio 18) Percent Forest Cover in a County Elevation	Use of multiple models with differing variables, lack of structured, consistent, and complete tick surveillance data along with the use of convenience data decreases model sensitivity. Host-related variables were not incorporated in the model, and ecological and environmental averages were used which may inaccurately represent regions with highly varied environments. Utilization of historical climate data, can impact model accuracy	**Statistical Methods** NA **Validation Methods** 10-fold cross-validation AUC **Predictive Accuracy** Testing datasets–ranges for all models: AUC: 0.85–0.86 Sensitivity: 75–77% Specificity: 77–78%
Chen et al., 2015[38]	Eastern Ontario (Canada) 2006–2012	Mechanistic (Simulation): Multi-criteria decision-making model	Tick data (Passive) Hydroclimatic data Ecological data	Evaluate the effectiveness of using a habitat suitability model for predicting the distribution of *I. scapularis* in eastern Ontario.	Host distribution	Data and variable limitations (sampling bias, temporal discrepancies, spatial granularity) could potentially undermine predictive accuracy. Limitations of deer harvest data may result in an inaccurate reflection of host abundance in these regions.	**Statistical Methods** Weight sum analysis using Rank sum method **Validation Methods** Not reported **Predictive Accuracy** Not reported

(Table 2 continues on next page)

Response variable focus: tick distribution (n = 22)

Studies	Location & time frame	Model framework (Type) & specific models	Predictor variable categories	Study objective(s)	Significant predictor(s)	Limitations	Statistical methods to assess outcomes, validation approaches & model predictive accuracy[a] (if provided)
(Continued from previous page)							
McPherson et al., 2017[39]	Northern Ontario Southern Ontario Nova Scotia (Canada) 1991–2008	Mechanistic (Simulation): General circulation model	Tick data (Proxy) Hydroclimatic data Ecological data	Investigate the uncertainties in predicted estimates of the basic reproduction number (R0) for *I. scapularis* by evaluating the impacts of various scenarios from the Representative Concentration Pathway (RCP) and climate model outputs.	Degree-Days Above 0 °C (DD > 0 °C)	Parameter variations and assumptions may lead to model prediction uncertainties. Insufficient resolution at the local scale is a limitation of the Global Climate Models (GCMs).	**Statistical Methods** Kolmogorov–Smirnov test **Validation Methods** Not reported **Predictive Accuracy** Not reported
Ogden et al., 2014[40]	Ontario, Quebec (Canada) Northeast and upper Midwest USA 1971–2010	Mechanistic (Compartmental): Ordinary differential equations (ODEs)	Tick data (Proxy) Hydroclimatic data Ecological data	Evaluate how the basic reproduction number (R0) of *I. scapularis* may be impacted by climate change.	Temperature	Limitation not given.	**Statistical Methods** NA **Validation Methods** Observed data **Predictive Accuracy** Not reported
Wu et al., 2013[41]	Ontario, Quebec (Canada) 1971–2000	Mechanistic (Compartmental): Periodic ordinary differential equations (ODEs)	Tick data (Proxy) Hydroclimatic data Ecological data	Determine the basic reproduction number (R0) by integrating climatic variability into a deterministic mathematical model for *I. scapularis* population distribution.	Temperature	Excludes random effects, owing to insufficient information for proper parameterization.	**Statistical Methods** NA **Validation Methods** Literature (Previous model) **Predictive Accuracy** Not reported
Cheng et al., 2017[42]	Eastern Ontario (Canada) 2000–2013	Mechanistic (Compartmental): Periodic ordinary differential equations (ODEs)	Tick data (Passive) Hydroclimatic data	Employ the basic reproduction number (R0) as a metric to investigate the impact of climate change on the dynamics of *I. scapularis* populations in the eastern region of Ontario.	Temperature	Model neglects host movement, and assumes that host density remains steady which may not reflect real-world dynamics. Tick data collected by health professionals may be impacted by reporting biases. Challenges in directly measuring the habitat range of *I. scapularis* prompt the use of proxies like sustainable climates.	**Statistical Methods** NA **Validation Methods** Field data **Predictive Accuracy** Not reported
Ogden et al., 2013[43]	Eastern and Central Canada 1991–2010	Mechanistic (Compartmental): (Susceptible-Infected) Ordinary differential equations (ODE)	Tick data (Passive) Hydroclimatic data Ecological data	Examine the dynamics and rate of *Borrelia burgdorferi* invasion subsequent to the establishmentof *Ixodes scapularis* ticks, investigate the associated time gap, analyze surveillance data for clusters with low prevalence of infection, and conduct simulations to identify influencing factors.	Host (Deer and *P. leucopus*) abundance	Model struggles with accurately simulating the establishment of tick populations, focusing on a singular scenario, that may not capture variability observed in the actual mergence zone.	**Statistical Methods** NA **Validation Methods** Observed data **Predictive Accuracy** Not reported
Illoldi-Rangel et al., 2012[44]	Mexico, Texas (USA) 2011	Data-driven (MaxEnt)	Tick data (Proxy) Hydroclimatic data Ecological data	Develop species distribution models for tick species (including *I. scapularis*) in Texas and Mexico using MaxEnt.	Low temperature High altitude Pine and oak forest	No limitation reported.	**Statistical Methods** NA **Validation Methods** AUC/ROC **Predictive Accuracy** AUC: 0.93
Slatculescu et al., 2020[45]	Ontario (Canada) 2015–2018	Data-driven (MaxEnt)	Tick data (Active, Passive) Ecological data Hydroclimatic data	Utilize Maxent to project *I. scapularis* and *Borrelia burgdorferi* distribution in the south-east region of Ontario and discern the underlying influential factors on Lyme disease risk.	Distance to coniferous or deciduous forest Elevation Annual cumulative degree days >0 °C	Absence of essential ecological variables from the model and sampling biases and limitations related to tick collection activities may have impacted results.	**Statistical Methods** NA **Validation Methods** 4-fold cross-validation AUC **Predictive Accuracy** AUC: 0.898 (Tick dragging) AUC: 0.727 (Public sources ticks)

(Table 2 continues on next page)

Response variable focus: tick distribution (n = 22)

Studies	Location & time frame	Model framework (Type) & specific models	Predictor variable categories	Study objective(s)	Significant predictor(s)	Limitations	Statistical methods to assess outcomes, validation approaches & model predictive accuracy[a] (if provided)
(Continued from previous page)							
Johnson et al., 2016[46]	Minnesota (USA) 2005–2014	Data-driven (MaxEnt):	Tick data (Active) Hydroclimatic data Ecological data	Construct a detailed subcounty-level distribution model for *I. scapularis* in Minnesota to determine the areas conducive to tick establishment, thus enhancing awareness of the geographic expansion of tick populations in the region.	Land cover Maximum temperature during the warmest month Precipitation of the wettest quarter Annual temperature range.	No limitation reported.	**Statistical Methods** NA **Validation Methods** 10-fold cross-validation AUC/ROC **Predictive Accuracy** AUC: 0.863
Feria-Arroyo et al., 2014[47]	Texas (USA) Tamaulipas, Nuevo León, and Coahuila (Mexico) 2011–2012	Data-driven (MaxEnt)	Tick data (Passive) Hydroclimatic data	Develop a robust distribution model using MaxEnt for *I. scapularis* in the non-endemic area of the U.S.-Mexico border, incorporating climate and habitat factors.	Isothermality Precipitation of the wettest quarter Maximum temp. of the warmest month Precipitation observed in the wettest month	The absence of fine spatial and geographical coordinates or accurate species identification in tick data is a challenge.	**Statistical Methods** NA **Validation Methods** AUC/ROC **Predictive Accuracy** AUC: 0.831
Zhang et al., 2022[48]	USA June 2021	Data-driven (MaxEnt)	Tick data (Proxy) Hydroclimatic data Ecological data	Forecast the possible global distribution of *I. scapularis* and identify key environmental factors in relation to potential climate change circumstances by using a MaxEnt model that incorporates existing distribution sites and environmental variables.	Precipitation in May (Prec 05) Precipitation in September (Prec 09) Precipitation of the driest month (Bio 14) Temperature seasonality (Bio 4) and Mean diurnal range (Bio 2)	The model does not incorporate soil characteristics, vegetation, nor the distribution of primary hosts.	**Statistical Methods** NA **Validation Methods** AUC/ROC **Predictive Accuracy** AUC: 0.950

Response variable focus: tick distribution and tick abundance (n = 5)

Studies	Location & time frame	Model framework (Type) & specific models	Predictor variable categories used	Study objective(s)	Significant predictor(s)	Limitations & model accuracy (if applicable)	Statistical methods to assess outcomes, validation approaches & model predictive accuracy[a] (if provided)
Dhingra et al., 2013[49]	Eastern region, USA 2001–2004	Mechanistic (Compartmental): Ordinary differential equations (ODEs) Dynamic population features modelling.	Tick data (Proxy) Hydroclimatic data Human data	Examine the dynamics of each *I. scapularis* life stage under present and future climate scenarios in the eastern region of the United States.	Temperature	Model did not incorporate host dynamics or key environmental variables (i.e., humidity). The response dynamics of *I. scapularis* populations vary spatially and across tick life stages.	**Statistical Methods** NA **Validation Methods** Spearman correlation coefficients AUC **Predictive Accuracy** AUC for Questing Adults (QA): 0.53–0.71 AUC for Questing Nymphs (QN): 0.54–0.69. AUC for Questing Larvae (QL): 0.52–0.69. AUC for QL (Wave angle and presence of ticks): 0.66–0.70

(Table 2 continues on next page)

Review

Response variable focus: tick distribution and tick abundance (n = 5)

Studies	Location & time frame	Model framework (Type) & specific models	Predictor variable categories used	Study objective(s)	Significant predictor(s)	Limitations & model accuracy (if applicable)	Statistical methods to assess outcomes, validation approaches & model predictive accuracy[a] (if provided)
(Continued from previous page)							
Johnson et al., 2018[50]	Minnesota (USA) May 31–June 30, 2015	Data-driven (Hybrid): Multivariable Zero-inflated negative binomial regression MaxEnt	Tick data (Active) Hydroclimatic data Ecological data	Predict *I. scapularis* nymph density in Minnesota using a model for acarological risk and assess tick habitat suitability in the region.	Mean diurnal temperature range. Elevation Annual temperature range Summer precipitation Agricultural land	The absence of tick data in far northern counties hampers predictive model accuracy. Use of a single year of tick sampling data neglects to account for multi-year tick population dynamics and may impact predictions.	**Statistical Methods** Difference in means (estimates, p-values) **Validation Methods** Scaled Pearson's residuals 5-fold cross-validation Mean absolute prediction error (MAPE) **Predictive Accuracy** Model forecasted nymph presence with 79% accuracy.
Ripoche et al., 2018b[51]	Montérégie, Southern Quebec (Canada) May–September 2013 May–September 2014	Data-driven (Regression): Multivariable Poisson/Negative binomial regression Multivariable Mixed-effect Poisson Regression	Tick data (Active) Hydroclimatic data Ecological data	Study the dispersal of *I. scapularis* nymphs in regions of southern Quebec that may be at risk for Lyme disease emergence by examining and comparing tick densities in wooded locations and diverse outdoor path proximities at various spatial scales.	Regional scale: Litter depth (abundance increases), Elevation (abundance decreases) Local scale: Relative humidity, proximity to trails	Depending solely on active surveillance with limited sampling may result in an incomplete understanding of tick distribution. Study did not consider host variables. Generalizing results to larger areas must be done carefully due to the small sample size.	**Statistical Methods** Difference in means (estimates, p-values) **Validation Methods** McFadden's Pseudo R-squared Mean absolute error (MAE) **Predictive Accuracy** MAE (Site scale): 11.26 MAE (Plot/Transect scale): 0.95
Simon et al., 2014[52]	Southern Québec (Canada) May–October 2011	Data-driven (Regression): Multivariable logistic regression Multivariable Generalized linear model (GLM) with negative binomial regression	Tick data (Active) Hydroclimatic data Ecological data	Model present and forecast future *I. scapularis* and rodent host distributions in southern Quebec to inform Lyme disease risk predictions through the use of climate and landscape models.	Landscape variables (including measures of patch area, connectivity, and land use) Temperature	Factors such as humidity, rainfall, and elevation, are not included in the climate model for ticks. Solely using temperature may lead to overestimation of tick habitat suitability ranges. Hosts aside from the white-footed mouse may also be important factors, and other ecological and behavioural factors may also impact tick and rodent distribution, but these are not incorporated in the model.	**Statistical Methods** Spearman's rank correlation Difference in means (estimates, p-values) **Validation Methods** Spearman's rank correlation **Predictive Accuracy** R^2 (Principal Component Analysis–first axis): 0.24
Burrows et al., 2022[53]	Kingston, Ontario. Ottawa, Ontario. Southern Ontario (Canada) May, June 2019	Data-driven (Hybrid): Simple Mixed-effect Negative binomial regression MaxEnt	Tick data (Active) Ecological data	Examine MaxEnt model capability in estimating present and future patterns of tick abundance and risk of Lyme disease in three Ontario districts.	Forest land cover	Using tick data from one year may limit the comprehensive understanding of long-term tick density patterns. The small sample size must be considered when interpreting results.	**Statistical Methods** Incidence Rate Ratio **Validation Methods** Spearman's correlation coefficient **Predictive Accuracy** Not reported.

Note that not all statistical methods reported and utilized in each study are included in this table. Here we include only those associated with tick density and distribution model validation and prediction (if reported). [a]Predictive accuracy values are associated with the model that best optimizes the validation methods reported.

Table 2: **Summary of key characteristics of included studies (n = 41) reported by authors associated with *I. scapularis* density and distribution.**

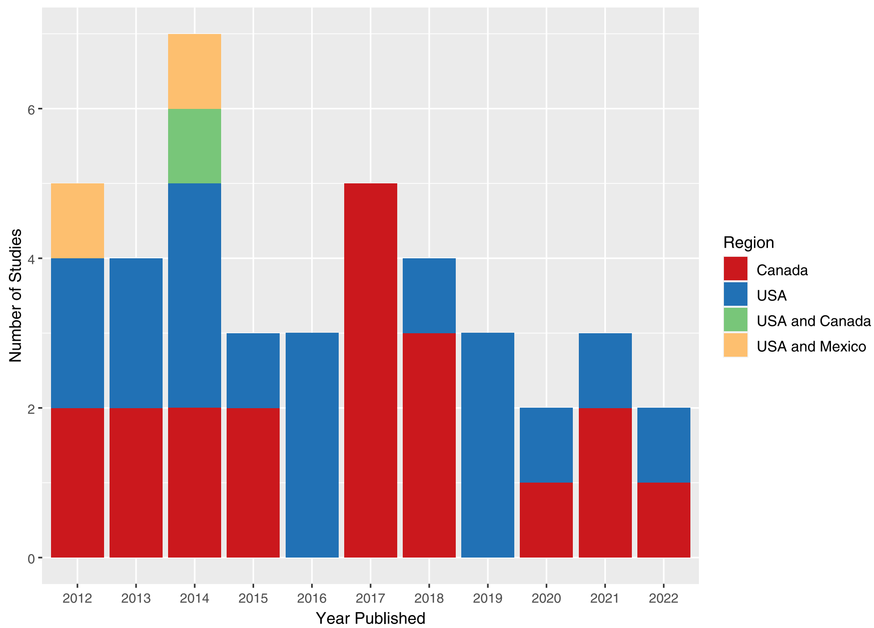

Fig 3: Timeline of included studies from the literature in North America.

14.3%),[36] multivariable logistic (n = 3, 42.8%),[30–32] mixed-effect multivariable logistic (n = 1, 14.3%)[7] and multivariable multinomial regression (n = 1,14.3%)[35] exploring diverse outcome categories (higher and lower risk exposure to infected ticks). Multivariable multinomial and simple LgR were utilized (n = 1, 14.3%), examining factors influencing outcomes and assessing the impact of individual ecological factors on the speed of *I. scapularis* spread (Fig. 6).[33]

Mechanistic models focusing on tick distribution (n = 6, 100%) utilized various approaches. A multi-criteria decision-making model (n = 1, 16.7%) explored the relationship between deer habitat and tick populations, identifying seven different factors that could affect deer

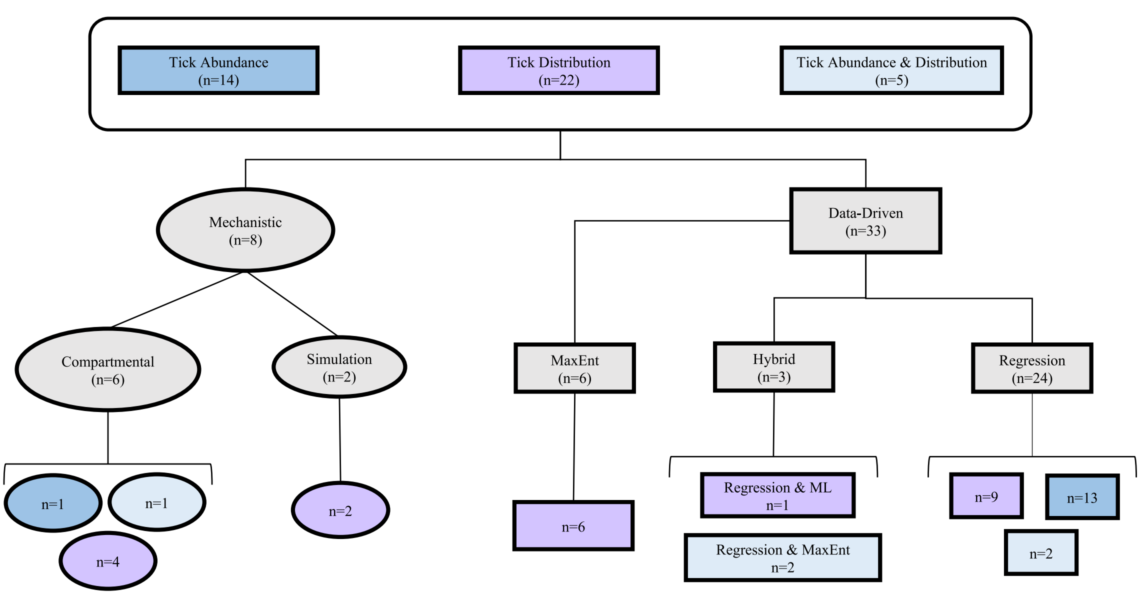

Fig 4: Tree diagram representing the hierarchical structure of the modelling frameworks and general model types used in the 41 included studies according to modelling output/response variable.

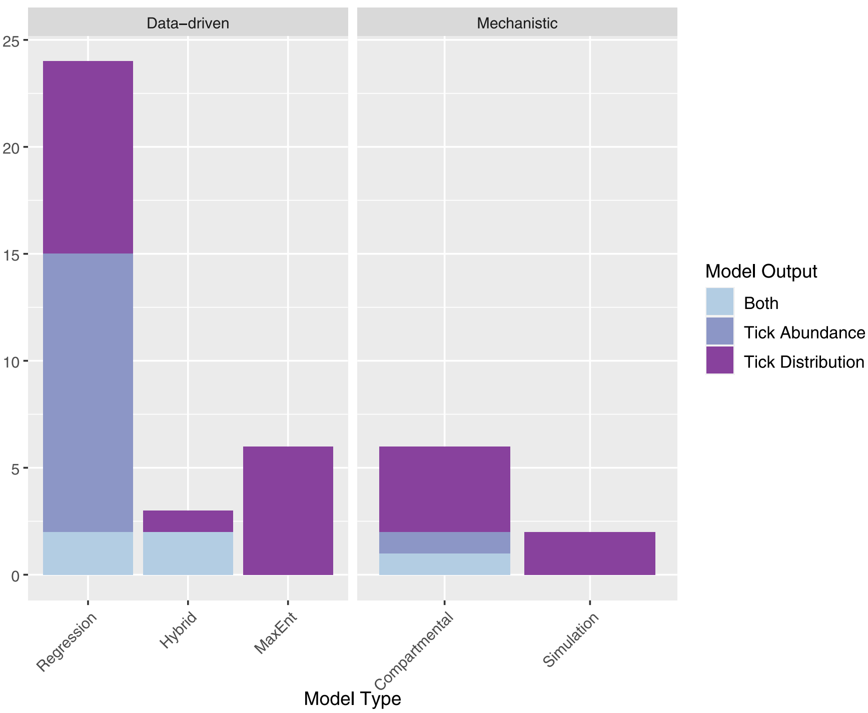

Fig 5: (a) Total number of included studies of all model types according to the resulting model output; (b) Total number of included studies of all model types according to the overarching model framework and associated model output.

habitat. An optimization technique, weighted sum analysis, was applied to combine these factors into a final habitat suitability mapping (Appendix D, Figure D2).[38] A General Circulation Model (GCM) estimated tick basic reproduction numbers (R0) under historical (1850–2005) and future (2006–2100) climate scenario simulations (n = 1, 16.7%).[39] Periodic ODEs were used to model tick population dynamics (e.g., development, questing activity) with the integration of climate factors and calculate R0 (n = 3, 50%) (Appendix D, Figures D3).[40–42] An ODE population model structure (n = 1, 16.6%) incorporated two susceptible-infected host sub-models for the white-footed mouse (*P. leucopus*) and an alternative mammal host (Appendix D, Figure D4).[43]

For both tick distribution and abundance, data-driven (n = 4, 80%) and mechanistic approaches (n = 1, 20%) were used. Among the data-driven models, studies utilized MaxEnt with multivariable zero-inflated and simple mixed-effect NB regression (n = 2, 50%).[50,53] One study employed multivariable mixed-effect Poisson and NB models (n = 1, 25%)[54] while another utilized multivariable LgR and multivariable GLM with a NB model, addressing distribution and abundance aspects (n = 1, 25%).[52] Furthermore, a dynamic temperature forced population modelling approach assessed temperature impacts of climate change on *I. scapularis* life cycle and population abundance across diverse regions (n = 1, 100%).[49]

All studies (n = 41) used tick data collected across a wide range of sampling areas, sources, and methods: active (n = 17, 41.4%), passive (n = 6, 14.6%), active and passive (n = 4, 9.7%), active and proxy (n = 3, 7.3%), and proxy (n = 11, 26.8%). The temporal scale ranged from months to several years or decades (Table 2). Hydroclimatic and ecological factors were included as predictor variables in many studies (Table 3, Appendix B, Fig. 7). Studies that used both hydroclimatic and ecological factors (n = 26, 63.4%) recognised their interconnected influence on tick dynamics. Surprisingly, only nine studies (21.9%) included host (deer) density (Fig. 7).

Additionally, 29 hydroclimatic and ecological databases were identified. WorldClim (n = 6, 20.7%) and the National Land Cover Database (n = 4, 13.8%) were commonly used for hydroclimatic and ecological data, respectively. Among these sources, the spatial resolution ranged from 1.82 m to 32 km, while the temporal scale varied from daily, weekly, and monthly values to long-term mean values (Appendix F).

Model accuracy was reported by nearly two-thirds of all included studies (n = 26, 63.4%) (Appendix E). Of these, accuracy was most frequently reported from studies incorporating both hydroclimatic and ecological predictor categories (n = 17, 65.4%) compared to studies using either predictor category alone (n = 3, 11.5%) or with human variables (n = 4; 15.4%). Comparison of

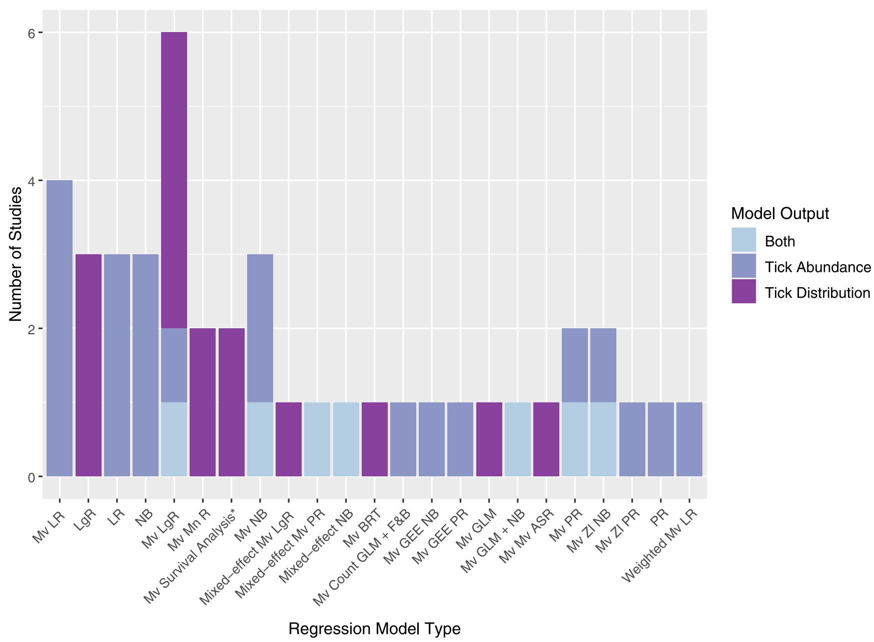

Fig 6: Types of regression models employed in the included studies according to the model output/response variable. Abbreviations: Simple logistic regression (LgR); Simple linear regression (LR); Simple mixed-effect NB regression (Mixed-effect NB); Mixed-effect multivariable logistic regression (Mmv LgR); Multivariable boosted regression tree (Mv BRT); Multivariable Count Generalized Linear Model with forward and backward stepwise selection (Mv Count GLM + F & B); Multivariable Generalized Estimating Equation negative binomial regression (Mv GEE NB); Multivariable Generalized Estimating Equation (GEE) Poisson regression (Mv GEE PR); Multivariable Generalized Linear Model (Mv GLM); Multivariable GLM with NB (Mv GLM + NB); Multivariable logistic regression (Mv LgR); Multivariable linear regression (Mv LR); Mixed-effect Multivariable Poisson regression (Mixed-effect Mv PR); Multivariable multinomial regression (Mv Mn R); Multivariable multivariate adaptive spline regression (Mv Mv ASR); Multivariable NB regression (Mv NB); Multivariable Poisson regression (Mv PR); Multivariable zero-inflated negative binomial regression (Mv ZI NB); Multivariable zero-inflated Poisson regression (Mv ZI PR); Simple negative binomial regression (NB); Simple Poisson regression (PR); Weighted multivariable linear regression (Weighted Mv LR); Multivariable survival analysis (Mv Survival Analysis*). *includes parametric (n = 1) and Cox regression (n = 1) methods.

these measures across studies is challenging, as results depend on data and models used, along with sample size limitations. It is also difficult to identify clear differences between variables used across model outputs (Fig. 7, Table 3).

Discussion

Our study represents a comprehensive assessment of the methodological approaches, variables, and data sources employed for *I. scapularis* population modelling in North America. While the scoping review by Kopsco and colleagues[11] examined species distribution models for medically significant ticks globally, our review extends and enriches the scope by prioritizing predictive modelling approaches for *I. scapularis* abundance and distribution in North America resulting in 41 included studies that utilize data-driven and mechanistic models. A summary of the features of identified model types is provided in Table 4.

Data-driven models identified key predictors associated with *I. scapularis* population dynamics (Tables 2 and 3), though complex relationships and contextual variations between these predictors are evident. For example, suitable tick habitats were predicted at higher altitudes and cooler temperatures in regions with consistently high temperatures (e.g., Mexico)[44] and at lower altitudes and warming temperatures in cool regions (e.g., Ontario).[45] Notably, one study found no significant association when a geographic area had suboptimal conditions (e.g., low temperature and host abundance),[31] while another showed no positive correlation between forest fragmentation and *I. scapularis* abundance despite high tick availability.[28] Absence of positive correlations between tick data and predictor variables may also be due to small sample sizes.[28,31]

Factors such as forest cover, habitat fragmentation, temperature and precipitation affect tick survival and host abundance.[5,60] Interestingly, despite the important role of white-tailed deer in the tick life cycle, host density was not included in many tick distribution studies. Lack of tick-host interaction data in models may inaccurately predict tick distribution by underestimating tick life stages.[10]

Significant predictor variables	Model output: Tick abundance	Tick distribution	Tick abundance & distribution
Ecological	**Landscape** (Finch et al., 2014[19]; Khatchikian et al., 2012[21]; Diuk-Wasser et al., 2012[22]; Werden et al., 2014[27]; Ferrell & Brinkerhoff, 2018[28]) **Vegetation** (Finch et al., 2014[19]; Khatchikian et al., 2012[21]; Diuk-Wasser et al., 2012[22]; Berger et al., 2013[26]; Werden et al., 2014[27]; Ferrell & Brinkerhoff, 2018[28]) **Location** (Khatchikian et al., 2012[21]; Brinkerhoff et al., 2014[24]) **Host density** (Werden et al., 2014[27])	**Landscape** (Gardner et al., 2020[1]; Soucy et al., 2018[15]; Leighton et al., 2012[34]; Hahn et al. 2016[37]; Illoldi-Rangel et al., 2012[44]; Slatculescu et al., 2020[45]) **Vegetation** (Gardner et al., 2020[1]; Clow et al., 2017a[7]; Soucy et al., 2018[15]; Gabriele-Rivet et al., 2017[32]; Sharareh et al., 2019[35]; Kotchi et al., 2021[36]; Hahn et al. 2016[37]; Illoldi-Rangel et al., 2012[44]; Slatculescu et al., 2020[45]; Johnson et al., 2016[46]) **Location** (Gardner et al., 2020[1]; Clow et al., 2017a[7]; Kotchi et al., 2021[36]) **Host density** (Sharareh et al., 2019[35]; Chen et al., 2015[38]; Ogden et al., 2013[43])	**Landscape** (Johnson et al., 2018[50]; Ripoche et al., 2018b[51]; Simon et al., 2014[52]) **Vegetation** (Johnson et al., 2018[50]; Ripoche et al., 2018b[51]; Burrow et al., 2022[52])
Hydroclimatic	**Temperature** (Khatchikian et al., 2012[21]; Diuk-Wasser et al., 2012[22]; Berger et al., 2013[26]; Werden et al., 2014[27]; Wallace et al., 2019[29]) **Precipitation** (Khatchikian et al., 2012[21]) **Humidity** (Diuk-Wasser et al., 2012[22]; Berger et al., 2014[23])	**Temperature** (Clow et al., 2017a[7]; Koffi et al., 2012[30]; Gabriele-Rivet et al., 2015[31]; Gabriele-Rivet et al., 2017[32]; Kotchi et al., 2021[36]; Leighton et al., 2012[34]; Hahn et al. 2016[37]; McPherson et al., 2017[39]; Ogden et al., 2014[40]; Wu et al., 2013[41]; Cheng et al., 2017[42]; Illoldi-Rangel et al., 2012[44]; Slatculescu et al., 2020[45]; Johnson et al., 2016[46]; Feria-Arroyo et al., 2014[47]; Zhang et al., 2022[48]) **Precipitation** (Gabriele-Rivet et al., 2017[32]; Leighton et al., 2012[34]; Hahn et al. 2016[37]; Johnson et al., 2016[46]; Feria-Arroyo et al., 2014[47]; Zhang et al., 2022[48])	**Temperature** (Dhingra et al., 2013[49]; Johnson et al., 2018[50]; Simon et al., 2014[52]) **Precipitation** (Johnson et al., 2018[50]) **Humidity** (Ripoche et al., 2018b[51])
Tick data	**Tick data** (Diuk-Wasser et al., 2012[22]; Ripoche et al., 2018a[54]; Tran et al., 2021[25])	**Tick Data** (Clow et al., 2017a[7]; Koffi et al., 2012[30]; Gabriele-Rivet et al., 2015[31]; Clow et al., 2017b[33])	
Human data	**Human behaviour** (Porter et al., 2019[18]; Tran et al., 2021[25])	**Human behaviour** (Sharareh et al., 2019[35]; Leighton et al., 2012[36])	

Note that these are dependent on the full set of predictor variables and model type used in each study. See Appendix B for specific details captured for each study.

Table 3: **General summary of significant predictor variable categories identified in each study.**

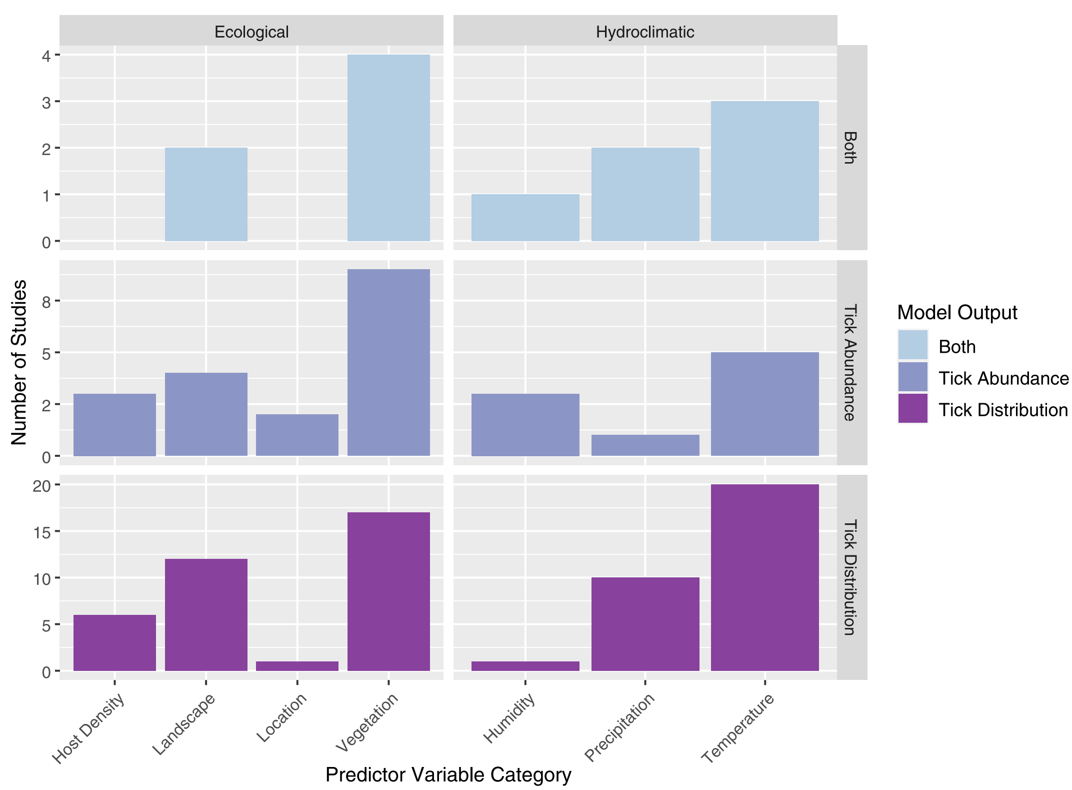

Fig 7: Hydroclimatic and ecological predictor variable categories used across all studies to model tick distribution, tick abundance, and both tick distribution and abundance. Specific predictor variables used in each study are found in Appendix B.

Model framework	Model framework description	Model type	Model type description	Strengths	Limitations
Data-driven	Use patterns and information from data to make predictions through computational algorithms.[8]	**Regression**	Statistical approach to model the relationship among variables.[33]	Simple and easy to interpret, providing a clear understanding of the impact of predictors on the outcome.[9] Offers statistical inference, allowing the testing of hypotheses and assessing the significance of predictors.[55]	Model accuracy may be compromised due to violation of assumptions.[56] Need sufficient sample size for reliable inference and prediction.[28,33,35,53]
		Maximum entropy (MaxEnt)	Machine learning technique that maximizes entropy.[8]	Flexible and can handle various types of data, including presence-only data, which is often the case in ecological studies.[44,48] Can incorporate multiple variables to predict species distributions, allowing for a more comprehensive understanding of the factors influencing the presence of a species.[8]	Sensitive to sample size and sampling bias.[15,45] Lack of mechanistic understanding by the model may lead to inaccurate representation of complex ecological processes and interactions.[8,57] Using presence-only data can introduce bias and affect predictive accuracy because of the model assumption that absence data is missing at random, which may not always be the case.[15]
Mechanistic	Based on underlying principles and mechanisms governing a system.[9]	**Compartmental**	Partitions a system into distinct compartments and monitors the flow between entities.[9]	Based on established theories and principles, providing a deeper understanding of the underlying mechanisms.[8,9] More robust to changes in underlying patterns, as these models are based on fundamental principles rather than specific data patterns[58] Adaptability of model to various tick species globally.[41]	Accuracy heavily depends on the correctness of underlying assumptions and known principles.[8,9,42] Adequate parameterization demands extensive, accurate data availability.[57] Challenges arise in interpreting complex models with numerous variables and interactions.[9]
		Simulation	Modelling a real-world system to examine and understand its behavior.[58]	Provide decision support by offering valuable insights into the potential consequences of various actions or interventions within a given system.[38] Cost-effective approach for evaluating strategies and scenarios within a controlled environment prior to real-world implementation.[38]	Developing and running simulations can be time-consuming and demand significant computational resources.[59] Model calibration can be challenging when dealing with complex systems.[57] Adequate parameterization demands extensive, accurate data availability.[39]

Table 4: **Summary of strengths and limitations of modelling techniques.**

MaxEnt models encounter challenges when predicting *I. scapularis* distribution in areas where they are not well established (e.g., northeastern and midwestern US, Canada)[46,50] or have low tick risk (e.g., Mexico).[44,47] Scarce occurrence data hinder accurate model training, and may result in unreliable predictions due to insufficient species distribution information.[47] Using presence-only data and entropy maximization may overestimate species occurrence probability in environmentally suitable areas, indicating upper bounds on abundance rather than actual densities. To assess the model effectively, spatial context should be considered and applied to regions anticipating tick presence.[61]

Alternatively, mechanistic models simulate interactions between factors impacting tick distribution and abundance and provide broader insights than data-driven models.[42,62] Their accuracy requires well-founded assumptions about species distribution drivers.[43] Using ordinary differential equations (ODEs), some mechanistic models calculate R0 to estimate questing activity, habitat suitability under climate variability, and vector population simulations[29,39,41–43,49] (Appendix D).

The highest R0 values in Canada were found in Lyme disease endemic areas.[41] R0 can demonstrate potential impacts of climate change on TBD transmission as increased temperature can expand suitable tick habitat ranges northward,[39] driving *I. scapularis* spread and establishment.[40,42] Tick development and reproduction are viable between 5 and 32 °C, however the optimal temperature range for *I. scapularis* is suggested to be 15–20 °C.[37,62,63] Some TBD risk measures show varying degrees of sensitivity to temperature changes[29] suggesting other influences. For example, infected tick abundance increases significantly with warming, while overall tick or host disease incidence increases relatively slowly, even during peak transmission.[29] Interestingly, some mechanistic and regression models suggest a time gap between tick invasion and TBD increase.[34,43]

While all studies included tick data, collection methods varied between active or passive surveillance and proxy data. Active tick surveillance data provides high geographical precision and rich information. Nonetheless, the cost and labour-intensive nature of this approach is restrictive, with data accuracy, completeness, and availability challenges. Data collection occurs over narrow time frames, limiting the ability to capture the complete 2–3-year tick life cycle, thus underestimating tick population abundance.[64] Conversely, passive surveillance requires fewer resources and provides rich data through continuous monitoring.[16,65] However, feasibility depends on large human populations willing to search for and submit ticks. Additionally, tick introduction by migratory birds (i.e., "adventitious" ticks)[66] yields low sensitivity for detecting emerging tick populations and low specificity for detecting areas with established tick populations, potentially leading to false positives.[60] Therefore, models must include both active[18] and passive tick surveillance to validate model predictions, leading to greater model robustness and accuracy; however they must control sampling bias especially for large scale data.[50] We suggest model repeatability will be improved.

Citizen science tick monitoring is gaining popularity due to its cost-effectiveness and community engagement.[18] However, this approach may underestimate human exposure to tick larvae and nymphs which are more difficult to detect than adult ticks. Training tick collectors to gather data correctly can enhance estimates of tick populations.[18]

Tick data availability and quality are critical for accurate predictive model development. Traditional regression models require large and balanced presence-absence datasets, while MaxEnt models can aptly handle sparse or incomplete data (e.g., presence-only data).[67–69] Both MaxEnt and LgR models provide habitat suitability probabilities rather than direct estimates of tick abundance, underscoring the importance of accurate and available tick data on model performance.[61–63] Mechanistic models capture the complex interplay between biotic and abiotic factors on tick population, thus highlighting population abundance trends and variability.

The WorldClim database[21,37,47] was used by many studies to inform tick habitat suitability assessments, but data accuracy varies by source, collection, and processing methods. The spatial resolution (1 km × 1 km) may be inferior to other finer resolution remote sensing data (e.g., temperature data <1 km) that can improve predictions and understand local factors impacting *I. scapularis* distribution.[42] Uncertainties in environmental data can hinder risk map accuracy,[36] so to avoid modelling bias, all relevant variables must be incorporated.

We identified some similar results to those of Kopsco and collaborators.[11] Both studies showed MaxEnt as a common approach for tick distribution models, and the WorldClim database as a common source for ecological and hydroclimatic data. However, while Kopsco and colleagues[11] focused only on studies using active or passive surveillance data, our scoping review included studies using citizen science tick data. Additionally, their models were purely data-driven,[11] whereas ours included both data-driven and mechanistic models. Our findings also reveal that host density data is rarely incorporated in predictive models, highlighting a unique and important contribution to the existing literature.

We believe our comprehensive review enhances the knowledge of *I. scapularis* population modelling in North America. However, we acknowledge some methodological caveats and limitations. Relevant analyses may have been omitted due to our exclusion of review articles, non-English publications, and studies focusing primarily on Lyme disease risk. In fact, while assessing the scoping review,[11] we identified an unintentional exclusion that occurred during abstract screening.[70] We believe this occurred because the study focused on Lyme disease risk using canine seroprevalence rather than tick distribution or abundance and human Lyme disease risk; however, an *I. scapularis* habitat suitability model was also developed. Additionally, as our review includes studies published between January 2012 and July 2022, any research published after the database search is excluded. For a wider perspective, future investigations should assess predictive methods and variables associated with other prevalent tick species and compare their consistency or variability with our results. Limitations identified in the included studies also provide opportunities for action. The frequency and coverage of *I. scapularis* and TBD surveillance activities must be expanded and standardized data collection protocols must be implemented to enhance the comprehensiveness, quality, and credibility of the data required for predictive model refinement and development. Furthermore, the use of appropriate and advanced modelling technologies and the inclusion of hydroclimatic and ecological data with finer scale resolution can contribute to robust modelling approaches.[36]

A clear gap is the absence of host population characteristics (e.g., deer density, small mammal abundance) and human factors[30,36] in model development. Additionally, climate change impacts on tick habitats and *I. scapularis* distribution patterns require comprehensive investigation. Incorporating fine-scale ecological variables can significantly enhance model predictive accuracy amid climate change, prioritizing tick–host relationships within their ecological contexts.

Future research efforts focusing on *I. scapularis* risk assessment and predictive modelling must adopt comprehensive One Health approaches that integrate acarological risk factors, biotic (host) and human factors, alongside climatic and ecological variables into dynamic models for meaningful and accurate public health responses.[31,36]

This review highlights the current complex relationship between abiotic and biotic factors affecting *I. scapularis* abundance and distribution in North America, highlighting significant predictor variables and tick population modelling methods. However, these results should be extrapolated cautiously to other tick species with consideration of ecological specificity. Overall, our results provide a first step in variable and method selection to inform future development of tick abundance and distribution prediction models.

Conclusion

Advances in modelling and data sources have significantly improved predictive capabilities. However, data quality and availability challenges continue to persist and impact model development. Furthermore, as data dimensions continue to grow, the delicate balance between predictive power and explainability adds complexity for both modelers and practitioners. Standardized protocols for running and validating predictive models are critical given the diverse approaches to estimating tick distributions and abundance.

To better understand the distribution and abundance of ticks and tick-borne pathogens, data challenges must be overcome, and standardized methodologies should be implemented. The integration of passive and active surveillance, hydroclimatic and ecological factors—especially host abundance–must occur to improve model reliability and predictive accuracy given TBD risk complexity. Overall, this scoping review highlights the urgency for proactive and timely action to protect population health and mitigate future climate-associated risk.

Contributors

BN, YS, and EL conceptualized and designed this study. YS and EL carried out the literature review, collected and analysed the data. Further analysis input was given by BN. The initial manuscript was written by YS and EL with feedback and supervision from BN in collaboration with JM and JK. Further edits and revisions were carried out by TP, YS, and EL following review and feedback from BN, SG, JM, JL, JK, and FH. Data visualization was completed by YS, EL, FH, and TP. The study and the revision process were mainly supervised by BN.

Declaration of interests

The authors declare no competing interests.

Acknowledgements

This study is funded by the Natural Sciences and Engineering Research Council of Canada (NSERC) Emerging Infectious Diseases Modelling Initiative (EIDM), awarded to the Mathematics for Public Health (MfPH) program led by the Fields Institute. The funder source had no role in the study design, data collection, analysis, interpretation, or writing of the report. The authors would also like to acknowledge the reviewers for their helpful comments and suggestions. Their input greatly improved the quality of this manuscript.

Appendix A. Supplementary data

Supplementary data related to this article can be found at https://doi.org/10.1016/j.lana.2024.100706.

References

1 Gardner AM, Pawlikowski NC, Hamer SA, et al. Landscape features predict the current and forecast the future geographic spread of Lyme disease. *Proc Biol Sci*. 2020;287(1941):20202278.
2 Jones KE, Patel NG, Levy MA, et al. Global trends in emerging infectious diseases. *Nature*. 2008;451(7181):990–993.
3 Couper LI, MacDonald AJ, Mordecai EA. Impact of prior and projected climate change on US Lyme disease incidence. *Glob Change Biol*. 2021;27(4):738–754.
4 Kim D, Maxwell S, Le Q. Spatial and temporal comparison of perceived risks and confirmed cases of lyme disease: an exploratory study of google trends. *Front Public Health*. 2020;8:395. [cited 2022 Aug 22] Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7456861/.
5 Ogden NH, Pang G, Ginsberg HS, et al. Evidence for geographic variation in life-cycle processes affecting phenology of the lyme disease vector Ixodes scapularis (Acari: Ixodidae) in the United States. *J Med Entomol*. 2018;55(6):1386–1401.
6 Embers ME, Grasperge BJ, Jacobs MB, Philipp MT. Feeding of ticks on animals for transmission and xenodiagnosis in Lyme disease research. *J Vis Exp*. 2013;78:50617.
7 Clow KM, Ogden NH, Lindsay LR, Michel P, Pearl DL, Jardine CM. The influence of abiotic and biotic factors on the invasion of Ixodes scapularis in Ontario, Canada. *Ticks Tick Borne Dis*. 2017;8(4):554–563.
8 Beery S, Cole E, Parker J, Perona P, Winner K. Species distribution modeling for machine learning practitioners: a review. In: *ACM SIGCAS conference on computing and sustainable societies*. 2021:329–348.
9 Vyhmeister E, Provan G, Doyle B, Bourke B, Castane GG, Reyes-Bozo L. Comparison of time series and mechanistic models of vector-borne diseases. *Spat Spatiotemporal Epidemiol*. 2022;41:100478.
10 Lippi CA, Ryan SJ, White AL, Gaff HD, Carlson CJ. Trends and opportunities in tick-borne disease geography. *J Med Entomol*. 2021;58(6):2021–2029.
11 Kopsco HL, Smith RL, Halsey SJ. A scoping review of species distribution modeling methods for tick vectors. *Front Ecol Evol*. 2022;10. Available from: https://www.frontiersin.org/articles/10.3389/fevo.2022.893016.
12 O'Neill X, White A, Gortázar C, Ruiz-Fons F. The impact of host abundance on the epidemiology of tick-borne infection. *Bull Math Biol*. 2023;85(4):30. https://doi.org/10.1007/s11538-023-01133-8.
13 Tricco AC, Lillie E, Zarin W, et al. PRISMA extension for scoping reviews (PRISMA-ScR): checklist and explanation. *Ann Intern Med*. 2018;169(7):467–473.
14 Falagas ME, Pitsouni EI, Malietzis GA, Pappas G. Comparison of PubMed, scopus, Web of science, and google scholar: strengths and weaknesses. *FASEB J*. 2008;22(2):338–342.
15 Soucy JPR, Slatculescu AM, Nyiraneza C, et al. High-resolution ecological niche modeling of Ixodes scapularis ticks based on passive surveillance data at the northern frontier of lyme disease emergence in North America. *Vector Borne Zoonotic Dis*. 2018;18(5):235–242.
16 Lewis J, Boudreau CR, Patterson JW, Bradet-Legris J, Lloyd VK. Citizen science and community engagement in tick surveillance—a Canadian case study. *Healthcare*. 2018;6(1):22 [cited 2023 Jun 30] Available from: https://www.mdpi.com/2227-9032/6/1/22.
17 Nelder MP, Russell CB, Dibernardo A, et al. Monitoring the patterns of submission and presence of tick-borne pathogens in Ixodes scapularis collected from humans and companion animals in Ontario, Canada (2011–2017). *Parasit Vectors*. 2021;14(1):260. https://doi.org/10.1186/s13071-021-04750-1.
18 Porter WT, Motyka PJ, Wachara J, et al. Citizen science informs human-tick exposure in the Northeastern United States. *Int J Health Geogr*. 2019;18(1):9. https://doi.org/10.1186/s12942-019-0173-0.
19 Finch C, Al-Damluji MS, Krause PJ, et al. Integrated assessment of behavioral and environmental risk factors for Lyme disease infection on block island, Rhode Island. *PLoS One*. 2014;9(1):e84758 [cited 2023 Feb 23] Available from: https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0084758.
20 Zolnik CP, Falco RC, Kolokotronis SO, Daniels TJ. No observed effect of landscape fragmentation on pathogen infection prevalence in blacklegged ticks (Ixodes scapularis) in the northeastern United States. *PLoS One*. 2015;10(10):e0139473.
21 Khatchikian CE, Prusinski M, Stone M, et al. Geographical and environmental factors driving the increase in the Lyme disease vector Ixodes scapularis. *Ecosphere*. 2012;3(10):art85. [cited 2023 Feb 23] Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3872055/.
22 Diuk-Wasser MA, Hoen AG, Cislo P, et al. Human risk of infection with Borrelia burgdorferi, the Lyme disease agent, in eastern United States. *Am J Trop Med Hyg*. 2012;86(2):320–327.
23 Berger KA, Ginsberg HS, Dugas KD, Hamel LH, Mather TN. Adverse moisture events predict seasonal abundance of Lyme disease

vector ticks (Ixodes scapularis). *Parasit Vectors*. 2014;7(1):181. https://doi.org/10.1186/1756-3305-7-181.

24 Brinkerhoff RJ, Gilliam WF, Gaines D. Lyme disease, Virginia, USA, 2000–2011. *Emerg Infect Dis*. 2014;20(10). [cited 2023 Jun 30] Available from: https://wwwnc.cdc.gov/eid/article/20/10/13-0782_article.

25 Tran T, Porter WT, Salkeld DJ, Prusinski MA, Jensen ST, Brisson D. Estimating disease vector population size from citizen science data. *J R Soc Interface*. 2021;18(184):20210610.

26 Berger KA, Wang Y, Mather TN. MODIS-derived land surface moisture conditions for monitoring blacklegged tick habitat in southern New England. *Int J Remote Sens*. 2013;34(1):73–85. [cited 2023 Feb 23] Available from: https://www.tandfonline.com/doi/full/10.1080/01431161.2012.705447.

27 Werden L, Barker IK, Bowman J, et al. Geography, deer, and host biodiversity shape the pattern of Lyme disease emergence in the Thousand Islands Archipelago of Ontario, Canada. *PLoS One*. 2014;9(1):e85640.

28 Ferrell AM, Brinkerhoff RJ. Using landscape analysis to test hypotheses about drivers of tick abundance and infection prevalence with Borrelia burgdorferi. *Int J Environ Res Public Health*. 2018;15(4):737. [cited 2023 Feb 23] Available from: https://www.mdpi.com/1660-4601/15/4/737.

29 Wallace D, Ratti V, Kodali A, et al. Effect of rising temperature on lyme disease: Ixodes scapularis population dynamics and Borrelia burgdorferi transmission and prevalence. *Can J Infect Dis Med Microbiol*. 2019;2019:9817930. https://doi.org/10.1155/2019/9817930.

30 Koffi JK, Leighton PA, Pelcat Y, et al. Passive surveillance for I. scapularis ticks: enhanced analysis for early detection of emerging Lyme disease risk. *J Med Entomol*. 2012;49(2):400–409.

31 Gabriele-Rivet V, Arsenault J, Badcock J, et al. Different ecological niches for ticks of public health significance in Canada. *PLoS One*. 2015;10(7):e0131282.

32 Gabriele-Rivet V, Koffi JK, Pelcat Y, et al. A risk model for the lyme disease vector Ixodes scapularis (Acari: Ixodidae) in the prairie provinces of Canada. *J Med Entomol*. 2017;54(4):862–868. https://doi.org/10.1093/jme/tjx036.

33 Clow KM, Leighton PA, Ogden NH, et al. Northward range expansion of Ixodes scapularis evident over a short timescale in Ontario, Canada. *PLoS One*. 2017;12(12):e0189393.

34 Leighton P, Koffi J, Pelcat Y, Lindsay R, Ogden N. Predicting the speed of tick invasion: an empirical model of range expansion for the Lyme disease vector Ixodes scapularis in Canada. *J Appl Ecol*. 2012;49.

35 Sharareh N, Behler RP, Roome AB, Shepherd J, Garruto RM, Sabounchi NS. Risk factors of Lyme disease: an intersection of environmental ecology and systems science. *Healthcare*. 2019;7(2):66. [cited 2023 Feb 23] Available from: https://www.mdpi.com/2227-9032/7/2/66.

36 Kotchi SO, Bouchard C, Brazeau S, Ogden NH. Earth observation-informed risk maps of the Lyme disease vector Ixodes scapularis in central and eastern Canada. *Remote Sens*. 2021;13(3):524. [cited 2023 Feb 23] Available from: https://www.mdpi.com/2072-4292/13/3/524.

37 Hahn MB, Jarnevich CS, Monaghan AJ, Eisen RJ. Modeling the geographic distribution of Ixodes scapularis and Ixodes pacificus (Acari: Ixodidae) in the contiguous United States. *J Med Entomol*. 2016;53(5):1176–1191.

38 Chen D, Wong H, Belanger P, Moore K, Peterson M, Cunningham J. Analyzing the correlation between deer habitat and the component of the risk for lyme disease in eastern Ontario, Canada: a GIS-based approach. *ISPRS Int J Geoinf*. 2015;4(1):105–123. [cited 2023 Feb 23] Available from: https://www.mdpi.com/2220-9964/4/1/105.

39 McPherson M, García-García A, Cuesta-Valero FJ, et al. Expansion of the lyme disease vector Ixodes scapularis in Canada inferred from CMIP5 climate projections. *Environ Health Perspect*. 2017;125(5):057008.

40 Ogden NH, Radojevic M, Wu X, Duvvuri VR, Leighton PA, Wu J. Estimated effects of projected climate change on the basic reproductive number of the Lyme disease vector Ixodes scapularis. *Environ Health Perspect*. 2014;122(6):631–638.

41 Wu X, Duvvuri VR, Lou Y, Ogden NH, Pelcat Y, Wu J. Developing a temperature-driven map of the basic reproductive number of the emerging tick vector of Lyme disease Ixodes scapularis in Canada. *J Theor Biol*. 2013;319:50–61. Available from: https://www.sciencedirect.com/science/article/pii/S0022519312005899.

42 Cheng A, Chen D, Woodstock K, Ogden NH, Wu X, Wu J. Analyzing the potential risk of climate change on Lyme disease in eastern Ontario, Canada using time series remotely sensed temperature data and tick population modelling. *Remote Sens*. 2017;9(6):609. [cited 2023 Feb 23] Available from: https://www.mdpi.com/2072-4292/9/6/609.

43 Ogden NH, Lindsay LR, Leighton PA. Predicting the rate of invasion of the agent of Lyme disease "Borrelia burgdorferi." *J Appl Ecol*. 2013;50(2):510–518. [cited 2023 Feb 23] Available from: https://www.jstor.org/stable/24031482.

44 Illoldi-Rangel P, Rivaldi CL, Sissel B, et al. Species distribution models and ecological suitability analysis for potential tick vectors of lyme disease in Mexico. *J Trop Med*. 2012;2012:959101. [cited 2023 Feb 23] Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3307011/.

45 Slatculescu AM, Clow KM, McKay R, et al. Species distribution models for the eastern blacklegged tick, Ixodes scapularis, and the Lyme disease pathogen, Borrelia burgdorferi, in Ontario, Canada. *PLoS One*. 2020;15(9):e0238126.

46 Johnson TL, Bjork JKH, Neitzel DF, Dorr FM, Schiffman EK, Eisen RJ. Habitat suitability model for the distribution of Ixodes scapularis (Acari: Ixodidae) in Minnesota. *J Med Entomol*. 2016;53(3):598–606. https://doi.org/10.1093/jme/tjw008.

47 Feria-Arroyo TP, Castro-Arellano I, Gordillo-Perez G, et al. Implications of climate change on the distribution of the tick vector Ixodes scapularis and risk for Lyme disease in the Texas-Mexico transboundary region. *Parasit Vectors*. 2014;7(1):199. https://doi.org/10.1186/1756-3305-7-199.

48 Zhang L, Ma D, Li C, Zhou R, Wang J, Liu Q. Projecting the potential distribution areas of Ixodes scapularis (Acari: Ixodidae) driven by climate change. *Biology*. 2022;11(1):107. [cited 2023 May 25] Available from: https://www.mdpi.com/2079-7737/11/1/107.

49 Dhingra R, Jimenez V, Chang HH, et al. Spatially-explicit simulation modeling of ecological response to climate change: methodological considerations in predicting shifting population dynamics of infectious disease vectors. *ISPRS Int J Geoinf*. 2013;2(3):645–664. [cited 2023 Feb 23] Available from: https://www.mdpi.com/2220-9964/2/3/645.

50 Johnson TL, Boegler KA, Clark RJ, et al. An acarological risk model predicting the density and distribution of host-seeking Ixodes scapularis nymphs in Minnesota. *Am J Trop Med Hyg*. 2018;98(6):1671–1682.

51 Ripoche M, Gasmi S, Adam-Poupart A, et al. Passive tick surveillance provides an accurate early signal of emerging lyme disease risk and human cases in southern Canada. *J Med Entomol*. 2018;55(4):1016–1026.

52 Simon JA, Marrotte RR, Desrosiers N, et al. Climate change and habitat fragmentation drive the occurrence of Borrelia burgdorferi, the agent of Lyme disease, at the northeastern limit of its distribution. *Evol Appl*. 2014;7(7):750–764.

53 Burrows H, Slatculescu AM, Feng CX, et al. The utility of a maximum entropy species distribution model for Ixodes scapularis in predicting the public health risk of Lyme disease in Ontario, Canada. *Ticks Tick Borne Dis*. 2022;13(5):101969.

54 Ripoche M, Lindsay LR, Ludwig A, Ogden NH, Thivierge K, Leighton PA. Multi-scale clustering of lyme disease risk at the expanding leading edge of the range of Ixodes scapularis in Canada. *Int J Environ Res Public Health*. 2018;15(4):603. [cited 2023 Feb 23] Available from: https://www.mdpi.com/1660-4601/15/4/603.

55 Alexopoulos EC. Introduction to multivariate regression analysis. *Hippokratia*. 2010;14(Suppl 1):23–28. [cited 2024 Jan 15] Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3049417/.

56 Harrell FE, Lee KL, Mark DB. Multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. *Stat Med*. 1996;15(4):361–387.

57 Quaghebeur W, Torfs E, De Baets B, Nopens I. Hybrid differential equations: integrating mechanistic and data-driven techniques for modelling of water systems. *Water Res*. 2022;213:118166.

58 Solle D, Hitzmann B, Herwig C, et al. Between the Poles of data-driven and mechanistic modeling for process operation. *Chem Ing Tech*. 2017;89(5):542–561. [cited 2024 Jan 15] Available from: https://onlinelibrary.wiley.com/doi/abs/10.1002/cite.201600175.

59 Robinson S. *Simulation: the practice of model development and use*. Bloomsbury Publishing; 2014.

60 Tufts DM, McClure M, Diuk-Wasser MA. Ixodes scapularis (Acari: Ixodidae) nymphal survival and host-finding success in the eastern United States. *J Med Entomol*. 2021;58(2):929–938.

61 Lobo JM, Jiménez-Valverde A, Real R. AUC: a misleading measure of the performance of predictive distribution models. *Glob Ecol Biogeogr*. 2008;17(2):145–151. [cited 2023 Aug 1] Available from: https://onlinelibrary.wiley.com/doi/10.1111/j.1466-8238.2007.00358.x.

62 Baker RE, Peña JM, Jayamohan J, Jérusalem A. Mechanistic models versus machine learning, a fight worth fighting for the biological community? *Biol Lett*. 2018;14(5):20170660. [cited 2023 Feb 24] Available from: https://royalsocietypublishing.org/doi/10.1098/rsbl.2017.0660.

63 Fieler AM, Rosendale AJ, Farrow DW, et al. Larval thermal characteristics of multiple ixodid ticks. *Comp Biochem Physiol Mol Integr Physiol*. 2021;257:110939.

64 Ripoche M, Bouchard C, Irace-Cima A, Leighton P, Thivierge K. Current and future distribution of Ixodes scapularis ticks in Québec: field validation of a predictive model. *PLoS One*. 2022;17(2):e0263243.

65 Johnson JL, Ginsberg HS, Zhioua E, et al. Passive tick surveillance, dog seropositivity, and incidence of human lyme disease. *Vector Borne Zoonotic Dis*. 2004;4(2):137–142.

66 Ogden NH, Trudel L, Artsob H, et al. Ixodes scapularis ticks collected by passive surveillance in Canada: analysis of geographic distribution and infection with Lyme borreliosis agent Borrelia burgdorferi. *J Med Entomol*. 2006;43(3):600–609.

67 Phillips SJ, Anderson RP, Schapire RE. Maximum entropy modeling of species geographic distributions. *Ecol Model*. 2006;190(3–4):231–259.

68 Syfert MM, Smith MJ, Coomes DA. The effects of sampling bias and model complexity on the predictive performance of MaxEnt species distribution models. *PLoS One*. 2013;8(2):e55158.

69 Wisz MS, Hijmans RJ, Li J, et al. Effects of sample size on the performance of species distribution models. *Divers Distrib*. 2008;14(5):763–773. [cited 2023 Jun 30] Available from: https://www.jstor.org/stable/20172033.

70 Lieske DJ, Lloyd VK. Combining public participatory surveillance and occupancy modelling to predict the distributional response of Ixodes scapularis to climate change. *Ticks Tick Borne Dis*. 2018;9(3):695–706.